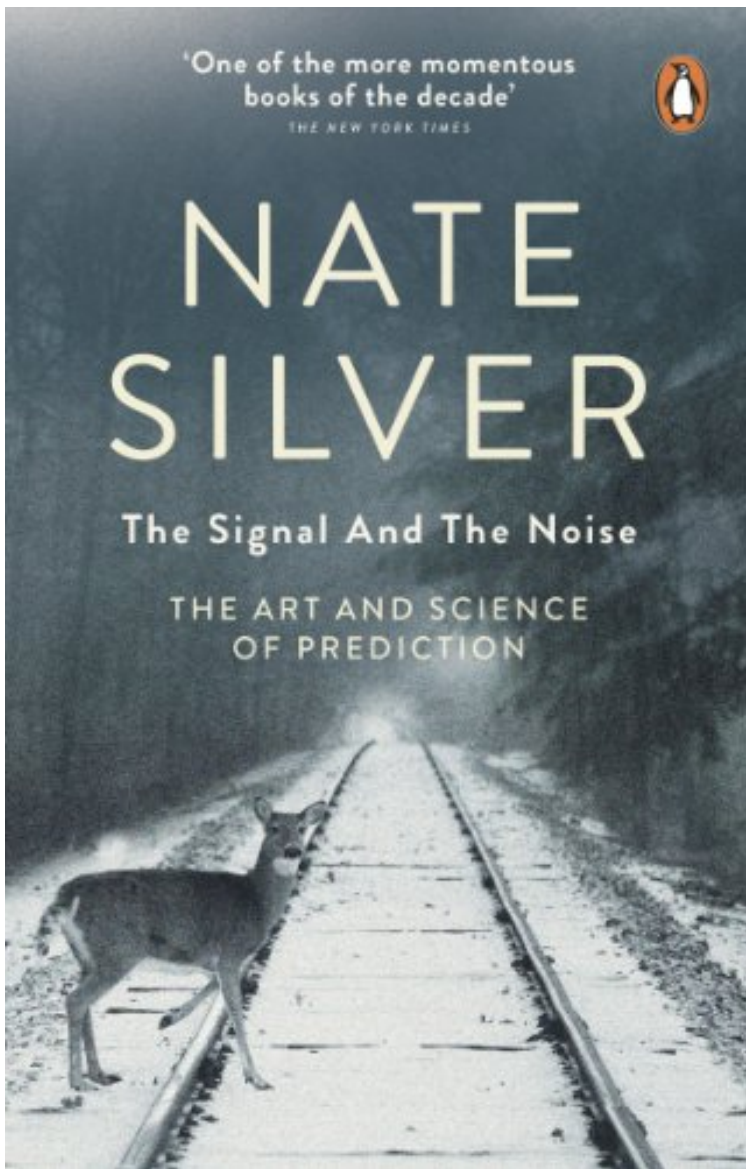


[Read free] File size: 16.Mb

The Signal and the Noise: The Art and Science of Prediction



Par Nate Silver
**Download PDF | ePub | DOC | audiobook | ebooks*

Dtails sur le produit Rang parmi les ventes : #24740 dans eBooksPubli le: 2012-09-27Sorti le: 2012-09-27Format: Ebook Kindle

[Read free] The Signal and the Noise: The Art and Science of Prediction

Par Nate Silver : The Signal and the Noise: The Art and Science of Prediction before purchasing it in order to gage whether or not it would be worth my time, and all praised The Signal and the Noise: The Art and Science of Prediction:

Download

Read Online

Description :

Prsentation de l'diteurEvery time we choose a route to work, decide whether to go on a second date, or set aside money for a rainy day, we are making a prediction about the future. Yet from the global financial crisis to 9/11 to the Fukushima disaster, we often fail to foresee hugely significant events. In The Signal and the Noise, the New York Times' political forecaster and statistics guru Nate Silver explores the art of prediction, revealing how we can all build a better crystal ball. In his quest to distinguish the true signal from a universe of noisy data, Silver visits hundreds of expert forecasters, in fields ranging from the stock market to the

poker table, from earthquakes to terrorism. What lies behind their success? And why do so many predictions still fail? By analysing the rare prescient forecasts, and applying a more quantitative lens to everyday life, Silver distils the essential lessons of prediction. We live in an increasingly data-driven world, but it is harder than ever to detect the true patterns amid the noise of information. In this dazzling insider's tour of the world of forecasting, Silver reveals how we can all develop better foresight in our everyday lives. Extrait At about the time *The Signal and the Noise* was first published in September 2012, Big Data was on its way becoming a Big Idea. Google searches for the term doubled over the course of a year,¹ as did mentions of it in the news media.² Hundreds of books were published on the subject. If you picked up any business periodical in 2013, advertisements for Big Data were as ubiquitous as cigarettes in an episode of *Mad Men*. But by late 2014, there was evidence that trend had reached its apex. The frequency with which Big Data was mentioned in corporate press releases had slowed down and possibly begun to decline.³ The technology research firm Gartner even declared that Big Data had passed the peak of its hype cycle.⁴ I hope that Gartner is right.

Coming to a better understanding of data and statistics is essential to help us navigate our lives. But as with most emerging technologies, the widespread benefits to science, industry, and human welfare will come only after the hype has died down.

FIGURE P-1: BIG DATA MENTIONS IN CORPORATE PRESS RELEASES
I worry that certain events in my life have contributed to the hype cycle. On November 6, 2012, the statistical model at my Web site FiveThirtyEight called the winner of the American presidential election correctly in all fifty states. I received a congratulatory phone call from the White House. I was hailed as lord and god of the algorithm by *The Daily Shows* Jon Stewart. My name briefly received more Google search traffic than the vice president of the United States. I enjoyed some of the attention, but I felt like an outlier even a fluke. Mostly I was getting credit for having pointed out the obvious and most of the rest was luck.* To be sure, it was reasonably clear by Election Day that President Obama was poised to win reelection. When voters went to the polls on election morning, FiveThirtyEight's statistical model put his chances of winning the Electoral College at about 90 percent.* A 90 percent chance is not quite a sure thing: Would you board a plane if the pilot told you it had a 90 percent chance of landing successfully? But when there's only reputation rather than life or limb on the line, it's a good bet. Obama needed to win only a handful of the swing states where he was tied or ahead in the polls; Mitt Romney would have had to win almost all of them. But getting every state right was a stroke of luck. In our Election Day forecast, Obama's chance of winning Florida was just 50.3 percent; the outcome was as random as a coin flip. Considering other states like Virginia, Ohio, Colorado, and North Carolina, our chances of going fifty-for-fifty were only about 20 percent.⁵ FiveThirtyEight's perfect forecast was fortuitous but contributed to the perception that statisticians are soothsayers only using computers rather than crystal balls. This is a wrongheaded and rather dangerous idea. American presidential elections are the exception to the rule of the few examples of a complex system in which outcomes are usually more certain than the conventional wisdom implies. (There are a number of reasons for this, not least that the conventional wisdom is often not very wise when it comes to politics.) Far more often, as this book will explain, we overrate our ability to predict the world around us.

With some regularity, events that are said to be certain fail to come to fruition or those that are deemed impossible turn out to occur. If all of this is so simple, why did so many pundits get the 2012 election wrong? It wasn't just on the fringe of the blogosphere that conservatives insisted that the polls were skewed toward President Obama. Thoughtful conservatives like George F. Will⁶ and Michael Barone⁷ also predicted a Romney win, sometimes by near-landslide proportions. One part of the answer is obvious: the pundits didn't have much incentive to make the right call. You can get invited back on television with a far worse track record than Barone's or Will's provided you speak with some conviction and have a viewpoint that matches the producers' goals. An alternative interpretation is slightly less cynical but potentially harder to swallow: human judgment is intrinsically fallible. It's hard for any of us (myself included) to recognize how much our relatively narrow range of experience can color our interpretation of the evidence. There's so much information out there today that none of us can plausibly consume all of it. We're constantly making decisions about what Web site to read, which television channel to watch, and where to focus our attention. Having a better understanding of statistics almost certainly helps. Over the past decade, the number of people employed as statisticians in the United States has increased by 35 percent⁸ even as the overall job market has stagnated. But it's a necessary rather than sufficient part of the solution. Some of the examples of failed predictions in this book concern people with exceptional intelligence and exemplary statistical training but whose biases still got in the way. These problems are not so simple and so this book does not promote simple answers to them. It makes some recommendations but they are philosophical as much as

technical. Once were getting the big stuff right coming to a better understanding of probably and uncertainty; learning to recognize our biases; appreciating the value of diversity, incentives, and experimentation will have the luxury of worrying about the finer points of technique. Gartners hype cycle ultimately has a happy ending. After the peak of inflated expectations theres a trough of disillusionment what happens when people come to recognize that the new technology will still require a lot of hard work. FIGURE P-2: GARTNERS HYPE CYCLE But right when views of the new technology have begun to lapse from healthy skepticism into overt cynicism, that technology can begin to pay some dividends. (Weve been through this before: after the computer boom in the 1970s and the Internet commerce boom of the late 1990s, among other examples.)

