

Overconfidence

The Illusion of Understanding

The trader-philosopher-statistician Nassim Taleb could also be considered a psychologist. In *The Black Swan*, Taleb introduced the notion of a *narrative fallacy* to describe how flawed stories of the past shape our views of the world and our expectations for the future. Narrative fallacies arise inevitably from our continuous attempt to make sense of the world. The explanatory stories that people find compelling are simple; are concrete rather than abstract; assign a larger role to talent, stupidity, and intentions than to luck; and focus on a few striking events that happened rather than on the countless events that failed to happen. Any recent salient event is a candidate to become the kernel of a causal narrative. Taleb suggests that we humans constantly fool ourselves by constructing flimsy accounts of the past and believing they are true.

Good stories provide a simple and coherent account >

A compelling narrative fosters an illusion of inevitability. Consider the story of how Google turned into a giant of the technology industry. Two creative graduate students in the computer science department at Stanford University come up with a superior way of searching information on the Internet. They seek and obtain funding to start a company and make a series of decisions that work out well. Within a few years, the company they started is one of the most valuable stocks in America, and the two former graduate students are among the richest people on the planet. On one memorable occasion, they were lucky, which makes the story even more compelling: a year after founding Google, they were willing to sell their company for less than \$1 million, but the buyer said the price was too high. Mentioning the single lucky incident actually makes it easier to underestimate the multitude of ways in which luck affected the outcome.

A detailed history would specify the decisions of Google's founders, but for our purposes it suffices to say that almost every choice they made had a good outcome. A more complete narrative would describe the actions of the firms that Google defeated. The hapless competitors would appear to be blind, slow, and altogether inadequate in dealing with the threat that eventually overwhelmed them.

I intentionally told this tale blandly, but you get the idea: there is a very good story here. Fleshed out in more detail, the story could give you the sense that you understand what made Google succeed; it would also make you feel that you have learned a valuable general lesson about what makes businesses succeed. Unfortunately, there is good reason to believe that your sense of understanding and learning from the Google story is largely illusory. The ultimate test of an explanation is whether it would have

made the event predictable in advance. No story of Google's unlikely success will meet that test, because no story can include the myriad of events that would have caused a different outcome. The human mind does not deal well with nonevents. The fact that many of the important events that did occur involve choices further tempts you to exaggerate the role of skill and underestimate the part that luck played in the outcome. Because every critical decision turned out well, the record suggests almost flawless prescience—but bad luck could have disrupted any one of the successful steps. The halo effect adds the final touches, lending an aura of invincibility to the heroes of the story.

Like watching a skilled rafter avoiding one potential calamity after another as he goes down the rapids, the unfolding of the Google story is thrilling because of the constant risk of disaster. However, there is for an instructive difference between the two cases. The skilled rafter has gone down rapids hundreds of times. He has learned to read the roiling water in front of him and to anticipate obstacles. He has learned to make the tiny adjustments of posture that keep him upright. There are fewer opportunities for young men to learn how to create a giant company, and fewer chances to avoid hidden rocks—such as a brilliant innovation by a competing firm. Of course there was a great deal of skill in the Google story, but luck played a more important role in the actual event than it does in the telling of it. And the more luck was involved, the less there is to be learned.

At work here is that powerful WY SIATI rule. You cannot help dealing with the limited information you have as if it were all there is to know. You build the best possible story from the information available to you, and if it is a good story, you believe it. Paradoxically, it is easier to construct a coherent story when you know little, when there are fewer pieces to fit into the puzzle. Our comforting conviction that the world makes sense rests on a secure foundation: our almost unlimited ability to ignore our ignorance.

I have heard of too many people who “knew well before it happened that the 2008 financial crisis was inevitable.” This sentence contains a highly objectionable word, which should be removed from our vocabulary in discussions of major events. The word is, of course, *knew*. Some people thought well in advance that there would be a crisis, but they did not know it. They now say they knew it because the crisis did in fact happen. This is a misuse of an important concept. In everyday language, we apply the word *know* only when what was known is true and can be shown to be true. We can know something only if it is both true and knowable. But the people who thought there would be a crisis (and there are fewer of them than now remember thinking it) could not conclusively show it at the time. Many

intelligent and well-informed people were keenly interested in the future of the economy and did not believe a catastrophe was imminent; I infer from this fact that the crisis was not knowable. What is perverse about the use of *know* in this context is not that some individuals get credit for prescience that they do not deserve. It is that the language implies that the world is more knowable than it is. It helps perpetuate a pernicious illusion.

The core of the illusion is that we believe we understand the past, which implies that the future also should be knowable, but in fact we understand the past less than we believe we do. *Know* is not the only word that fosters this illusion. In common usage, the words *intuition* and *premonition* also are reserved for past thoughts that turned out to be true. The statement “I had a premonition that the marriage would not last, but I was wrong” sounds odd, as does any sentence about an intuition that turned out to be false. To think clearly about the future, we need to clean up the language that we use in labeling the beliefs we had in the past.

The Social Costs of Hindsight

The mind that makes up narratives about the past is a sense-making organ. When an unpredicted event occurs, we immediately adjust our view of the world to accommodate the surprise. Imagine yourself before a football game between two teams that have the same record of wins and losses. Now the game is over, and one team trashed the other. In your revised model of the world, the winning team is much stronger than the loser, and your view of the past as well as of the future has been altered by your new perception. Learning from surprises is a reasonable thing to do, but it can have some dangerous consequences.

A general limitation of the human mind is its imperfect ability to reconstruct past states of knowledge, or beliefs that have changed. Once you adopt a new view of the world (or of any part of it), you immediately lose much of your ability to recall what you used to believe before your mind changed.

Many psychologists have studied what happens when people change their minds. Choosing a topic on which minds are not completely made up—say, the death penalty—the experimenter carefully measures people’s attitudes. Next, the participants see or hear a persuasive pro or con message. Then the experimenter measures people’s attitudes again; they usually are closer to the persuasive message they were exposed to. Finally, the participants report the opinion they held beforehand. This task turns out to be surprisingly difficult. Asked to reconstruct their former beliefs, people retrieve their current ones instead—an instance of

substitution—and many cannot believe that they ever felt differently.

Your inability to reconstruct past beliefs will inevitably cause you to underestimate the extent to which you were surprised by past events. Baruch Fischhoff first demonstrated this “I-knew-it-all-along” effect, or *hindsight bias*, when he was a student in Jerusalem. Together with Ruth Beyth (another of our students), Fischhoff conducted a survey before President Richard Nixon visited China and Russia in 1972. The respondents assigned probabilities to fifteen possible outcomes of Nixon’s diplomatic initiatives. Would Mao Zedong agree to meet with Nixon? Might the United States grant diplomatic recognition to China? After decades of enmity, could the United States and the Soviet Union agree on anything significant?

After Nixon’s return from his travels, Fischhoff and Beyth asked the same people to recall the probability that they had originally assigned to each of the fifteen possible outcomes. The results were clear. If an event had actually occurred, people exaggerated the probability that they had assigned to it earlier. If the possible event had not come to pass, the participants erroneously recalled that they had always considered it unlikely. Further experiments showed that people were driven to overstate the accuracy not only of their original predictions but also of those made by others. Similar results have been found for other events that gripped public attention, such as the O. J. Simpson murder trial and the impeachment of President Bill Clinton. The tendency to revise the history of one’s beliefs in light of what actually happened produces a robust cognitive illusion.

Hindsight bias has pernicious effects on the evaluations of decision makers. It leads observers to assess the quality of a decision not by whether the process was sound but by whether its outcome was good or bad. Consider a low-risk surgical intervention in which an unpredictable accident occurred that caused the patient’s death. The jury will be prone to believe, after the fact, that the operation was actually risky and that the doctor who ordered it should have known better. This outcome bias makes it almost impossible to evaluate a decision properly—in terms of the beliefs that were reasonable when the decision was made.

Hindsight is especially unkind to decision makers who act as agents for others—physicians, financial advisers, third-base coaches, CEOs, social workers, diplomats, politicians. We are prone to blame decision makers for good decisions that worked out badly and to give them too little credit for successful moves that appear obvious only after the fact. There is a clear *outcome bias*. When the outcomes are bad, the clients often blame their agents for not seeing the handwriting on the wall—forgetting that it was written in invisible ink that became legible only afterward. Actions that

seemed prudent in foresight can look irresponsibly negligent in hindsight. Based on an actual legal case, students in California were asked whether the city of Duluth, Minnesota, should have shouldered the considerable cost of hiring a full-time bridge monitor to protect against the risk that debris might get caught and block the free flow of water. One group was shown only the evidence available at the time of the city's decision; 24% of these people felt that Duluth should take on the expense of hiring a flood monitor. The second group was informed that debris had blocked the river, causing major flood damage; 56% of these people said the city should have hired the monitor, although they had been explicitly instructed not to let hindsight distort their judgment.

