SUPER-FORECASTING

THE ART & SCIENCE OF PREDICTION

'A manual for thinking clearly in an uncertain world. Read it.'
DANIEL KAHNEMAN, author of *THINKING, FAST AND SLOW*

PHILIP TETLOCK
& DAN GARDNER

SUPERFORECASTING

The Art and Science of Prediction

PHILIP TETLOCK

DAN GARDNER

BOOKS

Published by Random House Books in 2016

1 3 5 7 9 10 8 6 4 2

Random House Books
20 Vauxhall Bridge Road
London SW1V 2SA

Random House Books is part of the Penguin Random House
group of companies whose addresses can be found
at global.penguinrandomhouse.com

Penguin
Random House
UK

First published in the United States in 2015 by Crown Publishers,
an imprint of the Crown Publishing Group,
a division of Penguin Random House LLC, New York

First published in Great Britain by Random House Books in 2015

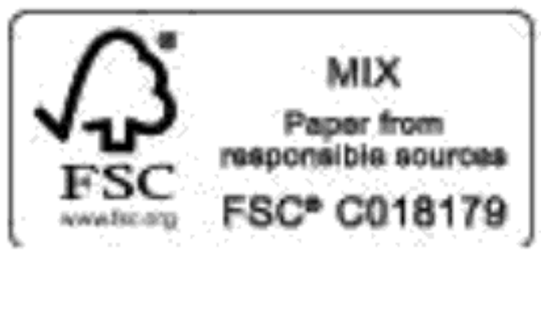
www.randomhouse.co.uk

A CIP catalogue record for this book is available from the British Library

ISBN 9781847947154

Printed and bound in Great Britain by Clays Ltd, St Ives plc

Penguin Random House is committed to a sustainable future for
our business, our readers and our planet. This book is made
from Forest Stewardship Council® certified paper



Contents

SUPERFORECASTING

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

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

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

weights 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 mean-

ing," 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 "free-style 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 reframe 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.

"They provide a clear account of what people believed they were doing, but almost none at all of whether they were right."[4] Did the ostrich egg poultices applied by ancient Egyptian physicians actually heal head fractures? In ancient Mesopotamia, did the treatments of the Keeper of the Royal Rectum actually keep royal rectums healthy? What about bloodletting? Everyone from the ancient Greeks to George Washington's doctors swore that it was wonderfully restorative, but did it work? The standard histories are usually mute on these scores, but when we use modern science to judge the efficacy of historical treatments, it becomes depressingly clear that most of the interventions were useless or worse. Until quite recently in historical terms, it was not unusual for a sick person to be better off if there was no physician available because letting an illness take its natural course was less dangerous than what a physician would inflict. And treatments seldom got better, no matter how much time passed. When George Washington fell ill in 1799, his esteemed physicians bled him relentlessly, dosed him with mercury to cause diarrhea, induced vomiting, and raised blood-filled blisters by applying hot cups to the old man's skin. A physician in Aristotle's Athens, or Nero's Rome, or medieval Paris, or Elizabethan London would have nodded at much of that hideous regimen.

Washington died. One might assume that such results would make physicians question their methods but, to be fair, the fact that Washington died proves nothing about the treatments beyond that they failed to prevent his death. It's possible that the treatments helped but not enough to overcome the disease that took Washington's life, or that they didn't help at all, or that the treatments even hastened Washington's death. It's impossible to know which of these conclusions is true merely by observing that one outcome. Even with many such observations, the truth can be difficult or impossible to tease out. There are just too many factors involved, too many possible explanations, too many unknowns. And if physicians are already

inclined to think the treatments work—which they are, or they wouldn't prescribe them—all that ambiguity is likely to be read in favor of the happy conclusion that the treatments really are effective. It takes strong evidence and more rigorous experimentation than the "bleed the patient and see if he gets better" variety to overwhelm preconceptions. And that was never done.

