Nate Silver, _The Signal and the Noise_

ch13

The most ominous signal of all was the silence. American intelligence officials had ingeniously succeeded in breaking PURPLE, the code that Japan used to encrypt its diplomatic messages, allowing us to decipher perhaps 97 percent of them.4 Our attempts to decode Japanese military transmissions were less successful. But even if we could not understan the messages, we heard them and could trace their location. The steady stream of click-clacks from Japan's fleet of aircra carriers ordinarily betrayed their whereabouts when they were out to sea.
From mid-November onward, however, there had been total radio silence; we had no idea where the carriers were. There were no global satellites in the 1940s, and only the primitive makings of radar. Air patrol reconnaissance missions were cost-prohibitive in the vast reaches of the Pacific and were carried out erratically at a distance of only three hundred or four hundred miles from the base.5 The radio transmissions were our best means of detection, and without them an entir fleet of these ships, each of them the size of six football fields, had disappeared.
Many in the intelligence community concluded that the carriers were close to their home waters where they could rely o alternate means of communication.6 The second possibility was that the fleet had ventured far into the Pacific, away fro American naval installations.7

But Rumsfeld was in a good mood, having scrutinized the detailed outline for this book that I had given to his young and able chief of staff, Keith Urbahn.11 I knew of Rumsfeld's interest in Pearl Harbor. He greeted me with a photocopy of the foreword to a remarkable book, Roberta Wohlstetter's 1962 Pearl Harbor: Warning and Decision, which outlined the myriad reasons why the Japanese attack had been such a surprise to our military and intelligence officers. Worse than being unprepared, we had mistaken our ignorance for knowledge and made ourselves more vulnerable as a result.
"In Pearl Harbor, what they prepared for were things that really didn't happen," Rumsfeld said. "They prepared for sabotage because they had so many Japanese descendants living in Hawaii. And so they stuck all the airplanes close together, so they could be protected. So of course the bombers came and they were enormously vulnerable, and they wer destroyed."
In advance of Pearl Harbor, as Rumsfeld mentioned, we had a theory that sabotage-attack from within-was the most likel means by which our planes and ships would be attacked. The concern over sabotage was pervasive in Hawaii.12 It was thought that the 80,000 Japanese nationals there might attack not just military bases but radio stations, pineapple farms, and dairy mills with little warning.* Any signals were interpreted in this context, logically or not, and we prepared for subterfuge.13 We stacked our planes wingtip to wingtip, and our ships stern to bow, on the theory that it would be easie to monitor one big target than several smaller ones.
Meanwhile we theorized that, if Japan seemed to be mobilizing for an attack, it would be against Russia or perhaps agai the Asian territorial possessions of the United Kingdom, Russia and the UK being countries that were already involved in the war. Why would the Japanese want to provoke the sleeping giant of the United States? We did not see that Japan believed our involvement in the war was inevitable,14 and they wanted to strike us when we were least prepared and the could cause the most damage to our Navy. The imperial Japanese government of the time was not willing to abandon its hopes for territorial expansion. We had not seen the conflict through the enemy's eyes.

The North American Aerospace Defense Command (NORAD) had actually proposed running a war game in which a hijacked airliner crashed into the Pentagon. But the idea was dismissed as being "too unrealistic."34 And in the unlikely event that such an attack were to occur, it was assumed, the plane would come from overseas and not from one of our domestic airports. (Ironically, this was the exact opposite of the mistake that we'd made before Pearl Harbor, where the possibility of an attack from abroad was dismissed because planners were concerned about sabotage.)
The possibility of a suicide attack may also have been hard to imagine. FAA policy was predicated on the idea that a hijacking would result in a tense standoff and perhaps a detour to some exotic airport in the Middle East. But it was assumed the terrorist would not want to destroy the plane, or to kill passengers other than as a negotiation tactic. Thus, cockpit doors were not tightly sealed and were often left entirely unlocked in practice.35

Power laws have some important properties when it comes to making predictions about the scale of future risks. In particular, they imply that disasters much worse than what society has experienced in the recent past are entirely possible, if infrequent. For instance, the terrorism power law predicts that a NATO country (not necessarily the United States) would experience a terror attack killing at least one hundred people about six times over the thirty-one-year perio from 1979 through 2009. (This is close to the actual figure: there were actually seven such attacks during this period.) Likewise, it implies that an attack that killed 1,000 people would occur about once every twenty-two years. And it sugges that something on the scale of September 11,48 which killed almost 3,000 people, would occur about once every forty years.
It's not that much of an accomplishment, however, to describe history in statistical terms. Sure, it's possible for a statistical model to accommodate an event like September 11 now that one has actually occurred. But what would Clauset's method have said about the possibility of such an attack before it happened?
September 11 certainly did shift the probabilities somewhat-just as the number of very large earthquakes in recent years implies that they are somewhat more common than we might have thought previously.49 Nevertheless, even before it occurred, the power-law method would have concluded that an attack on the scale of September 11 was a clear possibilit If the power-law process is applied to data collected entirely before 9/11-everything from the beginning of the modern wave of terrorism in 1979 through September 10, 2001-it implies that a September 11-scale attack would occur about once every eighty years in a NATO country, or roughly once in our lifetimes.50

Although Pakistan is ostensibly an ally of the United States, even the most generous interpretation would suggest that it represents a problem as well as a solution in the effort to contain terrorism. The country had initially been reluctant to cooperate with the United States after the September 11 attacks, and Pakistan's president later claimed that the U.S. ha

resorted to a threat to bomb it "back to the stone age" before it complied.62 Osama bin Laden had been living in Abbottabad, Pakistan, for as many as six years63 before he was killed. Meanwhile, Pakistan has roughly one hundred nuclear weapons and is building additional nuclear facilities and delivery systems at a rapid pace.64 The country now ranks seventh in the world in the Economist's Political Instability Index, up significantly from the recent past,65 meaning that the risk of a coup d'état or a revolution is quite high. A new regime could be openly hostile to the United States. All the conditions that a terrorist might need to acquire a nuclear weapon could then be in place.
Terrorist organizations are fundamentally weak and unstable: as is supposedly true of new restaurants, 90 percent of the fail within the first year.66 [Randy Borum, "Psychology of Terrorism," Encyclopedia of Peace Psychology (New York: Springer Science, 2010), p. 62. http://worlddefensereview.com/docs/PsychologyofTerrorism0707.pdf.]

The Gutenberg-Richter law dictates that, over the long term, the frequency of earthquakes is reduced about ten times for every one-point increase in magnitude. However, the energy released by earthquakes increases exponentially as a functi of magnitude. In particular, for every one-point increase in magnitude, an earthquake's energy release increases by abou thirty-two times. So a magnitude 6 earthquake releases around thirty-two times as much seismic energy as a magnitude while a magnitude 7 is close to 1,000 times more powerful.
The force released by earthquakes scales up at a faster rate than their frequency decreases. If there are ten magnitude 6 earthquakes for every magnitude 7, the magnitude 7 tremor will account for considerably more damage70 than all the magnitude 6s combined. Indeed, a mere handful of earthquakes are responsible for a very large fraction of their total seismic energy. In the one hundred years between 1906 and 2005, for instance, just three large earthquakes-the Chilean earthquake of 1960, the Alaskan earthquake of 1964, and the Great Sumatra Earthquake of 2004-accounted for almost half the total energy release of all earthquakes in the world over the entire century. So, seismologists and contingency planners are mostly concerned about very large earthquakes. A more modest earthquake in the wrong place at the wrong time can cause enormous damage (like the magnitude 7.0 earthquake in Haiti in 2010), but it's mostly the very high magnitude earthquakes that we have to worry about, even though they occur quite infrequently.

Although Israel is targeted by terrorists much more frequently than the United States, Israelis do not live in fear of terrorism. A 2012 survey of Israeli Jews found that only 16 percent described terrorism as their greatest fear81-no more than the number who said they were worried about Israel's education system.
No Israeli politician would say outright that he tolerates small-scale terrorism, but that's essentially what the country do It tolerates it because the alternative-having everyone be paralyzed by fear-is incapacitating and in line with the terrorist goals. A key element in the country's strategy is making life as normal as possible for people after an attack occurs. For instance, police typically try to clear the scene of an attack within four hours of a bomb going off,82 letting everyone get back to work, errands, or even leisure. Small-scale terrorism is treated more like crime than an existential threat.
What Israel certainly does not tolerate is the potential for large-scale terrorism (as might be made more likely, for instance, by one of their neighbors acquiring weapons of mass destruction). There is some evidence that their approach i successful: Israel is the one country that has been able to bend Clauset's curve. If we plot the fatality tolls from terrorist incidents in Israel using the power-law method (figure 13-8), we find that there have been significantly fewer large-scale terror attacks than the power-law would predict; no incident since 1979 has killed more than two hundred people. The fa that Israel's power-law graph looks so distinct is evidence that our strategic choices do make some difference.

Conclusion

The legendary shortstop Derek Jeter was a frequent subject of debate during the Moneyball era. Broadcasters and scouts noticed that Jeter seemed to make an especially large number of diving plays and concluded that he was an exceptional shortstop for that reason. Stat geeks crunched the numbers and detected a flaw in this thinking.1 Although Jeter was a terrific athlete, he often got a slow jump on the ball and dove because he was making up for lost time. In fact, the numbe suggested that Jeter was a fairly poor defensive shortstop, despite having won five Gold Glove awards. The plays that Jet had to dive for, a truly great defensive shortstop like Ozzie Smith might have made easily-perhaps receiving less credit fo them because he made them look routine.

One of the most spectacularly correct predictions in history was that of the English astronomer Edmund Halley, who in 1705 predicted that a great comet would return to the earth in 1758. Halley had many doubters, but the comet returned just in the nick of time.2 Comets, which in antiquity were regarded as being wholly unpredictable omens from the gods,3 are now seen as uncannily regular and predictable things.

5. Glenn Gunzelmann and Kevin A. Gluck, "Knowledge Tracing for Complex Training Applications: Beyond Bayesian Mastery Estimates," Air Force Research Laboratory,; Proceedings of the Thirteenth Conference on Behavior Representation in Modeling and Simulation, 2004, pp. 383-84. http://act-r.psy.cmu.edu/papers/710/gunzelmann_gluck-2004.pdf.
6. Sarah Lichtenstein and Baruch Fischhoff, "Training for Calibration," prepared for U.S. Army Research Institute for the Behavioral and Social Sciences, ARI Technical Report TR-78-A32; November 1978. http://www.dtic.mil/cgi-bin/GetTRDoc AD=ADA069703.
7. Christopher J. Gill, Lora Sabin and Christopher H. Schmidt, "Why Clinicians Are Natural Bayesians," British Medical Journal, vol. 330; May 7, 2005. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC557240/

Nate Silver 2012, _The Signal and the Noise_

ch12

Worse yet, the beer is expensive: the high taxes on alcohol and pretty much everything else in Denmark help to pay for a green-technology infrastructure that rivals almost anywhere in the world. Denmark consumes no more energy today than it did in the late 1960s,28 in part because it is environmentally friendly and in part because of its low population growth. (By contrast, the United States' energy consumption has roughly doubled over the same period.29) The implicit message seemed to be that an energy-efficient future would be cold, dark, and expensive.

It is little wonder, then, that the mood at Copenhagen's Bella Center ranged far beyond skepticism and toward outright cynicism. I had gone to the conference, somewhat naively, seeking a rigorous scientific debate about global warming. What I found instead was politics, and the differences seemed irreconcilable.

Delegates from Tuvalu, a tiny, low-lying Pacific island nation that would be among the most vulnerable to rising sea level roamed the halls, loudly protesting what they thought to be woefully inadequate targets for greenhouse-gas reduction. Meanwhile, the large nations that account for the vast majority of greenhouse-gas emissions were nowhere near agreement.

President Obama had arrived at the conference empty-handed, having burned much of his political capital on his health-care bill and his stimulus package. Countries like China, India, and Brazil, which are more vulnerable than the United States to climate change impacts because of their geography but are reluctant to adopt commitments that might impair their economic growth, weren't quite sure where to stand. Russia, with its cold climate and its abundance of fossil-fuel resources, was a wild card. Canada, also cold and energy-abundant, was another, unlikely to push for any deal that the United States lacked the willpower to enact.30

The criticisms that Armstrong and Green make about climate forecasts derive from their empirical study of disciplines lik economics in which there are few such physical models available49 and the causal relationships are poorly understood. Overly ambitious approaches toward forecasting have often failed in these fields, and so Armstrong and Green infer that they will fail in climate forecasting as well.

The goal of any predictive model is to capture as much signal as possible and as little noise as possible. Striking the right balance is not always so easy, and our ability to do so will be dictated by the strength of the theory and the quality and quantity of the data. In economic forecasting, the data is very poor and the theory is weak, hence Armstrong's argument that "the more complex you make the model the worse the forecast gets."

In climate forecasting, the situation is more equivocal: the theory about the greenhouse effect is strong, which supports more complicated models. However, temperature data is very noisy, which argues against them.

One of the more forthright early efforts to forecast temperature rise came in 1981, when James Hansen and six other scientists published a paper in the esteemed journal Science.72 These predictions, which were based on relatively simple statistical estimates of the effects of CO_2 and other atmospheric gases rather than a fully fledged simulation model, have done quite well. In fact, they very slightly underestimated the amount of global warming observed through 2011.73

Hansen is better known, however, for his 1988 congressional testimony as well as a related 1988 paper74 that he published in the Journal of Geophysical Research. This set of predictions did rely on a three-dimensional physical model o the atmosphere.

Hansen told Congress that Washington could expect to experience more frequent "hot summers." In his paper, he defined a hot summer as one in which average temperatures in Washington were in the top one-third of the summers observed from 1950 through 1980. He said that by the 1990s, Washington could expect to experience these summers 55 to 70 percent of the time, or roughly twice their 33 percent baseline rate.

In fact, Hansen's prediction proved to be highly prescient for Washington, DC. In the 1990s, six of the ten summers75 qualified as hot (figure 12-6), right in line with his prediction. About the same rate persisted in the 2000s and Washingto experienced a record heat wave in 2012. In his paper, Hansen had also made these predictions for three other cities: Omaha, Memphis, and New York. These results were more mixed and go to illustrate the regional variability of the climate. Just 1 out of 10 summers in Omaha in the 1990s qualified as "hot" by Hansen's standard, well below the historica average rate of 33 percent. But 8 out of 10 summers in New York did, according to observations at LaGuardia Airport.

Uncertainty in forecasts is not necessarily a reason not to act-the Yale economist William Nordhaus has argued instead that it is precisely the uncertainty in climate forecasts that compels action,86 since the high-warming scenarios could be quite bad. Meanwhile, our government spends hundreds of billions toward economic stimulus programs, or initiates wars in the Middle East, under the pretense of what are probably far more speculative forecasts than are pertinent in climate science.87

And in contrast to other fields in which poor predictions are quickly forgotten about, errors in forecasts about the climate are remembered for decades.

One common claim among climate critics is that there once had been predictions of global cooling and possibly a new ice age. Indeed, there were a few published articles that projected a cooling trend in the 1970s. They rested on a reasonable enough theory: that the cooling trend produced by sulfur emissions would outweigh the warming trend produced by carbon emissions.

These predictions were refuted in the majority of the scientific literature. [88. Thomas C. Peterson, William M. Connolley and John Fleck, "The Myth of the 1970s Global Cooling Scientific Consensus," Bulletin of the American Meteorological Society, September 2008. http://scienceblogs.com/stoat/Myth-1970-Global-Cooling-BAMS-2008.pdf.] This was less true in the news media. A Newsweek story in 1975 imagined that the River Thames and the Hudson River might freeze over and stated that there would be a "drastic decline" in food production89-implications drawn by the writer of the piece but not any of the scientists he spoke with.

