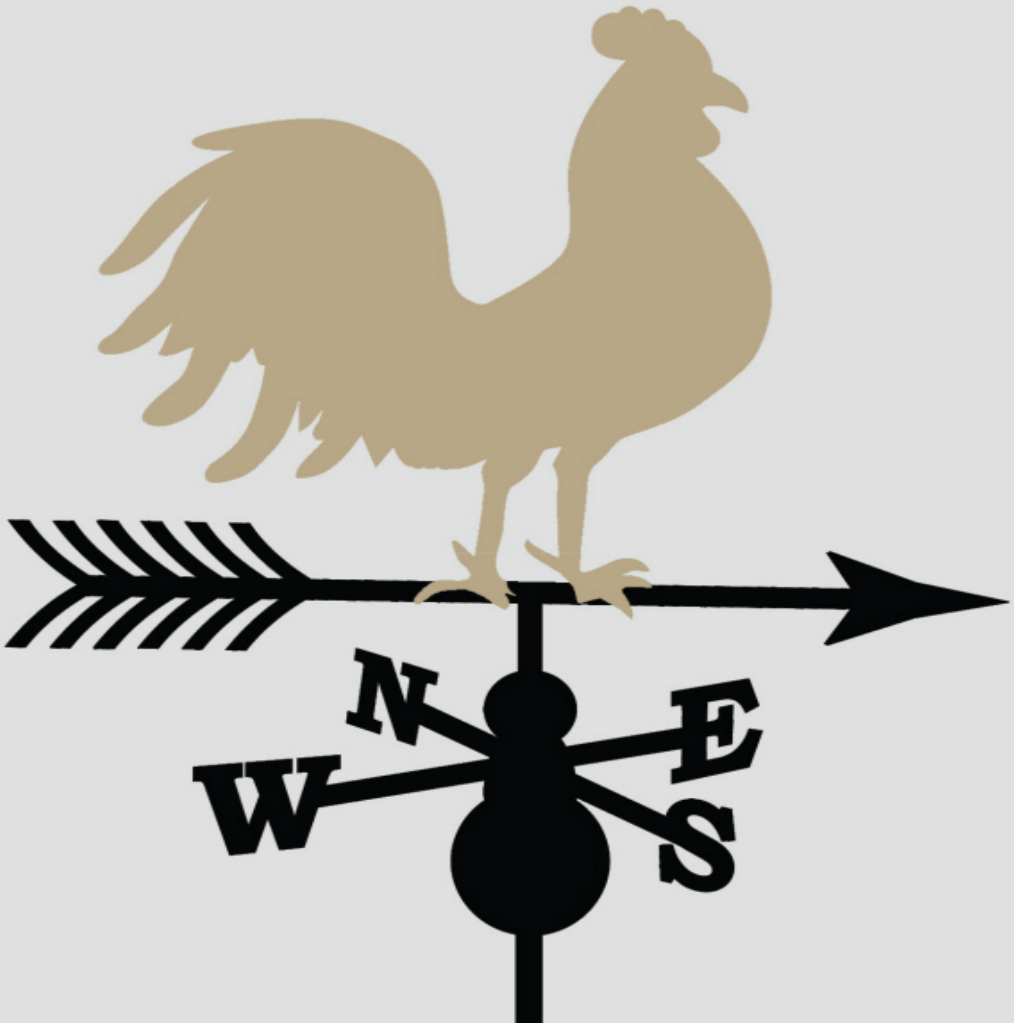
'A manual for thinking clearly in an uncertain world. Read it.'
Daniel Kahneman, author of *Thinking, Fast and Slow*

SUPERFORECASTING

THE ART & SCIENCE OF PREDICTION

PHILIP TETLOCK DAN GARDNER



Contents

About the Book

What if we could improve our ability to predict the future?

Everything we do involves forecasts about how the future will unfold. Whether buying a new house or changing job, designing a new product or getting married, our decisions are governed by implicit predictions of how things are likely to turn out. The problem is, we're not very good at it.

In a landmark, twenty-year study, Wharton professor Philip Tetlock showed that even the average expert was only slightly better at predicting the future than random guesswork. Tetlock's latest project – an unprecedented, government-funded forecasting tournament involving over a million individual predictions – has since shown that there are, however, some people with real, demonstrable foresight. These are ordinary people, from former ballroom dancers to retired computer programmers, who are nonetheless able to predict the future with a 60% greater degree of accuracy than regular forecasters. They are superforecasters.

In *Superforecasting*, Tetlock and his co-author Dan Gardner offer a fascinating insight into what we can learn from this elite group of people. They show the methods used by these superforecasters which enable them to outperform even professional intelligence analysts with access to classified data. And they offer practical advice on how we can all use these methods for our own benefit – whether in business, in international affairs, or in everyday life.

About the Authors

Philip E. Tetlock is the Annenberg University Professor at the University of Pennsylvania, with appointments in Wharton, psychology and political science. He is coleader of the Good Judgment Project, a multiyear forecasting study, and the author of *Expert Political Judgment* and, with Aaron Belkin, *Counterfactual Thought Experiments in World Politics*. He has published over two hundred articles in peer-reviewed journals and has won awards from many scientific societies, including the American Association for the Advancement of Science, the National Academy of Sciences, the American Psychological Association, and the American Political Science Association.