Consider Galen, the second-century physician to Roman emperors. No one has influenced more generations of physicians. Galen's writings were the indisputable source of medical authority for more than a thousand years. "It is I, and I alone, who has revealed the true path of medicine," Galen wrote with his usual modesty. And yet Galen never conducted anything resembling a modern experiment. Why should he? Experiments are what people do when they aren't sure what the truth is. And Galen was untroubled by doubt. Each outcome confirmed he was right, no matter how equivocal the evidence might look to someone less wise than the master. "All who drink of this treatment recover in a short time, except those whom it does not help, who all die," he wrote. "It is obvious, therefore, that it fails only in incurable cases."[5]

Galen is an extreme example but he is the sort of figure who pops up repeatedly in the history of medicine. They are men (always men) of strong conviction and a profound trust in their own judgment. They embrace treatments, develop bold theories for why they work, denounce rivals as quacks and charlatans, and spread their insights with evangelical passion. So it went from the ancient Greeks to Galen to Paracelsus to the German Samuel Hahnemann and the American Benjamin Rush. In the nineteenth century, American medicine saw pitched battles between orthodox physicians and a host of charismatic figures with curious new theories like Thomsonianism, which posited that most illness was due to an excess of cold in the body, or the orificial surgery of Edwin Hartley Pratt, whose fundamental insight was that, as one detractor put it, with only modest exag-

geration, "the rectum is the focus of existence, contains the essence of life, and performs the functions ordinarily ascribed to the heart and brain."[6] Fringe or mainstream, almost all of it was wrong, with the treatments on offer ranging from the frivolous to the dangerous. Some physicians feared as much but most carried on with business as usual. Ignorance and confidence remained defining features of medicine. As the surgeon and historian Ira Rutkow observed, physicians who furiously debated the merits of various treatments and theories were "like blind men arguing over the colors of the rainbow."[7]

The cure for this plague of certainty came tantalizingly close to discovery in 1747, when a British ship's doctor named James Lind took twelve sailors suffering from scurvy, divided them into pairs, and gave each pair a different treatment: vinegar, cider, sulfuric acid, seawater, a bark paste, and citrus fruit. It was an experiment born of desperation. Scurvy was a mortal threat to sailors on long-distance voyages and not even the confidence of physicians could hide the futility of their treatments. So Lind took six shots in the dark—and one hit. The two sailors given the citrus recovered quickly. But contrary to popular belief, this was not a eureka moment that ushered in the modern era of experimentation. "Lind was behaving in what sounds a modern way, but had no full understanding of what he was doing," noted Druin Burch. "He failed so completely to make sense of his own experiment that even he was left unconvinced of the exceptional benefits of lemons and limes."[8] For years thereafter, sailors kept getting scurvy and doctors kept prescribing worthless medicine.

Not until the twentieth century did the idea of randomized trial experiments, careful measurement, and statistical power take hold. "Is the application of the numerical method to the subject-matter of medicine a trivial and time-wasting ingenuity as some hold, or is it an important stage in the development of our art, as others proclaim it," the *Lancet* asked in 1921. The British statistician Austin Bradford Hill responded emphatically that it was the latter, and laid out a tem-

plate for modern medical investigation. If patients who were identical in every way were put into two groups, and the groups were treated differently, he wrote, we would know the treatment caused any difference in outcome. It seems simple but is impossible in practice because no two people are exactly alike, not even identical twins, so the experiment will be confounded by the differences among test subjects. The solution lay in statistics: Randomly assigning people to one group or the other would mean whatever differences there are among them should balance out if enough people participated in the experiment. Then we can confidently conclude that the treatment caused any differences in observed outcomes. It isn't perfect. There is no perfection in our messy world. But it beats wise men stroking their chins.

This seems stunningly obvious today. Randomized controlled trials are now routine. Yet it was revolutionary because medicine had never before been scientific. True, it had occasionally reaped the fruits of science like the germ theory of disease and the X-ray. And it dressed up as a science. There were educated men with impressive titles who conducted case studies and reported results in Latin-laden lectures at august universities. But it wasn't scientific.