If the media can draw false equivalences between "skeptics" and "believers" in the climate science debate, it can also sometimes cherry-pick the most outlandish climate change claims even when they have been repudiated by the bulk of a scientist's peers.

What is the baseline in the case of the climate? If the critique of global warming forecasts is that they are unrealistically complex, the alternative would be a simpler forecast, one grounded in strong theoretical assumptions but with fewer bell and whistles.

Suppose, for instance, that you had attempted to make a climate forecast based on an extremely simple statistical model one that looked solely at CO2 levels and temperatures, and extrapolated a prediction from these variables alone, ignoring sulfur and ENSO and sunspots and everything else. This wouldn't require a supercomputer; it could be calculated in a fe microseconds on a laptop. How accurate would such a prediction have been?

In fact, it would have been very accurate-quite a bit better, actually, than the IPCC's forecast. If you had placed the temperature record from 1850 through 1989 into a simple linear regression equation, along with the level of CO2 as measured in Antarctic ice cores93 and at the Mauna Loa Observatory in Hawaii, it would have predicted a global temperature increase at the rate of 1.5°C per century from 1990 through today, exactly in line with the actual figure (figure 12-9).

Another technique, only slightly more complicated, would be to use estimates that were widely available at the time abou the overall relationship between CO2 and temperatures. The common currency of any global warming forecast is a value that represents the effect on temperatures from a doubling (that is, a 100 percent increase) in atmospheric CO2. There has long been some agreement about this doubling value.94 From forecasts like those made by the British engineer G. S. Callendar in 193895 that relied on simple chemical equations, to those produced by today's supercomputers, estimates have congregated96 between 2°C and 3°C of warming from a doubling of CO2.

Given the actual rate of increase in atmospheric CO2, that simple conversion would have implied temperature rise at a rate of between 1.1°C and 1.7°C per century from 1990 through the present day. The actual warming pace of 0.015°C pe year or 1.5°C per century fits snugly within that interval.

James Hansen's 1981 forecasts, which relied on an approach much like this, did quite a bit better at predicting current tempertaures than his 1988 forecast, which relied on simulated models of the climate.

The Armstrong and Green critique of model complexity thus looks pretty good here. But the success of the more basic forecasting methods suggests that Armstrong's critique may have won the battle but not the war. He is asking some goo questions about model complexity, and the fact that the simple models do pretty well in predicting the climate is one pie of evidence in favor of his position that simpler models are preferable. However, since the simple methods correctly predicted a temperature increase in line with the rise in CO2, they are also evidence in favor of the greenhouse-effect hypothesis.

This type of framing can sometimes be made in bad faith. For instance, if you set the year 1998 as your starting point, which had record-high temperatures associated with the ENSO cycle, it will be easier to identify a cooling "trend." Conversely, the decadal "trend" from 2008 through 2018 will very probably be toward warming once it is calculated, sin 2008 was a relatively cool year. Statistics of this sort are akin to when the stadium scoreboard optimistically mentions th the shortstop has eight hits in his last nineteen at-bats against left-handed relief pitchers-ignoring the fact that he is batting .190 for the season.100

Neither Armstrong nor Schmidt was willing to hedge very much on their predictions about the temperature trend. "We d some simulations from 1850 up to 2007," Armstrong told me. "When we looked one hundred years ahead it was virtually certain that I would win that bet."101 Schmidt, meanwhile, was willing to offer attractive odds to anyone betting against his position that temperatures would continue to increase. "I could easily give you odds on the next decade being warme than this decade," he told me. "You want 100-to-1 odds, I'd give it to you."

The statistical forecasting methods that I outlined earlier can be used to resolve the dispute-and they suggest that neithe Armstrong nor Schmidt has it quite right. If you measure the temperature trend one decade at a time, it registers a warming trend about 75 percent of the time since 1900, but a cooling trend the other 25 percent of the time. As the growth rate of atmospheric CO2 increases, creating a stronger greenhouse signal, periods of flat or cooling temperature should become less frequent. Nevertheless, they are not impossible, nor are the odds anything like 100-to-1 against them Instead, if you assume that CO2 levels will increase at the current pace of about 2 ppm per year, the chance that there would be no net warming over the course of a given decade would be about 15 percent102 according to this method.

The street-fighter mentality, nevertheless, seems to be predicated on the notion that we are just on the verge of resolving our political problems, if only a few more people could be persuaded about the science. In fact, we are probably many years away. "There's a point when I come to the conclusion that we're going to have to figure out how to take the carbon out," Richard Rood told me in Copenhagen, anticipating that there was almost no way the 193 members of the United Nations would agree to mutually acceptable terms.

Meanwhile, the American public's confidence that global warming is occurring has decreased somewhat over the past several years.109 And even if there were 100 percent agreement on the effects of climate change, some states and some countries would make out better than others in any plan to mitigate carbon emissions. "We have some very progressive Democratic governors in coal states," I was told by the governor of Washington, Christine Gregoire. "Boy, are they nervo about all this."

Nate Silver 2012, _The Signal and the Noise_

ch11

In 2009, a year after a financial crisis had wrecked the global economy, American investors traded $8 million in stocks every second that the New York Stock Exchange was open for business. Over the course of the typical trading day, the volume grew to $185 billion, roughly as much as the economies of Nigeria, the Philippines or Ireland produce in an entir year. Over the course of the whole of 2009, more than $46 trillion1 in stocks were traded: four times more than the revenues of all the companies in the Fortune 500 put together.2
This furious velocity of trading is something fairly new. In the 1950s, the average share of common stock in an American company was held for about six years before being traded-consistent with the idea that stocks are a long-term investmen By the 2000s, the velocity of trading had increased roughly twelvefold. Instead of being held for six years, the same share of stock was traded after just six months.3

If there really were a Bayesland, then Justin Wolfers, a fast-talking, ponytailed polymath who is among America's best young economists, would be its chief of police, writing a ticket anytime he observed someone refusing to bet on their forecasts. Wolfers challenged me to a dinner bet after I wrote on my blog that I thought Rick Santorum would win the Iowa caucus, bucking the prediction market Intrade (as well as my own predictive model), which still showed Mitt Romne ahead. In that case, I was willing to commit to the bet, which turned out well for me after Santorum won by literally just few dozen votes after a weeks-long recount.* But there have been other times when I have been less willing to accept one of Wolfers' challenges. Presuming you are a betting man as I am, what good is a prediction if you aren't willing to put money on it?

Nevertheless, there is strong empirical and theoretical evidence that there is a benefit in aggregating different forecasts Across a number of disciplines, from macroeconomic forecasting to political polling, simply taking an average of everyone's forecast rather than relying on just one has been found to reduce forecast error,14 often by about 15 or 20 percent.
But before you start averaging everything together, you should understand three things. First, while the aggregate forecast will essentially always be better than the typical individual's forecast, that doesn't necessarily mean it will be good. For instance, aggregate macroeconomic forecasts are much too crude to predict recessions more than a few month in advance. They are somewhat better than individual economists' forecasts, however.
Second, the most robust evidence indicates that this wisdom-of-crowds principle holds when forecasts are made independently before being averaged together. In a true betting market (including the stock market), people can and do react to one another's behavior. Under these conditions, where the crowd begins to behave more dynamically, group behavior becomes more complex.
Third, although the aggregate forecast is better than the typical individual's forecast, it does not necessarily hold that it i better than the best individual's forecast. Perhaps there is some polling firm, for instance, whose surveys are so accurate that it is better to use their polls and their polls alone rather than dilute them with numbers from their less-accurate pee When this property has been studied over the long run, however, the aggregate forecast has often beaten even the very best individual forecast. A study of the Blue Chip Economic Indicators survey, for instance, found that the aggregate forecast was better over a multiyear period than the forecasts issued by any one of the seventy economists that made up the panel.15 Another study by Wolfers, looking at predictions of NFL football games, found that the consensus forecasts produced by betting markets were better than about 99.5 percent of those from individual handicappers.16 And this is certainly true of political polling; models that treat any one poll as the Holy Grail are more prone to embarrassing failures.17 Reducing error by 15 or 20 percent by combining forecasts may not sound all that impressive, but it's awfully hard to beat in a competitive market.

Also, while I accept the theoretical benefits of prediction markets, I don't know that political betting markets like Intrade are all that good right now-the standard of competition is fairly low. Intrade is becoming more popular, but it is still smal potatoes compared with the stock market or Las Vegas. In the weeks leading up to the Super Tuesday primaries in March 2012, for instance, about $1.6 million in shares were traded there;18 by contrast, $8 million is traded in the New York Stock Exchange in a single second. The biggest profit made by any one trader from his Super Tuesday bets was about $9,000, which is not enough to make a living, let alone to get rich. Meanwhile, Intrade is in a legal gray area and most of the people betting on American politics are from Europe or from other countries. There have also been some cases of market manipulation*19 or blatant irrational pricing20 there. And these markets haven't done very well at aggregating information in instances where there isn't much information worth aggregating, like in trying to guess the outcome of Supreme Court cases from the nebulous clues the justices provide to the public.
Could FiveThirtyEight and other good political forecasters beat Intrade if it were fully legal in the United States and its trading volumes were an order of magnitude or two higher? I'd think it would be difficult. Can they do so right now? My educated guess21 is that some of us still can, if we select our bets carefully.22

Efficient-market hypothesis is sometimes mistaken for an excuse for the excesses of Wall Street; whatever else those guy are doing, it seems to assert, at least they're behaving rationally. A few proponents of the efficient-market hypothesis might interpret it in that way. But as the theory was originally drafted, it really makes just the opposite case: the stock market is fundamentally and profoundly unpredictable. When something is truly unpredictable, nobody from your hairdresser to the investment banker making $2 million per year is able to beat it consistently.

Suppose that we looked at the daily closing price of the Dow Jones Industrial Average in the 10 years between 1966 and 1975-the decade just after Fama had published his thesis. Over this period, the Dow moved in the same direction from da to day-a gain was followed by a gain or a loss by a loss-58 percent of the time. It switched directions just 42 percent of th time. That seems nonrandom and it is: a standard statistical test38 would have claimed that there was only about a 1-in-

quintillion possibility (1 chance in 7,000,000,000,000,000) that this resulted from chance alone.
But statistical significance does not always equate to practical significance. An investor could not have profited from this trend.
Suppose that an investor had observed this pattern for ten years-gains tended to be followed by gains and losses by losse On the morning of January 2, 1976, he decided to invest $10,000 in an index fund39 which tracked the Dow Jones Industrial Average. But he wasn't going to be a passive investor. Instead he'd pursue what he called a Manic Momentum strategy to exploit the pattern. Every time the stock market declined over the day, he would pull all his money out, avoiding what he anticipated would be another decline the next day. He'd hold his money out of the market until he observed a day that the market rose, and then he would put it all back in. He would pursue this strategy for ten years, until the last trading day of 1985, at which point he would cash out his holdings for good, surely assured of massive profits.
How much money would this investor have at the end of the ten-year period? If you ignore dividends, inflation, and transaction costs, his $10,000 investment in 1976 would have been worth about $25,000 ten years later using the Manic Momentum strategy. By contrast, an investor who had adopted a simple buy-and-hold strategy during the same decade-buy $10,000 in stocks on January 2, 1976, and hold them for ten years, making no changes in the interim-would have onl about $18,000 at the end of the period. Manic Momentum seems to have worked! Our investor, using a very basic strateg that exploited a simple statistical relationship in past market prices, substantially beat the market average, seeming to disprove the efficient-market hypothesis in the process.
But there is a catch. We ignored this investor's transaction costs. This makes an enormous difference. Suppose that the investor had pursued the Manic Momentum strategy as before but that each time he cashes into or out of the market, he paid his broker a commission of 0.25 percent. Since this investor's strategy requires buying or selling shares hundreds o times during this period, these small costs will nickel-and-dime him to death. If you account for his transaction costs, in fact, the $10,000 investment in the Manic Momentum strategy would have been worth only about $1,100 ten years later, eliminating not only his profit but also almost all the money he put in originally. In this case, there is just a little bit of predictability in stock-market returns-but not nearly enough to make a profit from them, and so efficient-market hypothesis is not violated.
The other catch is that the pattern has since reversed itself. During the 2000s, the stock market changed direction from day to day about 54 percent of the time, just the opposite of the pattern from earlier decades. Had the investor pursued l Manic Momentum strategy for ten years beginning in January 2000, his $10,000 investment would have been whittled down to $4,000 by the end of the decade even before considering transaction costs.40 If you do consider transaction cost the investor would have had just $141 left over by the end of the decade, having lost almost 99 percent of his capital.

Some of the prices listed on the NADSAQ seemed to be plainly irrational. At one point during the dot-com boom, the market value of technology companies accounted for about 35 percent of the value of all stocks in the United States,41 implying they would soon come to represent more than a third of private-sector profits. What's interesting is that the technology itself has in some ways exceeded our expectations. Can you imagine what an investor in 2000 would have dor if you had shown her an iPad? And told her that, within ten years, she could use it to browse the Internet on an airplane flying 35,000 feet over Missouri and make a Skype call* to her family in Hong Kong? She would have bid Apple stock up infinity.
Nevertheless, ten years later, in 2010, technology companies accounted for only about 7 percent of economic activity.42 For every Apple, there were dozens of companies like Pets.com that went broke.

Identifying a bubble is of course much easier with the benefit of hindsight-but frankly, it does not seem all that challengi to do so in advance, as many economists did while the housing bubble was underway. Simply looking at periods when the stock market has increased at a rate much faster than its historical average can give you some inkling of a bubble. Of the eight times in which the S&P 500 increased in value by twice its long-term average over a five-year period,43 five cases were followed by a severe and notorious crash, such as the Great Depression, the dot-com bust, or the Black Monday cra of 1987.44

At various times, the P/E ratio for all companies in the S&P 500 ranged everywhere from about 5 (in 1921) to 44 (when Shiller published his book in 2000). Shiller found that these anomalies had predictable-seeming consequences for investors. When the P/E ratio is 10, meaning that stocks are cheap compared with earnings, they have historically produced a real return46 of about 9 percent per year, meaning that a $10,000 investment would be worth $22,000 ten years later. When the P/E ratio is 25, on the other hand, a $10,000 investment in the stock market has historically been worth just $12,000 ten years later. And when they are very high, above about 30-as they were in 1929 or 2000-the expected return has been negative.
However, these pricing patterns would not have been very easy to profit from unless you were very patient. They've become meaningful only in the long term, telling you almost nothing about what the market will be worth one month or one year later. Even looking several years in advance, they have only limited predictive power. Alan Greenspan first used the phrase "irrational exuberance" to describe technology stocks in December 1996,47 at which point the P/E ratio of the S&P 500 was 28-not far from the previous record of 33 in 1929 in advance of Black Tuesday and the Great Depression. T NASDAQ was more richly valued still. But the peak of the bubble was still more than three years away. An investor with perfect foresight, who had bought the NASDAQ on the day that Greenspan made his speech, could have nearly quadrupl his money if he sold out at exactly the right time. Instead, it's really only at time horizons ten or twenty years out that these P/E ratios have allowed investors to make reliable predictions.

But now consider what happens when the investor gets his bet wrong. This choice is much clearer.

• The trader buys but the market crashes. This is no fun: he's lost his firm a lot of money and there will be no big bonus and no new Lexus. But since he's stayed with the herd, most of his colleagues will have made the same mistake. Followin the last three big crashes on Wall Street, employment at securities firms decreased by about 20 percent.63 That means there is an 80 percent chance the trader keeps his job and comes out okay; the Lexus can wait until the next bull market.

A common experiment in economics classrooms, usually employed when the professor needs some extra lunch money, is hold an auction wherein students submit bids on the number of pennies in a jar.77 The student with the highest bid pays the professor and wins the pennies (or an equivalent amount in paper money if he doesn't like loose change). Almost invariably, the winning student will find that he has paid too much. Although some of the students' bids are too low and some are about right, it's the student who most overestimates the value of the coins in the jar who is obligated to pay for them; the worst forecaster takes the "prize." This is known as the "winner's curse."

