An Unfair Vote

It's very hard in military or in personal life to assure complete equality. Life is unfair.

—John Fitzgerald Kennedy

I just received the following wire from my generous daddy: "Dear Jack, Don't buy a single vote more than is necessary. I'll be damned if I'm going to pay for a landslide."

—John Fitzgerald Kennedy

Elections are inherently unfair. No matter what method a government uses to run an election, it can't be an equal contest, at least in a mathematical sense. It's an inescapable truth: all elections are flawed, and there's nothing we can do to fix them.

That's the bad news. The *really* bad news: a number of politicians and judges are making our flawed system of elections much, much more unfair than it already is.

Democracy is inherently an institution based upon a mathemat-

ical operation—that of counting votes. As we've seen, that operation is vulnerable to proofiness. Armed with bogus mathematical arguments and underhanded tactics, politicians and their judicial allies are working to stack the electoral deck to get their party into power and keep it there. They are succeeding.

Democracy is in danger, buckling under an assault from proofiness.

At the same time that Norm Coleman's final appeal was limping around the hallways of the Minnesota Supreme Court, Minnesota justices were deciding yet another lawsuit about Minnesota elections—one that had the potential to change the way elections in Minnesota are run. In theory, it could prevent fiascos like the Franken-Coleman race from ever happening again.

As an experiment, the city of Minneapolis dispensed with the standard vote-for-one-candidate "plurality" method for deciding certain elections. Instead, the city would use what's known as instant runoff voting, where each voter would rank the candidates in order of preference. Advocates argue that instant runoff voting makes elections more fair and more transparent. They have a point: had instant runoff voting been used in the 2008 Senate election, officials would almost certainly have been able to declare a victor within a matter of days.

After the recount, after all the legal challenges, the final results of the 2008 Minnesota Senate election were:

1,212,317 for Norm Coleman
1,212,629 for Al Franken
437,505 for Dean Barkley

Since this was a plurality election, with each voter casting a single vote in the Senate race, the winner was simply the person with the most votes—Al Franken, in this case. But because the election was so close (and the lawyers were so skilled), it took eight months and millions of dollars to determine the final winner.

An instant runoff vote, on the other hand, would probably have made the recount unnecessary. In such a vote, the ballots are slightly different from what they are in a plurality election. Instead of voting for a single candidate, each voter gets to rank the candidates, from least to greatest of all possible evils. Using those ranks, officials can figure out who wins the election. It's not quite as simple as merely counting votes; an instant runoff version of the 2008 Minnesota race would have had a much more complicated-looking result than a plurality election. It might have looked something like:

1,202,310 prefer Norm Coleman over Dean Barkley over Al Franken
10,007 prefer Norm Coleman over Al Franken over Dean Barkley
1,201,620 prefer Al Franken over Dean Barkley over Norm Coleman
11,009 prefer Al Franken over Norm Coleman over Dean Barkley
287,010 prefer Dean Barkley over Norm Coleman over Al Franken
150,495 prefer Dean Barkley over Al Franken over Norm Coleman

So . . . who wins? This requires a little number juggling. First, you look at everybody's first choice. In this case:

1,212,317 chose Norm Coleman as their first choice
1,212,629 chose Al Franken as their first choice
437,505 chose Dean Barkley as their first choice

. . . exactly the same as in the plurality election. But the election isn't over yet. In an instant runoff vote, you can only win if you get more than 50 percent of the votes. Neither Franken (at 42 percent) nor Coleman (also at 42 percent) managed to cross that threshold and win a majority of votes. When this happens, the "instant runoff" begins: the candidate at the bottom of the pack, Dean Barkley, is eliminated; it becomes a two-person race between Coleman and Franken. Barkley voters aren't disenfranchised, though. Officials count their second-choice votes in lieu of the now moot first-choice votes for Barkley. Using the above—hypothetical—numbers, the result of such an instant runoff would be:

1,499,327 votes for Norm Coleman (including 287,010 former Barkley voters)
1,363,124 votes for Al Franken (including 150,495 former Barkley voters)

Norm Coleman now has a solid majority so the election ends. As an added bonus, the race is no longer terribly close—the margin is nearly 5 percent of the votes cast—so there's no need for a recount. Instant runoff voting would likely have saved the state of Minnesota a whole lot of trouble.

However, when the city of Minneapolis proposed using instant runoff voting for some elections, opponents promptly sued, claiming that the scheme is unfair—more specifically, it's unconstitutional because it doesn't count every vote in precisely the same

way, an argument that the Minnesota Supreme Court promptly knocked down. However, it's absolutely true that instant runoff voting has disadvantages over plurality voting. Ballots in plurality elections are very simple—you simply vote for one person—yet as the Minnesota election showed, people screw them up all the time. Imagine all the ways that people would fill out a "rank three candidates in order of preference" ballot incorrectly. The error rate of the election would go through the roof.

Another drawback of instant runoff voting is that it doesn't always give the answer that the voting public thinks is best. Indeed, in the above example, Dean Barkley could argue that in fact he should be elected, because that choice would make the most people the least unhappy. To see this, take the instant runoff results above and modify them slightly. Every time a voter ranks a candidate in first place, give that candidate two points to represent the voter's strong desire to get that candidate elected. Every time a voter ranks a candidate second, the candidate gets one point, signifying that the candidate isn't the voter's ideal choice, but is not too horrible. Finally, every time a voter ranks a candidate in third place on a ballot, the candidate gets zero points for being the greatest of all evils. (This scheme is known as a *Borda count*, and, like instant runoff voting, it is often floated as an alternative to plurality votes.) Total up the points. All of a sudden, the election looks very different:

2,722,653 points for Norm Coleman
2,585,760 points for Al Franken
3,278,940 points for Dean Barkley

Dean Barkley wins handily—by virtue of being the least loathed candidate overall. Franken voters prefer Barkley to Coleman, and

Coleman voters prefer Barkley to Franken. Barkley is the compromise candidate that everybody can live with.