Dan Gardner is a journalist and the author of *Risk: The Science and Politics of Fear* and *Future Babble: Why Pundits Are Hedgehogs and Foxes Know Best*.

SUPERFORECASTING

The Art and Science of Prediction

PHILIP E. TETLOCK
DAN GARDNER

rh
BOOKS

Jenny, alive forever in the hearts of your mother and father,
as if that day were yesterday

1

An Optimistic Skeptic

WE ARE ALL forecasters. When we think about changing jobs, getting married, buying a home, making an investment, launching a product, or retiring, we decide based on how we expect the future will unfold. These expectations are forecasts. Often we do our own forecasting. But when big events happen—markets crash, wars loom, leaders tremble—we turn to the experts, those in the know. We look to people like Tom Friedman.

If you are a White House staffer, you might find him in the Oval Office with the president of the United States, talking about the Middle East. If you are a Fortune 500 CEO, you might spot him in Davos, chatting in the lounge with hedge fund billionaires and Saudi princes. And if you don't frequent the White House or swanky Swiss hotels, you can read his *New York Times* columns and bestselling books that tell you what's happening now, why, and what will come next.[1] Millions do.

Like Tom Friedman, Bill Flack forecasts global events. But there is a lot less demand for his insights.

For years, Bill worked for the US Department of Agriculture in Arizona—"part pick-and-shovel work, part spreadsheet"—but now he lives in Kearney, Nebraska. Bill is a native Cornhusker. He grew up in Madison, Nebraska, a farm town where his parents owned and published the *Madison Star-Mail*, a newspaper with lots of stories about

local sports and county fairs. He was a good student in high school and he went on to get a bachelor of science degree from the University of Nebraska. From there, he went to the University of Arizona. He was aiming for a PhD in math, but he realized it was beyond his abilities—“I had my nose rubbed in my limitations” is how he puts it—and he dropped out. It wasn’t wasted time, however. Classes in ornithology made Bill an avid bird-watcher, and because Arizona is a great place to see birds, he did fieldwork part-time for scientists, then got a job with the Department of Agriculture and stayed for a while.

Bill is fifty-five and retired, although he says if someone offered him a job he would consider it. So he has free time. And he spends some of it forecasting.

Bill has answered roughly three hundred questions like “Will Russia officially annex additional Ukrainian territory in the next three months?” and “In the next year, will any country withdraw from the eurozone?” They are questions that matter. And they’re difficult. Corporations, banks, embassies, and intelligence agencies struggle to answer such questions all the time. “Will North Korea detonate a nuclear device before the end of this year?” “How many additional countries will report cases of the Ebola virus in the next eight months?” “Will India or Brazil become a permanent member of the UN Security Council in the next two years?” Some of the questions are downright obscure, at least for most of us. “Will NATO invite new countries to join the Membership Action Plan (MAP) in the next nine months?” “Will the Kurdistan Regional Government hold a referendum on national independence this year?” “If a non-Chinese telecommunications firm wins a contract to provide Internet services in the Shanghai Free Trade Zone in the next two years, will Chinese citizens have access to Facebook and/or Twitter?” When Bill first sees one of these questions, he may have no clue how to answer it. “What on earth is the Shanghai Free Trade Zone?” he may think. But

he does his homework. He gathers facts, balances clashing arguments, and settles on an answer.

No one bases decisions on Bill Flack's forecasts, or asks Bill to share his thoughts on CNN. He has never been invited to Davos to sit on a panel with Tom Friedman. And that's unfortunate. Because Bill Flack is a remarkable forecaster. We know that because each one of Bill's predictions has been dated, recorded, and assessed for accuracy by independent scientific observers. His track record is excellent.

Bill is not alone. There are thousands of others answering the same questions. All are volunteers. Most aren't as good as Bill, but about 2% are. They include engineers and lawyers, artists and scientists, Wall Streeters and Main Streeters, professors and students. We will meet many of them, including a mathematician, a filmmaker, and some retirees eager to share their underused talents. I call them *superforecasters* because that is what they are. Reliable evidence proves it. Explaining why they're so good, and how others can learn to do what they do, is my goal in this book.

How our low-profile superforecasters compare with cerebral celebrities like Tom Friedman is an intriguing question, but it can't be answered because the accuracy of Friedman's forecasting has never been rigorously tested. Of course Friedman's fans and critics have opinions one way or the other—"he nailed the Arab Spring" or "he screwed up on the 2003 invasion of Iraq" or "he was prescient on NATO expansion." But there are no hard facts about Tom Friedman's track record, just endless opinions—and opinions on opinions.[2] And that is business as usual. Every day, the news media deliver forecasts without reporting, or even asking, how good the forecasters who made the forecasts really are. Every day, corporations and governments pay for forecasts that may be prescient or worthless or something in between. And every day, all of us—leaders of nations, corporate executives, investors, and voters—make critical

decisions on the basis of forecasts whose quality is unknown. Baseball managers wouldn't dream of getting out the checkbook to hire a player without consulting performance statistics. Even fans expect to see player stats on scoreboards and TV screens. And yet when it comes to the forecasters who help us make decisions that matter far more than any baseball game, we're content to be ignorant.[3]

In that light, relying on Bill Flack's forecasts looks quite reasonable. Indeed, relying on the forecasts of many readers of this book may prove quite reasonable, for it turns out that forecasting is not a "you have it or you don't" talent. It is a skill that can be cultivated. This book will show you how.