There is reason to suspect that of the various cognitive biases that investors suffer from, overconfidence is the most pernicious. Perhaps the central finding of behavioral economics is that most of us are overconfident when we make predictions. The stock market is no exception; a Duke University survey of corporate CFOs,78 whom you might expect to be fairly sophisticated investors, found that they radically overestimated their ability to forecast the price of the S&P 500 They were constantly surprised by large movements in stock prices, despite the stock market's long history of behaving erratically over short time periods.

The economist Terrance Odean of the University of California at Berkeley constructed a model in which traders had this flaw and this flaw only: they were overconfident in estimating the value of their information. Otherwise, they were perfectly rational.79 What Odean found was that overconfidence alone was enough to upset an otherwise rational market Markets with overconfident traders will produce extremely high trading volumes, increased volatility, strange correlation in stock prices from day to day, and below-average returns for active traders-all the things that we observe in the real world.

Say, for instance, that you had borrowed five hundred shares of the company InfoSpace on March 2, 1999, when they cost $27, promising to return them one year later. Borrowing these shares would have cost you about $13,400. One year later however, InfoSpace was trading at $482 per share, meaning that you would be obligated to return about $240,000-almost twenty times the initial value of your investment. Although this bet would have turned out to be brilliant in the end-InfoSpace later traded for as little as $1.40 per share-you would have taken a bath and your ability to make future investments would be crippled. In fact, the losses from shorting a stock are theoretically unlimited.

In practice, the investor loaning you the shares can demand them back anytime she wants, as she assuredly will if she thinks you are a credit risk. But this also means she can quit anytime she's ahead, an enormous problem since overvalued stocks often become even more overvalued before reverting back to fairer prices. Moreover, since the investor loaning you the stocks knows that you may have to dig into your savings to pay her back, she will charge you a steep interest rate for the privilege. Bubbles can take months or years to deflate. As John Maynard Keynes said, "The market can stay irrational longer than you can stay solvent."

...Few holders of Palm stock were willing to loan their shares out, and they had come to expect quite a premium for doing so: an interest rate of well over 100 percent per year.82 This pattern was common during the dot-com bubble:83 shorting dot-com stocks was prohibitively expensive when it wasn't literally impossible.

In practice, most everyday investors do not do even that well. Gallup and other polling organizations periodically survey Americans94 on whether they think it is a good time to buy stocks. Historically, there has been a strong relationship between these numbers and stock market performance-but the relationship runs in the exact opposite direction of what sound investment strategy would dictate. Americans tend to think it's a good time to buy when P/E ratios are inflated and stocks are overpriced. The highest figure that Gallup ever recorded in their survey was in January 2000, when a record high of 67 percent of Americans thought it was a good time to invest. Just two months later, the NASDAQ and other stock indices began to crash. Conversely, only 26 percent of Americans thought it was a good time to buy stocks in February 1990-but the S&P 500 almost quadrupled in value over the next ten years (figure 11-10).

Daniel Kahneman likens the problem to the Müller-Lyer illusion, a famous optical illusion involving two sets of arrows (figure 11-11). The arrows are exactly the same length. But in one case, the ends of the arrows outward, seem to signify expansion and boundless potential. In the other case, they point inward, making them seem self-contained and limited. T first case is analogous to how investors see the stock market when returns have been increasing; the second case is how they see it after a crash.

FIGURE 11-11: MÜLLER-LYER ILLUSION

"There's no way that you can control yourself not to have that illusion," Kahneman told me. "You look at them, and one of the arrows is going to look longer than the other. But you can train yourself to recognize that this is a pattern that causes an illusion, and in that situation, I can't trust my impressions; I've got to use a ruler."

Nate Silver 2012, _The Signal and the Noise_

ch10

The year 2003 was the start of the "poker boom," a sort of bubble economy in which the number of new and inexperience players was growing exponentially and even a modicum of poker skill could be parlayed into large profits. The phenomenon had two immediate and related causes. One was the 2003 World Series of Poker in Las Vegas, which was won by a twenty-seven-year-old amateur, a Nashville accountant with the auspicious name of Chris Moneymaker. Moneymaker was the literal embodiment of the poker everyman: a slightly pudgy office drone who, through a never-ending series of daring bluffs and lucky draws, had turned the $39 he'd paid to enter an online qualifying tournament int a $2.5 million purse.
ESPN turned Moneymaker's achievement into a six-part miniseries, played on nearly continuous repeat on weekday evenings until baseball season finally came along to fill the void. It was terrific advertising for the "sport" of poker, which until that time had a reputation for being seedy, archaic, and intimidating. Suddenly, every balding, five-foot-eight accountant who had long ago given up on his dream of being the next Michael Jordan or Derek Jeter could see in Moneymaker someone who looked just like him, who had a job just like his, and who in a matter of weeks had gone from rank amateur to the winner of the biggest poker tournament in the world.
But the ESPN broadcasts presented a highly sanitized version of what reality actually looks like at the poker table. For or thing, out of the necessity of compressing more than forty hours of play involving more than eight hundred players into s hours of broadcasts, they showed only a small fraction of the hands as they were actually played. What's more, because o the ingenious invention of the "hole cam"-pinhole-size cameras installed around the edge of the table beside each player- the cards of not just Moneymaker but those of each of his opponents were revealed to the home audience as the hand wa being played out, giving the audience the feeling of being clairvoyant. Poker is a pretty easy game if you know what card your opponent holds.
Moneymaker was cast as the protagonist who could do no wrong. Hands that a sober analysis might have concluded he'd played poorly were invariably praised by the announcers-rash bluffs became gutsy ones, premature folds became perceptive ones. Moneymaker was not some slightly-above-average schmoe getting the cards of his life*1 but a poker savant who was cunning enough to have developed into a world-class player almost overnight. [* Moneymaker has made "only" about $110,000 per year from poker tournaments since his World Series win, before accounting for his substantia entry fees into tournaments.]
The viewer was led to believe that poker is easy to learn, easy to profit from, and incredibly action-packed-none of which true. But that didn't stop many of them from concluding that only a ticket to Las Vegas separated them from life as the next Chris Moneymaker. The number of participants in the World Series of Poker's $10,000 main event exploded, from 8 the year that Moneymaker won it to 8,773 just three years later.
I was one of those people.2 I lived the poker dream for a while, and then it died.

I lost the initial $25 fairly quickly, but the players in the Pacific Poker games did not seem much more sophisticated than the mix of ex-convicts and septuagenarians who populated the games at the Soaring Eagle. So I deposited $100 of my ow Almost all professional poker players begin their careers on winning streaks-the ones that lose at first are usually sensib enough to quit-and I was no exception. My bankroll began to grow, by $50 or $100 a night at first and them sometimes b $500 or $1,000. After about three months, my winnings hit $5,000; I began staying up all night to play, taking a cab to work at the crack of dawn and faking my way through the workday. After six months and $15,000 in winnings, I quit my job, leaving the exciting world of international tax consulting behind to split my time between playing cards and working for Baseball Prospectus. It was liberating; I felt as though I'd hacked the system somehow.

Indeed, information is so hard to come by in Texas hold 'em that players begin to make estimates about their opponents' range of hands even before any of the cards are dealt. In online games, this is often done through data mining: you'll hav statistics on how loose or tight, how passive or aggressive, each opponent's play has been in previous games. In brick-an mortar casinos, it is done through players' past histories with one another-or, failing that, through what amounts to ethni profiling. Players from Sweden, Lebanon, and China, for instance, have a reputation for being more aggressive than thos from France, England, or India. Younger players are presumed to be looser and more aggressive than older ones. Men ar assumed to be more likely to bluff than women. These stereotypes, like any others, are not always true: at the hold 'em games I used to play in at the Bellagio in Las Vegas, the best players were very often women, and they were good in part because they were much more aggressive than their opponents assumed. But poker players don't have the time for political correctness. Even if the stereotype that women play more conservatively than men is false 45 percent of the tim the fact that it might be true 55 percent of the time gives them something to work with.

Dwan was once better known by his online screen name "durrrr," which he selected because he figured it would put the other players on tilt if they lost to him.

I mostly played limit hold 'em instead, where the betting increment is fixed on each round. (Until very recently, this was the most popular game outside of tournaments; ten years ago, there were often no more than two or three no-limit game running anywhere in the United States.15) Limit poker offers fewer opportunities for creativity. Still, until practice caug up with theory, I had a couple of very successful years by exploiting an aggressive approach. In both 2004 and 2005, I made an income from poker in the six figures, with my cumulative profits from the game peaking at about $400,000 overall.

The Pareto Principle of Prediction implies that the worst forecasters-those who aren't getting even the first 20 percent right-are much worse than the best forecasters are good. Put another way, average forecasters are closer to the top than to the bottom of the pool. I'm sure that I'd lose a ton of money if I played poker against Dwan. But I'd gladly play him if, a

part of the deal, I were also guaranteed a match for the same stakes against some random person I picked off the street, against whom I'd expect to make back my losses and then some.
We can test this hypothesis empirically by examining the statistical records of poker players. I evaluated the data from an online poker site, which consisted of a random sampling of no-limit hold 'em players over a period in 2008 and 2009. The statistics told me how much money the players won or lost per hand, relative to the stakes they were playing.17
Because near-term wins and losses are very much subject to luck, I applied a statistical procedure18 to estimate what the players' true long-term profitability was. I then ordered the players by their skill level and broke them down into ten equal size quadrants. The top quadrant-consisting of the top 10 percent of the player pool*-corresponds to the best player at a typical ten-person table.19 The bottom 10 percent, meanwhile, are the biggest fish.
Figure 10-8a represents my estimate of how skilled the players in each quadrant really are, measured as money won or lost per one hundred hands in a no-limit hold 'em game with $5/$10 blinds. The figures include both money won and lost to the other players and that lost to the casino, which either takes a small percentage of each pot (known as the rake) or charges an hourly fee for dealing the game.20
I estimate that the very best player at the table in one of these games is averaging a profit of about $110 per one hundred hands played over the long run. That's a nice wage in an online casino, where hands are dealt very quickly and you could get almost that many hands during an hour or two.* It's less attractive in a traditional casino, where it might take four hours to play the same number of hands, and translates to wage of $25 or $30 per hour.
The key insight, however, is that the worst players at the table are losing money much faster than even the best ones are making it. For instance, I estimate that the worst player in the game-the biggest fish-was losing at a rate of more than $400 per one hundred hands. This player is so poor that he would literally be better off folding every hand, which would cost him only $150 per one hundred hands instead.
...In the game I just described, the one fish was feeding a lot of hungry mouths. His presence was worth about $40 per 100 hands to the other players. That subsidy was enough that about half of them were making money, even after the house's cut. Poker abides by a "trickle up" theory of wealth: the bottom 10 percent of players are losing money quickly enough to support a relatively large middle class of break-even players.
But what happens when the fish-the sucker-busts out, as someone losing money at this rate is bound to do? Several of the marginally winning players turn into marginally losing ones (figure 10-8b). In fact, we now estimate that only the very best player at the table is still making money over the long run, and then less than he did before.
FIGURE 10-8B: ESTIMATED MONEY WON OR LOST PER 100 HANDS IN A $5/$10 NO-LIMIT HOLD 'EM GAME AFTER FISH BUSTS OUT
What's more, the subtraction of the fish from the table can have a cascading effect on the other players. The one who was formerly the next-to-worst player is now the sucker, and will be losing money at an even faster rate than before. So he may bust out too, in turn making the remaining players' task yet more challenging. The entire equilibrium of the poker ecosystem can be thrown out of balance.
How, in fact, do poker games sustain themselves if the worst players are a constant threat to go broke? Sometimes there are fishy players with bottomless pockets: PokerKingBlog.com has alleged that Guy Laliberté, the CEO of Cirque du Soleil lost as much as $17 million in online poker games in 2008,22 where he sought to compete in the toughest high-stakes games against opponents like Dwan. Whatever the number, Laliberté is a billionaire who was playing the game for the intellectual challenge and to him this was almost nothing, the equivalent of the average American losing a few hundred bucks at blackjack.
Much more commonly, the answer is that there is not just one fishy player who loses money in perpetuity but a steady stream of them who take their turn in the barrel, losing a few hundred or a few thousand dollars and then quitting. At a brick-and-mortar casino like the Bellagio, these players might wander in from the craps table, or from one of its nightclubs, or after going on a winning streak in a tournament or a smaller-stakes game.

Once Party Poker shut Americans out, however, and I shifted my play to tougher sites like PokerStars, I found that I wasn't winning anymore. In fact, I was losing-a lot: about $75,000 during the last few months of 2006, most of it in one horrible evening. I played through the first several months of 2007 and continued to lose-another $60,000 or so. At that point, no longer confident that I could beat the games, I cashed out the rest of my money and quit.
My conclusion at the time was that the composition of the player pool had changed dramatically. Many of the professional players, reliant on the game for income, had soldiered on and kept playing, but most of the amateurs withdrew their funds or went broke. The fragile ecology of the poker economy was turned upside down-without those weak players to prop the game up, the water level had risen, and some of the sharks turned into suckers.26
Meanwhile, even before the new law passed, my play had begun to deteriorate, or at least cease to improve. I had hit a wall, playing uncreative and uninspired poker. When I did play, I combined the most dangerous trait of the professional player-the sense that I was entitled to win money-with the bad habits of the amateur, playing late into the evening, sometimes after having been out with friends.
In retrospect, things worked out pretty fortunately for me. The extra time I had on my hands-and my increased interest in the political process following the passage of the UIGEA-eventually led to the development of FiveThirtyEight. And while it wasn't fun to lose a third of my winnings, it was better than losing all of them. Some players who continued in the game were not so lucky. In 2011, the "Black Friday" indictments filed by the Department of Justice shut down many of the online poker sites for good,27 some of which proved to be insolvent and did not let players cash out their bankrolls.
I've sometimes wondered what would have happened if I'd played on. Poker is so volatile that it's possible for a theoretically winning player to have a losing streak that persists for months, or even for a full year. The flip side of this is that it's possible for a losing player to go on a long winning streak before he realizes that he isn't much good.
...What this means is that even after literally tens of thousands of hands are played, a good player might wind up behind and a bad one might wind up ahead. In figure 10-11, I've modeled the potential profits and losses for a player with the statistics I just described. The bands in the chart show the plausible range of wins and losses for the player, enough to cover 95 percent of all possible cases. After he plays 60,000 hands-about as many as he'd get in if he played forty hours a week in a casino every week for a full year-the player could plausibly have made $275,000 or have lost $35,000. In essence, this player could go to work every day for a year and still lose money. This is why it is sometimes said that poker is a hard way to make an easy living.

...The Bayesian method described in the book The Mathematics of Poker, for instance, would suggest that a player who had made $30,000 in his first 10,000 hands at a $100/$200 limit hold 'em game was nevertheless more likely than not to be a long-term loser.

...Another player, Darse Billings, who developed a computer program that competed successfully33 against some of the world's best limit hold 'em players,* put it even more bluntly. "There is no other game that I know of where humans are smug, and think that they just play like wizards, and then play so badly," he told me. "Basically it's because they don't know anything, and they think they must be God-like, and the truth is that they aren't. If computer programs feed on human hubris, then in poker they will eat like kings."

