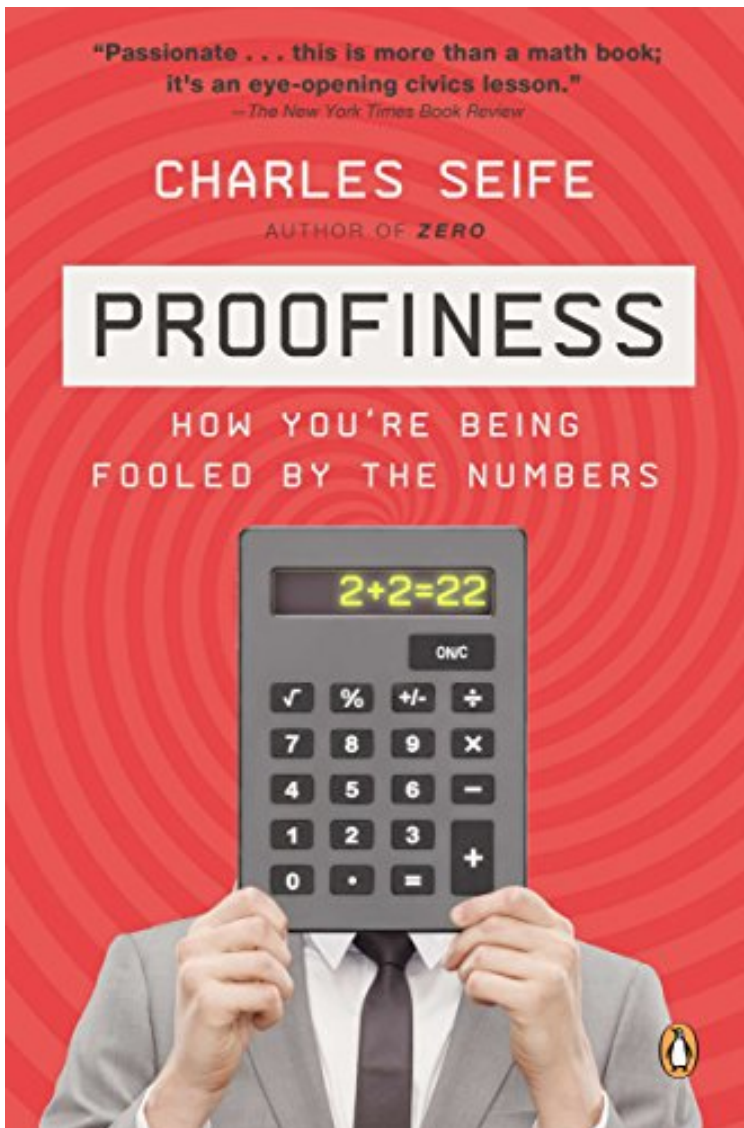


[Download] File size: 46.Mb

# Proofiness: How You're Being Fooled by the Numbers



Par Charles Seife  
ePub | \*DOC | audiobook | ebooks |  
Download PDF

Dtails sur le produit Rang parmi les ventes : #541069 dans eBooksPubli le: 2010-09-23Sorti le: 2010-09-23Format: Ebook Kindle

[Download] Proofiness: How You're Being Fooled by the Numbers

**Par Charles Seife : Proofiness: How You're Being Fooled by the Numbers** before purchasing it in order to gage whether or not it would be worth my time, and all praised Proofiness: How You're Being Fooled by the Numbers:

Download

Read Online

## Description :

Prsentation de l'diteurThe bestselling author of Zero shows how mathematical misinformation pervades-and shapes-our daily lives. According to MSNBC, having a child makes you stupid. You actually lose IQ points.

Good Morning America has announced that natural blondes will be extinct within two hundred years. Pundits estimated that there were more than a million demonstrators at a tea party rally in Washington, D.C., even though roughly sixty thousand were there. Numbers have peculiar powers-they can disarm skeptics, befuddle journalists, and hoodwink the public into believing almost anything. "Proofiness," as Charles Seife explains in this eye-opening book, is the art of using pure mathematics for impure ends, and he reminds

readers that bad mathematics has a dark side. It is used to bring down beloved government officials and to appoint undeserving ones (both Democratic and Republican), to convict the innocent and acquit the guilty, to ruin our economy, and to fix the outcomes of future elections. This penetrating look at the intersection of math and society will appeal to readers of *Freakonomics* and the books of Malcolm Gladwell.