THE ONE ABOUT THE CHIMP

I want to spoil the joke, so I'll give away the punch line: the average expert was roughly as accurate as a dart-throwing chimpanzee.

You've probably heard that one before. It's famous—in some circles, infamous. It has popped up in the *New York Times*, the *Wall Street Journal*, the *Financial Times*, the *Economist*, and other outlets around the world. It goes like this: A researcher gathered a big group of experts—academics, pundits, and the like—to make thousands of predictions about the economy, stocks, elections, wars, and other issues of the day. Time passed, and when the researcher checked the accuracy of the predictions, he found that the average expert did about as well as random guessing. Except that's not the punch line because "random guessing" isn't funny. The punch line is about a dart-throwing chimpanzee. Because chimpanzees are funny.

I am that researcher and for a while I didn't mind the joke. My study was the most comprehensive assessment of expert judgment in the scientific literature. It was a long

slog that took about twenty years, from 1984 to 2004, and the results were far richer and more constructive than the punch line suggested. But I didn't mind the joke because it raised awareness of my research (and, yes, scientists savor their fifteen minutes of fame too). And I myself had used the old "dart-throwing chimp" metaphor, so I couldn't complain too loudly.

I also didn't mind because the joke makes a valid point. Open any newspaper, watch any TV news show, and you find experts who forecast what's coming. Some are cautious. More are bold and confident. A handful claim to be Olympian visionaries able to see decades into the future. With few exceptions, they are not in front of the cameras because they possess any proven skill at forecasting. Accuracy is seldom even mentioned. Old forecasts are like old news—soon forgotten—and pundits are almost never asked to reconcile what they said with what actually happened. The one undeniable talent that talking heads have is their skill at telling a compelling story with conviction, and that is enough. Many have become wealthy peddling forecasting of untested value to corporate executives, government officials, and ordinary people who would never think of swallowing medicine of unknown efficacy and safety but who routinely pay for forecasts that are as dubious as elixirs sold from the back of a wagon. These people—and their customers—deserve a nudge in the ribs. I was happy to see my research used to give it to them.

But I realized that as word of my work spread, its apparent meaning was mutating. What my research had shown was that the average expert had done little better than guessing on many of the political and economic questions I had posed. "Many" does not equal all. It was easiest to beat chance on the shortest-range questions that only required looking one year out, and accuracy fell off the further out experts tried to forecast—approaching the dart-throwing-chimpanzee level three to five years out. That was an

important finding. It tells us something about the limits of expertise in a complex world—and the limits on what it might be possible for even superforecasters to achieve. But as in the children's game of "telephone," in which a phrase is whispered to one child who passes it on to another, and so on, and everyone is shocked at the end to discover how much it has changed, the actual message was garbled in the constant retelling and the subtleties were lost entirely. The message became "all expert forecasts are useless," which is nonsense. Some variations were even cruder—like "experts know no more than chimpanzees." My research had become a backstop reference for nihilists who see the future as inherently unpredictable and know-nothing populists who insist on preceding "expert" with "so-called."

So I tired of the joke. My research did not support these more extreme conclusions, nor did I feel any affinity for them. Today, that is all the more true.

There is plenty of room to stake out reasonable positions between the debunkers and the defenders of experts and their forecasts. On the one hand, the debunkers have a point. There are shady peddlers of questionable insights in the forecasting marketplace. There are also limits to foresight that may just not be surmountable. Our desire to reach into the future will always exceed our grasp. But debunkers go too far when they dismiss all forecasting as a fool's errand. I believe it is possible to see into the future, at least in some situations and to some extent, and that any intelligent, open-minded, and hardworking person can cultivate the requisite skills.

Call me an "optimistic skeptic."

THE SKEPTIC

To understand the "skeptic" half of that label, consider a young Tunisian man pushing a wooden handcart loaded with fruits and vegetables down a dusty road to a market in the

Tunisian town of Sidi Bouzid. When the man was three, his father died. He supports his family by borrowing money to fill his cart, hoping to earn enough selling the produce to pay off the debt and have a little left over. It's the same grind every day. But this morning, the police approach the man and say they're going to take his scales because he has violated some regulation. He knows it's a lie. They're shaking him down. But he has no money. A policewoman slaps him and insults his dead father. They take his scales and his cart. The man goes to a town office to complain. He is told the official is busy in a meeting. Humiliated, furious, powerless, the man leaves.

He returns with fuel. Outside the town office he douses himself, lights a match, and burns.

Only the conclusion of this story is unusual. There are countless poor street vendors in Tunisia and across the Arab world. Police corruption is rife, and humiliations like those inflicted on this man are a daily occurrence. They matter to no one aside from the police and their victims.