Nate Silver 2012, _The Signal and the Noise_

ch8

Consider a somber example: the September 11 attacks. Most of us would have assigned almost no probability to terrorist crashing planes into buildings in Manhattan when we woke up that morning. But we recognized that a terror attack was an obvious possibility once the first plane hit the World Trade Center. And we had no doubt we were being attacked once the second tower was hit. Bayes's theorem can replicate this result.
For instance, say that before the first plane hit, our estimate of the possibility of a terror attack on tall buildings in Manhattan was just 1 chance in 20,000, or 0.005 percent. However, we would also have assigned a very low probability t a plane hitting the World Trade Center by accident. This figure can actually be estimated empirically: in the previous 25,000 days of aviation over Manhattan39 prior to September 11, there had been two such accidents: one involving the Empire State Building in 1945 and another at 40 Wall Street in 1946. That would make the possibility of such an acciden about 1 chance in 12,500 on any given day. If you use Bayes's theorem to run these numbers (figure 8-5a), the probabilit we'd assign to a terror attack increased from 0.005 percent to 38 percent the moment that the first plane hit...And if you go through the calculation again, to reflect the second plane hitting the World Trade Center, the probability that we wer under attack becomes a near-certainty-99.99 percent. One accident on a bright sunny day in New York was unlikely enough, but a second one was almost a literal impossibility, as we all horribly deduced.

"In the last twenty years, with the exponential growth in the availability of information, genomics, and other technologie we can measure millions and millions of potentially interesting variables," Ioannidis told me. "The expectation is that we can use that information to make predictions work for us. I'm not saying that we haven't made any progress. Taking into account that there are a couple of million papers, it would be a shame if there wasn't. But there are obviously not a coupl of million discoveries. Most are not really contributing much to generating knowledge."
This is why our predictions may be more prone to failure in the era of Big Data. As there is an exponential increase in the amount of available information, there is likewise an exponential increase in the number of hypotheses to investigate. Fo instance, the U.S. government now publishes data on about 45,000 economic statistics. If you want to test for relationshi between all combinations of two pairs of these statistics-is there a causal relationship between the bank prime loan rate and the unemployment rate in Alabama?-that gives you literally one billion hypotheses to test.*

Even in the context of political polling, however, sampling error does not always tell the whole story. In the brief interval between the Iowa Democratic caucus and New Hampshire Democratic Primary in 2008, about 15,000 people were surveyed48 in New Hampshire-an enormous number in a small state, enough that the margin of error on the polls was theoretically just plus-or-minus 0.8 percent. The actual error in the polls was about ten times that, however: Hillary Clinton won the state by three points when the polls had her losing to Barack Obama by eight. Sampling error-the only type of error that frequentist statistics directly account for-was the least of the problem in the case of the New Hampshir polls.
Likewise, some polling firms consistently show a bias toward one or another party:49 they could survey all 200 million American adults and they still wouldn't get the numbers right. Bayes had these problems figured out 250 years ago. If you're using a biased instrument, it doesn't matter how many measurements you take-you're aiming at the wrong target.

Voulgaris soaks up as much basketball information as possible because everything could potentially shift his probability estimates. A professional sports bettor like Voulgaris might place a bet only when he thinks he has at least a 54 percent chance of winning it. This is just enough to cover the "vigorish" (the cut a sportsbook takes on a winning wager), plus the risk associated with putting one's money into play. And for all his skill and hard work-Voulgaris is among the best sports bettors in the world today-he still gets only about 57 percent of his bets right. It is just exceptionally difficult to do much better than that.

As an empirical matter, we all have beliefs and biases, forged from some combination of our experiences, our values, our knowledge, and perhaps our political or professional agenda. One of the nice characteristics of the Bayesian perspective that, in explicitly acknowledging that we have prior beliefs that affect how we interpret new evidence, it provides for a very good description of how we react to the changes in our world. For instance, if Fisher's prior belief was that there wa just a 0.00001 percent chance that cigarettes cause lung cancer, that helps explain why all the evidence to the contrary couldn't convince him otherwise. In fact, there is nothing prohibiting you under Bayes's theorem from holding beliefs tha you believe to be absolutely true. If you hold there is a 100 percent probability that God exists, or a 0 percent probability then under Bayes's theorem, no amount of evidence could persuade you otherwise.
I'm not here to tell you whether there are things you should believe with absolute and unequivocal certainty or not.* But perhaps we should be more honest about declaiming these. Absolutely nothing useful is realized when one person who holds that there is a 0 percent probability of something argues against another person who holds that the probability is 100 percent. Many wars-like the sectarian wars in Europe in the early days of the printing press-probably result from something like this premise.

ch9

Moreover, because the chess opening moves are more routine to players than positions they may encounter later on, humans can rely on centuries' worth of experience to pick the best moves. Although there are theoretically twenty moves that white might play to open the game, more than 98 percent of competitive chess games begin with one of the best four.19

Kasparov's goal, therefore, in his first game of his six-game match against Deep Blue in 1997, was to take the program ou of database-land and make it fly blind again. The opening move he played was fairly common; he moved his knight to the square of the board that players know as f3. Deep Blue responded on its second move by advancing its bishop to threaten Kasparov's knight-undoubtedly because its databases showed that such a move had historically reduced white's winning percentage* from 56 percent to 51 percent.

Those databases relied on the assumption, however, that Kasparov would respond as almost all other players had when faced with the position,22 by moving his knight back out of the way. Instead, he ignored the threat, figuring that Deep Blue was bluffing,23 and chose instead to move one of his pawns to pave the way for his bishop to control the center of the board.

Kasparov's move, while sound strategically, also accomplished another objective. He had made just three moves and Deep Blue had made just two, and yet the position they had now achieved (illustrated in figure 9-2) had literally occurred just once before in master-level competition24 out of the hundreds of thousands of games in Deep Blue's database.

In the final stage of a chess game, the endgame, the number of pieces on the board are fewer, and winning combinations are sometimes more explicitly calculable. Still, this phase of the game necessitates a lot of precision, since closing out a narrowly winning position often requires dozens of moves to be executed properly without any mistakes. To take an extreme case, the position illustrated in figure 9-4 has been shown to be a winning one for white no matter what black does, but it requires white to execute literally 262 consecutive moves correctly...However, just as chess computers have databases to cover the opening moves, they also have databases of these endgame scenarios. Literally all positions in which there are six or fewer pieces on the board have been solved to completion. Work on seven-piece positions is mostly complete-some of the solutions are intricate enough to require as many as 517 moves-but computers have memorized exactly which are the winning, losing, and drawing ones.

Nevertheless, there were some bugs in Deep Blue's inventory: not many, but a few. Toward the end of my interview with him, Campbell somewhat mischievously referred to an incident that had occurred toward the end of the first game in the 1997 match with Kasparov.

"A bug occurred in the game and it may have made Kasparov misunderstand the capabilities of Deep Blue," Campbell told me. "He didn't come up with the theory that the move that it played was a bug."

The bug had arisen on the forty-fourth move of their first game against Kasparov; unable to select a move, the program had defaulted to a last-resort fail-safe in which it picked a play completely at random. The bug had been inconsequential, coming late in the game in a position that had already been lost; Campbell and team repaired it the next day. "We had seen it once before, in a test game played earlier in 1997, and thought that it was fixed," he told me. "Unfortunately there was one case that we had missed."

In fact, the bug was anything but unfortunate for Deep Blue: it was likely what allowed the computer to beat Kasparov. In the popular recounting of Kasparov's match against Deep Blue, it was the second game in which his problems originated when he had made the almost unprecedented error of forfeiting a position that he could probably have drawn. But what had inspired Kasparov to commit this mistake? His anxiety over Deep Blue's forty-fourth move in the first game-the move in which the computer had moved its rook for no apparent purpose. Kasparov had concluded that the counterintuitive play must be a sign of superior intelligence. He had never considered that it was simply a bug.

Nate Silver 2012, _The Signal and the Noise_

ch7

On October 11, a report surfaced from Pittsburgh that three senior citizens had died shortly after receiving their flu shot so had two elderly persons in Oklahoma City; so had another in Fort Lauderdale.18 There was no evidence that any of th deaths were linked to the vaccinations-elderly people die every day, after all.19 But between the anxiety about the government's vaccination program and the media's dubious understanding of statistics,20 every death of someone who'd gotten a flu shot become a cause for alarm. Even Walter Cronkite, the most trusted man in America-who had broken fron his trademark austerity to admonish the media for its sensational handling of the story-could not calm the public down. Pittsburgh and many other cities shuttered their clinics.21
By late fall, another problem had emerged, this one far more serious. About five hundred patients, after receiving their shots, had begun to exhibit the symptoms of a rare neurological condition known as Guillain-Barré syndrome, an autoimmune disorder that can cause paralysis. This time, the statistical evidence was far more convincing: the usual incidence of Guillain-Barré in the general population is only about one case per million persons.22 In contrast, the rate in the vaccinated population had been ten times that-five hundred cases out of the roughly fifty million people who had bee administered the vaccine. Although scientists weren't positive why the vaccines were causing Guillain-Barré, manufacturing defects triggered by the rush production schedule were a plausible culprit,23 and the consensus of the medical community24 was that the vaccine program should be shut down for good, which the government finally did on December 16.
In the end, the outbreak of H1N1 at Fort Dix had been completely isolated; there was never another confirmed case anywhere in the country.25 Meanwhile, flu deaths from the ordinary A/Victoria strain were slightly below average in the winter of 1976-77.26 It had been much ado about nothing.
The swine flu fiasco-as it was soon dubbed-was a disaster on every level for President Ford, who lost his bid for another term to the Democrat Jimmy Carter that November.27 The drug makers had been absolved of any legal responsibility, leaving more than $2.6 billion in liability claims28 against the United States government. It seemed like every local pape had run a story about the poor waitress or schoolteacher who had done her duty and gotten the vaccine, only to have contracted Guillain-Barré. Within a couple of years, the number of Americans willing to take flu shots dwindled to only about one million,29 potentially putting the nation in grave danger had a severe strain hit in 1978 or 1979.30
Ford's handling of H1N1 was irresponsible on a number of levels. By invoking the likelihood of a 1918-type pandemic, he had gone against the advice of medical experts, who believed at the time that the chance of such a worst-case outcome was no higher than 35 percent and perhaps as low as 2 percent.31

The controversial 1968 book The Population Bomb, by the Stanford biologist Paul R. Ehrlich and his wife, Anne Ehrlich, made the opposite mistake, quite wrongly predicting that hundreds of millions of people would die from starvation in the 1970s.49 The reasons for this failure of prediction were myriad, including the Ehrlichs' tendency to focus on doomsday scenarios to draw attention to their cause. But one major problem was that they had assumed the record-high fertility rates in the free-love era of the 1960s would continue on indefinitely, meaning that there would be more and more hungr mouths to feed.* "When I wrote The Population Bomb I thought our interests in sex and children were so strong that it would be hard to change family size," Paul Ehrlich told me in a brief interview. "We found out that if you treat women decently and give them job opportunities, the fertility rate goes down." Other scholars who had not made such simplistic assumptions realized this at the time; population projections issued by the United Nations in the 1960s and 1970s generally did a good job of predicting what the population would look like thirty or forty years later.50
Extrapolation tends to cause its greatest problems in fields-including population growth and disease-where the quantity that you want to study is growing exponentially. In the early 1980s, the cumulative number of AIDS cases diagnosed in th United States was increasing in this exponential fashion:51 there were 99 cases through 1980, then 434 through 1981, and eventually 11,148 through 1984. You can put these figures into a chart, as some scholars did at the time,52 and seek to extrapolate the pattern forward. Doing so would have yielded a prediction that the number of AIDS cases diagnosed in the United States would rise to about 270,000 by 1995. This would not have been a very good prediction; unfortunately i was too low. The actual number of AIDS cases was about 560,000 by 1995, more than twice as high.
Perhaps the bigger problem from a statistical standpoint, however, is that precise predictions aren't really possible to begin with when you are extrapolating on an exponential scale. A properly applied version53 of this method, which accounted for its margin of error, would have implied that there could be as few as 35,000 AIDS cases through 1995 or a many as 1.8 million. That's much too broad a range to provide for much in the way of predictive insight. [53. The version applied here was to log-transform both the year variable and the AIDS-cases variable, then calculate the exponent via regression analysis. The 95 percent confidence interval on the exponent ran from about 2.2 to 3.7 by this method, with a most likely value of about 2.9. When applied ten years into the future, those relatively modest-seeming differences turn into an exceptionally broad range of possible outcomes.]

There are two major north-to-south routes through Manhattan: the West Side Highway, which borders the Hudson River, and the FDR Drive, which is on Manhattan's east side. Depending on her destination, a driver may not strongly prefer either thoroughfare. However, her GPS system will tell her which one to take, depending on which has less traffic-it is predicting which route will make for the shorter commute. The problem comes when a lot of other drivers are using the same navigation systems-all of a sudden, the route will be flooded with traffic and the "faster" route will turn out to be th slower one. There is already some theoretical66 and empirical67 evidence that this has become a problem on certain commonly used routes in New York, Boston, and London, and that these systems can sometimes be counterproductive.

The late 1990s and early 2000s were accompanied by a marked rise in unprotected sex in San Francisco's gay community,70 which had been devastated by the HIV/AIDS pandemic two decades earlier. Some researchers blamed this on increasing rates of drug use, particularly crystal methamphetamine, which is often associated with riskier sexual behavior. Others cited the increasing effectiveness of antiretroviral therapy-cocktails of medicine that can extend the live of HIV-positive patients for years or decades: gay men no longer saw an HIV diagnosis as a death sentence. Yet other theories focused on generational patterns-the San Francisco of the 1980s, when the AIDS epidemic was at its peak, was starting to feel like ancient history to a younger generation of gay men.71

The one thing the experts agreed on was that as unprotected sex increased, HIV infection rates were liable to do so as well.72

But that did not happen. Other STDs did increase: the number of new syphilis diagnoses among men who have sex with men (MSM)73-which had been virtually eradicated from San Francisco in the 1990s-rose substantially, to 502 cases in 2004 from 9 in 1998.74 Rates of gonorrhea also increased. Paradoxically, however, the number of new HIV cases did not rise. In 2004, when syphilis reached its highest level in years, the number of HIV diagnoses fell to their lowest figure sinc the start of the AIDS epidemic. This made very little sense to researchers; syphilis and HIV are normally strongly correlated statistically, and they also have a causal relationship, since having one disease can make you more vulnerable to acquiring the other one.75

The solution to the paradox, it now appears, is that gay men had become increasingly effective at "serosorting"-that is, they were choosing sex partners with the same HIV status that they had. How they were able to accomplish this is a subject of some debate, but it has been documented by detailed behavioral studies in San Francisco,76 Sydney,77 Londo and other cities with large gay populations. It may be that public health campaigns-some of which, wary of "condom fatigue," instead focused on the notion of "negotiated safety"-were having some positive effect. It may be that the Interne which to some extent has displaced the gay bar as the preferred place to pick up a sex partner, has different norms for disclosure: many men list their HIV status in their profiles, and it may be easier to ask tough questions (and to get hones responses) from the privacy of one's home than in the din of the dance hall.78

Nate Silver 2012, _The Signal and the Noise_

ch6

In April 1997, the Red River of the North flooded Grand Forks, North Dakota, overtopping the town's levees and spilling more than two miles into the city.*4 Although there was no loss of life, nearly all of the city's 50,000 residents had to be evacuated, cleanup costs ran into the billions of dollars,5 and 75 percent of the city's homes were damaged or destroyed. Unlike a hurricane or an earthquake, the Grand Forks flood may have been a preventable disaster. The city's floodwalls could have been reinforced using sandbags.7 It might also have been possible to divert the overflow into depopulated areas-into farmland instead of schools, churches, and homes.
Residents of Grand Forks had been aware of the flood threat for months. Snowfall had been especially heavy in the Great Plains that winter, and the National Weather Service, anticipating runoff as the snow melted, had predicted the waters o the Red River would crest to forty-nine feet, close to the all-time record.
There was just one small problem. The levees in Grand Forks had been built to handle a flood of fifty-one feet. Even a small miss in the forty-nine-foot prediction could prove catastrophic.
In fact, the river crested to fifty-four feet. The Weather Service's forecast hadn't been perfect by any means, but a five-fo miss, two months in advance of a flood, is pretty reasonable-about as well as these predictions had done on average historically. The margin of error on the Weather Service's forecast-based on how well their flood forecasts had done in th past-was about plus or minus nine feet. That implied about a 35 percent chance of the levees being overtopped.8
...Left to their own devices, many residents became convinced they didn't have anything to worry about. (Very few of the bought flood insurance.10) A prediction of a forty-nine-foot crest in the river, expressed without any reservation, seemed to imply that the flood would hit forty-nine feet exactly; the fifty-one-foot levees would be just enough to keep them safe. Some residents even interpreted the forecast of forty-nine feet as representing the maximum possible extent of the flood.11
An oft-told joke: a statistician drowned crossing a river that was only three feet deep on average.