In the Minnesota Senate election, three different voting schemes—plurality, instant runoff, and Borda—would have yielded three different victors, even though the voters' preferences were identical in all three hypothetical scenarios. You can take exactly the same ballots and look at them in three different ways and come up with a perfectly valid argument about why each candidate should be elected to the Senate.

This illustrates a central problem with voting. Reasonable people can look at the same pile of ballots and come to very different conclusions about who should win an election. While plurality voting, Borda voting, and instant runoff voting each have their advantages and disadvantages,* none can claim to be the fairest way of electing a politician; they're all flawed. It's a mathematical truism known as *Arrow's theorem*.

In the 1950s, economist Kenneth Arrow proved that it's impossible to have a perfectly fair election system. But what does "fair" mean in this context? Well, there are certain characteristics that you would expect a fair election to have. One seems ridiculously obvious: there can't be a dictator who determines the outcome of an election; there can't be an individual whose vote overrules everybody else's. A fair election implicitly follows this "no dictators" rule. Another obvious condition: a vote shouldn't flout the unanimous will of the people. If everybody in the nation votes for Ross Perot, then, by God, Ross Perot had better win the election. A fair election also implicitly follows this

* As do innumerable other voting methods that election reformers like to talk about, such as Condorcet voting, approval voting . . . the list goes on and on. In my view, plurality voting is the best system because its ballots are simplest and therefore minimize voter error.

"unanimous vote" rule. Finally, there's a third characteristic that's a little less obvious: the ranks of candidates in a perfectly fair election should be a faithful reflection of the voters' preference. If, in a head-to-head race, Ronald Reagan would always beat Jimmy Carter, it's clear that the population genuinely prefers Reagan to Carter. Thus, in a truly fair election, Ronald Reagan should always come out ahead of Jimmy Carter when the results are tallied, no matter what other candidates are running. Under no circumstances should Jimmy Carter ever be able to beat Ronald Reagan; if he did, that would be an imperfect reflection of the voters' true preference. Thus a perfectly fair election should implicitly follow this "faithful reflection" rule as well.

Arrow's theorem proves that these three conditions of a perfectly fair election—"no dictators," "unanimous vote," and "faithful reflection"—are mutually contradictory. It's mathematically impossible to have all three at the same time. This means that there is no such thing as a perfectly fair election.

In practice, elections fail the "faithful reflection" condition. They don't give a simple tally of voters' preference of one candidate over another; they're much messier than that. Even though people preferred Ronald Reagan to Jimmy Carter, Carter theoretically could have won the election if an attractive third-party candidate also ran for president in 1980, splitting the Republican vote. If this happened, the "faithful reflection" condition would have been violated, because Carter would have won the election despite the populace's preference of Reagan to Carter. In other words, elections are so complex that no matter what voting system we use, the "wrong" person might wind up being elected. This is just a fact of life, something that we've come to live with in a democracy. No matter what method we use to elect our officials, there is some level of unfairness inherent in the process. It's inescapable.

Even though Arrow's theorem ensures that no election can be perfectly fair, the level of unfairness is pretty mild. We're not too disturbed by the idea of third-party candidates mucking up an election. Indeed, it's part of what makes elections so unpredictable—and so interesting. However, Arrow's theorem accounts for just a tiny fraction of the unfairness that plagues elections in the United States and around the world. There are much, much more worrisome problems with our electoral process—problems that are quite purposeful. Politicians are trying very hard to turn a mildly unfair voting system into something that's mind-bogglingly unfair. Make no mistake: there are people who are attempting to undermine the very mechanisms of democracy in order to ensure that their ideological allies get elected—regardless of the will of the people. And they've got a powerful weapon in the struggle: proofiness.

People have been undermining democracy by tampering with our electoral system for years. In the United States, the tradition is almost as old as the nation itself. By the early 1800s, scheming politicians had already created a powerful—and legal—method for staying in office even when the public tried to vote them out. This method is named after a gentleman whose signature is on the right edge of the Declaration of Independence, not far below the signatures of his fellow Bostonians Sam and John Adams. His name was Elbridge Gerry.

Gerry became governor of Massachusetts in 1810, but his party, the Democratic-Republican, was becoming increasingly unpopular. He and his allies were losing ground to their rivals, the Federalists, and they were terrified of what the upcoming election

of 1812 would bring. If the voters had their way, the Democratic-Republicans would be kicked out of the statehouse, leaving the Federalists in power. From Gerry's point of view, this had to be avoided at all costs.

One of the powers of state government is to redistrict: to change the boundaries of voting precincts, altering which regions would be represented by which state senators. The Democratic-Republicans realized that if they got really creative with the way they redrew those boundaries, they could hold on to most of the seats in the state senate, even if the majority of voters in the state went Federalist.

When the Democratic-Republicans enacted the plan, it was extremely controversial. The bizarrely shaped new districts looked hideous and unnatural. A salamander-like district curled its body around Essex County, its belly to the west and its head snaking across the north. Wags promptly named it after its creator. The gerrymander was born.

As ugly as it was, the bizarre creature worked its magic for the Democratic-Republicans. Even though the Federalists got the majority of votes cast for state senator—50.8 percent of them, to be precise—the Democratic-Republicans won an overwhelming majority of the state senate seats: twenty-nine out of forty. By gerrymandering, they had turned what should have been an electoral defeat into a landslide victory. They held on to power not because of the will of the people, but in spite of it. They used gerrymandering to undermine the electoral process, annulling the votes of their opponents.