Eventually it matures to the point when there are fewer glossy advertisements but more gains in productivity it may even have become so commonplace that we take it for granted. I hope this book can accelerate the process, however slightly. This is a book about information, technology, and scientific progress. This is a book about competition, free markets, and the evolution of ideas. This is a book about the things that make us smarter than any computer, and a book about human error. This is a book about how we learn, one step at a time, to come to knowledge of the objective world, and why we sometimes take a step back. This is a book about prediction, which sits at the intersection of all these things. It is a study of why some predictions succeed and why some fail. My hope is that we might gain a little more insight into planning our futures and become a little less likely to repeat our mistakes. More Information, More Problems The original revolution in information technology came not with the microchip, but with the printing press. Johannes Gutenbergs invention in 1440 made information available to the masses, and the explosion of ideas it produced had unintended consequences and unpredictable effects. It was a spark for the Industrial Revolution in 1775,¹ a tipping point in which civilization suddenly went from having made almost no scientific or economic progress for most of its existence to the exponential rates of growth and change that are familiar to us today. It set in motion the events that would produce the European Enlightenment and the founding of the American Republic. But the printing press would first produce something else: hundreds of years of holy war. As mankind came to believe it could predict its fate and choose its destiny, the bloodiest epoch in human history followed.² Books had existed prior to Gutenberg, but they were not widely written and they were not widely read. Instead, they were luxury items for the nobility, produced one copy at a time by scribes.³ The going rate for reproducing a single manuscript was about one florin (a gold coin worth about \$200 in todays dollars) per five pages,⁴ so a book like the one youre 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. This made the accumulation of knowledge extremely difficult. It required heroic effort to prevent the volume of recorded knowledge from actually decreasing, since the books might decay faster than they could be reproduced. Various editions of the Bible survived, along with a small number of canonical texts, like from Plato and Aristotle. But an untold amount of wisdom was lost to the ages,⁵ and there was little incentive to record more of it to the page. The pursuit of knowledge seemed inherently futile, if not altogether vain. If today we feel a sense of impermanence because things are changing so rapidly, impermanence was a far more literal concern for the generations before us. There was nothing new under the sun, as the beautiful Bible verses in Ecclesiastes put it not so much because everything had been discovered but because everything would be forgotten.⁶ The printing press changed that, and did so permanently and profoundly. Almost overnight, the cost of producing a book decreased by about three hundred times,⁷ so a book that might have cost \$20,000 in todays dollars instead cost \$70. Printing presses spread very rapidly throughout Europe; from Gutenbergs 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. FIGURE I-1: EUROPEAN BOOK PRODUCTION As was the case during the early days of the World Wide Web, however, the quality of the information was highly varied. While the printing press paid almost immediate dividends in the production of higher quality maps,¹⁰ the bestseller list soon came to be dominated by heretical religious texts and pseudoscientific ones.¹¹ Errors could now be mass-produced, like in the so-called Wicked Bible, which committed the most unfortunate typo in history to the page: thou shalt commit adultery.¹² Meanwhile, exposure to so many new ideas was producing mass confusion. The amount of information was increasing much more rapidly than our understanding of what to do with it, or our ability to differentiate the useful information from the mistruths.¹³ Paradoxically, the result of having so much more shared knowledge was increasing isolation along national and religious lines.

The instinctual shortcut that we take when we have too much information is to engage with it selectively, picking out the parts we like and ignoring the remainder, making allies with those who have made the same choices and enemies of the rest. The most enthusiastic early customers of the printing press were those who used it to evangelize. Martin Luthers Ninety-five Theses were not that radical; similar sentiments had been debated many times over. What was revolutionary, as Elizabeth Eisenstein writes, is that Luthers theses did not stay tacked to the church door.¹⁴ Instead, they were reproduced at least three hundred thousand times by

Gutenbergs printing press¹⁵ a runaway hit even by modern standards. The schism that Luthers Protestant Reformation produced soon plunged Europe into war. From 1524 to 1648, there was the German Peasants War, the Schmalkaldic War, the Eighty Years War, the Thirty Years War, the French Wars of Religion, the Irish Confederate Wars, the Scottish Civil War, and the English Civil War many of them raging simultaneously. This is not to neglect the Spanish Inquisition, which began in 1480, or the War of the Holy League from 1508 to 1516, although those had less to do with the spread of Protestantism. The Thirty Years

War alone killed one-third of Germanys population,¹⁶ and the seventeenth century was possibly the bloodiest ever, with the early twentieth staking the main rival claim.¹⁷ But somehow in the midst of this, the printing press was starting to produce scientific and literary progress. Galileo was sharing his (censored) ideas, and Shakespeare was producing his plays. Shakespeares 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 ones fate seemed to have become part of the human consciousness by Shakespeares 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 elses 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 warns us through the voice of Ciceros good advice for anyone seeking to pluck through their newfound wealth of information. It was hard to tell the signal from the noise. The story the data tells us is often the one wed 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. The idea of man as master of his fate was gaining currency. The words predict and forecast are largely used interchangeably today, but in Shakespeares time, they meant different things. A prediction was what the soothsayer told you; a forecast was something more like Cassiuss idea. The term forecast came from English

Germanic roots,²⁰ unlike predict, which is from Latin.²¹ Forecasting reflected the new Protestant worldliness rather than the otherworldliness of the Holy Roman Empire. Making a forecast typically implied planning under conditions of uncertainty. It suggested having prudence, wisdom, and industriousness, more like the way we now use the word foresight. ²² The theological implications of this idea are complicated.²³

But they were less so for those hoping to make a gainful existence in the terrestrial world. These qualities were strongly associated with the Protestant work ethic, which Max Weber saw as bringing about capitalism and the Industrial Revolution.²⁴ This notion of forecasting was very much tied in to the notion of progress.

All that information in all those books ought to have helped us to plan our lives and profitably predict the worlds course. The Protestants who ushered in centuries of holy war were learning how to use their accumulated knowledge to change society. The Industrial Revolution largely began in Protestant countries and largely in those with a free press, where both religious and scientific ideas could flow without fear of censorship.²⁵ The importance of the Industrial Revolution is hard to overstate. Throughout essentially all of human history, economic growth had proceeded at a rate of perhaps 0.1 percent per year, enough to allow for

a very gradual increase in population, but not any growth in per capita living standards.²⁶ And then, suddenly, there was progress when there had been none. Economic growth began to zoom upward much faster than the growth rate of the population, as it has continued to do through to the present day, the occasional global financial meltdown notwithstanding.²⁷ FIGURE I-2: GLOBAL PER CAPITA GDP, 1000-2010

The explosion of information produced by the printing press had done us a world of good, it turned out. It had just taken 330 years and millions dead in battlefields around Europe for those advantages to take hold. The Productivity Paradox We face danger whenever information growth outpaces our understanding of

how to process it. The last forty years of human history imply that it can still take a long time to translate information into useful knowledge, and that if we are not careful, we may take a step back in the meantime. The term information age is not particularly new. It started to come into more widespread use in the late 1970s. The related term computer age was used earlier still, starting in about 1970.²⁸ It was at around this time that computers began to be used more commonly in laboratories and academic settings, even if they had not yet become common as home appliances. This time it did not take three hundred years before the growth in information technology began to produce tangible benefits to human society. But it did take fifteen to twenty. The 1970s were the high point for vast amounts of theory applied to extremely small amounts of data, as Paul Krugman put it to me. We had begun to use computers to produce models of the world, but it took us some time to recognize how crude and assumption laden they were, and that the precision that computers were capable of was no substitute for predictive accuracy. In fields ranging from economics to epidemiology, this was an era in which bold predictions were made, and equally often failed.

In 1971, for instance, it was claimed that we would be able to predict earthquakes within a decade,²⁹ a problem that we are no closer to solving forty years later. Instead, the computer boom of the 1970s and 1980s produced a temporary decline in economic and scientific productivity. Economists termed this the productivity paradox. You can see the computer age everywhere but in the productivity statistics, wrote the economist Robert Solow in 1987.³⁰ The United States experienced four distinct recessions between 1969 and 1982.³¹ The late 1980s were a stronger period for our economy, but less so for countries elsewhere in the world. Scientific progress is harder to measure than economic progress.³² But one mark of it is the number of patents produced, especially relative to the investment in research and development. If it has become cheaper to produce a new invention, this suggests that we are using our information wisely and are forging it into knowledge. If it is becoming more expensive, this suggests that we are seeing signals in the noise and wasting our time on false leads. In the 1960s the United States spent about \$1.5 million (adjusted for inflation³³) per patent application³⁴ by an American 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

FIGURE I-3: RESEARCH AND DEVELOPMENT EXPENDITURES PER PATENT APPLICATION

As we came to more realistic views of what that new technology could accomplish for us, our research productivity began to improve again in the 1990s. We wandered up fewer blind alleys; computers began to improve our everyday lives and help our economy. Stories of prediction are often those of long-term progress but short-term regress. Many things that seem predictable over the long run foil our best-laid plans in the meanwhile. The Promise and Pitfalls of Big Data

The fashionable term now is Big Data. IBM estimates that we are generating 2.5 quintillion bytes of data each day, more than 90 percent of which was created in the last two years.³⁶ This exponential growth in information is sometimes seen as a cure-all, as computers were in the 1970s. Chris Anderson, the editor of *Wired* magazine, wrote in 2008 that the sheer volume of data would obviate the need for theory, and even the scientific method.³⁷ This is an emphatically pro-science and pro-technology book, and I think of it as a very optimistic one. But it argues that these views are badly mistaken. The numbers have no way of speaking for themselves. We speak for them. We imbue them with meaning. Like Caesar, we may construe them in self-serving ways that are detached from their objective reality. Data-driven predictions can succeed and they can fail. It is when we deny our role in the process that the odds of failure rise. Before we demand more of our data, we need to demand more of ourselves. This attitude might seem surprising if you know my background. I have a reputation for working with data and statistics and using them to make successful predictions. In 2003, bored at a consulting job, I designed a system called PECOTA, which sought to predict the statistics of Major League Baseball players. It contained a number of innovations: its forecasts were probabilistic, for instance, outlining a range of possible outcomes for each player and we found that it outperformed competing systems when we compared their results. In 2008, I founded the Web site *FiveThirtyEight*, which sought to forecast the upcoming election. The *FiveThirtyEight* forecasts correctly predicted the winner of the presidential contest in forty-nine of fifty states as well as the winner of all thirty-five U.S. Senate races. After the election, I was approached by a number of publishers who wanted to capitalize on the success of books such as *Moneyball* and *Freakonomics* that told the story of nerds conquering the world. This book was conceived of along those lines as an investigation of data-driven predictions in fields ranging from baseball to finance to national security. But in speaking with well more than one hundred experts in more than a dozen fields over the course of four years, reading hundreds of journal articles and books, and traveling everywhere from Las Vegas to Copenhagen in pursuit of my investigation, I came to realize that prediction in the era of Big Data