As I mentioned, the economists in this survey thought that GDP would end up at about 2.4 percent in 2008, slightly belov its long-term trend. This was a very bad forecast: GDP actually shrank by 3.3 percent once the financial crisis hit. What may be worse is that the economists were extremely confident in their bad prediction. They assigned only a 3 percent chance to the economy's shrinking by any margin over the whole of 2008.15 And they gave it only about a 1-in-500 chanc of shrinking by at least 2 percent, as it did.16
Indeed, economists have for a long time been much too confident in their ability to predict the direction of the economy. figure 6-4, I've plotted the forecasts of GDP growth from the Survey of Professional Forecasters for the eighteen years between 1993 and 2010.17 The bars in the chart represent the 90 percent prediction intervals as stated by the economists.
A prediction interval is a range of the most likely outcomes that a forecast provides for, much like the margin of error in poll. A 90 percent prediction interval, for instance, is supposed to cover 90 percent of the possible real-world outcomes, leaving only the 10 percent of outlying cases at the tail ends of the distribution. If the economists' forecasts were as accurate as they claimed, we'd expect the actual value for GDP to fall within their prediction interval nine times out of te or all but about twice in eighteen years.
In fact, the actual value for GDP fell outside the economists' prediction interval six times in eighteen years, or fully one-third of the time. Another study,18 which ran these numbers back to the beginnings of the Survey of Professional Forecasters in 1968, found even worse results: the actual figure for GDP fell outside the prediction interval almost half th time. There is almost no chance19 that the economists have simply been unlucky; they fundamentally overstate the reliability of their predictions.
In reality, when a group of economists give you their GDP forecast, the true 90 percent prediction interval-based on how these forecasts have actually performed20 and not on how accurate the economists claim them to be-spans about 6.4 points of GDP (equivalent to a margin of error of plus or minus 3.2 percent).*
When you hear on the news that GDP will grow by 2.5 percent next year, that means it could quite easily grow at a spectacular rate of 5.7 percent instead. Or it could fall by 0.7 percent-a fairly serious recession. Economists haven't been able to do any better than that, and there isn't much evidence that their forecasts are improving. The old joke about economists' having called nine out of the last six recessions correctly has some truth to it; one actual statistic is that in th 1990s, economists predicted only 2 of the 60 recessions around the world a year ahead of time.21

The government produces data on literally 45,000 economic indicators each year.24 Private data providers track as many as four million statistics.25 The temptation that some economists succumb to is to put all this data into a blender and claim that the resulting gruel is haute cuisine. There have been only eleven recessions since the end of World War II.26 I you have a statistical model that seeks to explain eleven outputs but has to choose from among four million inputs to do s many of the relationships it identifies are going to be spurious. (This is another classic case of overfitting-mistaking noise for a signal-the problem that befell earthquake forecasters in chapter 5.)
Consider how creative you might be when you have a stack of economic variables as thick as a phone book. A once-famou "leading indicator" of economic performance, for instance, was the winner of the Super Bowl. From Super Bowl I in 1967 through Super Bowl XXXI in 1997, the stock market27 gained an average of 14 percent for the rest of the year when a team from the original National Football League (NFL) won the game.28 But it fell by almost 10 percent when a team fro the original American Football League (AFL) won instead.
Through 1997, this indicator had correctly "predicted" the direction of the stock market in twenty-eight of thirty-one year A standard test of statistical significance,29 if taken literally, would have implied that there was only about a 1-in-4,700,000 possibility that the relationship had emerged from chance alone.
It was just a coincidence, of course. And eventually, the indicator began to perform badly. In 1998, the Denver Broncos, a original AFL team, won the Super Bowl-supposedly a bad omen. But rather than falling, the stock market gained 28 percent amid the dot-com boom. In 2008, the NFL's New York Giants came from behind to upset the AFL's New England

Patriots on David Tyree's spectacular catch-but Tyree couldn't prevent the collapse of the housing bubble, which caused the market to crash by 35 percent. Since 1998, in fact, the stock market has done about 10 percent better when the AFL team won the Super Bowl, exactly the opposite of what the indicator was fabled to predict.

How does an indicator that supposedly had just a 1-in-4,700,000 chance of failing flop so badly? For the same reason tha even though the odds of winning the Powerball lottery are only 1 chance in 195 million,30 somebody wins it every few weeks. The odds are hugely against any one person winning the lottery-but millions of tickets are bought, so somebody is going to get lucky. Likewise, of the millions of statistical indicators in the world, a few will have happened to correlate especially well with stock prices or GDP or the unemployment rate. If not the winner of the Super Bowl, it might be chicken production in Uganda. But the relationship is merely coincidental.

...It's much harder to find something that identifies the signal; variables that are leading indicators in one economic cycle often turn out to be lagging ones in the next. Of the seven so-called leading indicators in a 2003 Inc. magazine article,33 all of which had been good predictors of the 1990 and 2001 recessions, only two-housing prices and temporary hiring-led the recession that began in 2007 to any appreciable degree. Others, like commercial lending, did not begin to turn downward until a year after the recession began.

Even the well-regarded Leading Economic Index, a composite of ten economic indicators published by the Conference Board, has had its share of problems. The Leading Economic Index has generally declined a couple of months in advance of recessions. But it has given roughly as many false alarms-including most infamously in 1984, when it sharply declined for three straight months,34 signaling a recession, but the economy continued to zoom upward at a 6 percent rate of growth. Some studies have even claimed that the Leading Economic Index has no predictive power at all when applied in real time.35

Historically, for instance, there has been a reasonably strong correlation between GDP growth and job growth. Economis refer to this as Okun's law. During the Long Boom of 1947 through 1999, the rate of job growth40 had normally been about half the rate of GDP growth, so if GDP increased by 4 percent during a year, the number of jobs would increase by about 2 percent.

The relationship still exists-more growth is certainly better for job seekers. But its dynamics seem to have changed. After each of the last couple of recessions, considerably fewer jobs were created than would have been expected during the Long Boom years. In the year after the stimulus package was passed in 2009, for instance, GDP was growing fast enough to create about two million jobs according to Okun's law.41 Instead, an additional 3.5 million jobs were lost during the period.

Economists often debate about what the change means. The most pessimistic interpretation, advanced by economists including Jeffrey Sachs of Columbia University, is that the pattern reflects profound structural problems in the American economy: among them, increasing competition from other countries, an imbalance between the service and manufacturin sectors, an aging population, a declining middle class, and a rising national debt. Under this theory, we have entered a new and unhealthy normal, and the problems may get worse unless fundamental changes are made. "We were underestimating the role of global change in causing U.S. change," Sachs told me. "The loss of jobs internationally to China and emerging markets have really jolted the American economy."

The bigger question is whether the volatility of the 2000s is more representative of the long-run condition of the economy perhaps the long boom years had been the outlier. During the Long Boom, the economy was in recession only 15 percent of the time. But the rate was more than twice that-36 percent-from 1900 through 1945.42

"I think the most interesting question is how little effort we actually put into forecasting, even on the things we say are important to us," Robin Hanson told me as the food arrived.

"In an MBA school you present this image of a manager as a great decision maker-the scientific decision maker. He's got his spreadsheet and he's got his statistical tests and he's going to weigh the various options. But in fact real managemen is mostly about managing coalitions, maintaining support for a project so it doesn't evaporate. If they put together a coalition to do a project, and then at the last minute the forecasts fluctuate, you can't dump the project at the last minute right?

Even academics aren't very interested in collecting a track record of forecasts-they're not very interested in making clea enough forecasts to score," he says later. "What's in it for them? The more fundamental problem is that we have a demar for experts in our society but we don't actually have that much of a demand for accurate forecasts."

Nate Silver 2012, _The Signal and the Noise_

ch5

Earthquakes kill more people than hurricanes, in fact,16 despite seeming like the rarer phenomenon.17 Perhaps that is because they are so seldom predicted successfully. Whereas the landfall position of hurricanes can be forecasted at least three times more accurately now than they were even twenty-five years ago, the science of earthquake forecasting seems barely to have evolved since the ninth century A.D., when the Japanese first claimed to be able to anticipate earthquakes by looking at the behavior of catfish.18 (Cows, pigs, eels, rats, parakeets, seagulls, turtles, goldfish, and snakes have also been reported at various times to behave unusually in advance of an earthquake.)

We all know that California is very seismically active; the USGS estimates that an earthquake of magnitude 6.8 or higher will hit San Francisco about once every thirty-five years. Many of you will also know that Alaska has many earthquakes-th second largest one in recorded history, magnitude 9.4, hit Anchorage in 1964.
But did you know about Charleston, South Carolina? It is seismically active too; indeed, it experienced a magnitude 7.3 earthquake in 1886. The USGS estimates that there will be another big earthquake there about once per six hundred years. If you live in Seattle, you should probably have an earthquake plan ready; it is more earthquake-prone than many parts of California, the USGS says. But you don't need one if you live in Denver, which is a safe distance away from any continental boundaries.

If you compare the frequencies of earthquakes with their magnitudes, you'll find that the number drops off exponentially as the magnitude increases. While there are very few catastrophic earthquakes, there are literally millions of smaller one about 1.3 million earthquakes measuring between magnitude 2.0 and magnitude 2.9 around the world every year.27 Mos of these earthquakes go undetected-certainly by human beings and often by seismometers.28 However, almost all earthquakes of magnitude 4.5 or greater are recorded today, however remote their location. Figure 5-3a shows the exponential decline in their frequencies, based on actual records of earthquakes from January 196429 through March 2012.30

According to the power law that Gutenberg and Richter uncovered, that means that an earthquake measuring between 6 and 6.9 should occur about once every thirty years in Tehran.
Furthermore, it follows that an earthquake that measured 7.0 or greater would occur about once every three hundred years near Tehran. This is the earthquake that Susan Hough fears. The Haiti earthquake of 2010, which measured magnitude 7.0 and killed 316,000,32 showed the apocalyptic consequences that earthquakes can produce in the developing world. Iran shares many of Haiti's problems-poverty, lax building codes, political corruption33-but it is much more densely populated. The USGS estimates, on the basis of high death tolls from smaller earthquakes in Iran, that between 15 and 30 percent of Tehran's population could die in the event of a catastrophic tremor there.34 Since there ar about thirteen million people in Tehran's metro area, that would mean between two and four million fatalities.

Large earthquakes are almost always followed by dozens or even thousands of aftershocks (the 2011 earthquake in Japan produced at least 1,200 of them). These aftershocks follow a somewhat predictable pattern.35 Aftershocks are more likel to occur immediately after an earthquake than days later, and more likely to occur days later than weeks after the fact. This, however, is not terribly helpful when it comes to saving lives. This is because aftershocks, by definition, are always less powerful than the initial earthquake. Usually, if a particular fault produces a sufficiently powerful earthquake, there will be a few aftershocks and then that'll be the end of the fireworks for a while. This isn't always the case, however. For example, the incredibly powerful earthquake that hit the New Madrid Fault on the Missouri-Tennessee border on December 16, 1811, evaluated by seismologists as magnitude 8.2, was followed just six hours later by another shock of about the same magnitude. And the fault was not yet quiesced: the December 16 quakes were succeeded by another magnitude 8.1 earthquake on January 23, and then yet another, even more powerful 8.3 earthquake on February 7. Whic ones were the foreshocks? Which ones were the aftershocks? Any interpretation is about as useless as any other.

One of the more infamous cases involved a geophysicist named Brian Brady, who had a Ph.D. from MIT and worked at Colorado School of Mines. Brady asserted that a magnitude 9.2 earthquake-one of the largest in recorded history-would l Lima, Peru, in 1981.40 His prediction initially had a fair amount of support in the seismological community-an early version of it had been coauthored with a USGS scientist. But as the theory became more elaborate-Brady would eventual invoke everything from the rock bursts he had observed in his studies of mines to Einstein's theory of relativity in suppor of it-colleagues had started telling him that theory was beyond their understanding:41 a polite way of saying that he was nuts. Eventually, he predicted that the magnitude 9.2 earthquake would be just one in a spectacular series in Peru, culminating in a magnitude 9.9 earthquake, the largest in recorded history, in August 1981.42
The prediction was leaked to the Peruvian media and terrified the population; this serious-seeming American scientist wa sure their capital city would be in ruins. Their fear only intensified when it was reported that the Peruvian Red Cross had requested 100,000 body bags to prepare for the disaster. Tourism and property values declined,43 and the U.S. government eventually dispatched a team of scientists and diplomats to Peru in an effort to calm nerves. It made front-page news when there was no Great Peruvian Earthquake in 1981 (or even a minor one).

In figure 5-7a, I've plotted the historical frequencies of earthquakes near the 2011 epicenter in Japan.63 The data include everything up through but not including the magnitude 9.1 earthquake on March 11. You'll see that the relationship almo follows the straight-line pattern that Gutenberg and Richter's method predicts. However, at about magnitude 7.5, there i a kink in the graph. There had been no earthquakes as large as a magnitude 8.0 in the region since 1964, and so the curv seems to bend down accordingly.
So how to connect the dots? If you go strictly by the Gutenberg-Richter law, ignoring the kink in the graph, you should st follow the straight line, as in figure 5-7b. Alternatively, you could go by what seismologists call a characteristic fit (figure

5-7c), which just means that it is descriptive of the historical frequencies of the earthquake in that area. In this case, that would mean that you took the kink in the historical data to be real-meaning, you thought there was some good reason wh earthquakes larger than about magnitude 7.6 were unlikely to occur in the region.

Here is another example where an innocuous-seeming choice of assumptions will yield radically distinct conclusions-in th case, about the probability of a magnitude 9 earthquake in this part of Japan. The characteristic fit suggests that such an earthquake was nearly impossible-it implies that one might occur about every 13,000 years. The Gutenberg-Richter estimate, on the other hand, was that you'd get one such earthquake every three hundred years. That's infrequent but hardly impossible-a tangible enough risk that a wealthy nation like Japan might be able to prepare for it.64

The characteristic fit matched the recent historical record from a bit more snugly. But as we've learned, this type of pattern-matching is not always a good thing-it could imply an overfit model, in which case it will do a worse job of matching the true relationship. In this case, an overfit model would dramatically underestimate the likelihood of a catastrophic earthquake in the area. The problem with the characteristic fit is that it relied on an incredibly weak signal. As I mentioned, there had been no earthquake of magnitude 8 or higher in this region in the forty-five years or so prior to Tohoku. However, these are rare events to begin with: the Gutenberg-Richter law posits that they might occur only abou once per thirty years in this area. It's not very hard at all for a once-per-thirty-year event to fail to occur in a forty-five-ye window,65 no more so than a .300 hitter having a bad day at the plate and going 0-for-5.66 Meanwhile, there were quite few earthquakes with magnitudes in the mid- to high 7's in this part of Japan. When such earthquakes had occurred in other parts of the world, they had almost always suggested the potential for larger ones. What justification was there to think that Japan would be a special case?

Actually, seismologists in Japan and elsewhere came up with a few rationalizations for that. They suggested, for instance that the particular composition of the seafloor in the region, which is old and relatively cool and dense, might prohibit the formation of such large earthquakes.67 Some seismologists observed that, before 2004, no magnitude 9 earthquake had occurred in a region with that type of seafloor.