It was cargo cult science, a term of mockery coined much later by the physicist Richard Feynman to describe what happened after American airbases from World War II were removed from remote South Pacific islands, ending the islanders' only contact with the outside world. The planes had brought wondrous goods. The islanders wanted more. So they "arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he's the controller—and they wait for the planes to land."[9] But the planes never returned. So cargo cult science has the outward form of science but lacks what makes it truly scientific.

What medicine lacked was doubt. "Doubt is not a fearful thing,"

Feynman observed, "but a thing of very great value."[10] It's what propels science forward.

> *When the scientist tells you he does not know the answer, he is an ignorant man. When he tells you he has a hunch about how it is going to work, he is uncertain about it. When he is pretty sure of how it is going to work, and he tells you, "This is the way it's going to work, I'll bet," he still is in some doubt. And it is of paramount importance, in order to make progress, that we recognize this ignorance and this doubt. Because we have the doubt, we then propose looking in new directions for new ideas. The rate of the development of science is not the rate at which you make observations alone but, much more important, the rate at which you create new things to test.*[11]

It was the absence of doubt—and scientific rigor—that made medicine unscientific and caused it to stagnate for so long.

PUTTING MEDICINE TO THE TEST

Unfortunately, this story doesn't end with physicians suddenly slapping themselves on their collective forehead and putting their beliefs to experimental tests. The idea of randomized controlled trials was painfully slow to catch on and it was only after World War II that the first serious trials were attempted. They delivered excellent results. But still the physicians and scientists who promoted the modernization of medicine routinely found that the medical establishment wasn't interested, or was even hostile to their efforts. "Too much that was being done in the name of health care lacked scientific validation," Archie Cochrane complained about medicine in the 1950s and 1960s, and the National Health Service—the British health care

that much more than intuition is needed before we draw firm conclusions. It is so obviously true. Why did people resist it? Why, specifically, did the specialist not think to wait to hear from the pathologist before removing a swath of Cochrane's flesh? Then there is the puzzle of Cochrane himself. Why did a man who stressed the importance of not rushing to judgment rush to judgment about whether he had terminal cancer?

THINKING ABOUT THINKING

It is natural to identify our thinking with the ideas, images, plans, and feelings that flow through consciousness. What else could it be? If I ask, "Why did you buy that car?" you can trot out reasons: "Good mileage. Cute style. Great price." But you can only share thoughts by introspecting; that is, by turning your attention inward and examining the contents of your mind. And introspection can only capture a tiny fraction of the complex processes whirling inside your head—and behind your decisions.

In describing how we think and decide, modern psychologists often deploy a dual-system model that partitions our mental universe into two domains. System 2 is the familiar realm of conscious thought. It consists of everything we choose to focus on. By contrast, System 1 is largely a stranger to us. It is the realm of automatic perceptual and cognitive operations—like those you are running right now to transform the print on this page into a meaningful sentence or to hold the book while reaching for a glass and taking a sip. We have no awareness of these rapid-fire processes but we could not function without them. We would shut down.

The numbering of the two systems is not arbitrary. System 1 comes first. It is fast and constantly running in the background. If a question is asked and you instantly know the answer, it sprang from

System 1. System 2 is charged with interrogating that answer. Does it stand up to scrutiny? Is it backed by evidence? This process takes time and effort, which is why the standard routine in decision making is this: first System 1 delivers an answer, and only then can System 2 get involved, starting with an examination of what System 1 decided.

Whether System 2 actually *will* get involved is another matter. Try answering this: "A bat and ball together cost $1.10. The bat costs a dollar more than the ball. How much does the ball cost?" If you are like just about everybody who has ever read this famous question, you instantly had an answer: "Ten cents." You didn't think carefully to get that. You didn't calculate anything. It just appeared. For that, you can thank System 1. Quick and easy, no effort required.

But is "ten cents" right? Think about the question carefully.