Phony Facts, Phony Figures Facts are stubborn things; and whatever may be our wishes, our inclinations, or the dictates of our passion, they cannot alter the state of facts and evidence. John Adams Facts are stupid things. Ronald Reagan If you want to get people to believe something really, really stupid, just stick a number on it. Even the silliest absurdities seem plausible the moment that they're expressed in numerical terms. Are blonds an endangered species? A few years ago, the media were all abuzz about a World Health Organization study proving that natural blonds would soon be a thing of the past. The BBC declared that people with blond hair will become extinct by 2202. Good Morning America told its viewers that natural blonds will vanish from the face of the earth within two hundred years because the blond gene is not as strong a gene as brunettes. The story was winging its way around the globe until the WHO issued an unusual statement: WHO wishes to clarify that it has never conducted research on this subject. Nor, to the best of its knowledge, has WHO issued a report predicting that natural blondes are likely to be extinct by 2202. WHO has no knowledge of how these news reports originated but would like to stress that we have no opinion on the future existence of blondes. It should have been obvious that the story was bogus, even before the WHO denial. One geneticist had even told the BBC as much. Genes don't die out unless there is a disadvantage to having that gene, he said. They don't disappear. But the BBC had suspended its faculties of disbelief. The reason, in part, was because of a phony number. The specificity, the seeming mathematical certainty of the prediction of when the last blond would be born, gave the story an aura of plausibility. It suckered journalists who should have known better. No matter how idiotic, how unbelievable an idea is, numbers can give it credibility. Fifty-eight percent of all the exercise done in America is broadcast on television, MSNBC host Deborah Norville declared in 2004, with a completely straight face. For instance, of the 3.5 billion sit-ups done during 2003, two million, three hundred thousand [sic] of them were on exercise shows. Without once pausing to think, Norville swallowed the bogus statistics and regurgitated them for her audience; just a moment's reflection should have revealed that the story was nonsense. (A few months later, perhaps unwilling to be outdone by his colleague, MSNBC host Keith Olbermann touted a five-year study just concluded at Indiana University which proved that upon the birth of their first child, 100 percent of parents lose at least 12 IQ points, and the average loss is 20. These numbers, too, are fiction.) The numbers had short-circuited Norville's brain, rendering her completely incapable of critical thought. It's typical. Numbers have that power over us, because in its purest form, a number is truth. The cold and crystalline world of numbers gives us the rarest of all things: absolute certainty. Two plus two is always four. It was always so, long before our species walked the earth, and it will be so long after the end of civilization. But there are numbers and there are numbers. Pure numbers are the domain of mathematicians—curious people who study numbers in the abstract, as Platonic ideals that reveal a higher truth. To a mathematician, a number is interesting in its own right. Not so for the rest of us. For a nonmathematician, numbers are interesting only when they give us information about the world. A number only takes on any significance in everyday life when it tells us how many pounds we've gained since last month or how many dollars it will cost to buy a sandwich or how many weeks are left before our taxes are due or how much money is left in our IRAs. We don't care about the properties of the number five. Only when that number becomes attached to a unit—the pounds or dollars or weeks that signify what real-world property the number represents—does it become interesting to a nonmathematician. A number without a unit is ethereal and abstract. With a unit, it acquires a meaning—but at the same time, it loses its purity. A number with a unit can no longer inhabit the Platonic realm of absolute truth; it becomes tainted with the uncertainties and imperfections of the real world. To mathematicians, numbers represent indisputable truths; to the rest of us, they come from inherently impure, imperfect measurements. This uncertainty is unavoidable. Every unit represents an implied measurement. Inches, for example, represent an implied measurement of length; when someone says that a coffee table is eighteen inches wide, he's saying that if we were to take the table and measure it with a ruler, the table would have the same length as eighteen of the little hash marks we call inches. When someone says he weighs 180 pounds, he's saying that if you measured him with a bathroom scale, the number on the dial would read 180. Every number that has a real-world meaning is tied, at least implicitly, to a measurement of some kind. Liters are tied to a measurement of volume. Acres imply a measurement of area. Watts imply a measurement of power. A measurement of speed is expressed in miles per hour or in knots. A measurement of wealth is in dollars or euros or yuan. If

someone says that he has five fingers, hes saying that if you count his digits and counting objects is a measurement too the answer will be five fingers. Its universal; behind every real-world number, theres a measurement. And because measurements are inherently error-prone (theyre performed by humans, after all, using instruments made by humans), they arent perfectly reliable. Even the simple act of counting objects is prone to error, as we shall see. As a result, every measurement and every real-world number is a little bit fuzzy, a little bit uncertain. It is an imperfect reflection of reality. A number is always impure: it is an admixture of truth, error, and uncertainty. Proofiness has power over us because were blind to this impurity.

Numbers, figures, and graphs all have an aura of perfection. They seem like absolute truth; they seem indisputable. But this is nothing but an illusion. The numbers of our everyday world the numbers we care about are flawed, and not just because measurements are imperfect. They can be changed and tinkered with, manipulated and spun and turned upside down. And because those lies are clad in the divine white garb of irrefutable fact, they are incredibly powerful. This is what makes proofiness so very dangerous. Its true: all measurements are imperfect. However, some are more imperfect than others. As a result, not all numbers are equally fallible. Some numbers, those that are based upon extremely reliable and objective measurements, can come very close to absolute truth. Others based on unreliable or subjective or nonsensical measurements come close to absolute falsehood. Its not always obvious which are which. Truthful numbers tend to come from good measurements. And a good measurement should be reproducible: repeat the measurement two or ten or five hundred times, you should get pretty much the same answer each time. A good measurement should also be objective. Even if different observers perform the measurement with different kinds of measuring devices, they should all agree about the outcome. A measurement of time or of length, for example, is objective and reproducible. If you hand stopwatches to a dozen people watching the same event say, the Kentucky Derby and ask them to time the race, theyll all come up with roughly the same answer (if theyre competent). A whole stadium full of people, each using different stopwatches and clocks and time-measuring devices, would agree that the race took, say, roughly one minute and fifty-nine seconds to complete, give or take a few fractions of a second. Similarly, ask a dozen people to measure an object like a pencil, and it doesnt matter whether they use a ruler or a tape measure or a laser to gauge its length. When they complete their measurements, theyll all agree that the pencil is, say, four and a half inches long, give or take a fraction of an inch. The result of the measurement doesnt depend on whos doing the measuring or what kind of equipments being used the answer is always roughly the same. This is an essential property of a good measurement. Bad measurements, on the other hand, deceive us into believing a falsehood sometimes by design. And there are lots of bad measurements. Luckily, there are warning signs that tell you when a measurement is rotten. One red flag is when a measurement attempts to gauge something thats ill-defined. For example, intelligence is a slippery concept nobody can nail down precisely what it means but that doesnt stop people from trying to measure it. Theres an entire industry devoted to trying to pin numbers to peoples brains; dozens and dozens of tests purport to measure intelligence, intellectual ability, or aptitude. (An applicant to Mensa, the high-IQ society, has his choice of some thirty-odd exams to prove his intellectual superiority.) Testing is just the tip of the multimillion-dollar iceberg. After measuring your intelligence, some companies sell you a set of exercises that help you improve your score on their tests, proving that youve become smarter. Dubious claims are everywhere: video games, DVDs, tapes, and books promise to make you more intelligent for a price. Even the British Broadcasting Company tried to cash in on the intelligence-enhancement fad. In 2006, a BBC program promised that you can become 40 percent cleverer within seven days by following diet advice and doing a few brain-teasers. Was there a sudden surge in the number of Britons understanding quantum physics? Unlikely. So long as researchers argue about what intelligence is, much less how to measure it, you can be assured that the 40 percent cleverer claim is worthless. In fact, I can personally guarantee that youll instantly be 63 percent smarter if you ignore all such statements. Even if a phenomenon has a reasonable definition that everybody can agree about, its not always easy to measure that phenomenon. Sometimes theres no settled-upon way to measure something reliably theres no measuring device or other mechanism that allows different observers to get the same numbers when trying to quantify the phenomenon which is another sign that the measurement is dubious. For example, its hard to measure pain or happiness. Theres no such thing as a painometer or a happinesscope that can give you a direct and repeatable reading about what a subject is feeling. (This doesnt stop scientists from trying. To measure the effectiveness of painkillers in mice, some scientists use a calibrated hotplate; they measure pain by timing how long it takes for the mouse to jump or otherwise react to the hot surface.) In lieu of devices that can measure these experiences directly, researchers are forced to use crude and unreliable