This was about like concluding that it was impossible for anyone from Pennsylvania to win the Powerball jackpot because no one had done so in the past three weeks. Magnitude 9 earthquakes, like lottery winners, are few and far between. Before 2004, in fact, only three of them had occurred in recorded history anywhere in the world. This wasn't nearly enough data to support such highly specific conclusions about the exact circumstances under which they might occur. No was Japan the first failure of such a theory; a similar one had been advanced about Sumatra68 at a time when it had experienced lots of magnitude 7 earthquakes69 but nothing stronger. Then the Great Sumatra Earthquake, magnitude 9.2,70 hit in December 2004.

The Gutenberg-Richter law would not have predicted the exact timing of the Sumatra or Japan earthquakes, but it would have allowed for their possibility.71 So far, it has held up remarkably well when a great many more elaborate attempts a earthquake prediction have failed.

Because they occur so rarely, it will take centuries to know what the true rate of magnitude 9 earthquakes is. It will take even longer to know whether earthquakes larger than magnitude 9.5 are possible. Hough told me that there may be som fundamental constraints on earthquake size from the geography of fault systems. If the largest continuous string of faults in the world ruptured together-everything from Tierra Del Fuego at the southern tip of South America all the way up through the Aleutians in Alaska-a magnitude 10 is about what you'd get, she said. But it is hard to know for sure.

Nate Silver, _The Signal and the Noise_

ch4

The public at large became more interested in weather forecasting after the Schoolhouse Blizzard of January 1888. On January 12 that year, initially a relatively warm day in the Great Plains, the temperature dropped almost 30 degrees in a matter of a few hours and a blinding snowstorm came.26 Hundreds of children, leaving school and caught unaware as th blizzard hit, died of hypothermia on their way home. As crude as early weather forecasts were, it was hoped that they might at least be able to provide some warning about an event so severe. So the National Weather Service was moved to the Department of Agriculture and took on a more civilian-facing mission.*

What is it, exactly, that humans can do better than computers that can crunch numbers at seventy-seven teraFLOPS? Th can see. Hoke led me onto the forecasting floor, which consisted of a series of workstations marked with blue overhangi signs with such legends as MARITIME FORECAST CENTER and NATIONAL CENTER. Each station was manned by one o two meterologists-accompanied by an armada of flat-screen monitors that displayed full-color maps of every conceivable type of weather data for every corner of the country. The forecasters worked quietly and quickly, with a certain amount o Grant's military precision.30
Some of the forecasters were drawing on these maps with what appeared to be a light pen, painstakingly adjusting the contours of temperature gradients produced by the computer models-fifteen miles westward over the Mississippi Delta, thirty miles northward into Lake Erie. Gradually, they were bringing them one step closer to the Platonic ideal they were hoping to represent.
The forecasters know the flaws in the computer models. These inevitably arise because-as a consequence of chaos theory even the most trivial bug in the model can have potentially profound effects. Perhaps the computer tends to be too conservative on forecasting nighttime rainfalls in Seattle when there's a low-pressure system in Puget Sound. Perhaps it doesn't know that the fog in Acadia National Park in Maine will clear up by sunrise if the wind is blowing in one direction but can linger until midmorning if it's coming from another. These are the sorts of distinctions that forecasters glean ove time as they learn to work around the flaws in the model, in the way that a skilled pool player can adjust to the dead spot on the table at his local bar.
...The NWS keeps two different sets of books: one that shows how well the computers are doing by themselves and anoth that accounts for how much value the humans are contributing. According to the agency's statistics, humans improve the accuracy of precipitation forecasts by about 25 percent over the computer guidance alone,31 and temperature forecasts by about 10 percent.32 Moreover, according to Hoke, these ratios have been relatively constant over time: as much progress as the computers have made, his forecasters continue to add value on top of it. Vision accounts for a lot.

When Hoke began his career, in the mid-'70s, the jokes about weather forecasters had some grounding in truth. On average, for instance, the NWS was missing the high temperature by about 6 degrees when trying to forecast it three day in advance (figure 4-4). That isn't much better than the accuracy you could get just by looking up a table of long-term averages. The partnership between man and machine is paying big dividends, however. Today, the average miss is about 3.5 degrees, meaning that almost half the inaccuracy has been stripped out.
Weather forecasters are also getting better at predicting severe weather. What are your odds of being struck-and killed-b lightning? Actually, this is not a constant number; they depend on how likely you are to be outdoors when lightning hits and unable to seek shelter in time because you didn't have a good forecast. In 1940, the chance of an American being killed by lightning in a given year was about 1 in 400,000.33 Today, it's just 1 chance in 11,000,000, making it almost thirty times less likely. Some of this reflects changes in living patterns (more of our work is done indoors now) and improvement in communications technology and medical care, but it's also because of better weather forecasts.
Perhaps the most impressive gains have been in hurricane forecasting. Just twenty-five years ago, when the National Hurricane Center tried to forecast where a hurricane would hit three days in advance of landfall, it missed by an average of 350 miles.34 That isn't very useful on a human scale. Draw a 350-mile radius outward from New Orleans, for instance, and it covers all points from Houston, Texas, to Tallahassee, Florida (figure 4-5). You can't evacuate an area that large. Today, however, the average miss is only about one hundred miles, enough to cover only southeastern Louisiana and the southern tip of Mississippi. The hurricane will still hit outside that circle some of the time, but now we are looking at a relatively small area in which an impact is even money or better-small enough that you could plausibly evacuate it sevent two hours in advance. In 1985, by contrast, it was not until twenty-four hours in advance of landfall that hurricane forecasts displayed the same skill. What this means is that we now have about forty-eight hours of additional warning tim before a storm hits-and as we will see later, every hour is critical when it comes to evacuating a city like New Orleans.*

What does bitterly cold mean? A chance of flurries? Just where is the dividing line between partly cloudy and mostly cloudy? The Weather Channel needs to figure this out, and it needs to establish formal rules for doing so, since it issues f too many forecasts for the verbiage to be determined on an ad hoc basis.
Sometimes the need to adapt the forecast to the consumer can take on comical dimensions. For many years, the Weather Channel had indicated rain on their radar maps with green shading (occasionally accompanied by yellow and red for severe storms). At some point in 2001, someone in the marketing department got the bright idea to make rain blue instead-which is, after all, what we think of as the color of water. The Weather Channel was quickly beseiged with phone calls from outraged-and occasionally terrified-consumers, some of whom mistook the blue blotches for some kind of heretofore unknown precipitation (plasma storms? radioactive fallout?). "That was a nuclear meltdown," Dr. Rose told me "Somebody wrote in and said, 'For years you've been telling us that rain is green-and now it's blue? What madness is this?'"

In 2002 an entrepeneur named Eric Floehr, a computer science graduate from Ohio State who was working for MCI, changed that. Floehr simply started collecting data on the forecasts issued by the NWS, the Weather Channel, and AccuWeather, to see if the government model or the private-sector forecasts were more accurate. This was mostly for his own edification at first-a sort of very large scale science fair project-but it quickly evolved into a profitable business, ForecastWatch.com, which repackages the data into highly customized reports for clients ranging from energy traders (f whom a fraction of a degree can translate into tens of thousands of dollars) to academics.

Floehr found that there wasn't any one clear overall winner. His data suggests that AccuWeather has the best precipitation forecasts by a small margin, that the Weather Channel has slightly better temperature forecasts, and the government's forecasts are solid all around. They're all pretty good.

But the further out in time these models go, the less accurate they turn out to be (figure 4-6). Forecasts made eight days advance, for example, demonstate almost no skill; they beat persistence but are barely better than climatology. And at intervals of nine or more days in advance, the professional forecasts were actually a bit worse than climatology. After a little more than a week, Loft told me, chaos theory completely takes over, and the dynamic memory of the atmopshere erases itself.

...Floehr's finding raises a couple of disturbing questions. It would be one thing if, after seven or eight days, the compute models demonstrated essentially zero skill. But instead, they actually display negative skill: they are worse than what you or I could do sitting around at home and looking up a table of long-term weather averages. How can this be? It is likely because the computer programs, which are hypersensitive to the naturally occurring feedbacks in the weather system, begin to produce feedbacks of their own. It's not merely that there is no longer a signal amid the noise, but that the noise is being amplified.

The bigger question is why, if these longer-term forecasts aren't any good, outlets like the Weather Channel (which publishes ten-day forecasts) and AccuWeather (which ups the ante and goes for fifteen) continue to produce them. Dr. Rose took the position that doing so doesn't really cause any harm; even a forecast based purely on climatology might be of some interest to their consumers.

The statistical reality of accuracy isn't necessarily the governing paradigm when it comes to commercial weather forecasting. It's more the perception of accuracy that adds value in the eyes of the consumer.

For instance, the for-profit weather forecasters rarely predict exactly a 50 percent chance of rain, which might seem wishy-washy and indecisive to consumers.41 Instead, they'll flip a coin and round up to 60, or down to 40, even though th makes the forecasts both less accurate and less honest.42

Floehr also uncovered a more flagrant example of fudging the numbers, something that may be the worst-kept secret in the weather industry. Most commercial weather forecasts are biased, and probably deliberately so. In particular, they ar biased toward forecasting more precipitation than will actually occur43-what meteorologists call a "wet bias." The furthe you get from the government's original data, and the more consumer facing the forecasts, the worse this bias becomes. Forecasts "add value" by subtracting accuracy.

...The National Weather Service's forecasts are, it turns out, admirably well calibrated46 (figure 4-7). When they say ther is a 20 percent chance of rain, it really does rain 20 percent of the time. They have been making good use of feedback, ar their forecasts are honest and accurate. The meteorologists at the Weather Channel will fudge a little bit under certain conditions. Historically, for instance, when they say there is a 20 percent chance of rain, it has actually only rained about percent of the time.47 In fact, this is deliberate and is something the Weather Channel is willing to admit to. It has to do with their economic incentives.

People notice one type of mistake-the failure to predict rain-more than another kind, false alarms. If it rains when it isn't supposed to, they curse the weatherman for ruining their picnic, whereas an unexpectedly sunny day is taken as a serendipitous bonus. It isn't good science, but as Dr. Rose at the Weather Channel acknolwedged to me: "If the forecast was objective, if it has zero bias in precipitation, we'd probably be in trouble."

Still, the Weather Channel is a relatively buttoned-down organization-many of their customers mistakenly think they are government agency-and they play it pretty straight most of the time. Their wet bias is limited to slightly exaggerating the probability of rain when it is unlikely to occur-saying there is a 20 percent chance when they know it is really a 5 or 10 percent chance-covering their butts in the case of an unexpected sprinkle. Otherwise, their forecasts are well calibrated (figure 4-8). When they say there is a 70 percent chance of rain, for instance, that number can be taken at face value.

...Kansas City ought to be a great market for weather forecasting-it has scorching-hot summers, cold winters, tornadoes, and droughts, and it is large enough to be represented by all the major networks. A man there named J. D. Eggleston began tracking local TV forecasts to help his daughter with a fifth-grade classroom project. Eggleston found the analysis so interesting that he continued it for seven months, posting the results to the Freakonomics blog.48

The TV meteorologists weren't placing much emphasis on accuracy. Instead, their forecasts were quite a bit worse than those issued by the National Weather Service, which they could have taken for free from the Internet and reported on th air. And they weren't remotely well calibrated. In Eggleston's study, when a Kansas City meteorologist said there was a 100 percent chance of rain, it failed to rain about one-third of the time (figure 4-9).

No people in New York City died from Hurricane Irene in 2011 despite massive media hype surrounding the storm, but three people did from flooding in landlocked Vermont52 once the TV cameras were turned off.

Evacuation decisions are not easy, in part because evacuations themselves can be deadly; a bus carrying hospital evacue from another 2005 storm, Hurricane Rita, burst into flames while leaving Houston, killing twenty-three elderly passengers.53

Studies from Katrina and other storms have found that having survived a hurricane makes one less likely to evacuate the next time one comes.57

Nate Silver 2012, _The Signal and the Noise_

ch3

Baseball, uniquely among the major American sports, has always been played on fields with nonstandard dimensions. It's much easier to put up a high batting average in snug and boxy Fenway Park, whose contours are shaped by compact New England street grids, than in the cavernous environs of Dodger Stadium, which is surrounded by a moat of parking lot. By observing how players perform both at home and on the road, we can develop "park factors" to account for the degree of difficulty that a player faces. (For example, Fred Lynn, an MVP with the Red Sox during the 1970s, hit .347 over the course of his career at Fenway Park but just .264 at every other stadium.) Likewise, by observing what happens to player who switch from the National League to the American League, we can tell quite a bit about which league is better and account for the strength of a player's competition.

Olympic gymnasts peak in their teens; poets in their twenties; chess players in their thirties11; applied economists in the forties,12 and the average age of a Fortune 500 CEO is 55.13 A baseball player, James found, peaks at age twenty-seven. Of the fifty MVP winners between 1985 and 2009, 60 percent were between the ages of twenty-five and twenty-nine, and 20 percent were aged twenty-seven exactly. This is when the combination of physical attributes and mental attributes needed to play the game well seem to be in the best balance.

The players in the PECOTA list had generated 546 wins for their major-league teams through 2011 (figure 3-3). But the players in Baseball America's list did better, producing 630 wins. Although the scouts' judgment is sometimes flawed, the were adding plenty of value: their forecasts were about 15 percent better than ones that relied on statistics alone. That might not sound like a big difference, but it really adds up. Baseball teams are willing to pay about $4 million per win on the free-agent market.30 The extra wins the scouts identified were thus worth a total of $336 million over this period.*
Although it would have been cool if the PECOTA list had gotten the better of the scouts, I didn't expect it to happen. As I wrote shortly after the lists were published:31
As much fun as it is to play up the scouts-versus-stats angle, I don't expect the PECOTA rankings to be as accurate as . the rankings you might get from Baseball America.The fuel of any ranking system is information-and being able to look both scouting and statistical information means that you have more fuel. The only way that a purely stat-based prospect list should be able to beat a hybrid list is if the biases introduced by the process are so strong that they overwhelm the benefit.
In other words, scouts use a hybrid approach. They have access to more information than statistics alone. Both the scout and PECOTA can look at what a player's batting average or ERA was; an unbiased system like PECOTA is probably a littl bit better at removing some of the noise from those numbers and placing them into context. Scouts, however, have acces to a lot of information that PECOTA has no idea about. Rather than having to infer how hard a pitcher throws from his strikeout total, for instance, they can take out their radar guns and time his fastball velocity. Or they can use their stopwatches to see how fast he runs the bases.
This type of information gets one step closer to the root causes of what we are trying to predict. In the minors, a pitcher with a weak fastball can rack up a lot of strikeouts just by finding the strike zone and mixing up his pitches; most of the hitters he is facing aren't much good, so he may as well challenge them. In the major leagues, where the batters are capable of hitting even a ninety-eight-mile-per-hour fastball out of the park, the odds are against the soft-tosser. PECOTA will be fooled by these false positives while a good scout will not be. Conversely, a scout may be able to identify players who have major-league talent but who have yet to harness it.

But statheads can have their biases too. One of the most pernicious ones is to assume that if something cannot easily be quantified, it does not matter. In baseball, for instance, defense has long been much harder to measure than batting or pitching. In the mid-1990s, Beane's Oakland A's teams placed little emphasis on defense, and their outfield was manned slow and bulky players, like Matt Stairs, who came out of the womb as designated hitters. As analysis of defense advance it became apparent that the A's defective defense was costing them as many as eight to ten wins per season,33 effectivel taking them out of contention no matter how good their batting statistics were. Beane got the memo, and his more recen and successful teams have had relatively good defenses.