You probably realized a couple of things. First, conscious thought is demanding. Thinking the problem through requires sustained focus and takes an eternity relative to the snap judgment you got with a quick look. Second, "ten cents" is wrong. It feels right. But it's wrong. In fact, it's obviously wrong—if you give it a sober second thought.

The bat-and-ball question is one item in an ingenious psychological measure, the Cognitive Reflection Test, which has shown that most people—including very smart people—aren't very reflective. They read the question, think "ten cents," and scribble down "ten cents" as their final answer without thinking carefully. So they never discover the mistake, let alone come up with the correct answer (five cents). That is normal human behavior. We tend to go with strong hunches. System 1 follows a primitive psycho-logic: if it feels true, it is.

In the Paleolithic world in which our brains evolved, that's not a bad way of making decisions. Gathering all evidence and mulling it over may be the best way to produce accurate answers, but a hunter-gatherer who consults statistics on lions before deciding whether to worry about the shadow moving in the grass isn't likely to live long

enough to bequeath his accuracy-maximizing genes to the next generation. Snap judgments are sometimes essential. As Daniel Kahneman puts it, "System 1 is designed to jump to conclusions from little evidence."[13]

So what about that shadow in the long grass? Should you worry? Well, can you recall a lion emerging from the grass and pouncing on someone? If that memory comes to you easily—it is not the sort of thing people tend to forget—you will conclude lion attacks are common. And then start to worry. Spelling out this process makes it sound ponderous, slow, and calculating but it can happen entirely within System 1—making it automatic, fast, and complete within a few tenths of a second. You see the shadow. Snap! You are frightened—and running. That's the "availability heuristic," one of many System 1 operations—or heuristics—discovered by Daniel Kahneman, his collaborator Amos Tversky, and other researchers in the fast-growing science of judgment and choice.

A defining feature of intuitive judgment is its insensitivity to the quality of the evidence on which the judgment is based. It has to be that way. System 1 can only do its job of delivering strong conclusions at lightning speed if it never pauses to wonder whether the evidence at hand is flawed or inadequate, or if there is better evidence elsewhere. It must treat the available evidence as reliable and sufficient. These tacit assumptions are so vital to System 1 that Kahneman gave them an ungainly but oddly memorable label: WYSIATI (What You See Is All There Is).[14]

Of course, System 1 can't conclude whatever it wants. The human brain demands order. The world must make sense, which means we must be able to explain what we see and think. And we usually can—because we are creative confabulators hardwired to invent stories that impose coherence on the world.

Imagine you're sitting at a table in a research lab, looking at rows of pictures. You pick one, a picture of a shovel. Why are you pointing

at that? Of course you can't answer without more information. But if you were actually at that table, with your finger pointing at a picture of a shovel, simply saying "I don't know" would be a lot harder than you might think. Sane people are expected to have sensible-sounding reasons for their actions. It is awkward to tell others, especially white-lab-coated neuroscientists, "I have no idea why—I just am."

In celebrated research, Michael Gazzaniga designed a bizarre situation in which sane people did indeed have no idea why they were doing what they were doing. His test subjects were "split-brain" patients, meaning that the left and right hemispheres of their brains could not communicate with each other because the connection between them, the corpus callosum, had been surgically severed (traditionally as a treatment for severe epilepsy). These people are remarkably normal, but their condition allows researchers to communicate directly with only one hemisphere of their brain—by showing an image to only the left or right field of vision—without sharing the communication with the other hemisphere. It's like talking to two different people. In this case, the left field of vision (which reports to the right hemisphere) was shown a picture of a snowstorm and the person was asked to point to the picture that related to it. So he quite reasonably pointed at the shovel. The right field of vision (which reports to the left hemisphere) was shown an image of a chicken claw—and the person was then asked why his hand was pointed at a shovel. The left hemisphere had no idea why. But the person didn't say "I don't know." Instead, he made up a story: "Oh, that's simple," one patient said. "The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed."[15]

This compulsion to explain arises with clocklike regularity every time a stock market closes and a journalist says something like "The Dow rose ninety-five points today on news that . . ." A quick check will often reveal that the news that supposedly drove the market came out well after the market had risen. But that minimal level of scrutiny

is seldom applied. It's a rare day when a journalist says, "The market rose today for any one of a hundred different reasons, or a mix of them, so no one knows." Instead, like a split-brain patient asked why he is pointing at a picture of a shovel when he has no idea why, the journalist conjures a plausible story from whatever is at hand.