was not going very well. I had been lucky on a few levels: first, in having achieved success despite having made many of the mistakes that I will describe, and second, in having chosen my battles well. 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 and 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 IBMs Deep Blue and took down a world chess champion. But these cases of progress in forecasting must be weighed against a series of failures. If there is one thing that defines Americans one thing that makes us exceptional it is our belief in Cassius's idea that we are in control of our own fates. Our country was founded at the dawn of the Industrial Revolution by religious rebels who had seen that the free flow of ideas had helped to spread not just their religious beliefs, but also those of science and commerce. Most of our strengths and weaknesses as a nation our ingenuity and our industriousness, our arrogance and our impatience stem from our unshakable belief in the idea that we choose our own course. But the new millennium got off to a terrible start for Americans. We had not seen the September 11 attacks coming. The problem was not want of information. As had been the case in the 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 naive 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 nevertheless 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 to prepare us for those that did. The Fukushima nuclear reactor had been designed to handle a magnitude 8.6 earthquake, in part because some seismologists concluded that anything larger was impossible. Then came Japan's horrible magnitude 9.1 earthquake in March 2011. There are entire disciplines in which predictions have been failing, often at great cost to society. Consider something like biomedical research. In 2005, an Athens-raised medical researcher named John P. Ioannidis published a controversial paper titled Why Most Published Research Findings Are False.³⁹ The paper studied positive findings documented in peer-reviewed journals: descriptions of successful predictions of medical hypotheses carried out in laboratory experiments. It concluded that most of these findings were likely to fail when applied in the real world. Bayer Laboratories recently confirmed Ioannidis's hypothesis. They could not replicate about two-thirds of the positive findings claimed in medical journals when they attempted the experiments themselves.⁴⁰ Big Data will produce progress eventually. How quickly it does, and whether we regress in the meantime, will depend on us. Why the Future Shocks Us Biologically, we are not very different from our ancestors. But some stone-age strengths have become information-age weaknesses. Human beings do not have very many natural defenses. We are not all that fast, and we are not all that strong. We do not have claws or fangs or body armor. We cannot spit venom. We cannot camouflage ourselves. And we cannot fly. Instead, we survive by means of our wits. Our minds are quick. We are wired to detect patterns and respond to opportunities and threats without much hesitation. This need of finding patterns, humans have this more than other animals, I was told by Tomaso Poggio, an MIT neuroscientist who studies how our brains process information. Recognizing objects in difficult situations means generalizing. A newborn baby can recognize the basic pattern of a face. It has been learned by evolution, not by the individual. The problem, Poggio says, is that these evolutionary instincts sometimes lead us to see patterns when there are none there. People have been doing that all the time, Poggio said. Finding patterns in

random noise. The human brain is quite remarkable; it can store perhaps three terabytes of information.⁴¹ And yet that is only about one one-millionth of the information that IBM says is now produced in the world each day. So we have to be terribly selective about the information we choose to remember. Alvin Toffler, writing in the book *Future Shock* in 1970, predicted some of the consequences of what he called information overload. He thought our defense mechanism would be to simplify the world in ways that confirmed our biases, even as the world itself was growing more diverse and more complex.⁴² Our biological instincts are not always very well adapted to the information-rich modern world. Unless we work actively to become aware of the biases we introduce, the returns to additional information may be minimal or diminishing. The information overload after the birth of the printing press produced greater sectarianism. Now those different religious ideas could be testified to with more information, more conviction, more proof and less tolerance for dissenting opinion. The same phenomenon seems to be occurring today. Political partisanship began to increase very rapidly in the United States beginning at about the time that Toffler wrote *Future Shock* and it may be accelerating even faster with the advent of the Internet.⁴³ These partisan beliefs can upset the equation in which more information will bring us closer to the truth. A recent study in *Nature* found that the more informed that strong political partisans were about global warming, the less they agreed with one another.⁴⁴ Meanwhile, if the quantity of information is increasing by 2.5 quintillion bytes per day, the amount of useful information almost certainly isn't. Most of it is just noise, and the noise is increasing faster than the signal. There are so many hypotheses to test, so many data sets to mine but a relatively constant amount of objective truth. The printing press changed the way in which we made mistakes. Routine errors of transcription became less common. But when there was a mistake, it would be reproduced many times over, as in the case of the Wicked Bible. Complex systems like the World Wide Web have this property. They may not fail as often as simpler ones, but when they fail they fail badly. Capitalism and the Internet, both of which are incredibly efficient at propagating information, create the potential for bad ideas as well as good ones to spread. The bad ideas may produce disproportionate effects. In advance of the financial crisis, the system was so highly levered that a single lax assumption in the credit ratings agencies models played a huge role in bringing down the whole global financial system. Regulation is one approach to solving these problems. But I am suspicious that it is an excuse to avoid looking within ourselves for answers. We need to stop, and admit it: we have a prediction problem. We love to predict things and we aren't very good at it.

Prediction Solution If prediction is the central problem of this book, it is also its solution. Prediction is indispensable to our lives. Every time we choose a route to work, decide whether to go on a second date, or set money aside for a rainy day, we are making a forecast about how the future will proceed and how our plans will affect the odds for a favorable outcome. Not all of these day-to-day problems require strenuous thought; we can budget only so much time to each decision. Nevertheless, you are making predictions many times every day, whether or not you realize it. For this reason, this book views prediction as a shared enterprise rather than as a function that a select group of experts or practitioners perform. It is amusing to poke fun at the experts when their predictions fail. However, we should be careful with our Schadenfreude. To say our predictions are no worse than the experts is to damn ourselves with some awfully faint praise. Prediction does play a particularly important role in science, however. Some of you may be uncomfortable with a premise that I have been hinting at and will now state explicitly: we can never make perfectly objective predictions. They will always be tainted by our subjective point of view. But this book is emphatically against the nihilistic viewpoint that there is no objective truth. It asserts, rather, that a belief in the objective truth and a commitment to pursuing it is the first prerequisite of making better predictions. The forecasters next commitment is to realize that she perceives it imperfectly. Prediction is important because it connects subjective and objective reality. Karl Popper, the philosopher of science, recognized this view.⁴⁵ For Popper, a hypothesis was not scientific unless it was falsifiable meaning that it could be tested in the real world by means of a prediction. What should give us pause is that the few ideas we have tested aren't doing so well, and many of our ideas have not or cannot be tested at all. In economics, it is much easier to test an unemployment rate forecast than a claim about the effectiveness of stimulus spending. In political science, we can test models that are used to predict the outcome of elections, but a theory about how changes to political institutions might affect policy outcomes could take decades to verify. I do not go as far as Popper in asserting that such theories are therefore unscientific or that they lack any value. However, the fact that the few theories we can test have produced quite poor results suggests that many of the ideas we haven't tested are very wrong as well. We are undoubtedly living with many delusions that we do not even realize. But there is a way forward. It is not a solution that relies on half-baked policy ideas particularly given that I have

come to view our political system as a big part of the problem. Rather, the solution requires an attitudinal change. This attitude is embodied by something called Bayes's theorem, which I introduce in chapter 8. Bayes's theorem is nominally a mathematical formula. But it is really much more than that. It implies that we must think differently about our ideas and how to test them. We must become more comfortable with probability and uncertainty. We must think more carefully about the assumptions and beliefs that we bring to a problem. The book divides roughly into halves. The first seven chapters diagnose the prediction problem while the final six explore and apply Bayes's solution. Each chapter is oriented around a particular subject and describes it in some depth. There is no denying that this is a detailed book in part because that is often where the devil lies, and in part because my view is that a certain amount of immersion in a topic will provide disproportionately more insight than an executive summary. The subjects I have chosen are usually those in which there is some publicly shared information. There are fewer examples of forecasters making predictions based on private information (for instance, how a company uses its customer records to forecast demand for a new product). My preference is for topics where you can check out the results for yourself rather than having to take my word for it.

A Short Road Map to the Book The book weaves between examples from the natural sciences, the social sciences, and from sports and games. It builds from relatively straightforward cases, where the successes and failures of prediction are more easily demarcated, into others that require slightly more finesse. Chapters 1 through 3 consider the failures of prediction surrounding the recent financial crisis, the successes in baseball, and the realm of political prediction where some approaches have worked well and others haven't. They should get you thinking about some of the most fundamental questions that underlie the prediction problem. How can we apply our judgment to the data without succumbing to our biases? When does market competition make forecasts better and how can it make them worse? How do we reconcile the need to use the past as a guide with our recognition that the future may be different? Chapters 4 through 7 focus on dynamic systems: the behavior of the earth's atmosphere, which brings about the weather; the movement of its tectonic plates, which can cause earthquakes; the complex human interactions that account for the behavior of the American economy; and the spread of infectious diseases. These systems are being studied by some of our best scientists. But dynamic systems make forecasting more difficult, and predictions in these fields have not always gone very well. Chapters 8 through 10 turn toward solutions first by introducing you to a sports bettor who applies Bayes's theorem more expertly than many economists or scientists do, and then by considering two other games, chess and poker. Sports and games, because they follow well-defined rules, represent good laboratories for testing our predictive skills. They help us to a better understanding of randomness and uncertainty and provide insight about how we might forge information into knowledge. 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 forecasters 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 have 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.

It was October 23, 2008. The stock market was in free fall, having plummeted almost 30 percent over the previous five weeks. Once-esteemed companies like Lehman Brothers had gone bankrupt. Credit markets had all but ceased to function. Houses in Las Vegas had lost 40 percent of their value.¹ Unemployment was skyrocketing. Hundreds of billions of dollars had been committed to failing financial firms. Confidence in government was the lowest that pollsters had ever measured.² The presidential election was less than two weeks away. Congress, normally dormant so close to an election, was abuzz with activity. The bailout bills it had passed were sure to be unpopular³ and it needed to create every impression that the wrongdoers would be punished. The House Oversight Committee had called the heads of the three major credit-rating agencies, Standard Poors (SP), Moodys, and Fitch Ratings, to testify before them. The ratings agencies were charged with assessing the likelihood that trillions of dollars in mortgage-backed securities would go into default. To put it mildly, it appeared they had blown the call.