Statistics, indeed, have been a part of the fabric of baseball since the very beginning. The first newspaper box score, whi included five categories of statistics for each player-runs, hits, putouts, assists, and errors-was published by Henry Chadwick in 1859,38 twelve years before the first professional league was established, in 1871. Many of the Moneyball-e debates concerned not whether statistics should be used, but which ones should be taken into account. On-base percentage (OBP), for instance, as analysts like James had been pointing out for years, is more highly correlated with scoring runs (and winning games) than batting average, a finding which long went underappreciated by traditionalists within the industry.39
...The further you get away from the majors-the more you are trying to predict a player's performance instead of measure it-the less useful statistics are. Statistics at the more advanced minor-league levels, like Double-A and Triple-A, have been shown to be almost as predictive as major-league numbers. But statistics at the lower minor-league levels are less reliabl and the numbers for college or high school players have very little predictive power.

Few professions, however, are as competitive as baseball. Among the thousands of professional baseball players, and the hundreds of thousands of amateurs, only 750 are able to play in the major leagues at any given time, and only a few doze of those will be All-Stars. Sanders's job is to search for those exceptional individuals who defy the odds. He has to work nearly as hard at his job as the players do, and he is still out on the road almost every day in his late sixties.
But [the scout] Sanders provides the Dodgers with the most valuable kind of information-the kind of information that oth people don't have.

As we've seen, baseball players do not become free agents until after six full seasons, which is usually not until they're at least thirty. As Bill James's analysis of the aging curve revealed, this often leads clubs to overspend on free agents-after all, their best years are usually behind them. But there is a flip side to this: before a player is thirty, he can provide tremendous value to his club. Moreover, baseball's economics are structured such that younger players can often be had for pennies on the dollar.42

If a baseball team is viewed, as with any other business, from a standpoint of profits and losses, almost all the value is created by the scouting and development process. If a team's forecasting system is exceptionally good, perhaps it can pa $10 million a year for a player whose real value is $12 million. But if its scouting is really good, it might be paying the same player just $400,000. That is how you compete in a small market like Oakland.

Indeed, the line between stats and scouting, and qualitative and quantitative information, has become very blurry in the baseball industry. Take, for example, the introduction of Pitch f/x, a system of three-dimensional cameras that have now been installed at every major-league stadium. Pitch f/x can measure not just how fast a pitch travels-that has been possib for years with radar guns-but how much it moves, horizontally and vertically, before reaching the plate. We can now say statistically, for instance, that Zack Greinke, a young pitcher with the Milwaukee Brewers who won the 2009 Cy Young Award as his league's best pitcher, has baseball's best slider,44 or that Mariano Rivera's cut fastball is really as good as reputed.45 Traditionally, these things were considered to be in the domain of scouting; now they're another variable that can be placed into a projection system.

We're not far from a point where we might have a complete three-dimensional recording of everything that takes place o a baseball field. We'll soon be able to measure exactly how good a jump Jacoby Ellsbury gets on a fly ball hit over his hea We'll know exactly how fast Ichiro Suzuki rounds the bases, or exactly how quickly Yadier Molina gets the ball down to second base when he's trying to throw out an opposing base-stealer.

This new technology will not kill scouting any more than Moneyball did, but it may change its emphasis toward the thing that are even harder to quantify and where the information is more exclusive, like a player's mental tools. Smart scouts like Sanders are already ahead of the curve.

Nate Silver 2012, _The Signal and the Noise_

ch2

There would be none of that on The McLaughlin Group when the same four panelists gathered again the following week.
The panel discussed the statistical minutiae of Obama’s win, his selection of Rahm Emanuel as his chief of staff, and his relations with Russian president Dmitry Medvedev. There was no mention of the failed prediction—made on national television in contradiction to essentially all available evidence. In fact, the panelists made it sound as though the outcom had been inevitable all along; Crowley explained that it had been a “change election year” and that McCain had run a terrible campaign—neglecting to mention that she had been willing to bet on that campaign just a week earlier.
Rarely should a forecaster be judged on the basis of a single prediction—but this case may warrant an exception. By the weekend before the election, perhaps the only plausible hypothesis to explain why McCain could still win was if there wa massive racial animus against Obama that had gone undetected in the polls.4 None of the panelists offered this hypothesis, however. Instead they seemed to be operating in an alternate universe in which the polls didn’t exist, the economy hadn’t collapsed, and President Bush was still reasonably popular rather than dragging down McCain.
Nevertheless, I decided to check to see whether this was some sort of anomaly. Do the panelists on The McLaughlin Grou —who are paid to talk about politics for a living—have any real skill at forecasting?
I evaluated nearly 1,000 predictions that were made on the final segment of the show by McLaughlin and the rest of the panelists. About a quarter of the predictions were too vague to be analyzed or concerned events in the far future. But I scored the others on a five-point scale ranging from completely false to completely true.
The panel may as well have been flipping coins. I determined 338 of their predictions to be either mostly or completely false. The exact same number—338—were either mostly or completely true.5
...Nor were any of the panelists—including Clift, who at least got the 2008 election right—much better than the others. F each panelist, I calculated a percentage score, essentially reflecting the number of predictions they got right. Clift and th three other most frequent panelists—Buchanan, the late Tony Blankley, and McLaughlin himself—each received almost identical scores ranging from 49 percent to 52 percent, meaning that they were about as likely to get a prediction right a wrong.7 They displayed about as much political acumen as a barbershop quartet.
The McLaughlin Group, of course, is more or less explicitly intended as slapstick entertainment for political junkies. It is holdover from the shouting match era of programs, such as CNN’s Crossfire, that featured liberals and conservatives endlessly bickering with one another. Our current echo chamber era isn’t much different from the shouting match era, except that the liberals and conservatives are confined to their own channels, separated in your cable lineup by a demilitarized zone demarcated by the Food Network or the Golf Channel.* This arrangement seems to produce higher ratings if not necessarily more reliable analysis.

As late as 1990, the CIA estimated—quite wrongly12—that the Soviet Union’s GDP was about half that of the United States13 (on a per capita basis, tantamount to where stable democracies like South Korea and Portugal are today). In fac more recent evidence has found that the Soviet economy—weakened by its long war with Afghanistan and the central government’s inattention to a variety of social problems—was roughly $1 trillion poorer than the CIA had thought and wa shrinking by as much as 5 percent annually, with inflation well into the double digits.

Big, bold, hedgehog-like predictions, in other words, are more likely to get you on television. Consider the case of Dick Morris, a former adviser to Bill Clinton who now serves as a commentator for Fox News. Morris is a classic hedgehog, an his strategy seems to be to make as dramatic a prediction as possible when given the chance. In 2005, Morris proclaime that George W. Bush’s handling of Hurricane Katrina would help Bush to regain his standing with the public.16 On the e of the 2008 elections, he predicted that Barack Obama would win Tennessee and Arkansas.17 In 2010, Morris predicted that the Republicans could easily win one hundred seats in the U.S. House of Representatives.18 In 2011, he said that Donald Trump would run for the Republican nomination—and had a “damn good” chance of winning it.19
All those predictions turned out to be horribly wrong. Katrina was the beginning of the end for Bush—not the start of a rebound. Obama lost Tennessee and Arkansas badly—in fact, they were among the only states in which he performed worse than John Kerry had four years earlier. Republicans had a good night in November 2010, but they gained sixty-thr seats, not one hundred. Trump officially declined to run for president just two weeks after Morris insisted he would do so But Morris is quick on his feet, entertaining, and successful at marketing himself—he remains in the regular rotation at Fox News and has sold his books to hundreds of thousands of people.

...liberals are not immune from the propensity to be hedgehogs. In my study of the accuracy of predictions made by McLaughlin Group members, Eleanor Clift—who is usually the most liberal member of the panel—almost never issued a prediction that would imply a more favorable outcome for Republicans than the consensus of the group. That may have served her well in predicting the outcome of the 2008 election, but she was no more accurate than her conservative counterparts over the long run.
Academic experts like the ones that Tetlock studied can suffer from the same problem. In fact, a little knowledge may be dangerous thing in the hands of a hedgehog with a Ph.D. One of Tetlock’s more remarkable findings is that, while foxes tend to get better at forecasting with experience, the opposite is true of hedgehogs: their performance tends to worsen a they pick up additional credentials. Tetlock believes the more facts hedgehogs have at their command, the more opportunities they have to permute and manipulate them in ways that confirm their biases. The situation is analogous to what might happen if you put a hypochondriac in a dark room with an Internet connection. The more time that you give him, the more information he has at his disposal, the more ridiculous the self-diagnosis he’ll come up with; before long he’ll be mistaking a common cold for the bubonic plague.

My interest in electoral politics had begun slightly earlier, however—and had been mostly the result of frustration rather any affection for the political process. I had carefully monitored the Congress’s attempt to ban Internet poker in 2006, which was then one of my main sources of income. I found political coverage wanting even as compared with something

like sports, where the “Moneyball revolution” had significantly improved analysis.
During the run-up to the primary I found myself watching more and more political TV, mostly MSNBC and CNN and Fox News. A lot of the coverage was vapid. Despite the election being many months away, commentary focused on the inevitability of Clinton’s nomination, ignoring the uncertainty intrinsic to such early polls. There seemed to be too much focus on Clinton’s gender and Obama’s race.24 There was an obsession with determining which candidate had “won the day” by making some clever quip at a press conference or getting some no-name senator to endorse them—things that 9 percent of voters did not care about.
Political news, and especially the important news that really affects the campaign, proceeds at an irregular pace. But new coverage is produced every day. Most of it is filler, packaged in the form of stories that are designed to obscure its unimportance.* Not only does political coverage often lose the signal—it frequently accentuates the noise. If there are a number of polls in a state that show the Republican ahead, it won’t make news when another one says the same thing. Bu if a new poll comes out showing the Democrat with the lead, it will grab headlines—even though the poll is probably an outlier and won’t predict the outcome accurately.
The bar set by the competition, in other words, was invitingly low. Someone could look like a genius simply by doing som fairly basic research into what really has predictive power in a political campaign. So I began blogging at the Web site Daily Kos, posting detailed and data-driven analyses on issues like polls and fundraising numbers. I studied which polling firms had been most accurate in the past, and how much winning one state—Iowa, for instance—tended to shift the numbers in another. The articles quickly gained a following, even though the commentary at sites like Daily Kos is usuall more qualitative (and partisan) than quantitative. In March 2008, I spun my analysis out to my own Web site, FiveThirtyEight

The further down the ballot you go, the more volatile the polls tend to be: polls of House races are less accurate than pol of Senate races, which are in turn less accurate than polls of presidential races. Polls of primaries, also, are considerably less accurate than general election polls. During the 2008 Democratic primaries, the average poll missed by about eight points, far more than implied by its margin of error. The problems in polls of the Republican primaries of 2012 may have been even worse.26 In many of the major states, in fact—including Iowa, South Carolina, Florida, Michigan, Washington, Colorado, Ohio, Alabama, and Mississippi—the candidate ahead in the polls a week before the election lost.
But polls do become more accurate the closer you get to Election Day. Figure 2-4 presents some results from a simplified version of the FiveThirtyEight Senate forecasting model, which uses data from 1998 through 2008 to infer the probabilit that a candidate will win on the basis of the size of his lead in the polling average. A Senate candidate with a five-point lead on the day before the election, for instance, has historically won his race about 95 percent of the time—almost a sur thing, even though news accounts are sure to describe the race as “too close to call.” By contrast, a five-point lead a yea before the election translates to just a 59 percent chance of winning—barely better than a coin flip.

Politicians and political observers, however, find this lack of clarity upsetting. In 2010, a Democratic congressman called me a few weeks in advance of the election. He represented a safely Democratic district on the West Coast. But given how well Republicans were doing that year, he was nevertheless concerned about losing his seat. What he wanted to know wa exactly how much uncertainty there was in our forecast. Our numbers gave him, to the nearest approximation, a 100 percent chance of winning. But did 100 percent really mean 99 percent, or 99.99 percent, or 99.9999 percent? If the latt —a 1 in 100,000 chance of losing—he was prepared to donate his campaign funds to other candidates in more vulnerable districts. But he wasn’t willing to take a 1 in 100 risk.
Political partisans, meanwhile, may misinterpret the role of uncertainty in a forecast; they will think of it as hedging your bets and building in an excuse for yourself in case you get the prediction wrong. That is not really the idea. If you forecas that a particular incumbent congressman will win his race 90 percent of the time, you’re also forecasting that he should lose it 10 percent of the time.28 The signature of a good forecast is that each of these probabilities turns out to be about right over the long run.

Few political analysts have a longer track record of success than the tight-knit team that runs the Cook Political Report. The group, founded in 1984 by a genial, round-faced Louisianan named Charlie Cook, is relatively little known outside th Beltway. But political junkies have relied on Cook’s forecasts for years and have rarely had reason to be disappointed wit their results.
Cook and his team have one specific mission: to predict the outcome of U.S. elections, particularly to the Congress. This means issuing forecasts for all 435 races for the U.S. House, as well as the 35 or so races for the U.S. Senate that take place every other year.
Predicting the outcome of Senate or gubernatorial races is relatively easy. The candidates are generally well known to voters, and the most important races attract widespread attention and are polled routinely by reputable firms. Under the circumstances, it is hard to improve on a good method for aggregating polls, like the one I use at FiveThirtyEight.
House races are another matter, however. The candidates often rise from relative obscurity—city councilmen or small-business owners who decide to take their shot at national politics—and in some cases are barely known to voters until ju days before the election. Congressional districts, meanwhile, are spread throughout literally every corner of the country, giving rise to any number of demographic idiosyncrasies. The polling in House districts tends to be erratic at best36 whe it is available at all, which it often isn’t.
But this does not mean there is no information available to analysts like Cook. Indeed, there is an abundance of it: in addition to polls, there is data on the demographics of the district and on how it has voted in past elections. There is data on overall partisan trends throughout the country, such as approval ratings for the incumbent president. There is data o fund-raising, which must be scrupulously reported to the Federal Elections Commission.
Other types of information are more qualitative, but are nonetheless potentially useful. Is the candidate a good public speaker? How in tune is her platform with the peculiarities of the district? What type of ads is she running? A political campaign is essentially a small business: How well does she manage people?
Of course, all of that information could just get you into trouble if you were a hedgehog who wasn’t weighing it carefully. But Cook Political has a lot of experience in making forecasts, and they have an impressive track record of accuracy.
Cook Political classifies races along a seven-point scale ranging from Solid Republican—a race that the Republican

candidate is almost certain to win—to Solid Democrat (just the opposite). Between 1998 and 2010, the races that Cook described as Solid Republican were in fact won by the Republican candidate on 1,205 out of 1,207 occasions—well over 9 percent of the time. Likewise, races that they described as Solid Democrat were won by the Democrat in 1,226 out of 1,229 instances.

Many of the races that Cook places into the Solid Democrat or Solid Republican categories occur in districts where the same party wins every year by landslide margins—these are not that hard to call. But Cook Political has done just about a well in races that require considerably more skill to forecast. Elections they've classified as merely "leaning" toward the Republican candidate, for instance, have in fact been won by the Republican about 95 percent of the time. Likewise, race they've characterized as leaning to the Democrat have been won by the Democrat 92 percent of the time.37 Furthermore the Cook forecasts have a good track record even when they disagree with quantitative indicators like polls.38

...His interview with Kapanke followed this template. Wasserman's knowledge of the nooks and crannies of political geography can make him seem like a local, and Kapanke was happy to talk shop about the intricacies of his district—just how many voters he needed to win in La Crosse to make up for the ones he'd lose in Eau Claire. But he stumbled over a series of questions on allegations that he had used contributions from lobbyists to buy a new set of lights for the Loggers ballpark.40

It was small-bore stuff; it wasn't like Kapanke had been accused of cheating on his wife or his taxes. But it was enough to dissuade Wasserman from changing the rating.41 Indeed, Kapanke lost his election that November by about 9,500 votes, even though Republicans won their races throughout most of the similar districts in the Midwest.