The worse the consequence, the greater the hindsight bias. In the case of a catastrophe, such as 9/11, we are especially ready to believe that the officials who failed to anticipate it were negligent or blind. On July 10, 2001, the Central Intelligence Agency obtained information that al-Qaeda might be planning a major attack against the United States. George Tenet, director of the CIA, brought the information not to President George W. Bush but to National Security Adviser Condoleezza Rice. When the facts later emerged, Ben Bradlee, the legendary executive editor of *The Washington Post*, declared, "It seems to me elementary that if you've got the story that's going to dominate history you might as well go right to the president." But on July 10, no one knew—or could have known—that this tidbit of intelligence would turn out to dominate history.

Because adherence to standard operating procedures is difficult to second-guess, decision makers who expect to have their decisions scrutinized with hindsight are driven to bureaucratic solutions—and to an extreme reluctance to take risks. As malpractice litigation became more common, physicians changed their procedures in multiple ways: ordered more tests, referred more cases to specialists, applied conventional treatments even when they were unlikely to help. These actions protected the physicians more than they benefited the patients, creating the potential for conflicts of interest. Increased accountability is a mixed blessing.

Although hindsight and the outcome bias generally foster risk aversion, they also bring undeserved rewards to irresponsible risk seekers, such as a general or an entrepreneur who took a crazy gamble and won. Leaders who have been lucky are never punished for having taken too much risk. Instead, they are believed to have had the flair and foresight to anticipate success, and the sensible people who doubted them are seen in hindsight as mediocre, timid, and weak. A few lucky gambles can crown a reckless leader with a halo of prescience and boldness.

Recipes for Success

The sense-making machinery of System 1 makes us see the world as more tidy, simple, predictable, and coherent than it really is. The illusion that one has understood the past feeds the further illusion that one can predict and control the future. These illusions are comforting. They reduce the anxiety that we would experience if we allowed ourselves to fully acknowledge the uncertainties of existence. We all have a need for the reassuring message that actions have appropriate consequences, and that success will reward wisdom and courage. Many business books are tailor-made to satisfy this need.

Do leaders and management practices influence the outcomes of firms in the market? Of course they do, and the effects have been confirmed by systematic research that objectively assessed the characteristics of CEOs and their decisions, and related them to subsequent outcomes of the firm. In one study, the CEOs were characterized by the strategy of the companies they had led before their current appointment, as well as by management rules and procedures adopted after their appointment. CEOs do influence performance, but the effects are much smaller than a reading of the business press suggests.

Researchers measure the strength of relationships by a correlation coefficient, which varies between 0 and 1. The coefficient was defined earlier (in relation to regression to the mean) by the extent to which two measures are determined by shared factors. A very generous estimate of the correlation between the success of the firm and the quality of its CEO might be as high as .30, indicating 30% overlap. To appreciate the significance of this number, consider the following question:

Suppose you consider many pairs of firms. The two firms in each pair are generally similar, but the CEO of one of them is better than the other. How often will you find that the firm with the stronger CEO is the more successful of the two?

In a well-ordered and predictable world, the correlation would be perfect (1), and the stronger CEO would be found to lead the more successful firm in 100% of the pairs. If the relative success of similar firms was determined entirely by factors that the CEO does not control (call them luck, if you wish), you would find the more successful firm led by the weaker CEO 50% of the time. A correlation of .30 implies that you would find the stronger CEO leading the stronger firm in about 60% of the pairs—an improvement of a mere 10 percentage points over random guessing, hardly grist for the

hero worship of CEOs we so often witness.

If you expected this value to be higher—and most of us do—then you should take that as an indication that you are prone to overestimate the predictability of the world you live in. Make no mistake: improving the odds of success from 1:1 to 3:2 is a very significant advantage, both at the racetrack and in business. From the perspective of most business writers, however, a CEO who has so little control over performance would not be particularly impressive even if her firm did well. It is difficult to imagine people lining up at airport bookstores to buy a book that enthusiastically describes the practices of business leaders who, on average, do somewhat better than chance. Consumers have a hunger for a clear message about the determinants of success and failure in business, and they need stories that offer a sense of understanding, however illusory.

In his penetrating book *The Halo Effect*, Philip Rosenzweig, a business school professor based in Switzerland, shows how the demand for illusory certainty is met in two popular genres of business writing: histories of the rise (usually) and fall (occasionally) of particular individuals and companies, and analyses of differences between successful and less successful firms. He concludes that stories of success and failure consistently exaggerate the impact of leadership style and management practices on firm outcomes, and thus their message is rarely useful.

To appreciate what is going on, imagine that business experts, such as other CEOs, are asked to comment on the reputation of the chief executive of a company. They popare keenly aware of whether the company has recently been thriving or failing. As we saw earlier in the case of Google, this knowledge generates a halo. The CEO of a successful company is likely to be called flexible, methodical, and decisive. Imagine that a year has passed and things have gone sour. The same executive is now described as confused, rigid, and authoritarian. Both descriptions sound right at the time: it seems almost absurd to call a successful leader rigid and confused, or a struggling leader flexible and methodical.

Indeed, the halo effect is so powerful that you probably find yourself resisting the idea that the same person and the same behaviors appear methodical when things are going well and rigid when things are going poorly. Because of the halo effect, we get the causal relationship backward: we are prone to believe that the firm fails because its CEO is rigid, when the truth is that the CEO appears to be rigid because the firm is failing. This is how illusions of understanding are born.

The halo effect and outcome bias combine to explain the extraordinary appeal of books that seek to draw operational morals from systematic examination of successful businesses. One of the best-known examples of

this genre is Jim Collins and Jerry I. Porras's *Built to Last*. The book contains a thorough analysis of eighteen pairs of competing companies, in which one was more successful than the other. The data for these comparisons are ratings of various aspects of corporate culture, strategy, and management practices. "We believe every CEO, manager, and entrepreneur in the world should read this book," the authors proclaim. "You can build a visionary company."

The basic message of *Built to Last* and other similar books is that good managerial practices can be identified and that good practices will be rewarded by good results. Both messages are overstated. The comparison of firms that have been more or less successful is to a significant extent a comparison between firms that have been more or less lucky. Knowing the importance of luck, you should be particularly suspicious when highly consistent patterns emerge from the comparison of successful and less successful firms. In the presence of randomness, regular patterns can only be mirages.

Because luck plays a large role, the quality of leadership and management practices cannot be inferred reliably from observations of success. And even if you had perfect foreknowledge that a CEO has brilliant vision and extraordinary competence, you still would be unable to predict how the company will perform with much better accuracy than the flip of a coin. On average, the gap in corporate profitability and stock returns between the outstanding firms and the less successful firms studied in *Built to Last* shrank to almost nothing in the period following the study. The average profitability of the companies identified in the famous *In Search of Excellence* dropped sharply as well within a short time. A study of *Fortune's* "Most Admired Companies" finds that over a twenty-year period, the firms with the worst ratings went on to earn much higher stock returns than the most admired firms.

You are probably tempted to think of causal explanations for these observations: perhaps the successful firms became complacent, the less successful firms tried harder. But this is the wrong way to think about what happened. The average gap must shrink, because the original gap was due in good part to luck, which contributed both to the success of the top firms and to the lagging performance of the rest. We have already encountered this statistical fact of life: regression to the mean.

Stories of how businesses rise and fall strike a chord with readers by offering what the human mind needs: a simple message of triumph and failure that identifies clear causes and ignores the determinative power of luck and the inevitability of regression. These stories induce and maintain an illusion of understanding, imparting lessons of little enduring value to

readers who are all too eager to believe them.

Speaking of Hindsight

“The mistake appears obvious, but it is just hindsight. You could not have known in advance.”

“He’s learning too much from this success story, which is too tidy. He has fallen for a narrative fallacy.”

“She has no evidence for saying that the firm is badly managed. All she knows is that its stock has gone down. This is an outcome bias, part hindsight and part halo effect.”

“Let’s not fall for the outcome bias. This was a stupid decision even though it worked out well.”