The explanatory urge is mostly a good thing. Indeed, it is the propulsive force behind all human efforts to comprehend reality. The problem is that we move too fast from confusion and uncertainty ("I have no idea why my hand is pointed at a picture of a shovel") to a clear and confident conclusion ("Oh, that's simple") without spending any time in between ("This is one possible explanation but there are others").

In 2011, when a massive car bomb killed eight people and injured over two hundred in Oslo, the capital of Norway, the first reaction was shock. This was Oslo, one of the most peaceful and prosperous cities on the planet. Speculation erupted on the Internet and cable news. It had to be radical Islamists. It was a car bomb intended to kill as many as possible. And the car had been parked outside the office tower where the prime minister works. It *had* to be Islamists. Just like the London, Madrid, and Bali bombings. Just like 9/11. People rushed to Google to see if they could find supporting information. They succeeded: Norway has soldiers in Afghanistan as part of the NATO mission; Norway has a poorly integrated Muslim community; a radical Muslim preacher had been charged with incitement the week before. Then came word that an even more shocking crime had been committed not long after the bombing. It was a mass shooting—dozens dead—at a summer camp for young people run by the ruling Labour Party. Everything fit. These were coordinated attacks by Islamist terrorists. No doubt about it. Whether the terrorists were homegrown or linked to al-Qaeda remained to be seen, but it was obvious the perpetrators had to be extremist Muslims.

As it turned out, there was only one perpetrator. His name is

it's called attribute substitution, but I call it *bait and switch*: when faced with a hard question, we often surreptitiously replace it with an easy one. "Should I worry about the shadow in the long grass?" is a hard question. Without more data, it may be unanswerable. So we substitute an easier question: "Can I easily recall a lion attacking someone from the long grass?" That question becomes a proxy for the original question and if the answer is yes to the second question, the answer to the first also becomes yes.

So the availability heuristic—like Kahneman's other heuristics—is essentially a bait-and-switch maneuver. And just as the availability heuristic is usually an unconscious System 1 activity, so too is bait and switch.[18]

Of course we aren't always oblivious to the machinations of our minds. If someone asks about climate change, we may say, "I have no training in climatology and haven't read any of the science. If I tried to answer based on what I know I'd make a mess of it. The knowledgeable people are the climatologists. So I'll substitute 'Do most climatologists think climate change is real?' for 'Is climate change real?'" An ordinary person told by an eminent cancer specialist that she has terminal cancer may engage in the same conscious bait and switch and just accept what the doctor says as true.

But Archie Cochrane was no ordinary fellow. He was a prominent physician. He knew that the pathologist hadn't reported. He knew better than anyone that physicians are often too sure of themselves and that this "God complex" can lead them to make terrible mistakes. And yet he immediately accepted that what the specialist said was the true and final word—because, I suspect, Cochrane unconsciously substituted "Is this the sort of person who should know if I have cancer?" for the question "Do I have cancer?" The answer was "Of course! He is an eminent cancer specialist. He saw the cancerous flesh with his own eyes. This is *exactly* the sort of person who should know whether I have cancer." So Cochrane acquiesced.

I realize I am rocking no one's mental universe by saying that people often jump to judgment. Anyone who has spent time around humans knows that. But that is telling. We *know* we should slow down and think before drawing firm conclusions. And yet, when we are faced with a problem, and a seemingly sensible solution springs to mind, we bypass System 2 and declare, "The answer is ten cents." No one is immune, not even skeptics like Archie Cochrane.