A gerrymander is a creature born from proofiness. At its core, gerrymandering is cherry-picking, with one key difference: the data

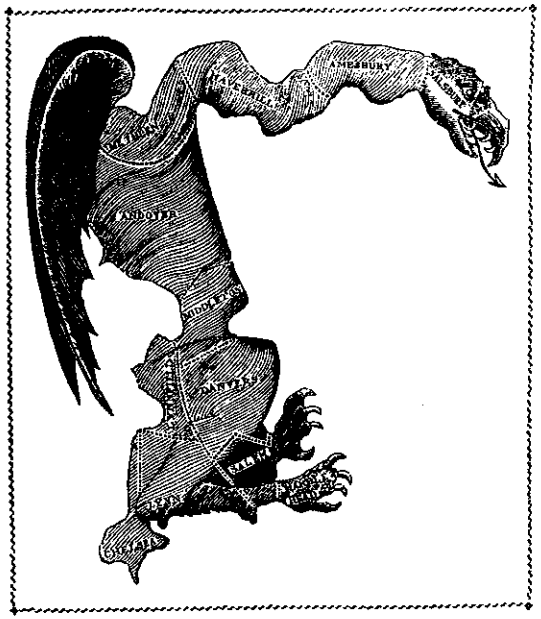

Figure 11. The original gerrymander.

being manipulated are votes. Indeed, the gerrymander is essentially a monster that allows politicians to carefully select votes, choosing those that they like and ignoring those that they don't.

Gerrymandering gets its power from two kinds of vote manipulation—two tricks that politicians have become extremely adept at over the years. These tricks are known as *packing* and *cracking*. Packing takes opposition votes and packs them tightly together, rendering most of them redundant. Cracking splits apart opposition strongholds, distributing their votes among multiple districts so that the enemy is not able to wield a majority in any one district.

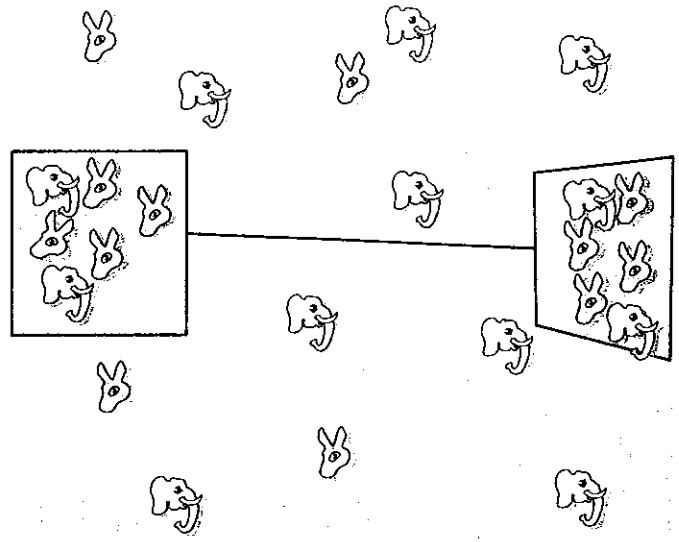

Figure 12. A fair division of districts.

As an example, imagine there's a county that is divided into four districts, two urban and two rural, each of which gets one representative to Congress. This county happens to be evenly divided between Democrats and Republicans; the two urban districts have a Democratic majority and the two rural districts have a Republican majority.

Each district sends a representative to the statehouse; the cities elect Democrats to represent them, while the two rural districts, naturally, elect Republicans. This is the way it should be; an evenly split electorate should split their representatives equally too. But imagine that the Republicans gain control of the statehouse, allowing them to redraw the boundaries of the districts. Their gerrymandering strategy will be to pack as many Democrats as they can into a single city district. They'll crack the other city apart, distributing

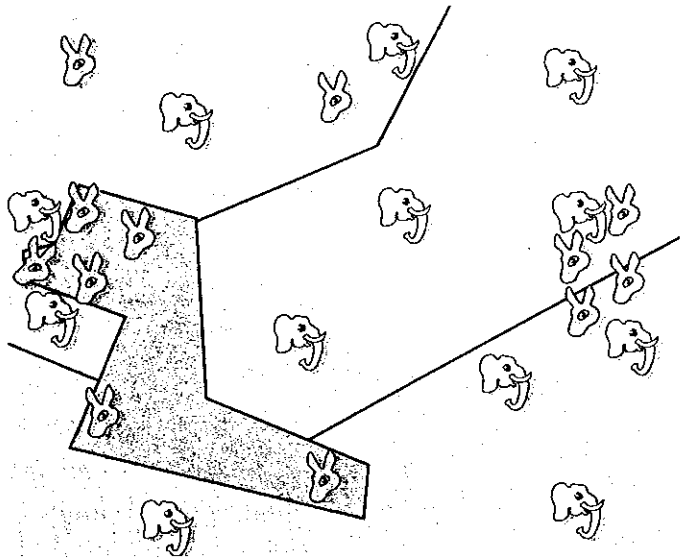

Figure 13. A pro-Republican gerrymander.

the city Democrats among Republican-heavy rural districts so that they're unable to muster a majority anywhere. The result is a bunch of districts that make a little less geographical sense—urban voters are mixed with rural voters—but gain the Republicans an extra seat in Congress.

Even though the electorate is split down the middle—there are exactly as many Democrats in the county as there are Republicans—the gerrymandering has allowed the Republicans to control 75 percent of the congressional seats. Conversely, if the Democrats had managed to gain control, they would do the exact same thing in reverse. They would pack Republicans into one rural district and crack the other, distributing Republicans among the three remaining districts so that they're firmly in the minority.