But this particular humiliation, on December 17, 2010, caused Mohamed Bouazizi, aged twenty-six, to set himself on fire, and Bouazizi's self-immolation sparked protests. The police responded with typical brutality. The protests spread. Hoping to assuage the public, the dictator of Tunisia, President Zine el-Abidine Ben Ali, visited Bouazizi in the hospital.

Bouazizi died on January 4, 2011. The unrest grew. On January 14, Ben Ali fled to a cushy exile in Saudi Arabia, ending his twenty-three-year kleptocracy.

The Arab world watched, stunned. Then protests erupted in Egypt, Libya, Syria, Jordan, Kuwait, and Bahrain. After three decades in power, the Egyptian dictator Hosni Mubarak was driven from office. Elsewhere, protests swelled into rebellions, rebellions into civil wars. This was the Arab Spring—and it started with one poor man, no different from

countless others, being harassed by police, as so many have been, before and since, with no apparent ripple effects.

It is one thing to look backward and sketch a narrative arc, as I did here, connecting Mohamed Bouazizi to all the events that flowed out of his lonely protest. Tom Friedman, like many elite pundits, is skilled at that sort of reconstruction, particularly in the Middle East, which he knows so well, having made his name in journalism as a *New York Times* correspondent in Lebanon. But could even Tom Friedman, if he had been present that fatal morning, have peered into the future and foreseen the self-immolation, the unrest, the toppling of the Tunisian dictator, and all that followed? Of course not. No one could. Maybe, given how much Friedman knew about the region, he would have mused that poverty and unemployment were high, the number of desperate young people was growing, corruption was rampant, repression was relentless, and therefore Tunisia and other Arab countries were powder kegs waiting to blow. But an observer could have drawn exactly the same conclusion the year before. And the year before that. Indeed, you could have said that about Tunisia, Egypt, and several other countries for decades. They may have been powder kegs but they never blew—until December 17, 2010, when the police pushed that one poor man too far.

In 1972 the American meteorologist Edward Lorenz wrote a paper with an arresting title: “Predictability: Does the Flap of a Butterfly’s Wings in Brazil Set Off a Tornado in Texas?” A decade earlier, Lorenz had discovered by accident that tiny data entry variations in computer simulations of weather patterns—like replacing 0.506127 with 0.506—could produce dramatically different long-term forecasts. It was an insight that would inspire “chaos theory”: in nonlinear systems like the atmosphere, even small changes in initial conditions can mushroom to enormous proportions. So, in principle, a lone butterfly in Brazil could flap its wings and set off a tornado in Texas—even though swarms of other

Brazilian butterflies could flap frantically their whole lives and never cause a noticeable gust a few miles away. Of course Lorenz didn't mean that the butterfly "causes" the tornado in the same sense that I cause a wineglass to break when I hit it with a hammer. He meant that if that particular butterfly hadn't flapped its wings at that moment, the unfathomably complex network of atmospheric actions and reactions would have behaved differently, and the tornado might never have formed—just as the Arab Spring might never have happened, at least not when and as it did, if the police had just let Mohamed Bouazizi sell his fruits and vegetables that morning in 2010.

Edward Lorenz shifted scientific opinion toward the view that there are hard limits on predictability, a deeply philosophical question.[4] For centuries, scientists had supposed that growing knowledge must lead to greater predictability because reality was like a clock—an awesomely big and complicated clock but still a clock—and the more scientists learned about its innards, how the gears grind together, how the weights and springs function, the better they could capture its operations with deterministic equations and predict what it would do. In 1814 the French mathematician and astronomer Pierre-Simon Laplace took this dream to its logical extreme:

> *We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.*

Laplace called his imaginary entity a "demon." If it knew everything about the present, Laplace thought, it could predict everything about the future. It would be omniscient.[5]

Lorenz poured cold rainwater on that dream. If the clock symbolizes perfect Laplacean predictability, its opposite is the Lorenzian cloud. High school science tells us that clouds form when water vapor coalesces around dust particles. This sounds simple but exactly how a particular cloud develops—the shape it takes—depends on complex feedback interactions among droplets. To capture these interactions, computer modelers need equations that are highly sensitive to tiny butterfly-effect errors in data collection. So even if we learn all that is knowable about how clouds form, we will not be able to predict the shape a particular cloud will take. We can only wait and see. In one of history's great ironies, scientists today know vastly more than their colleagues a century ago, and possess vastly more data-crunching power, but they are much less confident in the prospects for perfect predictability.

This is a big reason for the "skeptic" half of my "optimistic skeptic" stance. We live in a world where the actions of one nearly powerless man can have ripple effects around the world—ripples that affect us all to varying degrees. A woman living in a Kansas City suburb may think Tunisia is another planet, and her life has no connection to it, but if she were married to an air force navigator who flies out of the nearby Whiteman Air Force Base, she might be surprised to learn that one obscure Tunisian's actions led to protests, that led to riots, that led to the toppling of a dictator, that led to protests in Libya, that led to a civil war, that led to the 2012 NATO intervention, that led to her husband dodging antiaircraft fire over Tripoli. That's an easily traceable connection. Often the connections are harder to spot, but they are all around us, in things like the price we pay at the gas station or the layoffs down the street. In a world where a butterfly in Brazil can make the difference between just another sunny day in Texas and a tornado tearing through a town, it's misguided to think anyone can see very far into the future.[6]

THE OPTIMIST

But it is one thing to recognize the limits on predictability, and quite another to dismiss *all prediction* as an exercise in futility.