The Worst Prediction of a Sorry Lot The crisis of the late 2000s is often thought of as a failure of our political and financial institutions. It was obviously an economic failure of massive proportions. By 2011, four years after the Great Recession officially began, the American economy was still almost \$800 billion below its productive potential.⁴ I am convinced, however, that the best way to view the financial crisis is as a

failure of judgment a catastrophic failure of prediction. These predictive failures were widespread, occurring at virtually every stage during, before, and after the crisis and involving everyone from the mortgage brokers to the White House. The most calamitous failures of prediction usually have a lot in common. We focus on those signals that tell a story about the world as we would like it to be, not how it really is. We ignore the risks that are hardest to measure, even when they pose the greatest threats to our well-being. We make approximations and assumptions about the world that are much cruder than we realize. We abhor uncertainty, even when it is an irreducible part of the problem we are trying to solve. If we want to get at the heart of the financial crisis, we should begin by identifying the greatest predictive failure of all, a prediction that committed all these mistakes. The ratings agencies had given their AAA rating, normally reserved for a handful of the world's most solvent governments 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 Poors told investors, for instance, that when it rated a particularly complex type of security known as a collateralized debt obligation (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 next five years.⁶ This supposedly made it as safe as a AAA-rated corporate bond⁷ and safer than SP 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 SP'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 SP 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 were rated as being almost completely safe instead turned out to be almost completely unsafe. It was as if the weather forecast had been 86 degrees and sunny, and instead there was a blizzard.

FIGURE 1-1: FORECASTED AND ACTUAL 5-YEAR DEFAULT RATES FOR AAA-RATED CDO TRANCHES

When you make a prediction that goes so badly, you have a choice of how to explain it.

One path is to blame external circumstances what we might think of as bad luck. Sometimes this is a reasonable choice, or even the correct one. When the National Weather Service says there is a 90 percent chance of clear skies, but it rains instead and spoils your golf outing, you can't really blame them. Decades of historical data show that when the Weather Service says there is a 1 in 10 chance of rain, it really does rain about 10 percent of the time over the long run.* This explanation becomes less credible, however, when the forecaster does not have a history of successful predictions and when the magnitude of his error is larger. In these cases, it is much more likely that the fault lies with the forecaster's model of the world and not with the world itself. In the instance of CDOs, the ratings agencies had no track record at all: these were new and highly novel securities, and the default rates claimed by SP were not derived from historical data but instead were assumptions based on a faulty statistical model. Meanwhile, the magnitude of their error was enormous: AAA-rated CDOs were two hundred times more likely to default in practice than they were in theory. The ratings agencies shot at redemption would be to admit that the models had been flawed and the mistake had been theirs. But at the congressional hearing, they shirked responsibility and claimed to have been unlucky. They blamed an external contingency: the housing bubble. SP is not alone in having been taken by surprise by the extreme decline in the housing and mortgage markets. Deven Sharma, the head of Standard Poors, told Congress that October.¹² Virtually no one, be they homeowners, financial institutions, rating agencies, regulators or investors, anticipated what is coming. Nobody saw it coming. When you can't state your innocence, proclaim your ignorance: this is often the first line of defense when there is a failed forecast.¹³ But Sharma's statement was a lie, in the grand congressional tradition of I did not have sexual relations with that woman and I have never used steroids. 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 was 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 the 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 in reputable newspapers and periodicals.²¹ And yet, the ratings agencies whose job it is to measure risk in financial markets say that they missed it. It should tell you something that they seem to think of this as their best line of defense. The problems with their predictions ran very deep. I Don't Think They Wanted the Music to Stop None of the economists and investors I spoke with for this chapter had a favorable view of the ratings agencies. But they were divided on whether their bad ratings reflected avarice or ignorance: did they know any better? Jules Kroll is perhaps uniquely qualified to pass judgment on this question: he runs a ratings agency himself. Founded in 2009, Kroll Bond Ratings had just issued its first rating on a mortgage loan made to the builders of a gigantic shopping center in Arlington, Virginia when I met him at his office in New York in 2011. Kroll faults the ratings agencies most of all for their lack of surveillance. It is an ironic term coming from Kroll, who before getting into the ratings game had become modestly famous (and somewhat immodestly rich) from his original company, Kroll Inc., which acted as a sort of detective agency to patrol corporate fraud. They knew how to sniff out a scam such as the case of the kidnapers who took a hedge-fund billionaire hostage but foiled themselves by charging a pizza to his credit card.²² Kroll was sixty-nine years old when I met him, but his bloodhound instincts are keen and they were triggered when he began to examine what the ratings agencies were doing. Surveillance is a term of art in the ratings industry, Kroll told me. It means keeping investors informed as to what you're seeing. Every month you get a tape* of things like defaults on mortgages, prepayment of mortgages; you get a lot of data. That is the early warning: are things getting better or worse? The world expects you to keep them posted. The ratings agencies ought to have been just about the first ones to detect problems in the housing market, in other words. They had better information than anyone else: fresh data on whether thousands of borrowers were making their mortgage payments on time. But they did not begin to downgrade large batches of mortgage-backed securities until 2007 at which point the problems had become manifest and foreclosure rates had already doubled.²³ These are not stupid people, Kroll told me. They knew. I don't think they wanted the music to stop. Kroll Bond Ratings is one of ten registered NRSROs, or nationally recognized statistical rating organizations, firms that are licensed by the Securities and Exchange Commission to rate debt-backed securities. But Moodys, SP, and Fitch are three of the others, and they have had almost all the market share; SP and Moodys each rated almost 97 percent of the CDOs that were issued prior to the financial collapse.²⁴ One reason that SP and Moodys 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 SP and Moodys is often mandated by the bylaws of large pension funds,²⁵ about two-thirds of which²⁶ mention SP, Moodys, or both by name, requiring that they rate a piece of debt before the pension fund can purchase it.²⁷ SP and Moodys had taken advantage of their select status to build up exceptional profits despite picking rums out of Wall Street's reject pile.* Moodys²⁸ 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 Moodys to the highest profit margin of any company in the SP 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, Moodys 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 of their ratings until it was too late, the agencies had little incentive to compete on the basis of quality. The CEO of Moodys, Raymond McDaniel, explicitly told his board that ratings quality was the least important factor driving the company's profits.³² Instead their equation was simple. The ratings agencies were paid by the issuer of the CDO every time they rated one: the more CDOs, the more profit. A virtually unlimited number of CDOs could be created by combining different types of mortgages or when that got boring, combining different types of CDOs into derivatives of one another. Rarely did the ratings agencies turn down the opportunity to rate one. A government investigation later uncovered an instant-message exchange between two senior Moodys employees in which one claimed that a security could be structured by cows and Moodys would rate it.³³ In some cases, the ratings agencies went further still and abetted debt issuers in manipulating the ratings. In what it claimed was a nod to transparency,³⁴ SP provided the issuers with copies of their ratings software. This made it easy for the issuers to determine exactly how many bad mortgages they could add to the pool without seeing its rating decline.³⁵ The possibility of a housing bubble, and that it might burst, thus represented a threat to the ratings agencies' gravy train. Human beings have an

extraordinary capacity to ignore risks that threaten their livelihood, as though this will make them go away.

So perhaps Deven Sharmas claim isnt so implausibleperhaps the ratings agencies really had missed the housing bubble, even if others hadnt. In fact, however, the ratings agencies quite explicitly considered the possibility that there was a housing bubble. They concluded, remarkably, that it would be no big deal. A

memo provided to me by an SP spokeswoman, Catherine Mathis, detailed how SP had conducted a simulation in 2005 that anticipated a 20 percent decline in national housing prices over a two-year periodnot far from the roughly 30 percent decline in housing prices that actually occurred between 2006 and 2008. The

memo concluded that SPs existing 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 In some ways this is even more troubling than if the ratings agencies had missed the housing bubble entirely. In this book, Ill discuss the danger of unknown unknownsthe risks that we are not even aware of. Perhaps the only greater threat is the risks we think we have a handle on, but dont.* In these cases we not only fool ourselves, but our false confidence may be contagious. In the case of the ratings agencies, it helped to infect the entire financial system. The major difference between a thing that might go wrong and a thing that cannot possibly go wrong is that when a thing that cannot possibly go wrong goes wrong it usually turns out to be impossible to get at or repair, wrote Douglas Adams in *The Hitchhikers Guide to the Galaxy* series.³⁷ But how did the ratings agencies models, which had all the auspices of scientific precision, do such a poor job of describing

reality? How the Ratings Agencies Got It Wrong We have to dig a bit deeper to find the source of the problem. The answer requires a little bit of detail about how financial instruments like CDOs are structured, and a little bit about the distinction between uncertainty and risk. CDOs are collections of mortgage debt that

are broken into different pools, or tranches, some of which are supposed to be quite risky and others of which are rated as almost completely safe. My friend Anil Kashyap, who teaches a course on the financial crisis to students at the University of Chicago, has come up with a simplified example of a CDO, and Ill use a version of this example here. Imagine you have a set of five mortgages, each of which you assume has a 5 percent chance of defaulting. You can create a number of bets based on the status of these mortgages, each of which is progressively more risky. The safest of these bets, what Ill call the Alpha Pool, pays out unless all

five of the mortgages default. The riskiest, the Epsilon Pool, leaves you on the hook if any of the five mortgages defaults. Then there are other steps along the way. Why might an investor prefer making a bet on the Epsilon Pool to the Alpha Pool? Thats easybecause it will be priced more cheaply to account for the greater risk. But say youre a risk-averse investor, such as a pension fund, and that your bylaws prohibit you from investing in poorly rated securities. If youre going to buy anything, it will be the Alpha Pool, which will assuredly be rated AAA. The Alpha Pool consists of five mortgages, each of which has only a 5 percent chance of defaulting. You lose the bet only if all five actually do default. What is the risk of that

happening? Actually, that is not an easy questionand therein lies the problem. The assumptions and approximations you choose will yield profoundly different answers. If you make the wrong assumptions, your model may be extraordinarily wrong. One assumption is that each mortgage is independent of the others. In this scenario, your risks are well diversified: if a carpenter in Cleveland defaults on his mortgage, this will have no bearing on whether a dentist in Denver does. Under this scenario, the risk of losing your bet would be exceptionally smallthe equivalent of rolling snake eyes five times in a row. Specifically, it would