The Illusion of Validity

System 1 is designed to jump to conclusions from little evidence—and it is not designed to know the size of its jumps. Because of WYSIATI, only the evidence at hand counts. Because of confidence by coherence, the subjective confidence we have in our opinions reflects the coherence of the story that System 1 and System 2 have constructed. The amount of evidence and its quality do not count for much, because poor evidence can make a very good story. For some of our most important beliefs we have no evidence at all, except that people we love and trust hold these beliefs. Considering how little we know, the confidence we have in our beliefs is preposterous—and it is also essential.

The Illusion of Validity

Many decades ago I spent what seemed like a great deal of time under a scorching sun, watching groups of sweaty soldiers as they solved a problem. I was doing my national service in the Israeli Army at the time. I had completed an undergraduate degree in psychology, and after a year as an infantry officer was assigned to the army's Psychology Branch, where one of my occasional duties was to help evaluate candidates for officer training. We used methods that had been developed by the British Army in World War II.

One test, called the "leaderless group challenge," was conducted on an obstacle field. Eight candidates, strangers to each other, with all insignia of rank removed and only numbered tags to identify them, were instructed to lift a long log from the ground and haul it to a wall about six feet high. The entire group had to get to the other side of the wall without the log touching either the ground or the wall, and without anyone touching the wall. If any of these things happened, they had to declare itsignлб T and start again.

There was more than one way to solve the problem. A common solution was for the team to send several men to the other side by crawling over the pole as it was held at an angle, like a giant fishing rod, by other members of the group. Or else some soldiers would climb onto someone's shoulders and jump across. The last man would then have to jump up at the pole, held up at an angle by the rest of the group, shinny his way along its length as the others kept him and the pole suspended in the air, and leap safely to the other side. Failure was common at this point, which required them to start all over again.

As a colleague and I monitored the exercise, we made note of who took charge, who tried to lead but was rebuffed, how cooperative each soldier

was in contributing to the group effort. We saw who seemed to be stubborn, submissive, arrogant, patient, hot-tempered, persistent, or a quitter. We sometimes saw competitive spite when someone whose idea had been rejected by the group no longer worked very hard. And we saw reactions to crisis: who berated a comrade whose mistake had caused the whole group to fail, who stepped forward to lead when the exhausted team had to start over. Under the stress of the event, we felt, each man's true nature revealed itself. Our impression of each candidate's character was as direct and compelling as the color of the sky.

After watching the candidates make several attempts, we had to summarize our impressions of soldiers' leadership abilities and determine, with a numerical score, who should be eligible for officer training. We spent some time discussing each case and reviewing our impressions. The task was not difficult, because we felt we had already seen each soldier's leadership skills. Some of the men had looked like strong leaders, others had seemed like wimps or arrogant fools, others mediocre but not hopeless. Quite a few looked so weak that we ruled them out as candidates for officer rank. When our multiple observations of each candidate converged on a coherent story, we were completely confident in our evaluations and felt that what we had seen pointed directly to the future. The soldier who took over when the group was in trouble and led the team over the wall was a leader at that moment. The obvious best guess about how he would do in training, or in combat, was that he would be as effective then as he had been at the wall. Any other prediction seemed inconsistent with the evidence before our eyes.

Because our impressions of how well each soldier had performed were generally coherent and clear, our formal predictions were just as definite. A single score usually came to mind and we rarely experienced doubts or formed conflicting impressions. We were quite willing to declare, "This one will never make it," "That fellow is mediocre, but he should do okay," or "He will be a star." We felt no need to question our forecasts, moderate them, or equivocate. If challenged, however, we were prepared to admit, "But of course anything could happen." We were willing to make that admission because, despite our definite impressions about individual candidates, we knew with certainty that our forecasts were largely useless.

The evidence that we could not forecast success accurately was overwhelming. Every few months we had a feedback session in which we learned how the cadets were doing at the officer-training school and could compare our assessments against the opinions of commanders who had been monitoring them for some time. The story was always the same: our ability to predict performance at the school was negligible. Our forecasts were better than blind guesses, but not by much.

We weed ■re downcast for a while after receiving the discouraging news. But this was the army. Useful or not, there was a routine to be followed and orders to be obeyed. Another batch of candidates arrived the next day. We took them to the obstacle field, we faced them with the wall, they lifted the log, and within a few minutes we saw their true natures revealed, as clearly as before. The dismal truth about the quality of our predictions had no effect whatsoever on how we evaluated candidates and very little effect on the confidence we felt in our judgments and predictions about individuals.

What happened was remarkable. The global evidence of our previous failure should have shaken our confidence in our judgments of the candidates, but it did not. It should also have caused us to moderate our predictions, but it did not. We knew as a general fact that our predictions were little better than random guesses, but we continued to feel and act as if each of our specific predictions was valid. I was reminded of the Müller-Lyer illusion, in which we know the lines are of equal length yet still see them as being different. I was so struck by the analogy that I coined a term for our experience: the *illusion of validity*.

I had discovered my first cognitive illusion.

Decades later, I can see many of the central themes of my thinking—and of this book—in that old story. Our expectations for the soldiers' future performance were a clear instance of substitution, and of the representativeness heuristic in particular. Having observed one hour of a soldier's behavior in an artificial situation, we felt we knew how well he would face the challenges of officer training and of leadership in combat. Our predictions were completely nonregressive—we had no reservations about predicting failure or outstanding success from weak evidence. This was a clear instance of WYSIATI. We had compelling impressions of the behavior we observed and no good way to represent our ignorance of the factors that would eventually determine how well the candidate would perform as an officer.

Looking back, the most striking part of the story is that our knowledge of the general rule—that we could not predict—had no effect on our confidence in individual cases. I can see now that our reaction was similar to that of Nisbett and Borgida's students when they were told that most people did not help a stranger suffering a seizure. They certainly believed the statistics they were shown, but the base rates did not influence their judgment of whether an individual they saw on the video would or would not help a stranger. Just as Nisbett and Borgida showed, people are often

reluctant to infer the particular from the general.

Subjective confidence in a judgment is not a reasoned evaluation of the probability that this judgment is correct. Confidence is a feeling, which reflects the coherence of the information and the cognitive ease of processing it. It is wise to take admissions of uncertainty seriously, but declarations of high confidence mainly tell you that an individual has constructed a coherent story in his mind, not necessarily that the story is true.

The Illusion of Stock-Picking Skill

In 1984, Amos and I and our friend Richard Thaler visited a Wall Street firm. Our host, a senior investment manager, had invited us to discuss the role of judgment biases in investing. I knew so little about finance that I did not even know what to ask him, but I remember one exchange. "When you sell a stock," I asked, "who buys it?" He answered with a wave in the vague direction of the window, indicating that he expected the buyer to be someone else very much like him. That was odd: What made one person buy and the other sell? What did the sellers think they knew that the buyers did not?

Since then, my questions about the stock market have hardened into a larger puzzle: a major industry appears to be built largely on an *illusion of skill*. Billions of shares are traded every day, with many people buying each stock and others selling it to them. It is not unusual for more than 100 million shares of a single stock to change hands in one day. Most of the buyers and sellers know that they have the same information; they exchange the stocks primarily because they have different opinions. The buyers think the price is too low and likely to rise, while the sellers think the price is high and likely to drop. The puzzle is why buyers and sellers alike think that the current price is wrong. What makes them believe they know more about what the price should be than the market does? For most of them, that belief is an illusion.

In its broad outlines, the standard theory of how the stock market works is accepted by all the participants in the industry. Everybody in the investment business has read Burton Malkiel's wonderful book *A Random Walk Down Wall Street*. Malkiel's central idea is that a stock's price incorporates all the available knowledge about the value of the company and the best predictions about the future of the stock. If some people believe that the price of a stock will be higher tomorrow, they will buy more of it today. This, in turn, will cause its price to rise. If all assets in a market are correctly priced, no one can expect either to gain or to lose by trading.

Perfect prices leave no scope for cleverness, but they also protect fools from their own folly. We now know, however, that the theory is not quite right. Many individual investors lose consistently by trading, an achievement that a dart-throwing chimp could not match. The first demonstration of this startling conclusion was collected by Terry Odean, a finance professor at UC Berkeley who was once my student.

Odean began by studying the trading records of 10,000 brokerage accounts of individual investors spanning a seven-year period. He was able to analyze every transaction the investors executed through that firm, nearly 163,000 trades. This rich set of data allowed Odean to identify all instances in which an investor sold some of his holdings in one stock and soon afterward bought another stock. By these actions the investor revealed that he (most of the investors were men) had a definite idea about the future of the two stocks: he expected the stock that he chose to buy to do better than the stock he chose to sell.