Crank up the microscope on a day in the life of that woman living in a Kansas City suburb: At 6:30 in the morning, she drops papers into a briefcase, gets in her car, drives her usual route to work, and parks downtown. As she does every weekday morning, she walks past the statues of lions and into the Greek-inspired office building of the Kansas City Life Insurance Company. At her desk, she works on spreadsheets for a while, participates in a conference call at 10:30, spends a few minutes browsing on the Amazon website, and answers e-mails until 11:50. Then she walks to a little Italian restaurant to have lunch with her sister.

This woman's life is influenced by many unpredictable factors—from the lottery ticket in her purse to the Arab Spring that results in her husband flying missions over Libya to the fact that the price of gas just went up five cents a gallon because there was a coup in some country she's never heard of—but there is as much or more that is quite predictable. Why did she leave home at 6:30? She didn't want to get stuck in rush hour. Or to put that another way, she predicted that traffic would be much heavier later—and she was almost certainly right because rush hour is highly predictable. When she drove, she anticipated other drivers' behavior constantly: they will stop at the intersection when the light is red; they will stay in their lanes and signal before turning. She expected the people who said they would join the 10:30 conference call to do so, and she was right. She arranged to meet her sister at noon at the restaurant because the restaurant's posted hours indicated it would open then, and posted hours are a reliable guide.

We make mundane predictions like these routinely, while others just as routinely make predictions that shape our lives. When the woman turned her computer on, she increased electricity consumption in Kansas City by a bit, as every other worker bee did that morning, and collectively they caused a demand surge, as they do every nonholiday weekday morning around that time. But that didn't cause problems because electricity producers anticipate these surges and vary their output accordingly. When the woman went to Amazon, the website highlighted certain products it thought she would like, a forecast derived from her past purchases and browsing and that of millions of others. We constantly encounter predictive operations like that on the Internet—Google personalizes search results by putting what it thinks you will find most interesting on top—but they operate so smoothly we rarely notice. And then there's the woman's workplace. The Kansas City Life Insurance Company is in the business of forecasting disability and death, and it does a good job. That doesn't mean it knows precisely when I will die, but it does have a good idea of how long someone of my age and profile—sex, income, lifestyle—is likely to live. Kansas City Life was founded in 1895. If its actuaries weren't good forecasters, it would have gone bankrupt long ago.

So much of our reality is this predictable, or more so. I just Googled tomorrow's sunrise and sunset times for Kansas City, Missouri, and got them down to the minute. Those forecasts are reliable, whether they are for tomorrow, the day after, or fifty years from now. The same is true of tides, eclipses, and phases of the moon. All can be predicted from clocklike scientific laws with enough precision to satisfy Laplace's forecasting demon.

Of course each of these pockets of predictability can be abruptly punctured. A good restaurant is very likely to open its doors when it says it will, but it may not, for any number of reasons, from a manager sleeping late, to fire,

bankruptcy, pandemic, nuclear war, or a physics experiment accidentally creating a black hole that sucks up the solar system. The same is true of anything else. Even those fifty-year sunrise and sunset forecasts could be off somewhat if, sometime in the next fifty years, a massive space rock bumps Earth off its orbit around the sun. There are no certainties in life—not even death and taxes if we assign a nonzero probability to the invention of technologies that let us upload the contents of our brains into a cloud-computing network and the emergence of a future society so public-spirited and prosperous that the state can be funded with charitable donations.

So is reality clocklike or cloud-like? Is the future predictable or not? These are false dichotomies, the first of many we will encounter. We live in a world of clocks and clouds and a vast jumble of other metaphors. Unpredictability and predictability coexist uneasily in the intricately interlocking systems that make up our bodies, our societies, and the cosmos. How predictable something is depends on what we are trying to predict, how far into the future, and under what circumstances.

Look at Edward Lorenz's field. Weather forecasts are typically quite reliable, under most conditions, looking a few days ahead, but they become increasingly less accurate three, four, and five days out. Much beyond a week, we might as well consult that dart-throwing chimpanzee. So we can't say that weather is predictable or not, only that weather is predictable to some extent under some circumstances—and we must be very careful when we try to be more precise than that. Take something as seemingly simple as the relationship between time and predictability: it is generally true that the further we try to look into the future, the harder it is to see. But there can be prolonged exceptions to the rule. Predicting the continuation of a long bull market in stocks can prove profitable for many years—until it suddenly proves to be your undoing. And predicting

that dinosaurs would continue to preside at the top of the food chain was a safe bet for tens of millions of years—until an asteroid set off a cataclysm that opened up ecological niches for a tiny mammal that eventually evolved into a species that tries to predict the future. Laws of physics aside, there are no universal constants, so separating the predictable from the unpredictable is difficult work. There's no way around it.