be 5 percent taken to the fifth power, which is just one chance in 3,200,000. This supposed miracle of diversification is how the ratings agencies claimed that a group of subprime mortgages that had just a B+ credit rating on average³⁸ which would ordinarily imply³⁹ more than a 20 percent chance of default⁴⁰ had almost no chance of defaulting when pooled together. The other extreme is to assume that the mortgages, instead of being entirely independent of one another, will all behave exactly alike. That is, either all five mortgages will default or none will. Instead of getting five separate rolls of the dice, youre now staking your bet on the outcome of just one. Theres a 5 percent chance that you will roll snake eyes and all the mortgages

will defaultmaking your bet 160,000 times riskier than you had thought originally.⁴¹ Which of these assumptions is more valid will depend on economic conditions. If the economy and the housing market are healthy, the first scenariothe five mortgages have nothing to do with one another might be a reasonable approximation. Defaults are going to happen from time to time because of unfortunate rolls of the dice: someone gets hit with a huge medical bill, or they lose their job. However, one persons default risk wont have much to do with anothers. But suppose instead that there is some common factor that ties the fate of these homeowners together. For instance: there is a massive housing bubble that has caused home prices to rise by 80 percent without any tangible improvement in the fundamentals. Now youve got trouble: if one

borrower defaults, the rest might succumb to the same problems. The risk of losing your bet has increased by orders of magnitude. The latter scenario was what came into being in the United States beginning in 2007 (we'll conduct a short autopsy on the housing bubble later in this chapter). But it was the former assumption of largely uncorrelated risks that the ratings agencies had bet on. Although the problems with this assumption were understood in the academic literature⁴² and by whistle-blowers at the ratings agencies⁴³ long before the housing bubble burst, the efforts the ratings agencies made to account for it were feeble. Moodys, 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 ones 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 of highly geometric and nonlinear mistakes. Moodys 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 percent 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 a broader sense, the ratings agencies problem was in being unable or uninterested in appreciating the distinction between risk and uncertainty. Risk, as first articulated by the economist Frank H. Knight in 1921,⁴⁵ is something that you can put a price on. Say that you'll win a poker hand unless your opponent draws to an inside straight: the chances of that happening are exactly 1 chance in 11.46 This is risk. It is not pleasant when you take a bad beat in poker, but at least you know the odds of it and can account for it ahead of time. In the long run, you'll make a profit from your opponents making desperate draws with insufficient odds. Uncertainty, on the other hand, is risk that is hard to measure. You might have some vague awareness of the demons lurking out there. You might even be acutely concerned about them. But you have no real idea how many of them there are or when they might strike. Your back-of-the-envelope estimate might be off by a factor of 100 or by a factor of 1,000; there is no good way to know. This is uncertainty. Risk greases the wheels of a free-market economy; uncertainty grinds them to a halt. The alchemy that the ratings agencies performed was to spin uncertainty into what looked and felt like risk. They took highly novel securities, subject to an enormous amount of systemic uncertainty, and claimed the ability to quantify just how risky they were. Not only that, but of all possible conclusions, they came to the astounding one that these investments were almost risk-free. Too many investors mistook these confident conclusions for accurate ones, and too few made backup plans in case things went wrong. And yet, while the ratings agencies bear substantial responsibility for the financial crisis, they were not alone in making mistakes. The story of the financial crisis as a failure of prediction can be told in three acts. Act I: The Housing Bubble An American home has not, historically speaking, been a lucrative investment. 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 shift 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. The housing boom of the 1950s, however, had almost nothing in common with the housing bubble of the 2000s. The comparison helps to reveal why the 2000s became such a mess. The postwar years were associated with a substantial shift in living patterns. Americans had emerged from the war with a glut of savings⁴⁸ and into an age of prosperity. There was a great demand for larger living spaces. Between 1940 and 1960, the homeownership rate surged to 62 percent from 44 percent,⁴⁹ with most of the growth concentrated in the suburbs.⁵⁰ Furthermore, the housing boom was accompanied by the baby boom: the U.S. population was growing at a rate of about 20 percent per decade after the war, about twice its rate of growth during the 2000s. This meant that the number of homeowners increased by about 80 percent during the decade meeting or exceeding the increase in housing prices. In the 2000s, by contrast, homeownership rates increased only modestly: to a peak of about 69 percent in 2005 from 65 percent a decade earlier.⁵¹ Few Americans who hadn't already bought a home were in a position to afford one. The 40th percentile of household incomes increased by a nominal 15 percent between 2000 and 2006⁵² not enough to cover inflation, let alone a new home. Instead, the housing boom had been artificially

enhanced through speculators looking to flip homes and through ever more dubious loans to ever less creditworthy consumers. The 2000s were associated with record-low rates of savings: barely above 1 percent in some years. But a mortgage was easier to obtain than ever.⁵³ Prices had become untethered from supply and demand, as lenders, brokers, and the ratings agencies all of whom profited in one way or another from every home sale strove to keep the party going. 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 to 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 Shiller 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 inflation, or about 0.06 percent annually. These homeowners can perhaps be excused for their overconfidence in the housing market. The housing bubble had seeped into the culture to the point where two separate TV programs one named Flip This House and the other named Flip That House were launched within ten days of each other in 2005. Even home buyers who weren't counting on a huge return on investment may have been concerned about keeping up with the Joneses. I can remember twenty years ago, on the road to Sacramento, there were no traffic jams, I was told by George Akerlof, a frequent colleague of Shillers, whose office at the University of California at Berkeley sits at the epicenter of some of the worst declines in housing prices. Now there tend to be traffic stoppages a good share of the way. That's what people were thinking if I don't buy now then I'm gonna pay the same price in five years for a house that's ten miles up the road. Whether homeowners believed that they couldn't lose on a home or couldn't choose to defer the purchase, conditions were growing grimmer by the month. By late 2007 there were clear signs of trouble: home prices had declined over the year in seventeen of the twenty largest markets.⁵⁸ More ominous was the sharp decline in housing permits, a leading indicator of housing demand, which had fallen by 50 percent from their peak.⁵⁹ Creditors, meanwhile finally seeing the consequences of their lax standards in the subprime lending market were becoming less willing to make loans. Foreclosures had doubled by the end of 2007.⁶⁰ Policy makers first instinct was to reflate the bubble. Governor Charlie Crist of Florida, one of the worst-hit states, proposed a \$10,000 credit for new home buyers.⁶¹ A bill passed by the U.S. Congress in February 2008 went further, substantially expanding the lending capacity of Fannie Mae and Freddie Mac in that hope that more home sales might be spurred.⁶² Instead, housing prices continued their inexorable decline, falling a further 20 percent during 2008. Act II: Leverage, Leverage, Leverage 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 Professional 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 equity 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. The decline in consumer spending that resulted as consumers came to take a more realistic view of their finances what economists call a wealth effect is variously estimated at between about 1.5 percent⁶⁹ and 3.5 percent⁷⁰ of GDP per year, potentially enough to turn average growth into a recession. But a garden-variety recession is one thing. A global financial crisis is another, and the wealth effect does not suffice to explain how the housing bubble triggered one. In fact, the housing market is a fairly small part of the financial system. In

2007, the total volume of home sales in the United States was about \$1.7 trillion⁷¹ when compared with the \$40 trillion in stocks that are traded every year. But in contrast to the activity that was taking place on Main Street, Wall Street was making bets on housing at furious rates. In 2007, the total volume of trades in mortgage-backed securities was about \$80 trillion.⁷¹ That meant that for every dollar that someone was willing to put in a mortgage, Wall Street was making almost \$50 worth of bets on the side.⁷² Now we have the makings of a financial crisis: home buyers bets were multiplied fifty times over. The problem can be summed up in a single word: leverage. If you borrow \$20 to wager on the Redskins to beat the Cowboys, that is a leveraged bet.* Likewise, its leverage when you borrow money to take out a mortgage or when you borrow money to bet on a mortgage-backed security. Lehman Brothers, in 2007, had a leverage ratio of about 33 to 1,⁷³ meaning that it had about \$1 in capital for every \$33 in financial positions that it held. This meant that if there was just a 3 to 4 percent decline in the value of its portfolio, Lehman Brothers would have negative equity and would potentially face bankruptcy.⁷⁴ Lehman was not alone in being highly levered: the leverage ratio for other major U.S. banks was about 30 and had been increasing steadily in the run-up to the financial crisis.⁷⁵ Although historical data on leverage ratios for U.S. banks is spotty, an analysis by the Bank of England on United Kingdom banks suggests that the overall degree of leverage in the system was either near its historical highs in 2007 or was perhaps altogether unprecedented.⁷⁶ What particularly distinguished Lehman Brothers, however, was its voracious appetite for mortgage-backed securities. The \$85 billion it held in mortgage-backed securities in 2007 was about four times more than the underlying value of its capital, meaning that a 25 percent decline in their value would likely be enough to bankrupt the company.⁷⁷ Ordinarily, investors would have been extremely reluctant to purchase assets like these or at least they would have hedged their bets very carefully. 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 showed 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's 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 may 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 AAA ratings and helped to enable a market for them that might not otherwise have existed. The market was counting on them to be the Debbie Downer of the mortgage party but they were acting more like Robert Downey Jr. Lehman Brothers, in particular, could have used a designated driver. In a conference call in March 2007, Lehman CFO Christopher O'Meara told investors that the recent hiccup in the markets did not concern him and that Lehman hoped to do some bottom fishing from others who were liquidating their positions prematurely.⁷⁹ He explained that the credit quality in the mortgage market was very strong—a conclusion that could only have been reached by looking at the AAA ratings for the securities and not at the subprime quality of the collateral. Lehman had bought a lemon. One year later, as the housing bubble began to burst, Lehman was desperately trying to sell its position. But with the skyrocketing premiums that investors were demanding for credit default swap investments that pay you out in the event of a default and which therefore provide the primary means of insurance against one they were only able to reduce their exposure by about 20 percent.⁸⁰ It was too little and too late, and Lehman went bankrupt on September 14, 2008. *Intermission: Fear Is the New Greed* The precise sequence of events that followed the Lehman bankruptcy could fill its own book (and has been described in some excellent ones, like *Too Big to Fail*). It should suffice to remember that when a financial company dies, it can continue to haunt the economy through an afterlife of unmet obligations. If Lehman Brothers was no longer able to pay out on the losing bets that it had made, this meant that somebody else suddenly had a huge hole in his portfolio. Their problems, in turn, might affect yet other companies, with the effects cascading throughout the financial system. Investors and lenders, gawking at the accident but unsure about who owed what to