We could call this automatic, nearly effortless mode of thinking about the world the default setting, but that won't do. "Default" suggests we can flip the switch to something else. We can't. Like it or not, System 1 operations keep humming, nonstop, beneath the babbling brook of consciousness.

A better metaphor involves vision. The instant we wake up and look past the tip of our nose, sights and sounds flow to the brain and System 1 is engaged. This perspective is subjective, unique to each of us. Only you can see the world from the tip of your own nose. So let's call it the *tip-of-your-nose perspective*.

BLINKING AND THINKING

As imperfect as the view from the tip of your nose may be, you shouldn't discount it entirely.

Popular books often draw a dichotomy between intuition and analysis—"blink" versus "think"—and pick one or the other as the way to go. I am more of a thinker than a blinker, but blink-think is another false dichotomy. The choice isn't either/or, it is how to blend them in evolving situations. That conclusion is not as inspiring as a simple exhortation to take one path or the other, but it has the advantage of being true, as the pioneering researchers behind both perspectives came to understand.

While Daniel Kahneman and Amos Tversky were documenting

System 1's failings, another psychologist, Gary Klein, was examining decision making among professionals like the commanders of firefighting teams, and discovering that snap judgments can work astonishingly well. One commander told Klein about going to a routine kitchen fire and ordering his men to stand in the living room and hose down the flames. The fire subsided at first but roared back. The commander was baffled. He also noticed the living room was surprisingly hot given the size of the kitchen fire. And why was it so quiet? A fire capable of generating that much heat should make more noise. A vague feeling of unease came over the commander and he ordered everyone out of the house. Just as the firefighters reached the street, the floor in the living room collapsed—because the real source of the fire was in the basement, not the kitchen. How had the commander known they were in terrible danger? He told Klein he had ESP (extrasensory perception), but that was just a story he told himself to cover up the fact that he didn't know how he knew. He just knew—the hallmark of an intuitive judgment.

Drawing such seemingly different conclusions about snap judgments, Kahneman and Klein could have hunkered down and fired off rival polemics. But, like good scientists, they got together to solve the puzzle. "We agree on most of the issues that matter," they concluded in a 2009 paper.[19]

There is nothing mystical about an accurate intuition like the fire commander's. It's pattern recognition. With training or experience, people can encode patterns deep in their memories in vast number and intricate detail—such as the estimated fifty thousand to one hundred thousand chess positions that top players have in their repertoire.[20] If something doesn't fit a pattern—like a kitchen fire giving off more heat than a kitchen fire should—a competent expert senses it immediately. But as we see every time someone spots the Virgin Mary in burnt toast or in mold on a church wall, our pattern-recognition ability comes at the cost of susceptibility to false

positives. This, plus the many other ways in which the tip-of-your-nose perspective can generate perceptions that are clear, compelling, and wrong, means intuition can fail as spectacularly as it can work.

Whether intuition generates delusion or insight depends on whether you work in a world full of valid cues you can unconsciously register for future use. "For example, it is very likely that there are early indications that a building is about to collapse in a fire or that an infant will soon show obvious symptoms of infection," Kahneman and Klein wrote. "On the other hand, it is unlikely that there is publicly available information that could be used to predict how well a particular stock will do—if such valid information existed, the price of the stock would already reflect it. Thus, we have more reason to trust the intuition of an experienced fireground commander about the stability of a building, or the intuitions of a nurse about an infant, than to trust the intuitions of a stock broker."[21] Learning the cues is a matter of opportunity and effort. Sometimes learning the cues is easy. "A child does not need thousands of examples to learn to discriminate dogs from cats." But other patterns are much harder to master, like the estimated ten thousand hours of practice it takes to learn those fifty thousand to one hundred thousand chess patterns. "Without those opportunities to learn, a valid intuition can only be due to a lucky accident or to magic," Kahneman and Klein conclude, "and we do not believe in magic."[22]