The Democratic seat-grab is just as effective as the Republican

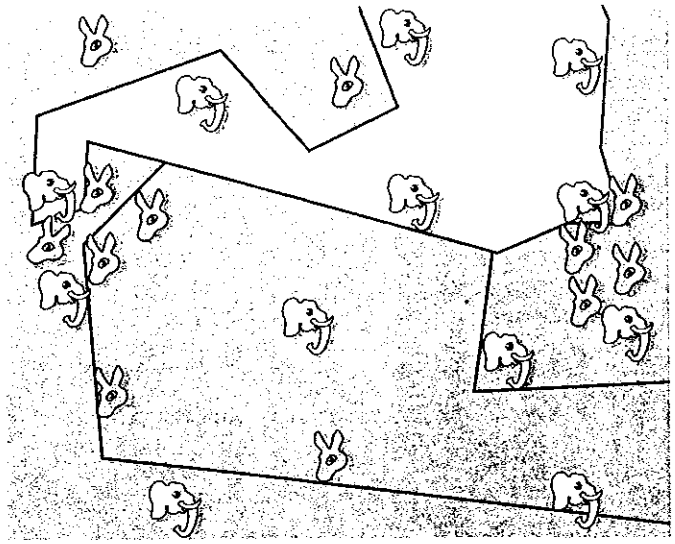

Figure 14. A pro-Democratic gerrymander.

version; they have 75 percent of the congressional representatives despite having only 50 percent of the vote.

In skillful hands, the gerrymander can give a party more power than the people want them to have; it can entrench an unpopular politician or dislodge a popular one; it can render some votes moot while investing others with great weight. Even worse, the practice of gerrymandering makes it more difficult for voters to punish the politicians who are robbing them of their votes—their redistricting plans make it extremely unlikely that the incumbents responsible for the gerrymander will lose the next election.

Gerrymandering is a direct affront to the democratic process; by allowing politicians to cherry-pick votes, it systematically undermines the validity of elections. But it's not easy to get rid of, because the issue is fiendishly complex. For many years, the federal

court system more or less decided not to touch the issue at all, declaring it something that should be fought over by politicians rather than by judges. But the courts couldn't ignore it indefinitely, because gerrymandering was gnawing at fundamental constitutional rights.

The U.S. Constitution dictates that each state gets a number of seats in the House of Representatives proportional to the population of each state. As the population grows and shifts, the representation has to change, so states need to change the boundaries of their voting districts in response to new census numbers every decade. (These census numbers are also subject to proofiness—more on this shortly.) However, the power to change those boundaries isn't absolute. It's constrained by the Fourteenth Amendment to the Constitution. Ratified shortly after the end of the Civil War, the amendment guarantees that all citizens are given "equal protection" by the laws of the land. That is, each citizen is seen as equal in the eyes of the law—which in theory means that each person's vote should have the same value. By the 1960s, the courts were forced to recognize that gerrymandering, in some circumstances, was making certain citizens' votes count less than others'—some people's votes were being diluted. Judges couldn't continue ignoring the issue, thanks largely to the civil rights movement.

Shortly after the end of the Civil War, former slaves were suddenly full citizens—and like other citizens, they were entitled to vote and given the same protection under the law as any other citizen. That's the way it was on paper. In reality, though, former slave-owning states used all sorts of tricks, such as poll taxes, that were designed to keep African Americans away from the voting booths. (More on this later in the chapter.) They also used gerrymandering, cracking populations of African Americans so that they didn't get

any representation in Congress. Despite constitutional guarantees, African Americans were being treated as second-class citizens.

In the mid-1960s, however, the civil rights movement changed the political landscape. Legislators and the courts finally attacked the problem head-on, outlawing the dirty tricks of racist politics. The Twenty-fourth Amendment to the Constitution outlawed poll taxes. A set of court decisions set the "one person, one vote" standard that dictates that each citizen's vote must be given roughly equal weight. The Voting Rights Act of 1965 outlawed procedures—including gerrymandering—intended to deny citizens their voting rights based on race or color. This made gerrymandering illegal, but only when it's done for racist motives.* Political gerrymandering—redistricting to gain political advantage—was still perfectly fine.

It would be simple except for the fact that race and politics can't be disentangled. In modern times, African-American voters are overwhelmingly Democratic. Latino voters also tend to support Democrats. So issues of race are always deeply political. Protecting the voting power of minorities is more or less tantamount to helping out the Democrats.† Making a distinction between racial gerrymandering and political gerrymandering is somewhat artificial. Even ignoring that fact, political gerrymandering is diluting citizens' votes just as surely as racial gerrymandering is.

Redistricting law, to put it politely, is a mess. The Supreme Court

---

* Confusingly, the Voting Rights Act had been interpreted as *encouraging* racial gerrymandering to give underrepresented minorities more power. However, Supreme Court cases in 1993 and 1995 functionally put an end to the practice.

† In the 1990s, Texas Republicans packed Democratic districts in a way that made a small number of African-American Democratic congressmen displace a larger number of white Democratic ones, so the gain for the African Americans was a loss for the Democratic Party. It was a fiendishly clever plan that its inventors dubbed "Project Ratfuck."

has been writhing in self-contradictory paroxysms trying to avoid addressing the issue. After years of pretending that political gerrymandering wasn't a topic suitable for lawsuits, in a 1986 decision the Court finally decided that gerrymandering for purely political purposes might theoretically fall afoul of Fourteenth Amendment protections. However, the Court gave no hints about what makes a partisan gerrymander cross the line into unconstitutionality. As a result, the ruling decided absolutely nothing; if anything, it made the matter more confusing. The first time the decision was tested in the Supreme Court—eighteen years later—the waters got even murkier. In a five-to-four party-line split, the Supreme Court declared a political gerrymander perfectly constitutional. Four of the majority justices went further. They declared that the Court was powerless to declare even the most obviously political gerrymanders unconstitutional, because there was no standard for determining whether a particular gerrymander was politically motivated or not. Anything goes.