This is, in fact, the more common occurrence; Wasserman will usually maintain the same rating after the interview. As hard as he works to glean new information from the candidates, it is often not important enough to override his prior tak on the race.

Wasserman's approach works because he is capable of evaluating this information without becoming dazzled by the candidate sitting in front of him. A lot of less-capable analysts would open themselves to being charmed, lied to, spun, or would otherwise get hopelessly lost in the narrative of the campaign. Or they would fall in love with their own spin about the candidate's interview skills, neglecting all the other information that was pertinent to the race.

Wasserman instead considers everything in the broader political context. A terrific Democratic candidate who aces her interview might not stand a chance in a district that the Republican normally wins by twenty points.

So why bother with the candidate interviews at all? Mostly, Wasserman is looking for red flags—like the time when the Democratic congressman Eric Massa (who would later abruptly resign from Congress after accusations that he sexually harassed a male staffer) kept asking Wasserman how old he was. The psychologist Paul Meehl called these "broken leg" cases—situations where there is something so glaring that it would be foolish not to account for it.42

Nate Silver 2012, _The Signal and the Noise_

ch1

The ratings agencies had given their AAA rating, normally reserved for a handful of the world’s most solvent government
and best-run businesses, to thousands of mortgage-backed securities, financial instruments that allowed investors to bet
on the likelihood of someone else defaulting on their home. The ratings issued by these companies are quite explicitly
meant to be predictions: estimates of the likelihood that a piece of debt will go into default.5 Standard & Poor’s told
investors, for instance, that when it rated a particularly complex type of security known as a collateralized debt obligatio
(CDO) at AAA, there was only a 0.12 percent probability—about 1 chance in 850—that it would fail to pay out over the ne
five years.6 This supposedly made it as safe as a AAA-rated corporate bond7 and safer than S&P now assumes U.S.
Treasury bonds to be.8 The ratings agencies do not grade on a curve.
In fact, around 28 percent of the AAA-rated CDOs defaulted, according to S&P’s internal figures.9 (Some independent
estimates are even higher.10) That means that the actual default rates for CDOs were more than two hundred times
higher than S&P had predicted.11
This is just about as complete a failure as it is possible to make in a prediction: trillions of dollars in investments that wer
rated as being almost completely safe instead turned out to be almost completely unsafe. It was as if the weather forecas
had been 86 degrees and sunny, and instead there was a blizzard.

What is remarkable about the housing bubble is the number of people who did see it coming—and who said so well in
advance. Robert Shiller, the Yale economist, had noted its beginnings as early as 2000 in his book Irrational
Exuberance.14 Dean Baker, a caustic economist at the Center for Economic and Policy Research, had written about the
bubble in August 2002.15 A correspondent at the Economist magazine, normally known for its staid prose, had spoken of
the “biggest bubble in history” in June 2005.16 Paul Krugman, the Nobel Prize–winning economist, wrote of the bubble
and its inevitable end in August 2005.17 “This was baked into the system,” Krugman later told me. “The housing crash w
not a black swan. The housing crash was the elephant in the room.”
Ordinary Americans were also concerned. Google searches on the term “housing bubble” increased roughly tenfold from
January 2004 through summer 2005.18 Interest in the term was heaviest in those states, like California, that had seen th
largest run-up in housing prices19—and which were about to experience the largest decline. In fact, discussion of the
bubble was remarkably widespread. Instances of the two-word phrase “housing bubble” had appeared in just eight news
accounts in 200120 but jumped to 3,447 references by 2005. The housing bubble was discussed about ten times per day
reputable newspapers and periodicals.21

One reason that S&P and Moody’s enjoyed such a dominant market presence is simply that they had been a part of the
club for a long time. They are part of a legal oligopoly; entry into the industry is limited by the government. Meanwhile, a
seal of approval from S&P and Moody’s is often mandated by the bylaws of large pension funds,25 about two-thirds of
which26 mention S&P, Moody’s, or both by name, requiring that they rate a piece of debt before the pension fund can
purchase it.27
S&P and Moody’s had taken advantage of their select status to build up exceptional profits despite picking résumés out o
Wall Street’s reject pile.* Moody’s28 revenue from so-called structured-finance ratings increased by more than 800
percent between 1997 and 2007 and came to represent the majority of their ratings business during the bubble years.29
These products helped Moody’s to the highest profit margin of any company in the S&P 500 for five consecutive years
during the housing bubble.30 (In 2010, even after the bubble burst and the problems with the ratings agencies had
become obvious, Moody’s still made a 25 percent profit.31)
With large profits locked in so long as new CDOs continued to be issued, and no way for investors to verify the accuracy
their ratings until it was too late, the agencies had little incentive to compete on the basis of quality. The CEO of Moody’s
Raymond McDaniel, explicitly told his board that ratings quality was the least important factor driving the company’s
profits.32
...A memo provided to me by an S&P spokeswoman, Catherine Mathis, detailed how S&P had conducted a simulation in
2005 that anticipated a 20 percent decline in national housing prices over a two-year period—not far from the roughly 30
percent decline in housing prices that actually occurred between 2006 and 2008. The memo concluded that S&P’s existin
models “captured the risk of a downturn” adequately and that its highly rated securities would “weather a housing
downturn without suffering a credit-rating downgrade.”36

Moody’s, for instance, went through a period of making ad hoc adjustments to its model44 in which it increased the
default probability assigned to AAA-rated securities by 50 percent. That might seem like a very prudent attitude: surely a
50 percent buffer will suffice to account for any slack in one’s assumptions?
It might have been fine had the potential for error in their forecasts been linear and arithmetic. But leverage, or
investments financed by debt, can make the error in a forecast compound many times over, and introduces the potential
highly geometric and nonlinear mistakes. Moody’s 50 percent adjustment was like applying sunscreen and claiming it
protected you from a nuclear meltdown—wholly inadequate to the scale of the problem. It wasn’t just a possibility that
their estimates of default risk could be 50 percent too low: they might just as easily have underestimated it by 500 perce
or 5,000 percent. In practice, defaults were two hundred times more likely than the ratings agencies claimed, meaning
that their model was off by a mere 20,000 percent.

In fact, according to an index developed by Robert Shiller and his colleague Karl Case, the market price of an American
home has barely increased at all over the long run. After adjusting for inflation, a $10,000 investment made in a home in
1896 would be worth just $10,600 in 1996. The rate of return had been less in a century than the stock market typically
produces in a single year.47
But if a home was not a profitable investment it had at least been a safe one. Prior to the 2000s, the most significant shif
in American housing prices had come in the years immediately following World War II, when they increased by about 60

percent relative to their nadir in 1942....If the United States had never experienced such a housing bubble before, however, other countries had—and results had been uniformly disastrous. Shiller, studying data going back hundreds of years in countries from the Netherlands to Norway, found that as real estate grew to unaffordable levels a crash almost inevitably followed.54 The infamous Japanese real estate bubble of the early 1990s forms a particularly eerie precedent t the recent U.S. housing bubble, for instance. The price of commercial real estate in Japan increased by about 76 percent over the ten-year period between 1981 and 1991 but then declined by 31 percent over the next five years, a close fit for the trajectory that American home prices took during and after the bubble55 (figure 1-4). Shiller uncovered another key piece of evidence for the bubble: the people buying the homes had completely unrealistic assumptions about what their investments might return. A survey commissioned by Case and Schiller in 2003 found that homeowners expected their properties to appreciate at a rate of about 13 percent per year.56 In practice, over that one-hundred-year period from 1896 through 199657 to which I referred earlier, sale prices of houses had increased by just 6 percent total after inflatio or about 0.06 percent annually.

While quite a few economists identified the housing bubble as it occurred, fewer grasped the consequences of a housing-price collapse for the broader economy. In December 2007, economists in the Wall Street Journal forecasting panel predicted only a 38 percent likelihood of a recession over the next year. This was remarkable because, the data would later reveal, the economy was already in recession at the time. The economists in another panel, the Survey of Profession Forecasters, thought there was less than a 1 in 500 chance that the economy would crash as badly as it did.63
There were two major factors that the economists missed. The first was simply the effect that a drop in housing prices might have on the finances of the average American. As of 2007, middle-class Americans64 had more than 65 percent of their wealth tied up in their homes.65 Otherwise they had been getting poorer—they had been using their household equity as ATMs.66 Nonhousehold wealth—meaning the sum total of things like savings, stocks, pensions, cash, and equit in small businesses—declined by 14 percent67 for the median family between 2001 and 2007.68 When the collapse of the housing bubble wiped essentially all their housing equity off the books, middle-class Americans found they were considerably worse off than they had been a few years earlier.

“If you’re in a market and someone’s trying to sell you something which you don’t understand,” George Akerlof told me, “you should think that they’re selling you a lemon.”
Akerlof wrote a famous paper on this subject called “The Market for Lemons”78—it won him a Nobel Prize. In the paper, he demonstrated that in a market plagued by asymmetries of information, the quality of goods will decrease and the market will come to be dominated by crooked sellers and gullible or desperate buyers.
Imagine that a stranger walked up to you on the street and asked if you were interested in buying his used car. He showe you the Blue Book value but was not willing to let you take a test-drive. Wouldn’t you be a little suspicious? The core problem in this case is that the stranger knows much more about the car—its repair history, its mileage—than you do. Sensible buyers will avoid transacting in a market like this one at any price. It is a case of uncertainty trumping risk. You know that you’d need a discount to buy from him—but it’s hard to know how much exactly it ought to be. And the lower the man is willing to go on the price, the more convinced you may become that the offer is too good to be true. There ma be no such thing as a fair price.
But now imagine that the stranger selling you the car has someone else to vouch for him. Someone who seems credible and trustworthy—a close friend of yours, or someone with whom you have done business previously. Now you might reconsider. This is the role that the ratings agencies played. They vouched for mortgage-backed securities with lots of AA ratings and helped to enable a market for them that might not otherwise have existed.

Once the housing bubble had burst, greedy investors became fearful ones who found uncertainty lurking around every corner. The process of disentangling a financial crisis—everyone trying to figure out who owes what to whom—can produce hangovers that persist for a very long time. The economists Carmen Reinhart and Kenneth Rogoff, studying volumes of financial history for their book This Time Is Different: Eight Centuries of Financial Folly, found that financial crises typically produce rises in unemployment that persist for four to six years.86 Another study by Reinhart, which focused on more recent financial crises, found that ten of the last fifteen countries to endure one had never seen their unemployment rates recover to their precrisis levels.87 This stands in contrast to normal recessions, in which there is typically above-average growth in the year or so following the recession88 as the economy reverts to the mean, allowing employment to catch up quickly. Yet despite its importance, many economic models made no distinction between the financial system and other parts of the economy.

Nate Silver 2012, _The Signal and the Noise_

Introduction

Books had existed prior to Gutenberg, but they were not widely written and they were not widely read. Instead, they wer luxury items for the nobility, produced one copy at a time by scribes.3 The going rate for reproducing a single manuscrip was about one florin (a gold coin worth about $200 in today’s dollars) per five pages,4 so a book like the one you’re reading now would cost around $20,000. It would probably also come with a litany of transcription errors, since it would be a copy of a copy of a copy, the mistakes having multiplied and mutated through each generation.
...The printing press changed that, and did so permanently and profoundly. Almost overnight, the cost of producing a boo decreased by about three hundred times,7 so a book that might have cost $20,000 in today’s dollars instead cost $70. Printing presses spread very rapidly throughout Europe; from Gutenberg’s Germany to Rome, Seville, Paris, and Basel by 1470, and then to almost all other major European cities within another ten years.8 The number of books being produced grew exponentially, increasing by about thirty times in the first century after the printing press was invented.9 The store of human knowledge had begun to accumulate, and rapidly.

Shakespeare’s plays often turn on the idea of fate, as much drama does. What makes them so tragic is the gap between what his characters might like to accomplish and what fate provides to them. The idea of controlling one’s fate seemed to have become part of the human consciousness by Shakespeare’s time—but not yet the competencies to achieve that end. Instead, those who tested fate usually wound up dead.18
These themes are explored most vividly in The Tragedy of Julius Caesar. Throughout the first half of the play Caesar receives all sorts of apparent warning signs—what he calls predictions19 (“beware the ides of March”)—that his coronation could turn into a slaughter. Caesar of course ignores these signs, quite proudly insisting that they point to someone else’s death—or otherwise reading the evidence selectively. Then Caesar is assassinated.
“[But] men may construe things after their fashion / Clean from the purpose of the things themselves,” Shakespeare warn us through the voice of Cicero—good advice for anyone seeking to pluck through their newfound wealth of information. I was hard to tell the signal from the noise. The story the data tells us is often the one we’d like to hear, and we usually make sure that it has a happy ending.
And yet if The Tragedy of Julius Caesar turned on an ancient idea of prediction—associating it with fatalism, fortune-telling, and superstition—it also introduced a more modern and altogether more radical idea: that we might interpret these signs so as to gain an advantage from them. “Men at some time are masters of their fates,” says Cassius, hoping to persuade Brutus to partake in the conspiracy against Caesar.

In the 1960s the United States spent about $1.5 million (adjusted for inflation33) per patent application34 by an America inventor. That figure rose rather than fell at the dawn of the information age, however, doubling to a peak of about $3 million in 1986.35

Baseball, for instance, is an exceptional case. It happens to be an especially rich and revealing exception, and the book considers why this is so—why a decade after Moneyball, stat geeks and scouts are now working in harmony.
The book offers some other hopeful examples. Weather forecasting, which also involves a melding of human judgment an computer power, is one of them. Meteorologists have a bad reputation, but they have made remarkable progress, being able to forecast the landfall position of a hurricane three times more accurately than they were a quarter century ago. Meanwhile, I met poker players and sports bettors who really were beating Las Vegas, and the computer programmers who built IBM’s Deep Blue and took down a world chess champion.
But these cases of progress in forecasting must be weighed against a series of failures....
We had not seen the September 11 attacks coming. The problem was not want of information. As had been the case in th Pearl Harbor attacks six decades earlier, all the signals were there. But we had not put them together. Lacking a proper theory for how terrorists might behave, we were blind to the data and the attacks were an “unknown unknown” to us.
There also were the widespread failures of prediction that accompanied the recent global financial crisis. Our naïve trust in models, and our failure to realize how fragile they were to our choice of assumptions, yielded disastrous results. On a more routine basis, meanwhile, I discovered that we are unable to predict recessions more than a few months in advance and not for lack of trying. While there has been considerable progress made in controlling inflation, our economic policy makers are otherwise flying blind.
The forecasting models published by political scientists in advance of the 2000 presidential election predicted a landslide 11-point victory for Al Gore.38 George W. Bush won instead. Rather than being an anomalous result, failures like these have been fairly common in political prediction. A long-term study by Philip E. Tetlock of the University of Pennsylvania found that when political scientists claimed that a political outcome had absolutely no chance of occurring, it neverthele happened about 15 percent of the time. (The political scientists are probably better than television pundits, however.)
There has recently been, as in the 1970s, a revival of attempts to predict earthquakes, most of them using highly mathematical and data-driven techniques. But these predictions envisaged earthquakes that never happened and failed t prepare us for those that did. The Fukushima nuclear reactor had been designed to handle a magnitude 8.6 earthquake, part because some seismologists concluded that anything larger was impossible. Then came Japan’s horrible magnitude 9.1 earthquake in March 2011.

Bayes’s theorem, however, can also be applied to more existential types of problems. Chapters 11 through 13 consider three of these cases: global warming, terrorism, and bubbles in financial markets. These are hard problems for forecaste and for society. But if we are up to the challenge, we can make our country, our economy, and our planet a little safer.
The world has come a long way since the days of the printing press. Information is no longer a scarce commodity; we hav more of it than we know what to do with. But relatively little of it is useful. We perceive it selectively, subjectively, and without much self-regard for the distortions that this causes. We think we want information when we really want knowledge.

The signal is the truth. The noise is what distracts us from the truth. This is a book about the signal and the noise.