whom, might become unable to distinguish the solvent companies from the zombies and unwilling to lend money at any price, preventing even healthy companies from functioning effectively. It is for this reason that governments at great cost to taxpayers as well as to their popularity sometimes bail out failing financial firms. But the Federal Reserve, which did bail out Bear Stearns and AIG, elected not to do so for Lehman Brothers, defying the expectations of investors and causing the Dow to crash by 500 points when it opened for business the next morning. Why the government bailed out Bear Stearns and AIG but not Lehman remains unclear. One explanation is that Lehman had been so irresponsible, and its financial position had become so decrepit, that the government wasn't sure what could be accomplished at what price and didn't want to chase good money after bad.⁸¹ Larry Summers, who was the director of the National Economic Council at the time that I met him in the White House in December 2009,⁸² told me that the United States might have had a modestly better outcome had it bailed out Lehman Brothers. But with the excess of leverage in the system, some degree of pain was inevitable. It was a self-denying prophecy, Summers told me of the financial crisis. Everybody leveraged substantially, and when everybody leverages substantially, there's substantial fragility, and their complacency proves to be unwarranted. Lehman was a burning cigarette in a very dry forest, he continued a little later. If that hadn't happened, it's quite likely that something else would have. Summers thinks of the American economy as consisting of a series of feedback loops. One simple feedback is between supply and demand. Imagine that you are running a lemonade stand.⁸³ You lower the price of lemonade and sales go up; raise it and they go down. If you're making lots of profit because it's 100 degrees outside and you're the only lemonade stand on the block, the annoying kid across the street opens his own lemonade stand and undercuts your price. Supply and demand is an example of a negative feedback: as prices go up, sales go down. Despite their name, negative feedbacks are a good thing for a market economy. Imagine if the opposite were true and as prices went up, sales went up. You raise the price of lemonade from 25 cents to \$2.50 but instead of declining, sales double.⁸⁴ Now you raise the price from \$2.50 to \$25 and they double again. Eventually, you're charging \$46,000 for a glass of lemonade—the average income in the United States each year—and all 300 million Americans are lined up around the block to get their fix. This would be an example of a positive feedback. And while it might seem pretty cool at first, you'd soon discover that everyone in the country had gone broke on lemonade. There would be nobody left to manufacture all the video games you were hoping to buy with your profits. Usually, in Summers's view, negative feedbacks predominate in the American economy, behaving as a sort of thermostat that prevents it from going into recession or becoming overheated. Summers thinks one of the most important feedbacks is between what he calls fear and greed. Some investors have little appetite for risk and some have plenty, but their preferences balance out: if the price of a stock goes down because a company's financial position deteriorates, the fearful investor sells his shares to a greedy one who is hoping to bottom-feed. Greed and fear are volatile quantities, however, and the balance can get out of whack. When there is an excess of greed in the system, there is a bubble. When there is an excess of fear, there is a panic. Ordinarily, we benefit from consulting our friends and neighbors before making a decision. But when their judgment is compromised, this means that ours will be too. People tend to estimate the prices of houses by making comparisons to other houses⁸⁵ if the three-bedroom home in the new subdivision across town is selling for \$400,000, the colonial home around the block suddenly looks like a steal at \$350,000. Under these circumstances, if the price of one house increases, it may make the other houses seem more attractive rather than less. Or say that you are considering buying another type of asset: a mortgage-backed security. This type of commodity may be even harder to value. But the more investors buy them and the more the ratings agencies vouch for them, the more confidence you might have that they are safe and worthwhile investments. Hence, you have a positive feedback and the potential for a bubble. A negative feedback did eventually rein in the housing market: there weren't any Americans left who could afford homes at their current prices. For that matter, many Americans who had bought homes couldn't really afford them in the first place, and soon their mortgages were underwater. But this was not until trillions of dollars in bets, highly leveraged and impossible to unwind without substantial damage to the economy, had been made on the premise that all the people buying these assets couldn't possibly be wrong. We had too much greed and too little fear, Summers told me in 2009. Now we have too much fear and too little greed.

Act III: This Time Wasn't Different

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. Reinhart and Rogoff's history lesson was one that the White House might have done more to heed. Soon, they would be responsible for their own notoriously bad prediction. In January 2009, as Barack Obama was about to take the oath of office, the White House's incoming economic team led by Summers and Christina Romer, the chair of the Council of Economic Advisers, were charged with preparing the blueprint for a massive stimulus package that was supposed to make up for the lack of demand among consumers and businesses. Romer thought that \$1.2 trillion in stimulus was called for.⁸⁹ Eventually, the figure was revised downward to about \$800 billion after objections from the White House's political team that a trillion-dollar price would be difficult to sell to Congress. To help pitch the Congress and the country on the stimulus, Romer and her colleagues prepared a memo⁹⁰ outlining the depth of the crisis and what the stimulus might do to ameliorate it. The memo prominently featured a graphic predicting how the unemployment rate would track with and without the stimulus. Without the stimulus, the memo said, the unemployment rate, which had been 7.3 percent when last reported in December 2008, would peak at about 9 percent in early 2010. But with the stimulus, employment would never rise above 8 percent and would begin to turn downward as early as July 2009. Congress passed the stimulus on a party-line vote in February 2009. But unemployment continued to rise to 9.5 percent in July and then to a peak of 10.1 percent in October 2009. This was much worse than the White House had projected even under the no stimulus scenario. Conservative bloggers cheekily updated Romer's graphic every month but with the actual unemployment rate superimposed on the too-cheery projections (figure 1-6).

FIGURE 1-6: WHITE HOUSE ECONOMIC PROJECTIONS, JANUARY 2009

People see this graphic now and come to different and indeed entirely opposite conclusions about it. Paul Krugman, who had argued from the start that the stimulus was too small,⁹¹ sees it as proof that the White House had dramatically underestimated the shortfall in demand. The fact that unemployment didn't come down much in the wake of this particular stimulus means that we knew we were facing one hell of a shock from the financial crisis, he told me. Other economists, of course, take the graph as evidence that the stimulus had completely failed.⁹² The White House can offer its version of SPs everyone else made the same mistake defense. Its forecasts were largely in line with those issued by independent economists at the time.⁹³ Meanwhile, the initial economic statistics had significantly underestimated the magnitude of the crisis.⁹⁴ The government's first estimate—the one available to Romer and Summers at the time the stimulus was being sold—was that GDP had declined at a rate of 3.8 percent in the fall of 2008.⁹⁵ In fact, the financial crisis had taken more than twice as large a bite out of the economy. The actual rate of GDP decline had been closer to 9 percent,⁹⁶ meaning that the country was about \$200 billion poorer than the government first estimated. Perhaps the White House's more inexcusable error was in making such a precise-seeming forecast and in failing to prepare the public for the eventuality that it might be wrong. No economist, whether in the White House or elsewhere, has been able to predict the progress of major economic indicators like the unemployment rate with much success. (I take a more detailed look at macroeconomic forecasting in chapter 6.) The uncertainty in an unemployment rate forecast⁹⁷ made during a recession had historically been about plus or minus 2 percent.⁹⁸ So even if the White House thought 8 percent unemployment was the most likely outcome, it might easily enough have wound up in the double digits instead (or it might have declined to as low as 6 percent). There is also considerable uncertainty about how effective stimulus spending really is. Estimates of the multiplier effect—how much each dollar in stimulus spending contributes to growth—vary radically from study to study,⁹⁹ with some claiming that \$1 in stimulus spending returns as much as \$4 in GDP growth and others saying the return is just 60 cents on the dollar. When you layer the large uncertainty intrinsic to measuring the effects of stimulus atop the large uncertainty intrinsic to making macroeconomic forecasts of any kind, you have the potential for a prediction that goes very badly.

What the Forecasting Failures Had in Common

There were at least four major failures of prediction that accompanied the financial crisis. The housing bubble can be thought of as a poor prediction. Homeowners and investors thought that rising prices implied that home values would continue to rise, when in fact history suggested this made them prone to decline. There was a failure on the part of the ratings agencies, as well as by banks like Lehman Brothers, to understand how risky mortgage-backed securities were. Contrary to the assertions they made