It was the Wild West for political gerrymandering, and even the few gentlemanly rules that seemed to hold people's ambitions in check were dissolving. In 2001, in response to the new census numbers, a split Texas legislature finally compromised on a redistricting plan. It was a long and bitter fight, but it was over until the next census. Or so the Democrats thought. When the Republicans won both houses of the legislature in 2002, they re-redistricted, flouting the once-a-decade tradition. The Democrats attempted to stop the re-redistricting, fleeing the state so that a vote couldn't be called—while the Republicans called in the Department of Homeland Security to track down the wayward legislators—but the plan eventually went through, giving Republicans control of twenty-two of thirty-two legislative districts in the state. The case—or more precisely, four cases—went to the Supreme Court, which decided that except for one predominantly Latino district that had been illegally cracked apart in violation of the Voting Rights Act, the gerrymandering was hunky-dory. Even though the redistricting was not tied to the census and was only carried out for bald political gain, it was just fine.

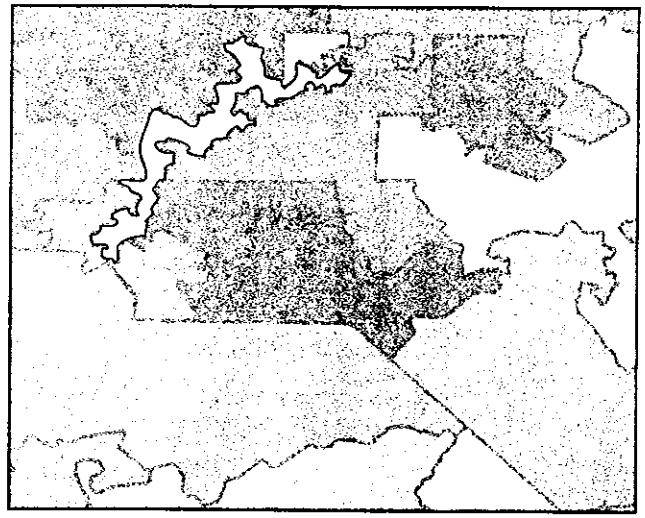

Figure 15. District 12 in North Carolina, one of many gerrymander monsters.

At the moment, gerrymandering, so long as it's not racially motivated, is perfectly respectable in the United States. Though it allows politicians to cherry-pick votes, functionally allowing them to dilute the unfavorable ones, the courts don't seem inclined to correct the problem.* As a result, many of our voting districts remain

* Gerrymandering has a solution. Mathematicians have ways of spotting politically mandated gerrymandering, so it's possible to set standards. And some states have external, nonpartisan committees that are in charge of redrawing boundaries in response to changing populations.

so twisted and distorted that they put Elbridge Gerry's original monstrosity to shame.

Until there's a major change in the way the Supreme Court views the practice, gerrymandering proofiness is here to stay.

Gerrymandering is just one mathematical threat to democracy. There's another form of proofiness that's even more dangerous because it's less overt. This method has enabled politicians to deprive opponents of their votes. Even more than that, it renders them nonexistent—turns them into nonpeople who don't have the right to be represented in Congress. And the worst part is that the authors of this scheme are the very people who are supposed to be a last check against the excesses of the politicians in the government: this particular brand of proofiness comes directly from the Supreme Court of the United States.

The proofiness in question is a form of voter suppression—keeping "undesirable" votes away from the polling places. There's a long history of voter suppression in the United States, and as with gerrymandering, the political monkey business is inseparably mixed up with racial nastiness. After the Civil War, former slave-owning states used all sorts of tricks to prevent emancipated slaves from exercising their right to vote. They would create barriers to voter registration that were particularly burdensome to African Americans, who tended to be in the poorer and less well-educated segments of society. For example, in a number of states, people weren't allowed to vote unless they had paid a small fee—a "poll tax"—which of course hit poor African Americans much harder than richer citizens. (Especially since many white voters were exempted from paying the tax because

of a "grandfather clause.")* As a result, the tax prevented many of them from voting.† The states pretended that the taxes had a legitimate purpose—they were intended to raise revenue, and they argued that people who paid state taxes became more interested in furthering the state's welfare. In reality, though, poll taxes were simply intended to keep African Americans from voting. Similarly, literacy tests were supposedly instituted to ensure that voters were able to make informed decisions; instead, they had the effect of barring less-educated African Americans from going to the polls. Legislators and judges eliminated these particular forms of voter suppression in the 1960s, but there are other more subtle forms that are still a problem. Voter ID laws, for example, are touted as a way to reduce voter fraud, as are periodic purges of voter registrations. However, there's very good reason to believe that these measures are being enacted because they have the indirect effect of reducing the number of African-American, Latino, and other minority voters.

However, none of the forms of voter suppression are more effective or more insidious than one engineered by the Supreme Court. In a series of decisions, the Court dressed a mathematical lie in the mantle of truth, wiping millions of people out of existence with the

* These sorts of clauses tended to exempt a person from the poll tax if he could prove that his grandfather had the right to vote—which white folk usually could and African Americans could not. Nowadays, shorn of its original racist heritage, a grandfather clause only refers to an exemption from a new law based upon prior circumstances.

† Interestingly, the "poll" in "poll tax" didn't specifically refer to voting, even though functionally it was a tax on going to the polls. The term comes from Middle English—*polle* meant "head," so a poll tax was in fact a tax put on each person's head (also known as a "capitation"). A polling place, on the other hand, is a place where your head is counted, so there's a shared etymology. Nevertheless, it's something of a coincidence that poll taxes were used to keep people from the polls.

stroke of a pen. Thanks to these rulings, more than 1 percent of the population of the United States consists of ghostlike disenfranchised creatures—citizens who in theory have the right to vote but are deliberately ignored. It's a stunning case of proofiness that goes right to the heart of what democracy is all about.