To determine whether those ideas were well founded, Odean compared the returns of the stock the investor had sold and the stock he had bought in its place, over the course of one year after the transaction. The results were unequivocally bad. On average, the shares that individual traders sold did better than those they bought, by a very substantial margin: 3.2 percentage points per year, above and beyond the significant costs of executing the two trades.

It is important to remember that this is a statement about averages: some individuals did much better, others did much worse. However, it is clear that for the large majority of individual investors, taking a shower and doing nothing would have been a better policy than implementing the ideas that came to their minds. Later research by Odean and his colleague Brad Barber supported this conclusion. In a paper titled "Trading Is Hazardous to Your Wealth," they showed that, on average, the most active traders had the poorest results, while the investors who traded the least earned the highest returns. In another paper, titled "Boys Will Be Boys," they showed that men acted on their useless ideas significantly more often than women, and that as a result women achieved better investment results than men.

Of course, there is always someone on the other side of each transaction; in general, these are financial institutions and professional investors, who are ready to take advantage of the mistakes that individual traders make in choosing a stock to sell and another stock to buy. Further research by Barber and Odean has shed light on these mistakes. Individual investors like to lock in their gains by selling "winners," stocks that have appreciated since they were purchased, and they hang on to their losers. Unfortunately for them, recent winners tend to do better than recent losers in the short run, so individuals sell the wrong stocks. They

also buy the wrong stocks. Individual investors predictably flock to companies that draw their attention because they are in the news. Professional investors are more selective in responding to news. These findings provide some justification for the label of “smart money” that finance professionals apply to themselves.

Although professionals are able to extract a considerable amount of wealth from amateurs, few stock pickers, if any, have the skill needed to beat the market consistently, year after year. Professional investors, including fund managers, fail a basic test of skill: persistent achievement. The diagnostic for the existence of any skill is the consistency of individual differences in achievement. The logic is simple: if individual differences in any one year are due entirely to luck, the ranking of investors and funds will vary erratically and the year-to-year correlation will be zero. Where there is skill, however, the rankings will be more stable. The persistence of individual differences is the measure by which we confirm the existence of skill among golfers, car salespeople, orthodontists, or speedy toll collectors on the turnpike.

Mutual funds are run by highly experienced and hardworking professionals who buy and sell stocks to achieve the best possible results for their clients. Nevertheless, the evidence from more than fifty years of research is conclusive: for a large majority of fund managers, the selection of stocks is more like rolling dice than like playing poker. Typically at least two out of every three mutual funds underperform the overall market in any given year.

More important, the year-to-year correlation between the outcomes of mutual funds is very small, barely higher than zero. The successful funds in any given year are mostly lucky; they have a good roll of the dice. There is general agreement among researchers that nearly all stock pickers, whether they know it or not—and few of them do—are playing a game of chance. The subjective experience of traders is that they are making sensible educated guesses in a situation of great uncertainty. In highly efficient markets, however, educated guesses are no more accurate than blind guesses.

Some years ago I had an unusual opportunity to examine the illusion of financial skill up close. I had been invited to speak to a group of investment advisers in a firm that provided financial advice and other services to very wealthy clients. I asked for some data to prepare my presentation and was granted a small treasure: a spreadsheet summarizing the investment outcomes of some twenty-five anonymous wealth advisers, for each of

eight consecutive years. Each adviser's scoof score for each year was his (most of them were men) main determinant of his year-end bonus. It was a simple matter to rank the advisers by their performance in each year and to determine whether there were persistent differences in skill among them and whether the same advisers consistently achieved better returns for their clients year after year.

To answer the question, I computed correlation coefficients between the rankings in each pair of years: year 1 with year 2, year 1 with year 3, and so on up through year 7 with year 8. That yielded 28 correlation coefficients, one for each pair of years. I knew the theory and was prepared to find weak evidence of persistence of skill. Still, I was surprised to find that the average of the 28 correlations was .01. In other words, zero. The consistent correlations that would indicate differences in skill were not to be found. The results resembled what you would expect from a dice-rolling contest, not a game of skill.

No one in the firm seemed to be aware of the nature of the game that its stock pickers were playing. The advisers themselves felt they were competent professionals doing a serious job, and their superiors agreed. On the evening before the seminar, Richard Thaler and I had dinner with some of the top executives of the firm, the people who decide on the size of bonuses. We asked them to guess the year-to-year correlation in the rankings of individual advisers. They thought they knew what was coming and smiled as they said "not very high" or "performance certainly fluctuates." It quickly became clear, however, that no one expected the average correlation to be zero.

Our message to the executives was that, at least when it came to building portfolios, the firm was rewarding luck as if it were skill. This should have been shocking news to them, but it was not. There was no sign that they disbelieved us. How could they? After all, we had analyzed their own results, and they were sophisticated enough to see the implications, which we politely refrained from spelling out. We all went on calmly with our dinner, and I have no doubt that both our findings and their implications were quickly swept under the rug and that life in the firm went on just as before. The illusion of skill is not only an individual aberration; it is deeply ingrained in the culture of the industry. Facts that challenge such basic assumptions—and thereby threaten people's livelihood and self-esteem—are simply not absorbed. The mind does not digest them. This is particularly true of statistical studies of performance, which provide base-rate information that people generally ignore when it clashes with their personal impressions from experience.

The next morning, we reported the findings to the advisers, and their response was equally bland. Their own experience of exercising careful

judgment on complex problems was far more compelling to them than an obscure statistical fact. When we were done, one of the executives I had dined with the previous evening drove me to the airport. He told me, with a trace of defensiveness, "I have done very well for the firm and no one can take that away from me." I smiled and said nothing. But I thought, "Well, I took it away from you this morning. If your success was due mostly to chance, how much credit are you entitled to take for it?"

What Supports the Illusions of Skill and Validity?

Cognitive illusions can be more stubborn than visual illusions. What you learned about the Müller-Lyer illusion did not change the way you see the lines, but it changed your behavior. You now know that you cannot trust your impression of the length of lines that have fins appended to them, and you also know that in the standard Müller-Lyer display you cannot trust what you see. When asked about the length of the lines, you will report your informed belief, not the illusion that you continue to see. In contrast, when my colleagues and I in the army learned that our leadership assessment tests had low validity, we accepted that fact intellectually, but it had no impact on either our feelings or our subsequent actions. The response we encountered in the financial firm was even more extreme. I am convinced that the message that Thaler and I delivered to both the executives and the portfolio managers was instantly put away in a dark corner of memory where it would cause no damage.

Why do investors, both amateur and professional, stubbornly believe that they can do better than the market, contrary to an economic theory that most of them accept, and contrary to what they could learn from a dispassionate evaluation of their personal experience? Many of the themes of previous chapters come up again in the explanation of the prevalence and persistence of an illusion of skill in the financial world.

The most potent psychological cause of the illusion is certainly that the people who pick stocks are exercising high-level skills. They consult economic data and forecasts, they examine income statements and balance sheets, they evaluate the quality of top management, and they assess the competition. All this is serious work that requires extensive training, and the people who do it have the immediate (and valid) experience of using these skills. Unfortunately, skill in evaluating the business prospects of a firm is not sufficient for successful stock trading, where the key question is whether the information about the firm is already incorporated in the price of its stock. Traders apparently lack the skill to answer this crucial question, but they appear to be ignorant of their

ignorance. As I discovered from watching cadets on the obstacle field, subjective confidence of traders is a feeling, not a judgment. Our understanding of cognitive ease and associative coherence locates subjective confidence firmly in System 1.

Finally, the illusions of validity and skill are supported by a powerful professional culture. We know that people can maintain an unshakable faith in any proposition, however absurd, when they are sustained by a community of like-minded believers. Given the professional culture of the financial community, it is not surprising that large numbers of individuals in that world believe themselves to be among the chosen few who can do what they believe others cannot.

The Illusions of Pundits

The idea that the future is unpredictable is undermined every day by the ease with which the past is explained. As Nassim Taleb pointed out in *The Black Swan*, our tendency to construct and believe coherent narratives of the past makes it difficult for us to accept the limits of our forecasting ability. Everything makes sense in hindsight, a fact that financial pundits exploit every evening as they offer convincing accounts of the day's events. And we cannot suppress the powerful intuition that what makes sense in hindsight today was predictable yesterday. The illusion that we understand the past fosters overconfidence in our ability to predict the future.