Meteorologists know that better than anyone. They make large numbers of forecasts and routinely check their accuracy—which is why we know that one-and two-day forecasts are typically quite accurate while eight-day forecasts are not. With these analyses, meteorologists are able to sharpen their understanding of how weather works and tweak their models. Then they try again. Forecast, measure, revise. Repeat. It's a never-ending process of incremental improvement that explains why weather forecasts are good and slowly getting better. There may be limits to such improvements, however, because weather is the textbook illustration of nonlinearity. The further out the forecaster tries to look, the more opportunity there is for chaos to flap its butterfly wings and blow away expectations. Big leaps in computing power and continued refinement of forecasting models may nudge the limits a little further into the future but those advances gradually get harder and the payoffs shrink toward zero. How good can it get? No one knows. But knowing the current limits is itself a success.

In so many other high-stakes endeavors, forecasters are groping in the dark. They have no idea how good their forecasts are in the short, medium, or long term—and no idea how good their forecasts could become. At best, they have vague hunches. That's because the forecast-measure-revise procedure operates only within the rarefied confines of high-tech forecasting, such as the work of macroeconomists at central banks or marketing and

financial professionals in big companies or opinion poll analysts like Nate Silver.[7] More often forecasts are made and then … nothing. Accuracy is seldom determined after the fact and is almost never done with sufficient regularity and rigor that conclusions can be drawn. The reason? Mostly it's a demand-side problem: The consumers of forecasting—governments, business, and the public—don't demand evidence of accuracy. So there is no measurement. Which means no revision. And without revision, there can be no improvement. Imagine a world in which people love to run, but they have no idea how fast the average person runs, or how fast the best could run, because runners have never agreed to basic ground rules—stay on the track, begin the race when the gun is fired, end it after a specified distance—and there are no independent race officials and timekeepers measuring results. How likely is it that running times are improving in this world? Not very. Are the best runners running as fast as human beings are physically capable? Again, probably not.

"I have been struck by how important measurement is to improving the human condition," Bill Gates wrote. "You can achieve incredible progress if you set a clear goal and find a measure that will drive progress toward that goal. … This may seem basic, but it is amazing how often it is not done and how hard it is to get right."[8] He is right about what it takes to drive progress, and it is surprising how rarely it's done in forecasting. Even that simple first step—setting a clear goal—hasn't been taken.

You might think the goal of forecasting is to foresee the future accurately, but that's often not the goal, or at least not the sole goal. Sometimes forecasts are meant to entertain. Think of CNBC's Jim Cramer with his "booyah!" shtick, or John McLaughlin, the host of *The McLaughlin Group*, bellowing at his panelists to predict the likelihood of an event "on a scale from zero to ten, with zero representing zero possibility and ten representing complete metaphysical

certitude!” Sometimes forecasts are used to advance political agendas and galvanize action—as activists hope to do when they warn of looming horrors unless we change our ways. There is also dress-to-impress forecasting—which is what banks deliver when they pay a famous pundit to tell wealthy clients about the global economy in 2050. And some forecasts are meant to comfort—by assuring the audience that their beliefs are correct and the future will unfold as expected. Partisans are fond of these forecasts. They are the cognitive equivalent of slipping into a warm bath.

This jumble of goals is seldom acknowledged, which makes it difficult to even start working toward measurement and progress. It’s a messy situation, which doesn’t seem to be getting better.

And yet this stagnation is a big reason why I am an *optimistic* skeptic. We know that in so much of what people want to predict—politics, economics, finance, business, technology, daily life—predictability exists, to some degree, in some circumstances. But there is so much else we do not know. For scientists, not knowing is exciting. It’s an opportunity to discover; the more that is unknown, the greater the opportunity. Thanks to the frankly quite amazing lack of rigor in so many forecasting domains, this opportunity is huge. And to seize it, all we have to do is set a clear goal—accuracy!—and get serious about measuring.

I’ve been doing that for much of my career. The research that produced the dart-throwing-chimpanzee result was phase one. Phase two started in the summer of 2011, when my research (and life) partner Barbara Mellers and I launched the Good Judgment Project and invited volunteers to sign up and forecast the future. Bill Flack responded. So did a couple of thousand others that first year, and thousands more in the four years that followed. Cumulatively, more than twenty thousand intellectually curious laypeople tried to figure out if protests in Russia

would spread, the price of gold would plummet, the Nikkei would close above 9,500, war would erupt on the Korean peninsula, and many other questions about complex, challenging global issues. By varying the experimental conditions, we could gauge which factors improved foresight, by how much, over which time frames, and how good forecasts could become if best practices were layered on each other. Laid out like that, it sounds simple. It wasn't. It was a demanding program that took the talents and hard work of a multidisciplinary team based at the University of California, Berkeley, and the University of Pennsylvania.