methods to try to get a handle on the degree of pain or happiness that a subject is feeling. They use questionnaires to gauge how much pain someone is in (circle the frowny face that corresponds with your level of pain) or how good someone feels (circle the number that represents how happy you are). Making matters even more difficult, pain and joy are subjective experiences. People feel them differently; some people are extremely tolerant to pain and some are very sensitive; some are emotional and regularly climb up towering peaks of bliss while others are more even-keeled. This means that even if a scientist could somehow devise an experiment where people would experience exactly the same amount of pain or joy, they would almost certainly give different answers on the questionnaire because their perceptions are different. A swift kick to the shins will elicit a super-duper frowny face from someone who has a low pain tolerance, while a more stoic person would barely deviate from a mild grimace. When rational people will come up with different answers to a question how painful a blow to the head is, how beautiful a person in a photo is, how easy a book is to read, how good a movie is the measurement can have some value, but the number is certainly far from the realm of absolute truth. But it's not the farthest away. That honor goes to numbers that are tied to phony measurements—measurements that are fake or meaningless or even nonexistent. Numbers like these are everywhere, but product labels seem to be their favorite habitat. Just try to imagine what kind of measurement L'Oréal made to determine that its Extra Volume Collagen Mascara gives lashes twelve times more impact. (Perhaps they had someone blink and listened to how much noise her eyelashes made when they clunked together.) How much diligence do you think Vaseline put into its research that allowed it to conclude its new moisturizer delivers 70 percent more moisture in every drop. (Presumably it would deliver less moisture than water, which is 100 percent moisture, after all.) No matter how ridiculous an idea, putting it into numerical form makes it sound respectable, even if the implied measurement is transparently absurd. This is why paranormal researchers feel compelled to claim, without giggling, that 29 percent of Christian saints had exhibited psychic powers. Making up scientific-sounding measurements is a grand old tradition; cigarette companies used to excel at the practice, the better to fill their ads with a thick haze of nonsense. From first puff to last, Chesterfield gives you a smoke measurably smoother . . . cooler . . . best for you! read one advertisement from 1955. You can't measure the smoothness and coolness of a cigarette any more than you can measure the impact of an eyelash. Even if people tried to quantify impact or smoothness or coolness, the results would be worthless. These are phony measurements. They're like actors dressed up in lab coats—they appear to be scientific, but they're fake through and through. As a result, the numbers associated with these measurements are utterly devoid of meaning. They are fabricated statistics: Potemkin numbers. According to legend, Prince Grigory Potemkin didn't want the empress of Russia to know that a region in the Crimea was a barren wasteland. Potemkin felt he had to convince the empress that the area was thriving and full of life, so he constructed elaborate façades along her route—crudely painted wooden frameworks meant to look like villages and towns from afar. Even though these Potemkin villages were completely empty—a closer inspection would reveal them to be mere imitations of villages rather than real ones—they were good enough to fool the empress, who breezed by them without alighting from her carriage. Potemkin numbers are the mathematical equivalent of Potemkin villages. They're numerical façades that look like real data. Meaningful real-world numbers are tied to a reasonably solid measurement of some sort, at least implicitly. Potemkin numbers aren't meaningful because either they are born out of a nonsensical measurement or they're not tied to a genuine measurement at all, springing forth fully formed from a fabricator's head. For example, on October 16, 1995, Louis Farrakhan, leader of the Nation of Islam, held an enormous rally: the Million Man March. Of course, the gathering was named long before anyone knew whether a million men would actually attend the rally. When buses and trains and planes filled with men began converging on the National Mall in Washington, D.C., it was a huge event, but did it really live up to its name? Quite naturally, Farrakhan said that it did; his unofficial count of the crowd topped one million people. However, Farrakhan's count was a Potemkin number. It was a foregone conclusion that the organizers of the Million Man March would declare that a million men were in attendance, regardless of how many actually attended—anything else would be embarrassing. The Park Service, which was in charge of giving an official estimate of the crowd size, was already feeling the pressure to inflate the numbers. If we say [the crowd] was 250,000, we'll be told it was a half-million, a U.S. Park Police officer told the Washington Post shortly before the rally. If we say it was a half-million, we'll be told it was a million. Anything short of a million, and you can probably bet we'll take some heat for it. Nevertheless, the Park Service dutifully peered at aerial photos and counted heads in an attempt to size up the crowd. As predicted, when the official tally came in—400,000 people, give or take 20 percent or so—a furious Farrakhan threatened to

sue. (Though the Park Services estimate was probably the best out there, some scholarly estimates went as high as roughly 900,000 people still short of a million, but more than double the official estimates. This just goes to show that even the seemingly simple act of counting objects can be difficult and politically sensitive.) As a result, the Park Service stopped estimating crowd sizes for more than a decade, giving rally organizers free rein to make up Potemkin-number crowd estimates without fear of contradiction from a reliable source.