The often-used image of the "march of history" implies order and direction. Marches, unlike strolls or walks, are not random. We think that we should be able to explain the past by focusing on either large social movements and cultural and technological developments or the intentions and abilities of a few great men. The idea that large historical events are determined by luck is profoundly shocking, although it is demonstrably true. It is hard to think of the history of the twentieth century, including its large social movements, without bringing in the role of Hitler, Stalin, and Mao Zedong. But there was a moment in time, just before an egg was fertilized, when there was a fifty-fifty chance that the embryo that became Hitler could have been a female. Compounding the three events, there was a probability of one-eighth of a twentieth century without any of the three great villains and it is impossible to argue that history would have been roughly the same in their absence. The fertilization of these three eggs had momentous consequences, and it makes a joke of the idea that long-term developments are predictable.

Yet the illusion of valid prediction remains intact, a fact that is exploited by people whose business is prediction—not only financial experts but

pundits in business and politics, too. Television and radio stations and newspapers have their panels of experts whose job it is to comment on the recent past and foretell the future. Viewers and readers have the impression that they are receiving information that is somehow privileged, or at least extremely insightful. And there is no doubt that the pundits and their promoters genuinely believe they are offering such information. Philip Tetlock, a psychologist at the University of Pennsylvania, explained these so-called expert predictions in a landmark twenty-year study, which he published in his 2005 book *Expert Political Judgment: How Good Is It? How Can We Know?* Tetlock has set the terms for any future discussion of this topic.

Tetlock interviewed 284 people who made their living “commenting or offering advice on political and economic trends.” He asked them to assess the probabilities that certain events would occur in the not too distant future, both in areas of the world in which they specialized and in regions about which they had less knowledge. Would Gorbachev be ousted in a coup? Would the United States go to war in the Persian Gulf? Which country would become the next big emerging market? In all, Tetlock gathered more than 80,000 predictions. He also asked the experts how they reached their conclusions, how they reacted when proved wrong, and how they evaluated evidence that did not support their positions. Respondents were asked to rate the probabilities of three alternative outcomes in every case: the persistence of the status quo, more of something such as political freedom or economic growth, or less of that thing.

The results were devastating. The experts performed worse than they would have if they had simply assigned equal probabilities to each of the three potential outcomes. In other words, people who spend their time, and earn their living, studying a particular topic produce poorer predictions than dart-throwing monkeys who would have distributed their choices evenly over the options. Even in the region they knew best, experts were not significantly better than nonspecialists.

Those who know more forecast very slightly better than those who know less. But those with the most knowledge are often less reliable. The reason is that the person who acquires more knowledge develops an enhanced illusion of her skill and becomes unrealistically overconfident. “We reach the point of diminishing marginal predictive returns for knowledge disconcertingly quickly,” Tetlock writes. “In this age of academic hyperspecialization, there is no reason for supposing that contributors to top journals—distinguished political scientists, area study specialists, economists, and so on—are any better than journalists or attentive readers

of *The New York Times* in 'reading¶17; emerging situations." The more famous the forecaster, Tetlock discovered, the more flamboyant the forecasts. "Experts in demand," he writes, "were more overconfident than their colleagues who eked out existences far from the limelight."

Tetlock also found that experts resisted admitting that they had been wrong, and when they were compelled to admit error, they had a large collection of excuses: they had been wrong only in their timing, an unforeseeable event had intervened, or they had been wrong but for the right reasons. Experts are just human in the end. They are dazzled by their own brilliance and hate to be wrong. Experts are led astray not by what they believe, but by how they think, says Tetlock. He uses the terminology from Isaiah Berlin's essay on Tolstoy, "The Hedgehog and the Fox." Hedgehogs "know one big thing" and have a theory about the world; they account for particular events within a coherent framework, bristle with impatience toward those who don't see things their way, and are confident in their forecasts. They are also especially reluctant to admit error. For hedgehogs, a failed prediction is almost always "off only on timing" or "very nearly right." They are opinionated and clear, which is exactly what television producers love to see on programs. Two hedgehogs on different sides of an issue, each attacking the idiotic ideas of the adversary, make for a good show.

Foxes, by contrast, are complex thinkers. They don't believe that one big thing drives the march of history (for example, they are unlikely to accept the view that Ronald Reagan single-handedly ended the cold war by standing tall against the Soviet Union). Instead the foxes recognize that reality emerges from the interactions of many different agents and forces, including blind luck, often producing large and unpredictable outcomes. It was the foxes who scored best in Tetlock's study, although their performance was still very poor. They are less likely than hedgehogs to be invited to participate in television debates.

It is Not the Experts' Fault—The World is Difficult

The main point of this chapter is not that people who attempt to predict the future make many errors; that goes without saying. The first lesson is that errors of prediction are inevitable because the world is unpredictable. The second is that high subjective confidence is not to be trusted as an indicator of accuracy (low confidence could be more informative).

Short-term trends can be forecast, and behavior and achievements can be predicted with fair accuracy from previous behaviors and achievements. But we should not expect performance in officer training

and in combat to be predictable from behavior on an obstacle field—behavior both on the test and in the real world is determined by many factors that are specific to the particular situation. Remove one highly assertive member from a group of eight candidates and everyone else's personalities will appear to change. Let a sniper's bullet move by a few centimeters and the performance of an officer will be transformed. I do not deny the validity of all tests—if a test predicts an important outcome with a validity of .20 or .30, the test should be used. But you should not expect more. You should expect little or nothing from Wall Street stock pickers who hope to be more accurate than the market in predicting the future of prices. And you should not expect much from pundits making long-term forecasts—although they may have valuable insights into the near future. The line that separates the possibly predictable future from the unpredictable distant future is in ■ yet to be drawn.

Speaking of Illusory Skill

“He knows that the record indicates that the development of this illness is mostly unpredictable. How can he be so confident in this case? Sounds like an illusion of validity.”

“She has a coherent story that explains all she knows, and the coherence makes her feel good.”

“What makes him believe that he is smarter than the market? Is this an illusion of skill?”

“She is a hedgehog. She has a theory that explains everything, and it gives her the illusion that she understands the world.”

“The question is not whether these experts are well trained. It is whether their world is predictable.”

Intuitions vs. Formulas

Paul Meehl was a strange and wonderful character, and one of the most versatile psychologists of the twentieth century. Among the departments in which he had faculty appointments at the University of Minnesota were psychology, law, psychiatry, neurology, and philosophy. He also wrote on religion, political science, and learning in rats. A statistically sophisticated researcher and a fierce critic of empty claims in clinical psychology, Meehl was also a practicing psychoanalyst. He wrote thoughtful essays on the philosophical foundations of psychological research that I almost memorized while I was a graduate student. I never met Meehl, but he was one of my heroes from the time I read his *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*.

In the slim volume that he later called "my disturbing little book," Meehl reviewed the results of 20 studies that had analyzed whether *clinical predictions* based on the subjective impressions of trained professionals were more accurate than *statistical predictions* made by combining a few scores or ratings according to a rule. In a typical study, trained counselors predicted the grades of freshmen at the end of the school year. The counselors interviewed each student for forty-five minutes. They also had access to high school grades, several aptitude tests, and a four-page personal statement. The statistical algorithm used only a fraction of this information: high school grades and one aptitude test. Nevertheless, the formula was more accurate than 11 of the 14 counselors. Meehl reported generally similar results across a variety of other forecast outcomes, including violations of parole, success in pilot training, and criminal recidivism.

Not surprisingly, Meehl's book provoked shock and disbelief among clinical psychologists, and the controversy it started has engendered a stream of research that is still flowing today, more than fifty years after its publication. The number of studies reporting comparisons of clinical and statistical predictions has increased to roughly two hundred, but the score in the contest between algorithms and humans has not changed. About 60% of the studies have shown significantly better accuracy for the algorithms. The other comparisons scored a draw in accuracy, but a tie is tantamount to a win for the statistical rules, which are normally much less expensive to use than expert judgment. No exception has been convincingly documented.

The range of predicted outcomes has expanded to cover medical variables such as the longevity of cancer patients, the length of hospital stays, the diagnosis of cardiac disease, and the susceptibility of babies to

sudden infant death syndrome; economic measures such as the prospects of success for new businesses, the evaluation of credit risks by banks, and the future career satisfaction of workers; questions of interest to government agencies, including assessments of the suitability of foster parents, the odds of recidivism among juvenile offenders, and the likelihood of other forms of violent behavior; and miscellaneous outcomes such as the evaluation of scientific presentations, the winners of football games, and the future prices of Bordeaux wine. Each of these domains entails a significant degree of uncertainty and unpredictability. We describe them as “low-validity environments.” In every case, the accuracy of experts was matched or exceeded by a simple algorithm.