Big as it was, the Good Judgment Project (GJP) was only part of a much larger research effort sponsored by the Intelligence Advanced Research Projects Activity (IARPA). Don't be put off by the bland name. IARPA is an agency within the intelligence community that reports to the director of National Intelligence and its job is to support daring research that promises to make American intelligence better at what it does. And a big part of what American intelligence does is forecast global political and economic trends. By one rough estimate, the United States has twenty thousand intelligence analysts assessing everything from minute puzzles to major events such as the likelihood of an Israeli sneak attack on Iranian nuclear facilities or the departure of Greece from the eurozone.[9] How good is all this forecasting? That is not easily answered because the intelligence community, like so many major producers of forecasting, has never been keen on spending money to figure that out. There are various reasons for that reluctance, some more respectable than others, but we'll get into that later. What matters is that this forecasting is critical to national security and yet little can be said with any confidence about how good it is, or even whether it's as good as a multibillion-dollar operation with twenty thousand people should be. To change that, IARPA created a forecasting tournament in which five scientific teams led by

top researchers in the field would compete to generate accurate forecasts on the sorts of tough questions intelligence analysts deal with every day. The Good Judgment Project was one of those five teams. Each team would effectively be its own research project, free to improvise whatever methods it thought would work, but required to submit forecasts at 9 a.m. eastern standard time every day from September 2011 to June 2015. By requiring teams to forecast the same questions at the same time, the tournament created a level playing field—and a rich trove of data about what works, how well, and when. Over four years, IARPA posed nearly five hundred questions about world affairs. Time frames were shorter than in my earlier research, with the vast majority of forecasts extending out more than one month and less than one year. In all, we gathered over one million individual judgments about the future.

In year 1, GJP beat the official control group by 60%. In year 2, we beat the control group by 78%. GJP also beat its university-affiliated competitors, including the University of Michigan and MIT, by hefty margins, from 30% to 70%, and even outperformed professional intelligence analysts with access to classified data. After two years, GJP was doing so much better than its academic competitors that IARPA dropped the other teams.[10]

I'll delve into details later, but let's note two key conclusions that emerge from this research. One, foresight is real. Some people—people like Bill Flack—have it in spades. They aren't gurus or oracles with the power to peer decades into the future, but they do have a real, measurable skill at judging how high-stakes events are likely to unfold three months, six months, a year, or a year and a half in advance. The other conclusion is what makes these superforecasters so good. It's not really who they are. It is what they do. Foresight isn't a mysterious gift bestowed at birth. It is the product of particular ways of thinking, of

gathering information, of updating beliefs. These habits of thought can be learned and cultivated by any intelligent, thoughtful, determined person. It may not even be all that hard to get started. One result that particularly surprised me was the effect of a tutorial covering some basic concepts that we'll explore in this book and are summarized in the Ten Commandments appendix. It took only about sixty minutes to read and yet it improved accuracy by roughly 10% through the entire tournament year. Yes, 10% may sound modest, but it was achieved at so little cost. And never forget that even modest improvements in foresight maintained over time add up. I spoke about that with Aaron Brown, an author, a Wall Street veteran, and the chief risk manager at AQR Capital Management, a hedge fund with over $100 billion in assets. "It's so hard to see because it's not dramatic," he said, but if it is sustained "it's the difference between a consistent winner who's making a living, or the guy who's going broke all the time."[11] A world-class poker player we will meet soon could not agree more. The difference between heavyweights and amateurs, she said, is that the heavyweights know the difference between a 60⁄40 bet and a 40⁄60 bet.

And yet, if it's possible to improve foresight simply by measuring, and if the rewards of improved foresight are substantial, why isn't measuring standard practice? A big part of the answer to that question lies in the psychology that convinces us we know things we really don't—things like whether Tom Friedman is an accurate forecaster or not. I'll explore this psychology in chapter 2. For centuries, it hobbled progress in medicine. When physicians finally accepted that their experience and perceptions were not reliable means of determining whether a treatment works, they turned to scientific testing—and medicine finally started to make rapid advances. The same revolution needs to happen in forecasting.

It won’t be easy. Chapter 3 examines what it takes to test forecasting as rigorously as modern medicine tests treatments. It’s a bigger challenge than it may appear. In the late 1980s I worked out a methodology and conducted what was, at the time, the biggest test of expert political forecasting accuracy ever. One result, delivered many years later, was the punch line that now makes me squirm. But another discovery of that research didn’t receive nearly as much attention even though it was far more important: one group of experts had modest but real foresight. What made the difference between the experts with foresight and those who were so hopeless they dragged the average down to the level of a dart-throwing chimp? It wasn’t some mystical gift or access to information others didn’t have. Nor was it any particular set of beliefs. Indeed, within a quite wide range of views, *what* they thought didn’t matter. It was *how* they thought.

Inspired in part by that insight, IARPA created its unprecedented forecasting tournament. Chapter 4 is the story of how that happened—and the discovery of superforecasters. Why are they so good? That question runs through chapters 5 through 9. When you meet them it’s hard not to be struck by how smart they are, so you might suspect it’s intelligence that makes all the difference. It’s not. They’re also remarkably numerate. Like Bill Flack, many have advanced degrees in mathematics and science. So is the secret arcane math? No. Even superforecasters who are card-carrying mathematicians rarely use much math. They also tend to be newsjunkies who stay on top of the latest developments and regularly update their forecasts, so you might be tempted to attribute their success to spending endless hours on the job. Yet that too would be a mistake.

Superforecasting does require minimum levels of intelligence, numeracy, and knowledge of the world, but anyone who reads serious books about psychological research probably has those prerequisites. So what is it that

elevates forecasting to superforecasting? As with the experts who had real foresight in my earlier research, what matters most is *how* the forecaster thinks. I'll describe this in detail, but broadly speaking, superforecasting demands thinking that is open-minded, careful, curious, and—above all—self-critical. It also demands focus. The kind of thinking that produces superior judgment does not come effortlessly. Only the determined can deliver it reasonably consistently, which is why our analyses have consistently found commitment to self-improvement to be the strongest predictor of performance.