This particular scheme has to do with manipulating the U.S. Census. This may not seem like such a sinister plot, but at its root, a democracy is a government based on counting—on counting its citizens and their votes. The founding fathers of the United States recognized the importance of counting to their new government. Indeed, only five paragraphs into the U.S. Constitution, there's a passage that dictates that the government must perform an "actual enumeration" of its citizens every ten years. This decennial census is crucial to the functioning of the Republic, because it determines how much power different groups get to wield in the House of Representatives.

The 435 representatives in the House are divided (roughly) equally among the citizens of the United States—each representative nowadays votes by proxy for a block of roughly 700,000 people. The more citizens that a state has, the more representatives it gets, and the more power it wields in Congress. As the population shifts, political power (and money) follows. As the Northeast of the country atrophies, New York and Pennsylvania have been losing their preeminence to California and Texas. The political fortunes of a region—and of those who live in that region—hinge upon the results of the decennial census.

The government spends an ungodly amount of time and money to make an accurate count of its citizens; in 2000, the census cost roughly $6.5 billion—more than twenty dollars for each man, woman, and child in the United States. It's an incredible undertaking, and it's

about as accurate as such a measurement can be. Unfortunately, the census, like any measurement, is fallible. And since the 1940s, statisticians have been forced to admit a depressing fact: no matter how hard census workers try, there's a systematic error that they can't get rid of. They can't count everybody.

In some ways, the census is like a monstrous government-run poll, but there's one very important difference. Instead of querying a sample of households and extending those results to the entire population, the census attempts to reach every single household in the United States; in theory, there's no extrapolation needed. So, just as in the case of voting, there's no statistical error. There isn't any worry that a statistical fluke makes the census sample look different from the entire population because the sample *is* the entire population. The census workers only have to worry about systematic errors. And there are quite a few to worry about.

Every poll relies upon the cooperation of its subjects—a poll can't record the opinions of people who toss their reply card in the trash or who slam down the phone when they hear the voice of a pollster. As a result, all polls are subject to "volunteer bias" that can inject an enormous amount of error into the poll. The census is no different. Every single household in the United States gets a census form, and the majority fill it out, but quite a few don't. In 2000, roughly one in three households didn't bother to return their questionnaire. To get an accurate count of the population, the Census Bureau still has to count the citizens in the households that refused to respond. This is where the big spending comes in. The bureau dispatches thousands of census workers who spend months going from household to household tracking down nonrespondents. The harder it is to get a household to respond, the more money is spent to try to contact the people in that household, but the bureau keeps

trying until they run out of time and are required, by law, to give Congress the results. By the end of the process, the bureau manages to wring data out of all but about 2 percent of the population. It's an extraordinary effort. But it still is full of errors.

Not only has the census failed to reach 2 percent of the population; it accidentally double-counts about 1 percent. This means that for all that effort, the census is only good to within about ten million people, plus or minus. This plus or minus is enormously important, politically; these ten million people would be entitled to roughly fourteen representatives in the House. It's incredibly disheartening; all that time and money spent, and errors in the census are still huge. These errors are impossible to correct by ordinary means. The government could theoretically stake out the homes of every single nonrespondent, but that would cost astronomical amounts of money, and even this wouldn't manage to catch everybody. Even with double its current budget, the Census Bureau can't do much better with its measurements than they already are. However, the situation isn't hopeless. There is a way to reduce these errors enormously by using a set of statistical tricks known collectively as *sampling*.

The best way to understand sampling is through an example. Imagine that there's a shallow pond that's full of trout and minnows. The government has hired you to count how many fish the pond contains. You row gently from one end of the pond to the other, counting the fish that you see along the way. You come up with a count of 599 trout and 301 minnows. Your grand total is 900 fish in the pond, about 67 percent of which are trout and 33 percent minnows.

As you can probably guess, the answer is off because your count is error-prone. One source of error is that the fish are constantly moving about, making it all but certain that you'll count some fish

twice and others not at all. Another source of error is that minnows are harder to spot than trout. They're tiny and timid; they tend to hide when the boat comes nearby. So it's quite likely that you're undercounting minnows—and no matter how many times you count from your boat, minnows are likely to be underrepresented. Conversely, big, visible trout are more likely to be double-counted.

You can correct for these errors, but to do so, you have to make another measurement to figure out how bad they really are. After you've done your initial survey of aquatic life, you do another boat count of a small, representative section of the pond and record the numbers of fish that you find (say, 30 trout and 15 minnows). Then you make a more careful (and more invasive) count of that small section. Net off that little region of the pond, dredge up every single fish in that area, and pull them into the boat. Counting them as you toss them one by one back into the pond yields an incredibly accurate count: say you find that there are really 28 trout and 19 minnows.

This new information tells you how accurate your boat count really was. The data tell you that you did in fact overcount trout (you counted 30 from the boat, but there were really 28) and undercounted minnows (you counted 15 from the boat, but there were really 19). And now that the data tell you the nature of your measurement errors, you can correct for them. You now know that your original count of 599 trout is too large and should be adjusted downward—to about 560—to compensate for your tendency to overcount trout. Similarly, your count of 301 minnows is too small and should be adjusted upward—to about 380—to account for timid minnows that you were unable to see from the boat. Your new, adjusted total is 940 fish in the pond, about 60 percent of which are trout and 40 percent of which are minnows.

The new numbers aren't perfect by any means. It's possible that the small netted-off section of the pond was not truly representative of the entire pond. There might have been a particularly dense and hard-to-spot concentration of minnows in the area, for example.* Also, since you're extending your observations about a small number of fish to the entire pond, you have to worry about statistical errors that would be irrelevant in a direct count of the entire population. However, the increase in statistical error is more than compensated for by the decrease in systematic error—your measurement allows a dramatic reduction in the problems caused by miscounting certain segments of the population. In short, you're trading large, systematic errors for (hopefully) smaller, mostly statistical errors—and the result is a better, more accurate count.