But there is a catch. As Kahneman and Klein noted, it's often hard to know when there are enough valid cues to make intuition work. And even where it clearly can, caution is advisable. "Often, I cannot explain a certain move, only know that it feels right, and it seems that my intuition is right more often than not," observed the Norwegian prodigy Magnus Carlsen, the world chess champion and the highest-ranked player in history. "If I study a position for an hour then I am usually going in loops and I'm probably not going to come up with something useful. I usually know what I am going to

do after 10 seconds; the rest is double-checking."[23] Carlsen respects his intuition, as well he should, but he also does a lot of "double-checking" because he knows that sometimes intuition can let him down and conscious thought can improve his judgment.

That's excellent practice. The tip-of-your-nose perspective can work wonders but it can also go terribly awry, so if you have the time to think before making a big decision, do so—and be prepared to accept that what seems obviously true now may turn out to be false later.

It is hard to argue with advice that feels about as controversial as a fortune-cookie platitude. But tip-of-your-nose illusions are often so convincing that we bypass the advice and go with our gut. Consider a forecast made by Peggy Noonan—the *Wall Street Journal* columnist and former speechwriter for Ronald Reagan—the day before the presidential election of 2012. It will be a Romney victory, Noonan wrote. Her conclusion was based on the big numbers turning out to Romney rallies. The candidate "looks happy and grateful," Noonan observed. And someone who attended a campaign stop had told Noonan about "the intensity and joy of the crowd." Add it up, Noonan concluded, and "the vibrations are right." It's easy to mock Noonan's vibrations. But who among us hasn't felt mistaken certainty that an election, or some other event, was going to break one way or another because it just felt that way? You may not have said "the vibrations are right," but the thinking is the same.[24]

That's the power of the tip-of-your-nose perspective. It is so persuasive that for thousands of years physicians did not doubt their convictions, causing unnecessary suffering on a gargantuan scale. Progress only really began when physicians accepted that the view from the tip of their nose was not enough to determine what works.

All too often, forecasting in the twenty-first century looks too much like nineteenth-century medicine. There are theories, assertions, and arguments. There are famous figures, as confident as they

in general? All these unanswered questions add up to a big problem. The first step in learning what works in forecasting, and what doesn't, is to judge forecasts, and to do that we can't make assumptions about what the forecast means. We have to know. There can't be any ambiguity about whether a forecast is accurate or not and Ballmer's forecast is ambiguous. Sure, it looks wrong. It feels wrong. There is a strong case to be made that it *is* wrong. But is it wrong beyond all reasonable doubt?

I don't blame the reader for thinking this is too lawyerly, a tad too reminiscent of Bill Clinton's infamous "it depends on what the meaning of the word 'is' is."[3] After all, Ballmer's meaning seems plain, even if a literal reading of his words doesn't support that. But consider his full statement, in context, in an April 2007 interview with *USA Today*: "There's no chance that the iPhone is going to get any significant market share. No chance. It's a $500 subsidized item. They may make a lot of money. But if you actually look at the 1.3 billion phones that get sold, I'd prefer to have our software in 60% or 70% or 80% of them, than I would to have 2% or 3%, which is what Apple might get." That clarifies some things. For one, Ballmer was clearly referring to the global mobile phone market, so it's wrong to measure his forecast against US or global smartphone market share. Using data from the Gartner IT consulting group, I calculated that the iPhone's share of global mobile phone sales in the third quarter of 2013 was roughly 6%.[4] That's higher than the "2% or 3%" Ballmer predicted, but unlike the truncated version so often quoted, it's not laugh-out-loud wrong. Note also that Ballmer didn't say the iPhone would be a bust for Apple. Indeed, he said, "They may make a lot of money." But still there is ambiguity: how much more than 2% or 3% of the global mobile phone market would the iPhone have to capture to be deemed "significant"? Ballmer didn't say. And how much money was he talking about when he said Apple could earn "a lot of money"? Again, he didn't say.