As Meehl pointed out with justified pride thirty years after the publication of his book, “There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one.”

The Princeton economist and wine lover Orley Ashenfelter has offered a compelling demonstration of the power of simple statistics to outdo world-renowned experts. Ashenfelter wanted to predict the future value of fine Bordeaux wines from information available in the year they are made. The question is important because fine wines take years to reach their peak quality, and the prices of mature wines from the same vineyard vary dramatically across different vintages; bottles filled only twelve months apart can differ in value by a factor of 10 or more. An ability to forecast future prices is of substantial value, because investors buy wine, like art, in the anticipation that its value will appreciate.

It is generally agreed that the effect of vintage can be due only to variations in the weather during the grape-growing season. The best wines are produced when the summer is warm and dry, which makes the Bordeaux wine industry a likely beneficiary of global warming. The industry is also helped by wet springs, which increase quantity without much effect on quality. Ashenfelter converted that conventional knowledge into a statistical formula that predicts the price of a wine—for a particular property and at a particular age—by three features of the weather: the average temperature over the summer growing season, the amount of rain at harvest-time, and the total rainfall during the previous winter. His formula provides accurate price forecasts years and even decades into the future. Indeed, his formula forecasts future prices much more accurately than the current prices of young wines do. This new example of a “Meehl pattern” challenges the abilities of the experts whose opinions help shape the early price. It also challenges economic theory, according to which prices should reflect all the available information, including the weather. Ashenfelter’s formula is extremely accurate—the correlation between his predictions and

actual prices is above .90.

Why are experts inferior to algorithms? One reason, which Meehl suspected, is that experts try to be clever, think outside the box, and consider complex combinations of features in making their predictions. Complexity may work in the odd case, but more often than not it reduces validity. Simple combinations of features are better. Several studies have shown that human decision makers are inferior to a prediction formula even when they are given the score suggested by the formula! They feel that they can overrule the formula because they have additional information about the case, but they are wrong more often than not. According to Meehl, there are few circumstances under which it is a good idea to substitute judgment for a formula. In a famous thought experiment, he described a formula that predicts whether a particular person will go to the movies tonight and noted that it is proper to disregard the formula if information is received that the individual broke a leg today. The name "broken-leg rule" has stuck. The point, of course, is that broken legs are very rare—as well as decisive.

Another reason for the inferiority of expert judgment is that humans are incorrigibly inconsistent in making summary judgments of complex information. When asked to evaluate the same information twice, they frequently give different answers. The extent of the inconsistency is often a matter of real concern. Experienced radiologists who evaluate chest X-rays as "normal" or "abnormal" contradict themselves 20% of the time when they see the same picture on separate occasions. A study of 101 independent auditors who were asked to evaluate the reliability of internal corporate audits revealed a similar degree of inconsistency. A review of 41 separate studies of the reliability of judgments made by auditors, pathologists, psychologists, organizational managers, and other professionals suggests that this level of inconsistency is typical, even when a case is reevaluated within a few minutes. Unreliable judgments cannot be valid predictors of anything.

The widespread inconsistency is probably due to the extreme context dependency of System 1. We know from studies of priming that unnoticed stimuli in our environment have a substantial influence on our thoughts and actions. These influences fluctuate from moment to moment. The brief pleasure of a cool breeze on a hot day may make you slightly more positive and optimistic about whatever you are evaluating at the time. The prospects of a convict being granted parole may change significantly during the time that elapses between successive food breaks in the parole judges' schedule. Because you have little direct knowledge of what goes on in your mind, you will never know that you might have made a different

judgment or reached a different decision under very slightly different circumstances. Formulas do not suffer from such problems. Given the same input, they always return the same answer. When predictability is poor—which it is in most of the studies reviewed by Meehl and his followers—inconsistency is destructive of any predictive validity.

The research suggests a surprising conclusion: to maximize predictive accuracy, final decisions should be left to formulas, especially in low-validity environments. In admission decisions for medical schools, for example, the final determination is often made by the faculty members who interview the candidate. The evidence is fragmentary, but there are solid grounds for a conjecture: conducting an interview is likely to diminish the accuracy of a selection procedure, if the interviewers also make the final admission decisions. Because interviewers are overconfident in their intuitions, they will assign too much weight to their personal impressions and too little weight to other sources of information, lowering validity. Similarly, the experts who evaluate the quality of immature wine to predict its future have a source of information that almost certainly makes things worse rather than better: they can taste the wine. In addition, of course, even if they have a good understanding of the effects of the weather on wine quality, they will not be able to maintain the consistency of a formula.

The most important development in the field since Meehl's original work is Robyn Dawes's famous article "The Robust Beauty of Improper Linear Models in Decision Making." The dominant statistical practice in the social sciences is to assign weights to the different predictors by following an algorithm, called multiple regression, that is now built into conventional software. The logic of multiple regression is unassailable: it finds the optimal formula for putting together a weighted combination of the predictors. However, Dawes observed that the complex statistical algorithm adds little or no value. One can do just as well by selecting a set of scores that have some validity for predicting the outcome and adjusting the values to make them comparable (by using standard scores or ranks). A formula that combines these predictors with equal weights is likely to be just as accurate in predicting new cases as the multiple-regression formula that was optimal in the original sample. More recent research went further: formulas that assign equal weights to all the predictors are often superior, because they are not affected by accidents of sampling.

The surprising success of equal-weighting schemes has an important practical implication: it is possible to develop useful algorithms without any prior statistical research. Simple equally weighted formulas based on

existing statistics or on common sense are often very good predictors of significant outcomes. In a memorable example, Dawes showed that marital stability is well predicted by a formula:

frequency of lovemaking minus frequency of quarrels

You don't want your result to be a negative number.

The important conclusion from this research is that an algorithm that is constructed on the back of an envelope is often good enough to compete with an optimally weighted formula, and certainly good enough to outdo expert judgment. This logic can be applied in many domains, ranging from the selection of stocks by portfolio managers to the choices of medical treatments by doctors or patients.

A classic application of this approach is a simple algorithm that has saved the lives of hundreds of thousands of infants. Obstetricians had always known that an infant who is not breathing normally within a few minutes of birth is at high risk of brain damage or death. Until the anesthesiologist Virginia Apgar intervened in 1953, physicians and midwives used their clinical judgment to determine whether a baby was in distress. Different practitioners focused on different cues. Some watched for breathing problems while others monitored how soon the baby cried. Without a standardized procedure, danger signs were often missed, and many newborn infants died.

One day over breakfast, a medical resident asked how Dr. Apgar would make a systematic assessment of a newborn. "That's easy," she replied. "You would do it like this." Apgar jotted down five variables (heart rate, respiration, reflex, muscle tone, and color) and three scores (0, 1, or 2, depending on the robustness of each sign). Realizing that she might have made a breakthrough that any delivery room could implement, Apgar began rating infants by this rule one minute after they were born. A baby with a total score of 8 or above was likely to be pink, squirming, crying, grimacing, with a pulse of 100 or more—in good shape. A baby with a score of 4 or below was probably bluish, flaccid, passive, with a slow or weak pulse—in need of immediate intervention. Applying Apgar's score, the staff in delivery rooms finally had consistent standards for determining which babies were in trouble, and the formula is credited for an important contribution to reducing infant mortality. The Apgar test is still used every day in every delivery room. Atul Gawande's recent *A Checklist Manifesto* provides many other examples of the virtues of checklists and simple rules.

The Hostility to Algorithms

From the very outset, clinical psychologists responded to Meehl's ideas with hostility and disbelief. Clearly, they were in the grip of an illusion of skill in terms of their ability to make long-term predictions. On reflection, it is easy to see how the illusion came about and easy to sympathize with the clinicians' rejection of Meehl's research.

The statistical evidence of clinical inferiority contradicts clinicians' everyday experience of the quality of their judgments. Psychologists who work with patients have many hunches during each therapy session, anticipating how the patient will respond to an intervention, guessing what will happen next. Many of these hunches are confirmed, illustrating the reality of clinical skill.

The problem is that the correct judgments involve short-term predictions in the context of the therapeutic interview, a skill in which therapists may have years of practice. The tasks at which they fail typically require long-term predictions about the patient's future. These are much more difficult, even the best formulas do only modestly well, and they are also tasks that the clinicians have never had the opportunity to learn properly—they would have to wait years for feedback, instead of receiving the instantaneous feedback of the clinical session. However, the line between what clinicians can do well and what they cannot do at all well is not obvious, and certainly not obvious to them. They know they are skilled, but they don't necessarily know the boundaries of their skill. Not surprisingly, then, the idea that a mechanical combination of a few variables could outperform the subtle complexity of human judgment strikes experienced clinicians as obviously wrong.