This is sampling in a nutshell. By looking extremely carefully at a sample of the population, the Census Bureau can generate data that allow it to correct for the systematic undercounts and overcounts in the census. From a statistician's point of view, it's a no-brainer. A corrected count would produce a much more accurate depiction of the population of the United States than a count-every-head census ever could. Instead of having censuses that are good to within a few percent, it would be possible to reduce the errors down to a fraction of a percent. The most accurate tally of the population of the United States would not come from a straight head count; instead, it should be a census that is corrected by sampling. As an added bonus, a census that uses sampling is cheaper than a straight head count. Instead of spending billions of dollars to try to chase down that recalcitrant last few percent who don't respond to census

* Due, no doubt, to the presence of that most feared of aquatic creatures, the statistical fluke.

workers, the bureau can spend a few tens of millions doing the same thing, even more exhaustively, in a small number of communities and use that data to correct for the undercount. Sampling is more accurate and it's cheaper. So every politician should be in favor of it, right?

Not quite. Unfortunately, sampling is caught up in the racial politics of voter suppression. The citizens who tend to be undercounted by the census tend to be poorer people who rent their homes rather than own them. A disproportionate number don't speak English and are distrustful of government authorities (including the Census Bureau). They tend to be minorities—and they tend to vote Democratic. Conversely, the overcounted tend to be white and affluent, and are more likely than not to vote Republican. If the United States were a pond, minorities would be the minnows while whites would be the trout. The moment you use sampling to correct for the undercount, you suddenly add several million more minorities—Democrats—into your count of the population. It's something that Republicans want to prevent so badly that they are forced to take an idiotic stance: they insist the proper way to conduct a census is the least accurate and most expensive method.*

The Census Bureau was reduced to reporting two population numbers to Congress every decade: a sampling-corrected number that statisticians and population experts use because they need precise data to estimate everyday population trends, such as poverty

* The opposition to sampling, as with other forms of voter suppression, doesn't run ideologically deep in the Republican Party any more than pro-minority, pro-voting-rights sentiments run deep in the thoughts of mainstream Democrats. If the roles were reversed—if it were primarily Republicans who were being undercounted—there's little doubt that Democrats would be trying to suppress sampling while Republicans would be championing it. It's all petty-minded scrabbling to gain a political advantage.

rates, incomes, and household sizes; and the highly error-ridden head count that Congress insisted on using to reapportion House seats. When in the late 1990s the Census Bureau finally proposed presenting only its best, sampling-corrected number to Congress, Republicans in the House of Representatives promptly sued. Using sampling to correct the numbers, the Republicans argued, was unconstitutional; the "actual enumeration" required by the Constitution had to be a simple head count unspoiled by any statistical mumbo-jumbo that might make it more accurate.

The case went all the way up to the Supreme Court. The American Statistical Association filed a brief with the court that made the case pretty clear: "Properly designed sampling is often a better and more accurate method of gaining such knowledge than an inevitably incomplete attempt to survey all members of such a population. . . . There are no sound scientific grounds for rejecting all use of statistical sampling in the 2000 census." But the Supreme Court disagreed. In a five-to-four decision—the five most conservative judges versus the four most liberal—the Court determined that sampling was illegal. The apportionment of House seats, by law, had to be based upon flawed, highly error-prone population numbers that undercounted minority voters.

Even though the ruling evaded the question about whether the use of statistics ran contrary to the Constitution—whether the "actual enumeration" clause referred to a head count and nothing else—there's no question that the conservative majority was hostile to the whole concept of sampling. In a concurring opinion penned by Antonin Scalia, one of the most conservative justices of the Court, Scalia strongly implied that any use of statistical techniques would make the founding fathers spin in their graves.

Scalia turned to eighteenth-century dictionaries to show that

the phrase "actual enumeration" had to mean counting each individual person, one by one; he cited the 1773 Samuel Johnson dictionary, for example, which defined "enumerate" as "To reckon up singly; to count over distinctly; to number." Aha! To reckon up singly! Scalia pounced: by using the term "enumerate," the founding fathers meant to count each person, one by one. Thus the census "requires an actual counting, and not just an estimation of number." Further, using statistical techniques will "give the party controlling Congress the power to distort representation in its own favor." Only head counts, as inaccurate as they are, are free from manipulation. Thus sampling is unconstitutional.

This is a specious argument on several counts. First, even if Johnson's dictionary was the key to the founding fathers' intent, the definition "to number" is just as valid as "to reckon singly." Besides, the word "enumeration" was merely an accident. It was inserted during the drafting of the U.S. Constitution by the Committee on Style, a group that made minor, nonmeaningful changes to the document to correct grammar and clarity. They changed the word "census" to "enumeration" for reasons unknown.* So to dwell upon the finer points of the dictionary definition of "enumeration" is ignoring the fact that the founding fathers called for a "census," and that "enumeration" was substituted for reasons of style.

More important, there's no bright-line distinction between "actual counting" and "estimation." As we've seen, counting is a measurement like any other, and is thus subject to error. This error is unavoidable; it turns even the best count into nothing more than an

* One theory is that people were afraid of taking a census that wasn't approved by God. According to the Bible, when King David called for an unsanctioned census, God punished him by sending a pestilence that killed 70,000 Israelites in three days.

estimate, an approximation of the truth. And if there are other measurement techniques that give you a better approximation of the truth—such as statistical sampling—they deserve the title of "actual" more than counting does. Pretending otherwise is to place too much faith in the error-prone numbers that come from a head count: it is an act of disestimation.