before Congress, the problem was not that the ratings agencies failed to see the housing bubble. Instead, their forecasting models were full of faulty assumptions and false confidence about the risk that a collapse in housing prices might present. There was a widespread failure to anticipate how a housing crisis could trigger a global financial crisis. It had resulted from the high degree of leverage in the market, with \$50 in side bets staked on every \$1 that an American was willing to invest in a new home. Finally, in the immediate aftermath of the financial crisis, there was a failure to predict the scope of the economic problems that it might create. Economists and policy makers did not heed Reinhart and Rogoffs finding that financial crises typically produce very deep and long-lasting recessions. There is a common thread among these failures of prediction. In each case, as people evaluated the data, they ignored a key piece of context: The confidence that homeowners had about housing prices may have stemmed from the fact that there had not been a substantial decline in U.S. housing prices in the recent past. However, there had never before been such a widespread increase in U.S. housing prices like the one that preceded the collapse. The confidence that the banks had in Moodys and SPs ability to rate mortgage-backed securities may have been based on the fact that the agencies had generally performed competently in rating other types of financial assets. However, the ratings agencies had never before rated securities as novel and complex as credit default options. The confidence that economists had in the ability of the financial system to withstand a housing crisis may have arisen because housing price fluctuations had generally not had large effects on the financial system in the past. However, the financial system had probably never been so highly leveraged, and it had certainly never made so many side bets on housing before. The confidence that policy makers had in the ability of the economy to recuperate quickly from the financial crisis may have come from their experience of recent recessions, most of which had been associated with rapid, V-shaped recoveries. However, those recessions had not been associated with financial crises, and financial crises are different. There is a technical term for this type of problem: the events these forecasters were considering were out of sample. When there is a major failure of prediction, this problem usually has its fingerprints all over the crime scene. What does the term mean? A simple example should help to explain it. Out of Sample, Out of Mind: A Formula for a Failed Prediction Suppose that youre a very good driver. Almost everyone thinks theyre a good driver, 100 but you really have the track record to prove it: just two minor fender benders in thirty years behind the wheel, during which time you have made 20,000 car trips. Youre also not much of a drinker, and one of the things youve absolutely never done is driven drunk. But one year you get a little carried away at your office Christmas party. A good friend of yours is leaving the company, and youve been under a lot of stress: one vodka tonic turns into about twelve. Youre blitzed, three sheets to the wind. Should you drive home or call a cab? That sure seems like an easy question to answer: take the taxi. And cancel your morning meeting. But you could construct a facetious argument for driving yourself home that went like this: out of a sample of 20,000 car trips, youd gotten into just two minor accidents, and gotten to your destination safely the other 19,998 times. Those seem like pretty favorable odds. Why go through the inconvenience of calling a cab in the face of such overwhelming evidence? The problem, of course, is that of those 20,000 car trips, none occurred when you were anywhere near this drunk. Your sample size for drunk driving is not 20,000 trips but zero, and you have no way to use your past experience to forecast your accident risk. This is an example of an out-of-sample problem. As easy as it might seem to avoid this sort of problem, the ratings agencies made just this mistake. Moodys estimated the extent to which mortgage defaults were correlated with one another by building a model from past data specifically, they looked at American housing data going back to about the 1980s.¹⁰¹ The problem is that from the 1980s through the mid-2000s, home prices were always steady or increasing in the United States. Under these circumstances, the assumption that one homeowners mortgage has little relationship to anothers was probably good enough. But nothing in that past data would have described what happened when home prices began to decline in tandem. The housing collapse was an out-of-sample event, and their models were worthless for evaluating default risk under those conditions. The Mistakes That Were Made and What We Can Learn from Them Moodys was not completely helpless, however. They could have come to some more plausible estimates by expanding their horizons. The United States had never experienced such a housing crash before but other countries had, and the results had been ugly. Perhaps if Moodys had looked at default rates after the Japanese real estate bubble, they could have had some more realistic idea about the precariousness of mortgage-backed securities and they would not have stamped their AAA rating on them. But forecasters often resist considering these out-of-sample problems. When we expand our sample to include events further apart from us in time and space, it often means that we will encounter cases in which the relationships we are studying did not hold up as well as we are

accustomed to. The model will seem to be less powerful. It will look less impressive in a PowerPoint presentation (or a journal article or a blog post). We will be forced to acknowledge that we know less about the world than we thought we did. Our personal and professional incentives almost always discourage us from doing this. We forget we willfully ignore that our models are simplifications of the world. We figure that if we make a mistake, it will be at the margin. In complex systems, however, mistakes are not measured in degrees but in whole orders of magnitude. SP and Moodys underestimated the default risk associated with CDOs by a factor of two hundred. Economists thought there was just a 1 in 500 chance of a recession as severe as what actually occurred. One of the pervasive risks that we face in the information age, as I wrote in the introduction, is that even if the amount of knowledge in the world is increasing, the gap between what we know and what we think we know may be widening. This syndrome is often associated with very precise-seeming predictions that are not at all accurate. Moodys carried out their calculations to the second decimal place but they were utterly divorced from reality. This is like claiming you are a good shot because your bullets always end up in about the same place even though they are nowhere near the target (figure 1-7).

FIGURE 1-7: ACCURACY VERSUS PRECISION Financial crises and most other failures of prediction stem from this false sense of confidence. Precise forecasts masquerade as accurate ones, and some of us get fooled and double-down our bets. It is exactly when we think we have overcome the flaws in our judgment that something as powerful as the American economy can be brought to a screeching halt. For many people, political prediction is synonymous with the television program *The McLaughlin Group*, a political roundtable that has been broadcast continually each Sunday since 1982 and parodied by *Saturday Night Live* for nearly as long. The show, hosted by John McLaughlin, a cantankerous octogenarian who ran a failed bid for the United States Senate in 1970, treats political punditry as sport, cycling through four or five subjects in the half hour, with McLaughlin barking at his panelists for answers on subjects from Australian politics to the prospects for extraterrestrial intelligence. At the end of each edition of *The McLaughlin Group*, the program has a final segment called *Predictions*, in which the panelists are given a few seconds to weigh in on some matter of the day. Sometimes, the panelists are permitted to pick a topic and make a prediction about anything even vaguely related to politics. At other times, McLaughlin calls for a forced prediction, a sort of pop quiz that asks them their take on a specific issue. Some of McLaughlin's questions say, to name the next Supreme Court nominee from among several plausible candidates are difficult to answer. But others are softballs. On the weekend before the 2008 presidential election, for instance, McLaughlin asked his panelists whether John McCain or Barack Obama was going to win.¹ That one ought not to have required very much thought. Barack Obama had led John McCain in almost every national poll since September 15, 2008, when the collapse of Lehman Brothers had ushered in the worst economic slump since the Great Depression. Obama also led in almost every poll of almost every swing state: in Ohio and Florida and Pennsylvania and New Hampshire and even in a few states that Democrats don't normally win, like Colorado and Virginia. Statistical models like the one I developed for *FiveThirtyEight* suggested that Obama had in excess of a 95 percent chance of winning the election. Betting markets were slightly more equivocal, but still had him as a 7 to 1 favorite.² But McLaughlin's first panelist, Pat Buchanan, dodged the question. The undecideds will decide this weekend, he remarked, drawing guffaws from the rest of the panel. Another guest, the *Chicago Tribune's* Clarence Page, said the election was too close to call. Fox News Monica Crowley was bolder, predicting a McCain win by half a point. Only *Newsweek's* Eleanor Clift stated the obvious, predicting a win for the Obama-Biden ticket. The following Tuesday, Obama became the president-elect with 365 electoral votes to John McCain's 173 almost exactly as polls and statistical models had anticipated. While not a landslide of historic proportions, it certainly hadn't been too close to call: Obama had beaten John McCain by nearly ten million votes. Anyone who had rendered a prediction to the contrary had some explaining to do. There would be none of that on *The McLaughlin Group* when the same four panelists gathered again the following week.³ 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 outcome 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 was 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 Group 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. For each panelist, I calculated a percentage score, essentially reflecting the number of predictions they got right. Clift and the 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 as 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 a 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. But what about those who are paid for the accuracy and thoroughness of their scholarship rather than the volume of their opinions? Are political scientists, or analysts at Washington think tanks, any better at making predictions? Are Political Scientists Better Than Pundits? The disintegration of the Soviet Union and other countries of the Eastern bloc occurred at a remarkably fast pace and all things considered, in a remarkably orderly way.* On June 12, 1987, Ronald Reagan stood at the Brandenburg Gate and implored Mikhail Gorbachev to tear down the Berlin Wall an applause line that seemed as audacious as John F. Kennedy's pledge to send a man to the moon. Reagan was prescient; less than two years later, the wall had fallen. On November 16, 1988, the parliament of the Republic of Estonia, a nation about the size of the state of Maine, declared its independence from the mighty USSR. Less than three years later, Gorbachev parried a coup attempt from hard-liners in Moscow and the Soviet flag was lowered for the last time before the Kremlin; Estonia and the other Soviet Republics would soon become independent nations. If the fall of the Soviet empire seemed predictable after the fact, however, almost no mainstream political scientist had seen it coming. The few exceptions were often the subject of ridicule.⁸ If political scientists couldn't predict the downfall of the Soviet Union perhaps the most important event in the latter half of the twentieth century then what exactly were they good for? Philip Tetlock, a professor of psychology and political science, then at the University of California at Berkeley,⁹ was asking some of the same questions. As it happened, he had undertaken an ambitious and unprecedented experiment at the time of the USSR's collapse. Beginning in 1987, Tetlock started collecting predictions from a broad array of experts in academia and government on a variety of topics in domestic politics, economics, and international relations.¹⁰ Political experts had difficulty anticipating the USSR's collapse, Tetlock found, because a prediction that not only forecast the regime's demise but also understood the reasons for it required different strands of argument to be woven together. There was nothing inherently contradictory about these ideas, but they tended to emanate from people on different sides of the political spectrum,¹¹ and scholars firmly entrenched in one ideological camp were unlikely to have embraced them both. On the one hand, Gorbachev was clearly a major part of the story his desire for reform had been sincere. Had Gorbachev chosen to become an accountant or a poet instead of entering politics, the Soviet Union might have survived at least a few years longer. Liberals were more likely to hold this sympathetic view of Gorbachev. Conservatives were less trusting of him, and some regarded his talk of glasnost as little more than posturing. Conservatives, on the other hand, were more instinctually critical of communism. They were quicker to understand that the USSR's economy was failing and that life was becoming increasingly difficult for the average citizen. 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 fact, 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 was shrinking by as much as 5 percent annually, with inflation well into the double digits. Take these two factors together, and the Soviet Union's collapse is fairly easy to envision. By opening the country's media and its markets and giving his citizens greater democratic authority, Gorbachev had provided his people with the mechanism to catalyze a regime change. And because of the dilapidated state of the country's economy, they were happy to take him up on his offer. The center was too weak to hold: not only were Estonians sick of Russians, but Russians were nearly as sick of Estonians, since the satellite republics contributed less to the Soviet economy than they received in subsidies from Moscow.¹⁴ Once the dominoes began falling in Eastern Europe—Czechoslovakia, Poland, Romania, Bulgaria, Hungary, and East Germany were all in the midst of revolution by the end of 1989—there was little Gorbachev or anyone else could do to prevent them from caving the country in. A lot of Soviet scholars understood parts of the problem, but few experts had put all the puzzle pieces together, and almost no one had forecast the USSR's sudden collapse. Tetlock, inspired by the example of the Soviet Union, began to take surveys of expert opinion in other areas, asking the experts to make predictions about the Gulf War, the Japanese real-estate bubble, the potential secession of Quebec from Canada, and almost every other major event of the 1980s and 1990s. Was the failure to predict the collapse of the Soviet Union an anomaly, or does expert political analysis rarely live up to its billing? His studies, which spanned more than fifteen years, were eventually published in the 2005 book *Expert Political Judgment*. Tetlock's conclusion was damning. The experts in his survey—regardless of their occupation, experience, or subfield—had done barely any better than random chance, and they had done worse than even rudimentary statistical methods at predicting future political events. They were grossly overconfident and terrible at calculating probabilities: about 15 percent of events that they claimed had no chance of occurring in fact happened, while about 25 percent of those that they said were absolutely sure things in fact failed to occur.¹⁵ It didn't matter whether the experts were making predictions about economics, domestic politics, or international affairs; their judgment was equally bad across the board.