Whenever there's a big public event, someone has a vested interest in making up a number that makes the crowd look huge. Pretty much every year, the organizers of Pasadena's Rose Parade announce (presumably for the ears of sponsors) that they estimate a crowd of one million spectators; the real number is probably half that. Estimates of the crowd at Barack Obama's inauguration topped five million at one point probably between double and triple the number of people actually in attendance. The right is just as guilty as the left of fudging crowd numbers. In September 2009, a number of right-wing commentators claimed that ABC News had estimated that anti-administration tea party protests in D.C. had drawn more than a million people; in fact, ABC had pegged the crowd at between 60,000 and 70,000 attendees. Two months later, Republican representative Michele Bachmann bragged on the conservative Sean Hannity Show that a protest she organized had attracted between 20,000 and 45,000 angry citizens and a video montage of swirling crowds seemed to support her assertions. However, the estimate was way off the Washington Post estimated that the crowd was about 10,000. Worse yet, some of The Sean Hannity Shows footage was recycled from the much larger September protest, making the crowd look much more substantial than it actually was. (After the deception was exposed by comedian Jon Stewart on The Daily Show, Sean Hannity promptly admitted that he had made an error: We screwed up. It was an inadvertent mistake, but it was a mistake nonetheless. Then, smirking, he thanked Stewart for watching the show.) Creators of Potemkin numbers care little about whether their numbers are grounded in any sort of reality. From afar, however, they seem convincing, and a Potemkin number can be a powerful tool to prop up a sagging argument or to bludgeon a stubborn opponent.

Even the flimsiest of them can do tremendous damage. Joe McCarthy's famous claim to know of 205 communists in the State Department, for example, was transparently false, yet it made him a national figure. Using Potemkin numbers is the most overt form of proofiness. It takes the least skill to pull off. After all, it's incredibly easy to fabricate a number that suits whatever argument you are trying to make. (This is why Potemkin numbers are so common: 78 percent of all statistics are made up on the spot and 36 percent of readers will actually believe that statistic.) However, the power of Potemkin numbers is limited. It begins to evaporate as soon as someone is thoughtful enough to examine the numbers more carefully. There are other more subtle and dangerous forms of proofiness to watch out for. There's an anecdote about an aging guide at a natural history museum. Every day, the guide gives tours of the exhibits, without fail ending with the most spectacular sight in the museum. It's a skeleton of a fearsome dinosaur, a tyrannosaurus rex that towers high over the wide-eyed tour group. One day, a teenager gestures at the skeleton and asks the guide, How old is it? Sixty-five million and thirty-eight years old, the guide responds proudly. How could you possibly know that? the teenager shoots back. Simple! On the very first day that I started working at the museum, I asked a scientist the very same question. He told me that the skeleton was sixty-five million years old. That was thirty-eight years ago. It's not a hilarious anecdote, but it illustrates an important point. The number the museum guide gives, sixty-five million and thirty-eight years old, is absurd. When the guide tacks on that thirty-eight years, he's committed a numerical sin that turns his answer into a falsehood. He's committed an act of proofiness. When the scientist told the guide that the dinosaur was sixty-five million years old, the answer was tied, implicitly, to a set of measurements. Paleontologists somehow measured the age of the dinosaur perhaps by looking at the decay of radioactive materials in its skeleton or by looking at the age of the rock strata that the fossils were embedded in, yielding an answer of sixty-five million years. But that number is only an approximation; fossil dates can be off by a few hundred thousand or even a few million years. The sixty-five-million-year date isn't perfectly accurate. Measurements never are. They always contain errors that obscure the truth. By being very careful with your measurements, by using sophisticated and precise tools, you can drive that error down. A typical ruler makes errors on the order of half of a millimeter. If you use one to measure, say, a pencil, and you're very diligent, closing one eye to make sure that you're lining the pencil up with the hash marks properly, you can be confident that the answer you get (150 millimeters) is no more than half a millimeter from the truth. This isn't the final word, though. You can get a better answer if you want. You just have to shell out money for fancier equipment. Perhaps you can use a costly engineering ruler with ultra-precise hash marks or, even better, a set of calipers or a micrometer that can measure the same object with much more precision. With that kind of equipment, you might drive the

error down to even a thousandth or a ten-thousandth of a millimeter. Where an old wooden ruler says that a pencil is 150 millimeters long, give or take half a millimeter, an expensive set of calipers might be able to declare that the pencil is 150.112 millimeters long, give or take 0.001 millimeters. By virtue of their smaller error, the calipers can get you closer to the truth than the ruler can. In the laboratory, researchers using lasers or atom interferometers or other devices can make measurements that are even more precise still, 0.0000001 millimeters and beyond. In theory, someone might be able to discover that the pencil is 150.1124835 millimeters long, give or take a tiny, tiny fraction of a millimeter. The more precise (and expensive) your measurements, the more you reduce the error inherent to those measurements, getting you closer and closer to the truth. (Assuming that the truth exists, which isn't always so obvious. These are atomic-scale distances, and the concept of length stops having a fixed meaning when you're looking at measurements this precise. The atoms in the pencil are constantly moving, so the pencil is expanding and contracting so the pencil no longer really has a given length for more than an instant. After a certain point, nature conspires to keep you from finding out the true length of a pencil by making the truth itself rather fuzzy. But the general point still holds: there's always a level of error in any measurement that obscures reality.) However, that error never completely disappears. No matter how careful you are, the measurement is always, in some sense, a mere approximation of reality, an imperfect reflection of truth. Error obscures reality, injecting a touch of falsehood and uncertainty into our measurements. Whether that error is half a millimeter or 0.001 millimeters or 0.000001 millimeters, it is still there, obscuring the true answer just a little. Error ensures that any measurement is true only up to a point dictated by the nature of the device that you're using to make the measurement. (We'll get to the pollsters favorite slippery convenience, margin of error, in chapter 4.) Trying to push beyond the precision of the instrument is pointless. Even the fanciest ruler will only allow you to measure a pencil with a precision of a quarter or a tenth of a millimeter or so. No matter how careful you are, no matter how many times you measure and re-measure, no matter how carefully you squint through a magnifying glass when you read the hash marks, there's no way you can use a ruler to get an answer much better than roughly 150.1 millimeters for the length of a pencil. That's as precise as you can get with a ruler. If you listen carefully enough, numbers tell you that they're only approximations. They reveal their limitations better yet, they tell you how far to trust them. This information is encoded in the way we talk about numbers; it's already second nature to you, even though you might not recognize it. When someone declares that a pencil is 150.112 millimeters long, you automatically assume that the measurement is extremely precise. On the other hand, if he says it is 150 millimeters long, you would assume that the measurement is much rougher. Nice round numbers are sending a subliminal signal that their associated measurements have large errors; the numbers are announcing that you can't trust them very far because they're crude approximations. Long, ultra-specific numbers send exactly the opposite message: that they come from measurements that are more trustworthy and closer to absolute truth. All real-world numbers behave like this. When someone tells you that his car cost \$15,000, you automatically assume that there's quite a bit of slop in the figure; the real cost was somewhere in the ballpark of fifteen grand, give or take a few hundred dollars. Yet if someone says that his car cost \$15,232, you then assume that this was the precise amount he paid, give or take a few pennies. Similarly, if someone tells you that he's eighteen years old, you expect that he's somewhere between eighteen and nineteen years of age. If he says that he's eighteen years, two months, and three days old, you know that his answer is good to within a few hours and that he's probably a bit obsessive. The roundness of a number gives you clues about how precise the number is, and how seriously you can take it. This is the key to the dinosaur anecdote. When a scientist says that a dinosaur skeleton is sixty-five million years old, it's a signal that the number is a fairly rough approximation; the measurement error is on the order of tens or hundreds of thousands of years. (A great deal of scientific effort goes into increasing the precision of measurements to get closer to the truth by adding an extra decimal place or two. For example, in the past two decades, scientists' estimate of the age of the universe went from about 15 billion years to 14 billion years to 13.7 billion years. This seemingly subtle change represents an extraordinary almost revolutionary advance in our knowledge about the cosmos.) In reality, the skeleton might be 64,963,211 years old; perhaps it's 65,031,844 years old. However, the paleontologists' measurements weren't precise enough to reveal that truth. When he said that the skeleton was sixty-five million years old, he was admitting that his measurement had large errors; it was sixty-five million years old, give or take tens or hundreds of thousands or even millions of years. The museum guide screwed up when he took the sixty-five million-year figure too literally. He ignored the errors inherent to the measurements of the dinosaurs; he signaled the errors by the roughness of the figure and instead assumed that the skeleton was