In the final chapters, I'll resolve an apparent contradiction between the demands of good judgment and effective leadership, respond to what I think are the two strongest challenges to my research, and conclude—appropriately for a book about forecasting—with a consideration of what comes next.

A FORECAST ABOUT FORECASTING

But maybe you think this is all hopelessly outdated. After all, we live in an era of dazzlingly powerful computers, incomprehensible algorithms, and Big Data. At its core, the forecasting I study involves subjective judgment: it is people thinking and deciding, nothing more. Isn't the time for such sloppy guesswork drawing to a close?

In 1954, a brilliant psychologist, Paul Meehl wrote a small book that caused a big stir.[12] It reviewed twenty studies showing that well-informed experts predicting outcomes—whether a student would succeed in college or a parolee would be sent back to prison—were not as accurate as simple algorithms that added up objective indicators like ability test scores and records of past conduct. Meehl's claim upset many experts, but subsequent research—now more than two hundred studies—has shown that in most cases statistical algorithms beat subjective judgment, and in

the handful of studies where they don't, they usually tie. Given that algorithms are quick and cheap, unlike subjective judgment, a tie supports using the algorithm. The point is now indisputable: when you have a well-validated statistical algorithm, use *it*.

This insight was never a threat to the reign of subjective judgment because we so rarely have well-validated algorithms for the problem at hand. It was just impractical for math to displace plain old thinking—in 1954 and even today.

But spectacular advances in information technology suggest we are approaching a historical discontinuity in humanity's relationship with machines. In 1997 IBM's Deep Blue beat chess champion Garry Kasparov. Now, commercially available chess programs can beat any human. In 2011 IBM's Watson beat *Jeopardy!* champions Ken Jennings and Brad Rutter. That was a vastly tougher computing challenge, but Watson's engineers did it. Today, it's no longer impossible to imagine a forecasting competition in which a supercomputer trounces superforecasters and superpundits alike. After that happens, there will still be human forecasters, but like human *Jeopardy!* contestants, we will only watch them for entertainment.

So I spoke to Watson's chief engineer, David Ferrucci. I was sure that Watson could easily field a question about the present or past like "Which two Russian leaders traded jobs in the last ten years?" But I was curious about his views on how long it will take for Watson or one of its digital descendants to field questions like "Will two top Russian leaders trade jobs in the next ten years?"

In 1965 the polymath Herbert Simon thought we were only twenty years away from a world in which machines could do "any work a man can do," which is the sort of naively optimistic thing people said back then, and one reason why Ferrucci—who has worked in artificial

intelligence for thirty years—is more cautious today.[13] Computing is making enormous strides, Ferrucci noted. The ability to spot patterns is growing spectacularly. And machine learning, in combination with burgeoning human-machine interactions that feed the learning process, promises far more fundamental advances to come. “It’s going to be one of these exponential curves that we’re kind of at the bottom of now,” Ferrucci said.

But there is a vast difference between “Which two Russian leaders traded jobs?” and “Will two Russian leaders trade jobs again?” The former is a historical fact. The computer can look it up. The latter requires the computer to make an informed guess about the intentions of Vladimir Putin, the character of Dmitri Medvedev, and the causal dynamics of Russian politics, and then integrate that information into a judgment call. People do that sort of thing all the time, but that doesn’t make it easy. It means the human brain is wondrous—because the task is staggeringly hard. Even with computers making galloping advances, the sort of forecasting that superforecasters do is a long way off. And Ferrucci isn’t sure we will ever see a human under glass at the Smithsonian with a sign saying “subjective judgment.”

Machines may get better at “mimicking human meaning,” and thereby better at predicting human behavior, but “there’s a difference between mimicking and reflecting meaning and originating meaning,” Ferrucci said. That’s a space human judgment will always occupy.

In forecasting, as in other fields, we will continue to see human judgment being displaced—to the consternation of white-collar workers—but we will also see more and more syntheses, like “freestyle chess,” in which humans with computers compete as teams, the human drawing on the computer’s indisputable strengths but also occasionally overriding the computer. The result is a combination that can (sometimes) beat both humans and machines. To re-frame the man-versus-machine dichotomy, combinations of

Garry Kasparov and Deep Blue may prove more robust than pure-human or pure-machine approaches.

What Ferrucci does see becoming obsolete is the guru model that makes so many policy debates so puerile: “I’ll counter your Paul Krugman polemic with my Niall Ferguson counterpolemic, and rebut your Tom Friedman op-ed with my Bret Stephens blog.” Ferrucci sees light at the end of this long dark tunnel: “I think it’s going to get stranger and stranger” for people to listen to the advice of experts whose views are informed only by their subjective judgment. Human thought is beset by psychological pitfalls, a fact that has only become widely recognized in the last decade or two. “So what I want is that human expert paired with a computer to overcome the human cognitive limitations and biases.”[14]

If Ferrucci is right—I suspect he is—we will need to blend computer-based forecasting and subjective judgment in the future. So it’s time we got serious about both.