The debate about the virtues of clinical and statistical prediction has always had a moral dimension. The statistical method, Meehl wrote, was criticized by experienced clinicians as "mechanical, atomistic, additive, cut and dried, artificial, unreal, arbitrary, incomplete, dead, pedantic, fractionated, trivial, forced, static, superficial, rigid, sterile, academic, pseudoscientific and blind." The clinical method, on the other hand, was lauded by its proponents as "dynamic, global, meaningful, holistic, subtle, sympathetic, configural, patterned, organized, rich, deep, genuine, sensitive, sophisticated, real, living, concrete, natural, true to life, and understanding."

This is an attitude we can all recognize. When a human competes with a machine, whether it is John Henry a-hammerin' on the mountain or the chess genius Garry Kasparov facing off against the computer Deep Blue, our sympathies lie with our fellow human. The aversion to algorithms

making decisions that affect humans is rooted in a strong preference that many people have for the ornatural over the synthetic or artificial. Asked whether they would rather eat an organic or a commercially grown apple, most people prefer the “all natural” one. Even after being informed that the two apples taste the same, have identical nutritional value, and are equally healthful, a majority still prefer the organic fruit. Even the producers of beer have found that they can increase sales by putting “All Natural” or “No Preservatives” on the label.

The deep resistance to the demystification of expertise is illustrated by the reaction of the European wine community to Ashenfelter’s formula for predicting the price of Bordeaux wines. Ashenfelter’s formula answered a prayer: one might thus have expected that wine lovers everywhere would be grateful to him for demonstrably improving their ability to identify the wines that later would be good. Not so. The response in French wine circles, wrote *The New York Times*, ranged “somewhere between violent and hysterical.” Ashenfelter reports that one oenophile called his findings “ludicrous and absurd.” Another scoffed, “It is like judging movies without actually seeing them.”

The prejudice against algorithms is magnified when the decisions are consequential. Meehl remarked, “I do not quite know how to alleviate the horror some clinicians seem to experience when they envisage a treatable case being denied treatment because a ‘blind, mechanical’ equation misclassifies him.” In contrast, Meehl and other proponents of algorithms have argued strongly that it is unethical to rely on intuitive judgments for important decisions if an algorithm is available that will make fewer mistakes. Their rational argument is compelling, but it runs against a stubborn psychological reality: for most people, the cause of a mistake matters. The story of a child dying because an algorithm made a mistake is more poignant than the story of the same tragedy occurring as a result of human error, and the difference in emotional intensity is readily translated into a moral preference.

Fortunately, the hostility to algorithms will probably soften as their role in everyday life continues to expand. Looking for books or music we might enjoy, we appreciate recommendations generated by soft ware. We take it for granted that decisions about credit limits are made without the direct intervention of any human judgment. We are increasingly exposed to guidelines that have the form of simple algorithms, such as the ratio of good and bad cholesterol levels we should strive to attain. The public is now well aware that formulas may do better than humans in some critical decisions in the world of sports: how much a professional team should pay for particular rookie players, or when to punt on fourth down. The expanding list of tasks that are assigned to algorithms should eventually

reduce the discomfort that most people feel when they first encounter the pattern of results that Meehl described in his disturbing little book.

Learning from Meehl

In 1955, as a twenty-one-year-old lieutenant in the Israeli Defense Forces, I was assigned to set up an interview system for the entire army. If you wonder why such a responsibility would be forced upon someone so young, bear in mind that the state of Israel itself was only seven years old at the time; all its institutions were under construction, and someone had to build them. Odd as it sounds today, my bachelor's degree in psychology probably qualified me as the best-trained psychologist in the army. My direct supervisor, a brilliant researcher, had a degree in chemistry.

An interview routine was already in place when I was given my mission. Every soldier drafted into the army completed a battery of psychometric tests, and each man considered for combat duty was interviewed for an assessment of personality. The goal was to assign the recruit a score of general fitness for combat and to find the best match of his personality among various branches: infantry, artillery, armor, and so on. The interviewers were themselves young draftees, selected for this assignment by virtue of their high intelligence and interest in dealing with people. Most were women, who were at the time exempt from combat duty. Trained for a few weeks in how to conduct a fifteen- to twenty-minute interview, they were encouraged to cover a range of topics and to form a general impression of how well the recruit would do in the army.

Unfortunately, follow-up evaluations had already indicated that this interview procedure was almost useless for predicting the future success of recruits. I was instructed to design an interview that would be more useful but would not take more time. I was also told to try out the new interview and to evaluate its accuracy. From the perspective of a serious professional, I was no more qualified for the task than I was to build a bridge across the Amazon.

Fortunately, I had read Paul Meehl's "little book," which had appeared just a year earlier. I was convinced by his argument that simple, statistical rules are superior to intuitive "clinical" judgments. I concluded that the then current interview had failed at least in part because it allowed the interviewers to do what they found most interesting, which was to learn about the dynamics of the interviewee's mental life. Instead, we should use the limited time at our disposal to obtain as much specific information as possible about the interviewee's life in his normal environment. Another lesson I learned from Meehl was that we should abandon the procedure in

which the interviewers' global evaluations of the recruit determined the final decision. Meehl's book suggested that such evaluations should not be trusted and that statistical summaries of separately evaluated attributes would achieve higher validity.

I decided on a procedure in which the interviewers would evaluate several relevant personality traits and score each separately. The final score of fitness for combat duty would be computed according to a standard formula, with no further input from the interviewers. I made up a list of six characteristics that appeared relevant to performance in a combat unit, including "responsibility," "sociability," and "masculine pride." I then composed, for each trait, a series of factual questions about the individual's life before his enlistment, including the number of different jobs he had held, how regular and punctual he had been in his work or studies, the frequency of his interactions with friends, and his interest and participation in sports, among others. The idea was to evaluate as objectively as possible how well the recruit had done on each dimension.

By focusing on standardized, factual questions, I hoped to combat the halo effect, where favorable first impressions influence later judgments. As a further precaution against halos, I instructed the interviewers to go through the six traits in a fixed sequence, rating each trait on a five-point scale before going on to the next. And that was that. I informed the interviewers that they need not concern themselves with the recruit's future adjustment to the military. Their only task was to elicit relevant facts about his past and to use that information to score each personality dimension. "Your function is to provide reliable measurements," I told them. "Leave the predicok tive validity to me," by which I meant the formula that I was going to devise to combine their specific ratings.

The interviewers came close to mutiny. These bright young people were displeased to be ordered, by someone hardly older than themselves, to switch off their intuition and focus entirely on boring factual questions. One of them complained, "You are turning us into robots!" So I compromised. "Carry out the interview exactly as instructed," I told them, "and when you are done, have your wish: close your eyes, try to imagine the recruit as a soldier, and assign him a score on a scale of 1 to 5."

Several hundred interviews were conducted by this new method, and a few months later we collected evaluations of the soldiers' performance from the commanding officers of the units to which they had been assigned. The results made us happy. As Meehl's book had suggested, the new interview procedure was a substantial improvement over the old one. The sum of our six ratings predicted soldiers' performance much more accurately than the global evaluations of the previous interviewing method, although far from perfectly. We had progressed from "completely

useless” to “moderately useful.”

The big surprise to me was that the intuitive judgment that the interviewers summoned up in the “close your eyes” exercise also did very well, indeed just as well as the sum of the six specific ratings. I learned from this finding a lesson that I have never forgotten: intuition adds value even in the justly derided selection interview, but only after a disciplined collection of objective information and disciplined scoring of separate traits. I set a formula that gave the “close your eyes” evaluation the same weight as the sum of the six trait ratings. A more general lesson that I learned from this episode was do not simply trust intuitive judgment—your own or that of others—but do not dismiss it, either.

Some forty-five years later, after I won a Nobel Prize in economics, I was for a short time a minor celebrity in Israel. On one of my visits, someone had the idea of escorting me around my old army base, which still housed the unit that interviews new recruits. I was introduced to the commanding officer of the Psychological Unit, and she described their current interviewing practices, which had not changed much from the system I had designed; there was, it turned out, a considerable amount of research indicating that the interviews still worked well. As she came to the end of her description of how the interviews are conducted, the officer added, “And then we tell them, ‘Close your eyes.’”