exactly sixty-five million years old when he began work at the museum. Only then would his hyper-precise figure of 65,000,038 years make sense. But since the errors in measurement absolutely dwarf the time he spent working at the museum, his figure of 65,000,038 years is ridiculous. The skeleton is still sixty-five million years old as it will be a hundred or a thousand years in the future. The guide erred because he trusted the measurement beyond the point where it should be trusted. He committed an act of disestimation.

Disestimation is the act of taking a number too literally, understating or ignoring the uncertainties that surround it. Disestimation imbues a number with more precision than it deserves, dressing a measurement up as absolute fact instead of presenting it as the error-prone estimate that it really is. Its a subtle form of proofiness: it makes a number look more truthful than it actually is and the result can be as silly and meaningless as the museum guides 65,000,038-year-old dinosaur. Every few years, public officials and the news media perform a ritual form of disestimation when a population clock reaches a big milestone. Population experts at the Census Bureau and around the world are constantly estimating the populations of each nation. Their estimates are pretty good, predicting when, say, the worlds population reaches six billion they might even be able to guess when the six billionth person is born to within a few hours. Thats about as good as any possible measurement of population can get. Populations constantly fluctuate, with people dying and being born at irregular intervals, often far from the eyes of people who count such things, so its impossible to know at any given moment the true number of people alive on earth. Nevertheless, on October 13, 1999, as flashbulbs popped around him, UN secretary-general Kofi Annan held a young Bosnian boy, welcoming him into the world as the six billionth person on earth. (The UN insisted that Annans presence in Sarajevo was a complete coincidence. It was just a lucky break that the six billionth person was born in the city where Annan happened to be visiting.) Theres no way that anyone could pinpoint which baby became the six billionth person living on earth. The uncertainties in measurement are simply too huge. You wouldnt know, probably to within several thousand, whether a baby is number 6,000,000,000 or 5,999,998,346 or 6,000,001,954. Only by disestimating, by ignoring the uncertainty in population numbers, could anyone claim to know for certain who was the six billionth living person. Yet at every population milestone, world officials and the news media go through the same bizarre pantomime. In 2006, the Chicago Sun-Times declared a local baby Alyzandra Ruiz to be the 300 millionth resident of the United States. (They cleverly jumped the gun on everybody, making the arbitrary call almost an hour before the official Census Bureau population estimate reached the 300 million mark.) And when the world population reaches seven billion, probably in early 2012, officials will declare some lucky baby to be the seven billionth living person, completely indifferent to the fact that its a lie. Disestimates have much more staying power than Potemkin numbers. While a Potemkin number is purely fanciful and intended to deceive, a disestimate has its origin in a real, meaningful, good-faith measurement the problem is that we dont take the resulting number with a big enough grain of salt. Its a rather subtle problem. As a result, disestimates can be difficult to spot. And they dont wither under scrutiny like Potemkin numbers do. Once a disestimate is believed by the public, it can be devilishly hard to debunk. As an example, ask yourself: what body temperature is normal? If you live in the United States one of the few countries left that still use the antiquated Fahrenheit scale your answer almost certainly is 98.6 degrees Fahrenheit. If you have an (analog) medical thermometer in your medicine cabinet, it probably has a little arrow pointing to 98.6F. When you see the little line of liquid creep beyond that arrow, you probably conclude that youve got a fever. What you might not know, though, is that 98.6F is a disestimate. The idea that normal body temperature is 98.6F comes from research done in the late 1860s by the German physician Carl Wunderlich. Even though this number seems very precise and official, that precision is an illusion. There are quite a few reasons not to take the 98.6F number literally. Wunderlich may have been faking his data. He made the (rather dubious) claim to have measured a million body temperatures with unlikely precision. After taking those temperatures, he came to the conclusion that normal temperature was 37 degrees Celsius a nice round number in the temperature scale used by most of the world. Converting the nice round 37-degree Celsius number into Fahrenheit yields 98.6 degrees F, automatically making the number seem more precise than it actually is. Furthermore, Wunderlich took the body temperatures in his subjects armpits, so even if his measurements were valid and precise, his definition of normal wouldnt apply to measurements of body temperature taken from the mouth or other orifices, as these have slightly different temperatures. Body temperature isnt uniform the answer you get depends on where you make the measurement. This is a huge source of error that most people dont take into account. Neither do they seem to compensate for the fact that body temperatures can change dramatically throughout the day, and that normal is very different from person to person. There is no hard-and-fast definition of normal, much less one thats