The Right Attitude for Making Better Predictions: Be Foxy While the experts' performance was poor in the aggregate, however, Tetlock found that some had done better than others. On the losing side were those experts whose predictions were cited most frequently in the media. The more interviews that an expert had done with the press, Tetlock found, the worse his predictions tended to be. Another subgroup of experts had done relatively well, however. Tetlock, with his training as a psychologist, had been interested in the experts' cognitive styles—how they thought about the world. So he administered some questions lifted from personality tests to all the experts. On the basis of their responses to these questions, Tetlock was able to classify his experts along a spectrum between what he called hedgehogs and foxes. The reference to hedgehogs and foxes comes from the title of an Isaiah Berlin essay on the Russian novelist Leo Tolstoy: *The Hedgehog and the Fox*. Berlin had in turn borrowed his title from a passage attributed to the Greek poet Archilochus: *The fox knows many little things, but the hedgehog knows one big thing*. Unless you are a fan of Tolstoy or of flowery prose, you'll have no particular reason to read Berlin's essay. But the basic idea is that writers and thinkers can be divided into two broad categories: Hedgehogs are type A personalities who believe in Big Ideas in governing principles about the world that behave as though they were physical laws and undergird virtually every interaction in society. Think Karl Marx and class struggle, or Sigmund Freud and the unconscious. Or Malcolm Gladwell and the tipping point. Foxes, on the other hand, are scrappy creatures who believe in a plethora of little ideas and in taking a multitude of approaches toward a problem. They tend to be more tolerant of nuance, uncertainty, complexity, and dissenting opinion. If hedgehogs are hunters, always looking out for the big kill, then foxes are gatherers. Foxes, Tetlock found, are considerably better at forecasting than hedgehogs. They had come closer to the mark on the Soviet Union, for instance. Rather than seeing the USSR in highly ideological terms—as an intrinsically evil empire, or as a relatively successful (and perhaps even admirable) example of a Marxist economic system—they instead saw it for what it was: an increasingly dysfunctional nation that was in danger of coming apart at the seams. Whereas the hedgehogs' forecasts were barely any better than random chance, the foxes demonstrated predictive skill.

FIGURE 2-2: ATTITUDES OF FOXES AND HEDGEHOGS Why Hedgehogs Make Better Television Guests I met Tetlock for lunch one winter afternoon at the Hotel Durant, a stately and sunlit property just off the Berkeley campus. Naturally enough, Tetlock revealed himself to be a fox: soft-spoken and studious, with a habit of pausing for twenty or thirty seconds before answering my questions (lest he provide me with too incautiously considered a response). What are the incentives for a public intellectual? Tetlock asked me. There are some academics who are quite content to be relatively anonymous. But there are other people who aspire to be public intellectuals, to be pretty bold and to attach nonnegligible probabilities to fairly dramatic

change. That's much more likely to bring you attention. 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, and his strategy seems to be to make as dramatic a prediction as possible when given the chance. In 2005, Morris proclaimed that George W. Bush's handling of Hurricane Katrina would help Bush to regain his standing with the public.¹⁶ On the eve 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-three 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. Foxes sometimes have more trouble fitting into type A cultures like television, business, and politics. Their belief that many problems are hard to forecast and that we should be explicit about accounting for these uncertainties may be mistaken for a lack of self-confidence. Their pluralistic approach may be mistaken for a lack of conviction; Harry Truman famously demanded a one-handed economist, frustrated that the foxes in his administration couldn't give him an unqualified answer. But foxes happen to make much better predictions. They are quicker to recognize how noisy the data can be, and they are less inclined to chase false signals. They know more about what they don't know. If you're looking for a doctor to predict the course of a medical condition or an investment adviser to maximize the return on your retirement savings, you may want to entrust a fox. She might make more modest claims about what she is able to achieve, but she is much more likely to actually realize them. Why Political Predictions Tend to Fail: Fox-like attitudes may be especially important when it comes to making predictions about politics. There are some particular traps that can make suckers of hedgehogs in the arena of political prediction and which foxes are more careful to avoid. One of these is simply partisan ideology.

Morris, despite having advised Bill Clinton, identifies as a Republican and raises funds for their candidates, and his conservative views fit in with those of his network, Fox News. But 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 a 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 as 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. But while Tetlock found that left-wing and right-wing hedgehogs made especially poor predictions, he also found that foxes of all political persuasions were more immune from these effects.²⁰ Foxes may have emphatic convictions about the way the world ought to be. But they can usually separate that from their analysis of the way that the world actually is and how it is likely to be in the near future. Hedgehogs, by contrast, have more trouble distinguishing their rooting interest from their analysis. Instead, in Tetlock's words, they create a blurry fusion between facts and values all lumped together. They take a prejudicial view toward the evidence, seeing what they want to see and not what is really there. You can apply Tetlock's test to diagnose whether you are a hedgehog: Do your predictions improve when you have access to more information? In theory, more information should give your predictions a wind at their back; you can always ignore the information if it doesn't seem to be helpful. But hedgehogs often trap themselves in the briar patch. Consider the case of the National Journal Political Insiders Poll, a survey of roughly 180 politicians, political consultants, pollsters, and pundits. The survey is divided between Democratic and Republican partisans, but

both groups are asked the same questions. Regardless of their political persuasions, this group leans hedgehog: political operatives are proud of their battle scars, and see themselves as locked in a perpetual struggle against the other side of the cocktail party. A few days ahead of the 2010 midterm elections, National Journal asked its panelists whether Democrats were likely to retain control of both the House and the Senate.²¹ There was near-universal agreement on these questions: Democrats would keep the Senate but Republicans would take control of the House (the panel was right on both accounts). Both the Democratic and Republican insiders were also almost agreed on the overall magnitude of Republican gains in the House; the Democratic experts called for them to pick up 47 seats, while Republicans predicted a 53-seat gain—a trivial difference considering that there are 435 House seats. National Journal, however, also asked its panelists to predict the outcome of eleven individual elections, a mix of Senate, House, and gubernatorial races. Here, the differences were much greater. The panel split on the winners they expected in the Senate races in Nevada, Illinois, and Pennsylvania, the governors race in Florida, and a key House race in Iowa. Overall, Republican panelists expected Democrats to win just one of the eleven races, while Democratic panelists expected them to win 6 of the 11. (The actual outcome, predictably enough, was somewhere in the middle—Democrats won three of the eleven races that National Journal had asked about.²²) Obviously, partisanship plays some role here: Democrats and Republicans were each rooting for the home team. That does not suffice to explain, however, the unusual divide in the way that the panel answered the different types of questions. When asked in general terms about how well Republicans were likely to do, there was almost no difference between the panelists. They differed profoundly, however, when asked about specific cases—these brought the partisan differences to the surface.²³ Too much information can be a bad thing in the hands of a hedgehog. The question of how many seats Republicans were likely to gain on Democrats overall is an abstract one: unless you'd studied all 435 races, there was little additional detail that could help you to resolve it. By contrast, when asked about any one particular race—say, the Senate race in Nevada—the panelists had all kinds of information at their disposal: not just the polls there, but also news accounts they'd read about the race, gossip they'd heard from their friends, or what they thought about the candidates when they saw them on television. They might even know the candidates or the people who work for them personally. Hedgehogs who have lots of information construct stories—stories that are neater and tidier than the real world, with protagonists and villains, winners and losers, climaxes and denouements—and, usually, a happy ending for the home team. The candidate who is down ten points in the polls is going to win, goddamnit, because I know the candidate and I know the voters in her state, and maybe I heard something from her press secretary about how the polls are tightening—and have you seen her latest commercial?

Revue de presse Outstanding... fun to read... Silver has produced a signal that is a pleasure to follow (Tim Harford Financial Times) Engagingly written... wholly satisfying... one of the more momentous books of the decade (The New York Times Book) Fascinating... Statisticians are to our age what engineers were to the Victorians, the makers of the particular forms of truth we value and crave. Nate Silver, to pursue the analogy, is being tipped to be our age's Brunel (Bryan Appleyard Sunday Times) Balanced, intelligent and erudite (Spectator) Is there anything now that Nate Silver could tell us that we wouldn't believe? (Jonathan Freedland) In this important book, Nate Silver explains why the performance of experts varies from prescient to useless and why we must plan for the unexpected. Must reading for anyone who cares about what might happen next (Richard Thaler, author of Nudge) The inhabitants of Westminster are speed-reading The Signal and the Noise by Nate Silver, the New York Times statistician who called the election with cool accuracy. They will find the book remarkable and rewarding (Sunday Telegraph Matthew d'Ancona) The Galileo of number crunchers (Independent) A surprisingly accessible peek into the world of mathematical probability (Daily Telegraph) A whirlwind tour of the success and failure of predictions in a wide variety of fields... Mr. Silver's breezy style makes even the most difficult statistical material accessible. What is more, his arguments and examples are painstakingly researched (Wall Street Journal) Engaging... Silver displays a knack not just for mining data but for explaining his thinking in an accessible manner (Bloomberg) A supremely valuable resource for anyone who wants to make good guesses about the future, or who wants to assess the guesses made by others. In other words, everyone (The Boston Globe) Engaging and sophisticated... [An] entertaining popularization of a subject that scares many people off (Slate) Here's a prediction: after you read The Signal and the Noise, you'll have much more insight into why some models work well—and also why many don't. You'll learn to pay more attention to weather forecasts for the coming week—and none at all for weather forecasts beyond that. Nate Silver takes a complex, difficult subject and makes it fun, interesting, and relevant (Peter Orszag, Bloomberg columnist and former Director of the Office

of Management and Budget under President Barack Obama)Nate Silver is a new kind of political superstar. One who actually knows what he's talking about...he's singlehandedly shown that most political punditry is about as effective a method of truth-seeking as the ducking stool (Observer New)