Finally, the idea that head counts are free from manipulation is wrong. In fact, even the purest, most pristine head count that the U.S. Census can possibly perform is subject to statistical tinkering. It has to be. Census Bureau workers have to be able to interpret and even alter the data using statistical tools, otherwise the census would be utterly meaningless.

Imagine, for example, that some joker in Sitka, Alaska, fills out his census form to say that there are 300 million people living in his household. If the Census Bureau were to take him seriously, it would mean that Alaska would suddenly be the most populous state in the Union by a huge margin; indeed, half of the representatives in the House would be representing this gentleman's household. Luckily, no census worker is stupid enough to believe him. It's obvious that the guy is lying—he gave the census a bad piece of data. But what can the Census Bureau do about it? The only choice is to clean up the datum somehow—and doing this means that they must use a statistical technique known as *imputation*.

In an imputation, a Census Bureau statistician picks out a datum that looks wrong. (Anyone who says that he has seventy-seven children or is 175 years old, for example, is probably lying.) Then the statistician wipes out the questionable answer and replaces it with census data from similar-looking households. The replacement number is a guess, but an educated one—and it's certainly closer to

the truth than the phony datum. And in fact, there's really no alternative. Wiping out the datum or, more drastically, tossing out the entire census form is also imputation. The act of wiping out a datum is a substitution: the worker is still replacing someone's answer (seventy-seven children) with another answer (zero children); a null answer is still an answer. Similarly, tossing out a census form is equivalent to imputing that a dwelling is vacant. Instead of making bad imputations by simply wiping out dubious results, the Census Bureau prefers to make an educated guess from the freshest census data it has, a process known as "hot-deck imputation."* It's more likely to be approximately correct, so it does less violence to the validity of the census results. The only other option—the only way to avoid imputation entirely—is to take every single census form at face value. You have to duly record the responses of every 175-year-old woman, every man with seventy-seven children, and, yes, the gentleman in Sitka who has 300 million people in his household. Without imputation, the results of the census become worthless.

The Supreme Court decision about sampling was effectively a ban on using statistical mumbo-jumbo, but imputation is a form of statistical mumbo-jumbo that wasn't addressed by the previous decision. So when the (sampling-free!) 2000 census results were released, the state of Utah, which was denied an extra representative in Congress, sued. They argued that the bureau's use of imputation was illegal, and in 2002 this case worked its way up to the Supreme Court.

Utah's case put the court in a bind. If the justices ruled that

* The "deck" was a deck of punch cards, which shows how long this technique has been in use—since the first half of the twentieth century.

imputation was unconstitutional, it would have rendered the Census Bureau powerless to correct spurious data; the one joker from Sitka could theoretically render the entire count meaningless. However, if the Court decided that imputation was permissible, it had to split hairs to explain why one statistical technique—imputation—was acceptable while another—sampling—was illegal.

In another five-to-four decision—the liberals were joined by the usually conservative chief justice, William Rehnquist—the Court decided to take the latter course. In a shining example of how many justices it takes to split a hair, the Census Bureau was allowed to continue using imputation, at least for the moment. However, the minority lobbed grenades at the decision, accusing the bureau of using illegal, and perhaps unconstitutional, statistical witchcraft. Sandra Day O'Connor wrote that imputation was simply a form of sampling and should thus be banned. Clarence Thomas essentially repeated Scalia's argument from the earlier sampling case (even recycling the dictionary definitions of "enumeration") but turned up the volume a little bit. He insisted that the founding fathers were "well familiar with methods of estimation," so had explicitly rejected the sophisticated techniques used by the Census Bureau. (Which, as those techniques were developed in the twentieth century, would be quite a feat.) And Thomas repeated the canard that "actual counting" is fundamentally different from "estimation" even though counting *is* an estimate. Once again, the justices marshaled proofiness to justify using bad numbers instead of good ones—and to try to ensure that certain people are robbed of representation by virtue of a purposefully inaccurate census.

As this book goes to press, the Census Bureau is beginning its 2010 census. The Supreme Court's two rulings leave census law in a

complete shambles; there's a sense that some statistical techniques are kosher while others are illegal, and there's no real basis for telling which are which. Republican legislators are already challenging the validity of the new census. They threatened to sue long before the first census form went out in the mail, and a few have gone beyond mere threats. In October 2009, Louisiana senator David Vitter—probably best known for his use of a high-end D.C. escort agency—tried to force the Census Bureau to rewrite its census forms to ask respondents to declare whether or not they were U.S. citizens. If his measure had passed, not only would it have cost enormous amounts of money (425 million forms would have had to be thrown out) and delayed the start of the census, but it would also certainly have scared noncitizens, particularly illegal immigrants, from participating in the census.* It was a shameless play to try to make the undercount even worse. Predictably, it was defeated by a strict party-line vote, with the Democrats opposing and the Republicans backing it.

However, open season will really begin once the results of the new census are in. It's almost certain that the new decade will bring a fresh cluster of lawsuits about census methods, and, given the conservative makeup of the Supreme Court, it's quite possible that the first census of the twenty-first century will be forced to divest itself of all mathematical techniques that were developed after the eighteenth.

Proofiness has undermined the very foundations of our democracy—the mechanisms that we use to count our citizens and ensure that they are justly represented in the Republic. Gerryman-

* The Constitution requires that *all* persons be counted, not just U.S. citizens or legal immigrants.

dering for political gain is deemed acceptable, even though it clearly dilutes the votes of some of our citizens. Statistical sampling is deemed unacceptable, even though rejecting it forces the government to use numbers that it knows are inaccurate. No matter how many intellectual backflips legislators and judges go through to justify their positions, the fact remains: bad mathematics is being used to deny our citizens—mostly our minorities—their rightful vote. In a democracy, there can be no graver sin.