precise to within a tenth of a degree as the 98.6F number seems to be. Yet its a fiction that we still cling to even medical dictionaries sometimes define a fever as a body temperature above 98.6F. We all imbue the highly precise number with tremendous importance, even though in truth the definition of normal temperature is imprecise, fuzzy, and somewhat arbitrary. Its a disestimation, yet one thats persisted for a century and a half. Because of disestimations subtlety and longevity, it can be a particularly nasty form of proofiness. Even though disestimates arent complete nonsense in the way that Potemkin numbers are, they mix a whole lot of fiction in with their fact. Failure to recognize the inherent limitations of a measurement can be extremely dangerous, because it can potentially create an authentic-sounding number that is in fact far removed from the realm of truth. There are many roads that lead to proofiness. Potemkin numbers create meaningless statistics. Disestimation distorts numbers, turning them into falsehoods by ignoring their inherent limitations. A third method, fruit-packing, is slightly different. In fruit-packing, its not the individual numbers that are false; its the presentation of the data that creates the proofiness. Supermarkets select their fruit and arrange it just so and package it so that even mediocre produce looks delectable.

Similarly, numerical fruit packers select data and arrange them and dress them up so that they look unassailable, even when theyre questionable. The most skilled fruit packers can make numbers, even solid ones, lie by placing them in the wrong context. Its a surprisingly effective technique. A particularly powerful weapon in the fruit packers arsenal is whats known as cherry-picking. Cherry-picking is the careful selection of data, choosing those that support the argument you wish to make while underplaying or ignoring data that undermine it. Since real-world numbers are fuzzy, answers to numerical questions arent always clear-cut.

Measuring the same thing in different ways can give different answers; some of the numbers will be too high, some will be too low, and, with luck, others will be reasonably close to the right answer. The best way to figure out where the truth lies is to look at all of the data together, figuring out the advantages and disadvantages of each kind of measurement so that you get as close to the truth as possible. A cherry picker, on the other hand, selects the data that support his argument and presents only them, willfully excluding numbers that are less supportive, even if those numbers may be closer to the truth. Cherry-picking is lying by exclusion to make an argument seem more compelling. And its extremely common, especially in the political world. (Its also very common in the scientific world, thanks to a phenomenon known as publication bias. Peer-reviewed journals cherry-pick the most exciting papers, selecting them for publication. This means that papers with spectacular results are published in high-profile journals while less sexy ones (including negative results) are relegated to lesser journals or arent published at all. Publication bias distorts science, making new drugs, for example, seem more effective than they actually are.) Every politician is guilty of it, at least to some extent. Al Gore is guilty of cherry-picking in his film *An Inconvenient Truth*. At the heart of the 2006 movie is a breathtaking and disturbing sequence where he shows computer simulations of what global warming will do to the surface of the earth. In a series of maps, he shows the worlds coastlines disappearing under the rising oceans. Much of Florida and Louisiana will be submerged, and most of Bangladesh will sink beneath the waves. The animations are stunning, leaving viewers with little doubt that global warming will dramatically reshape our planet. However, those animations are based upon a cherry-picked number: Gores pictures assume that melting ice will drive the sea level up by twenty feet. Lots of scientists have tried to model the effects of global warming, and most have come to a very different conclusion. They tend to agree that global warming is real, that human activities are responsible for a sizable portion of that warming, and that the sea level will indeed rise over the next century. (Despite my singling out Al Gore for cherry-picking, there are unambiguous data that show that global warming is occurring. Its just that sea levels arent going to rise twenty feet anytime soon.) Theres an outside chance that the sea level will rise by twenty feet or more if a very worst-case scenario occurs (such as the near-complete melting of the ice sheets in Greenland or West Antarctica). However, most serious estimates project a sea level rise much lower than what Gore used. Some climatologists say the oceans will rise two feet or so in the next century; some go as high as four feet these are the best scientific guesses right now. Yet Gore ignores these more modest estimates and picks the most extreme model of sea level risethe twenty-footerso he can flash his dire graphics on the screen. It wowed the audiences, but it was cherry-picking. George W. Bush is just as guilty of cherry-picking as his erstwhile opponent. Like every president, he put the best possible spin on all of his pet projects. No Child Left Behind, for example, was the name for a shift in educational strategy early in his administration; the act, signed by Bush in 2002, offered money to states in return for mandatory testing and other concessions. It was a controversial move. Several years later, in his State of the Union address, Bush declared that No Child Left Behind was working wonders in Americas schools. Five years

ago, we rose above partisan differences to pass the No Child Left Behind Act . . . and because we acted, students are performing better in reading and math. This statement was a rare instance of a double cherry-pick. First, when Bush flatly declared that students are doing better in math and reading, he had to do a bit of cherry-picking. His data came from the Department of Education, which periodically sponsors a national set of assessment tests to determine how well the nations students are doing in various subjects. The data show that fourth- and eighth-grade students reading and math scores have in fact improved since No Child Left Behind started. But there are other data that he ignored. Twelfth-grade students reading scores declined over the same period. And though its a little more complicated (the test changed form, making the trend harder to figure out), twelfth-grade math scores also seem to have declined slightly. So saying that students are performing better in reading and math is only true if you ignore the twelfth-grade results that say otherwise.

Accentuate the positive; eliminate the negative. Cherry-pick number one. Second, pretending that the improvement in math and reading scores is due to the No Child Left Behind Act requires some cherry-picking too. If you look at the scores carefully, you see that fourth- and eighth-grade math scores have been improving at roughly the same rate since the 1990s, long before the act was passed. Similarly, fourth- and eighth-grade reading scores have been improving at roughly the same (very modest) rate in the same time period. By ignoring data from before 2002, Bush was able to pretend that No Child Left Behind was responsible for the improving scores, even when its clear that the trend is essentially unchanged over the years. No Child Left Behind only seems responsible for the improved scores if you fail to present earlier data that put the scores in the proper context. Cherry-pick number two. (Theres a third kind of cherry-pick here, in fact. Even if you accept that math and reading scores are important, schools teach a lot more subjects: writing, science, history, and more. Data from these disciplines show either mild improvement or, in some cases, decline, particularly in upper grades. Concentrating on reading and, particularly, math as indicators of a schools improvement is only looking at part of the picture.) Voil. Bush can declare No Child Left Behind a successeven if it isnt. Education statistics are a hotbed of fruit-packing. Its really hard to improve the school system; it requires lots of money and effort and time to make a change in a huge bureaucracy with such enormous inertia. Even more disturbing for a politician, it takes years before an administration can reap the benefits of making good educational policy; you might be out of office long before citizens realize that youve improved the school system. Fruit-packing provides a shortcutit allows a politician to reap the benefits of improving the schools without ever having to do the hard work of changing policy. In New York, scores on the states reading and math tests have risen sharply since 2005. Any politician who has anything to do with education policy in the state basks in the glow of the rising scores each year. In 2008, for example, New York City mayor Michael Bloomberg declared that the dramatic upward trend in state test scores showed that the citys public schools were in a different league than when he took office. However, soon after the state tests were administered in 2005, teachers told reporters that the test was much easier than the year before. What a difference from the 2004 test, a principal of a Bronx school told the New York Times. I was so happy for the kidsthey felt good after they took the 2005 test. Scores on the state tests climbed year after year, rising dramatically and improbably. On national tests, though, New York didnt seem like such a success story. In New York City, for example, scores on the national tests stayed more or less unchanged. Its pretty clear that New York State had been tinkering with the difficulty of the test. By making the test easier year after year, it artificially makes students test scores rise. It seems as if the children are performing better on the tests, but in fact the rise in scores is meaningless; the 2004 test score doesnt mean the same thing as the 2006 test score, which doesnt mean the same thing as the 2008 test score. By pretending that these tests are equivalent, New York is engaging in another form of fruit-packing: comparing apples to oranges. This particular trick is a game of units. As mentioned earlier in this chapter, every real-world number has a unit attached to ita little tag like feet or seconds or kilograms or dollars that tells you what kind of measurement the number is tied to. When you compare two numbers to see which ones bigger than the other, its important to ensure that the units are the same, otherwise the comparison is meaningless. Sometimes this is obvious (which is longer, 55 seconds or 6.3 feet?), but sometimes its a little tricky to tell that the units arent quite the same. Which is better: a 50 percent score on test A or a 70 percent score on test B? Its a meaningless comparison unless you have some way of converting a test A score into a test B score and vice versa. Theres no value to making a direct comparison of test scores from year to year unless the test makers ensure that the value of a given score always stays the same, yet this is precisely what New York State did, exploiting this apples-and-oranges problem to make it look as though their educational system were improving. The effect is very similar to cherry-picking; when New York compared apples to oranges, they distorted the meaning of

numbers, making the statistics appear to support an argument that they don't. Apple-orange comparisons can be really tough to spot, because units can be fluid creatures. Some of them change their meaning over time.

The dollar, for example, is the unit that Americans use to measure money. But as a unit of wealth, it is always changing. Flip through an old magazine and look at the ads. The changing value of the dollar will hit you on the head. A December 1970 copy of *Esquire* that I happen to have in my office shows that a low-end two-door car cost \$1,899. Name-brand carry-on luggage would set you back \$17. A pair of mass-produced mens shoes is worth \$19. Right now (2010), the equivalent low-end two-door car costs about \$12,000. The

name-brand carry-on luggage will set you back \$130. The pair of mass-produced mens shoes is worth roughly \$100. Even though all of these numbers seem to have the same unit symbol next to them, a dollar in 1970 is a very different unit than a dollar in 2010. In 1970, a dollar bought more than 5 percent of a pair of shoes. Now it buys less than 1 percent. Back then, \$ had a lot more value than it does today. If you look at the prices in that 1970s magazine carefully, you should quickly come to the conclusion that a 1970 dollar had somewhere between five and seven times more purchasing power than the 2010 dollar does. (Purchasing

a car costs 6.3 times as many 2010 dollars as it did 1970 dollars, for example.) A 1970 dollar is a very different unit from a 2010 dollar, just as surely as a gallon is different from a quart. As you probably know, this change is caused by inflation. Goods and services get a touch typically around 3 percent more expensive every year. Year after year, it takes more dollars to buy a gallon of paint, and it also takes more dollars to hire someone to paint your house, more dollars to take the bus to get to work, and more dollars to do just about everything you need money to do. Thus the almighty dollar gets less and less valuable as the years tick by. This makes it a little bit difficult to compare spending over time; you've got to adjust for inflation, taking into account the changing value of the dollar. However, sometimes people conveniently forget to make that

conversion to make their arguments seem stronger. In November 2005, a group of House Democrats (the

Blue Dog Coalition) attacked George W. Bush with this little tidbit: Throughout the first 224 years (1776-2000) of our nation's history, 42 U.S. presidents borrowed a combined \$1.01 trillion from foreign governments and financial institutions according to the U.S. Treasury Department. In the past four years alone (2001-2005), the Bush Administration has borrowed a staggering \$1.05 trillion. These figures are all true, and, yes, Bush was quite the deficit spender. However, this comparison is utterly meaningless because its comparing apples to oranges. 2005 and 2001 dollars are very different from 1776 and 1803 and 1926 dollars. The Louisiana Purchase cost \$15 million; Alaska was a bargain at roughly half that. Back then, those were enormous expenditures that only wealthy states could afford. Nowadays the same sum is the cost of a fancy Manhattan penthouse. Past presidents who drove up the nation's debt did it in much smaller dollar figures simply because those dollar figures represented a lot more purchasing power than they do today.

Comparing apples and oranges can be quite powerful; used skillfully, it can make the false seem true and the true seem false. For example, in 2005, the director of the National Science Foundation gently bragged about the agency's budget request for the following year. The NSF, he said, would get \$132 million, or 2.4 percent, more than in [fiscal year] 2005. This modest increase allows us to assume new responsibilities, meet our ongoing commitments, and employ more staff. . . . At first glance, a \$132 million increase seems like something to celebrate. But the number was a fruit packer's fantasy. It came from comparing 2006 dollars to

2005 dollars without taking inflation into account. The 2006 dollar was worth less than the 2005 dollar, so the increased budget in 2006 would in fact be worth less than the budget in 2005. If you crunch the numbers properly, that \$132 million increase was a disaster; in fact, the agency would be losing about \$30 million (in 2006 dollars). There wouldn't be any extra money for new responsibilities or new staff. The NSF director was lying, turning a defeat into a victory. By comparing apples with oranges, a skilled official can make a decrease look like an increase, up look like down, and black look like white. Its Orwellian comparing apples to oranges can make fiction of fact and fact of fiction. (Budgets are always subject to proofiness. The people

making them have a vested interest in making expenditures seem tiny; those criticizing them are trying to make those same numbers seem large. Though the details are too wonky for this book, it's worth mentioning that in the United States, the government keeps certain expenditures (like Social Security spending) off the official budget, counting them separately. Because of this, politicians can cherry-pick, ignoring off-budget expenditures when it suits them.) Yet another variety of fruit-packing, apple-polishing, is used to put the finishing touches on data, manipulating them so they appear more favorable than they actually are. Just as

greengrocers employ subtle artifices to make their produce look fresher and tastier than it actually is waxing and polishing apples to make them look fresher, gassing tomatoes to make them turn red, piling cantaloupes so that their blemishes are hidden mathematical fruit packers tweak their data, subtly polishing the numbers

to make them look as appealing as possible. There are endless ways to polish mathematical apples; it would be impossible to describe them all, especially since inventive fruit packers are inventing new ones all the time. But there are a few common tricks worth mentioning. Graphs visual depictions of data are particularly vulnerable to apple-polishing. A fruit packer can choose to display data in an endless variety of ways, fiddling with the look so that the graph makes the data look more impressive than they actually are. Take the case of Quaker Oats. It's a bland and relatively unappetizing product not easy to come up with an ad campaign for. Yet people will eat anything that they think will improve their health, so ad executives launched a blitz to make the barely digestible oat fiber appear to be a medicinal vacuum cleaner, sucking the cholesterol right out of your bloodstream. They emphasized the point with a graph. (See page 35 for graph) The message was clear: eat Quaker Oats, and within a few weeks your cholesterol levels will drop dramatically. However, if you look carefully, you will discover that this graph is deceptive. We normally assume that the line at the bottom of the chart represents zero cholesterol the little oat-fiber machines have gobbled up every single dollop of cholesterol in your blood. But if you examine the vertical axis of the chart, you see that the bottom isn't zero, but 196. This makes the data seem much more dramatic than they actually are, as you can easily see if you look at a more honest graph. (See page 36 for graph) By tweaking the scale of the chart just so, Quaker Oats made it look as if oatmeal was having a huge effect when it wasn't. (After receiving complaints, Quaker withdrew the chart.) Of the many ways to manipulate data in graphs, this is probably the most common. Another form of apple-polishing exploits the term average to make numbers seem smaller or larger than they really are. Most people think that average means typical that if, say, the average salary at a company is \$100,000, then each employee earns \$100,000, more or less. In fact, that's often not the case. The average of a set of numbers more precisely, the mean has a precise mathematical meaning: you add everything together and then divide by the number of data points that you added together. For example, if you had a company of ten people, each of whom earned roughly \$100,000, you add those ten salaries together ( $\$100,000 + \$101,000 + \$98,500 + \$99,700 + \$103,200 + \$100,300 + \$99,000 + \$96,800 + \$100,000 + \$101,500 = \$1,000,000$ ) and then divide by the number of salaries ( $\$1,000,000 / 10 = \$100,000$ ). In this case, the average, \$100,000, does in fact represent a typical salary. However, consider a company where the CEO earns \$999,991 per year, and there are nine interns who each earn \$1. The mean, again, is the sum of those salaries ( $\$999,991 + \$1 + \$1 + \$1 + \$1 + \$1 + \$1 + \$1 + \$1 + \$1 = \$1,000,000$ ) divided by the number of salaries ( $\$1,000,000 / 10 = \$100,000$ ). So here too the average salary is \$100,000. However, \$100,000 is not a typical salary in any meaningful way. If you were to pick a person at random from the company, you'd probably find that he earns a measly \$1. So in this case, it's deceptive to pretend that average is typical. (In cases like this, it's often better to use a construct known as the median to figure out what a typical salary should be. To calculate the median, you line the numbers up from lowest to highest and pick the one in the middle. Here, the median salary would be \$1, clearly a better representation of typical than the mean would be.) If the CEO were to recruit new employees by highlighting the company's average salary of \$100,000, he would be apple-polishing. The new hire would be shocked when he gets his first paycheck. Whenever a politician announces a tax cut, it's almost guaranteed that he'll pull the exact same stunt to make the tax breaks look larger than they actually are. He'll give a speech that talks about the average refund the mean tax break and make his constituents extremely happy. However, the average is usually far from typical. Most people will be disappointed when they receive their refund checks in the mail. For example, George W. Bush made tax cuts a central thrust of his administration and always polished apples when describing them. A typical incident occurred at the end of his first term when he said, The tax relief we passed, 11 million taxpayers this year will save \$1,086 off their taxes. (The White House quickly corrected the figures; Bush meant to say 111 million taxpayers and an average of \$1,586 in savings.) As it happens, though, both figures were deceptive. The typical family didn't see anywhere near \$1,586 or \$1,086 in tax breaks; most received less than \$650. The reason was the same as the greedy-CEO example: a relatively small number of people received very large refunds, making the average very atypical. The rosy numbers coming out of the White House were technically true, but they were functionally lies they were apple-polished to make them look much larger than they should. Apple-polishing, cherry-picking, comparing apples to oranges all the tricks of the fruit packer present numbers in a misleading manner, distorting them to the point of falsehood. Potemkin numbers dress up nonsense in the guise of meaningful data. Disestimates stretch numbers beyond their breaking points, turning even valid measurements into lies. All of these techniques are forms of proofiness; all of them allow an unscrupulous person to make falsehoods look like numerical fact. And because we humans tend to think of numbers as representing absolute truth, we are hard-wired to accept

them without question. Proofiness has such a hold over us because our minds are primed to accept mathematical falsehoods. Because of the way our brains work, certain kinds of numbers make them malfunction. As a result, we humans believe some absurd and embarrassing lies. Revue de presse "A delightful and remarkably revealing book that should be required reading for . . . well, for everyone." -

Booklist (Starred review)