Do It Yourself

The message of this chapter is readily applicable to tasks other than making manpower decisions for an army. Implementing interview procedures in the spirit of Meehl and Dawes requires relatively little effort but substantial discipline. Suppose that you need to hire a sales representative for your firm. If you are serious about hiring the best possible person for the job, this is what you should do. First, select a few traits that are prerequisites for success in this position (technical proficiency, engaging personality, reliability, and so on). Don’t overdo it—six dimensions is a good number. The traits you choose should be as independent as possible from each other, and you should feel that you can assess them reliably by asking a few factual questions. Next, make a list of those questions for each trait and think about how you will score it, say on a 1–5 scale. You should have an idea of what you will call a “very weak” or “very strong.”

These preparations should take you half an hour or so, a small investment that can make a significant difference in the quality of the people you hire. To avoid halo effects, you must collect the information on

one trait at a time, scoring each before you move on to the next one. Do not skip around. To evaluate each candidate, add up the six scores. Because you are in charge of the final decision, you should not do a “close your eyes.” Firmly resolve that you will hire the candidate whose final score is the highest, even if there is another one whom you like better—try to resist your wish to invent broken legs to change the ranking. A vast amount of research offers a promise: you are much more likely to find the best candidate if you use this procedure than if you do what people normally do in such situations, which is to go into the interview unprepared and to make choices by an overall intuitive judgment such as “I looked into his eyes and liked what I saw.”

Speaking of Judges vs. Formulas

“Whenever we can replace human judgment by a formula, we should at least consider it.”

“He thinks his judgments are complex and subtle, but a simple combination of scores could probably do better.”

“Let’s decide in advance what weight to give to the data we have on the candidates’ past performance. Otherwise we will give too much weight to our impression from the interviews.”

Expert Intuition: When Can We Trust It?

Professional controversies bring out the worst in academics. Scientific journals occasionally publish exchanges, often beginning with someone's critique of another's research, followed by a reply and a rejoinder. I have always thought that these exchanges are a waste of time. Especially when the original critique is sharply worded, the reply and the rejoinder are often exercises in what I have called sarcasm for beginners and advanced sarcasm. The replies rarely concede anything to a biting critique, and it is almost unheard of for a rejoinder to admit that the original critique was misguided or erroneous in any way. On a few occasions I have responded to criticisms that I thought were grossly misleading, because a failure to respond can be interpreted as conceding error, but I have never found the hostile exchanges instructive. In search of another way to deal with disagreements, I have engaged in a few "adversarial collaborations," in which scholars who disagree on the science agree to write a jointly authored paper on their differences, and sometimes conduct research together. In especially tense situations, the research is moderated by an arbiter.

My most satisfying and productive adversarial collaboration was with Gary Klein, the intellectual leader of an association of scholars and practitioners who do not like the kind of work I do. They call themselves students of Naturalistic Decision Making, or NDM, and mostly work in organizations where they often study how experts work. The NDMers adamantly reject the focus on biases in the heuristics and biases approach. They criticize this model as overly concerned with failures and driven by artificial experiments rather than by the study of real people doing things that matter. They are deeply skeptical about the value of using rigid algorithms to replace human judgment, and Paul Meehl is not among their heroes. Gary Klein has eloquently articulated this position over many years.

This is hardly the basis for a beautiful friendship, but there is more to the story. I had never believed that intuition is always misguided. I had also been a fan of Klein's studies of expertise in firefighters since I first saw a draft of a paper he wrote in the 1970s, and was impressed by his book *Sources of Power*, much of which analyzes how experienced professionals develop intuitive skills. I invited him to join in an effort to map the boundary that separates the marvels of intuition from its flaws. He was intrigued by the idea and we went ahead with the project—with no certainty that it would succeed. We set out to answer a specific question: When can you trust an experienced professional who claims to have an intuition? It was obvious

that Klein would be more disposed to be trusting, and I would be more skeptical. But could we agree on principles for answering the general question?

Over seven or eight years we had many discussions, resolved many disagreements, almost blew up more than once, wrote many drafts, became friends, and eventually published a joint article with a title that tells the story: "Conditions for Intuitive Expertise: A Failure to Disagree." Indeed, we did not encounter real issues on which we disagreed—but we did not really agree.

Marvels and Flaws

Malcolm Gladwell's bestseller *Blink* appeared while Klein and I were working on the project, and it was reassuring to find ourselves in agreement about it. Gladwell's book opens with the memorable story of art experts faced with an object that is described as a magnificent example of a kouros, a sculpture of a striding boy. Several of the experts had strong visceral reactions: they felt in their gut that the statue was a fake but were not able to articulate what it was about it that made them uneasy. Everyone who read the book—millions did—remembers that story as a triumph of intuition. The experts agreed that they knew the sculpture was a fake without knowing how they knew—the very definition of intuition. The story appears to imply that a systematic search for the cue that guided the experts would have failed, but Klein and I both rejected that conclusion. From our point of view, such an inquiry was needed, and if it had been conducted properly (which Klein knows how to do), it would probably have succeeded.

Although many readers of the kouros example were surely drawn to an almost magical view of expert intuition, Gladwell himself does not hold that position. In a later chapter he describes a massive failure of intuition: Americans elected President Harding, whose only qualification for the position was that he perfectly looked the part. Square jawed and tall, he was the perfect image of a strong and decisive leader. People voted for someone who looked strong and decisive without any other reason to believe that he was. An intuitive prediction of how Harding would perform as president arose from substituting one question for another. A reader of this book should expect such an intuition to be held with confidence.

Intuition as Recognition

The early experiences that shaped Klein's views of intuition were starkly

different from mine. My thinking was formed by observing the illusion of validity in myself and by reading Paul Meehl's demonstrations of the inferiority of clinical prediction. In contrast, Klein's views were shaped by his early studies of fireground commanders (the leaders of firefighting teams). He followed them as they fought fires and later interviewed the leader about his thoughts as he made decisions. As Klein described it in our joint article, he and his collaborators

investigated how the commanders could make good decisions without comparing options. The initial hypothesis was that commanders would restrict their analysis to only a pair of options, but that hypothesis proved to be incorrect. In fact, the commanders usually generated only a single option, and that was all they needed. They could draw on the repertoire of patterns that they had compiled during more than a decade of both real and virtual experience to identify a plausible option, which they considered first. They evaluated this option by mentally simulating it to see if it would work in the situation they were facing.... If the course of action they were considering seemed appropriate, they would implement it. If it had shortcomings, they would modify it. If they could not easily modify it, they would turn to the next most plausible option and run through the same procedure until an acceptable course of action was found.

Klein elaborated this description into a theory of decision making that he called the recognition-primed decision (RPD) model, which applies to firefighters but also describes expertise in other domains, including chess. The process involves both System 1 and System 2. In the first phase, a tentative plan comes to mind by an automatic function of associative memory—System 1. The next phase is a deliberate process in which the plan is mentally simulated to check if it will work—an operation of System 2. The model of intuitive decision making as pattern recognition develops ideas presented some time ago by Herbert Simon, perhaps the only scholar who is recognized and admired as a hero and founding figure by all the competing clans and tribes in the study of decision making. I quoted Herbert Simon's definition of intuition in the introduction, but it will make more sense when I repeat it now: "The situation has provided a cue; this cue has given the expert access to information stored in memory, and the information provides the answer. Intuition is nothing more and nothing less than recognition."

This strong statement reduces the apparent magic of intuition to the everyday experience of memory. We marvel at the story of the firefighter

who has a sudden urge to escape a burning house just before it collapses, because the firefighter knows the danger intuitively, “without knowing how he knows.” However, we also do not know how we immediately know that a person we see as we enter a room is our friend Peter. The moral of Simon’s remark is that the mystery of knowing without knowing is not a distinctive feature of intuition; it is the norm of mental life.

Acquiring Skill

How does the information that supports intuition get “stored in memory”? Certain types of intuitions are acquired very quickly. We have inherited from our ancestors a great facility to learn when to be afraid. Indeed, one experience is often sufficient to establish a long-term aversion and fear. Many of us have the visceral memory of a single dubious dish that still leaves us vaguely reluctant to return to a restaurant. All of us tense up when we approach a spot in which an unpleasant event occurred, even when there is no reason to expect it to happen again. For me, one such place is the ramp leading to the San Francisco airport, where years ago a driver in the throes of road rage followed me from the freeway, rolled down his window, and hurled obscenities at me. I never knew what caused his hatred, but I remember his voice whenever I reach that point on my way to the airport.

My memory of the airport incident is conscious and it fully explains the emotion that comes with it. On many occasions, however, you may feel uneasy in a particular place or when someone uses a particular turn of phrase without having a conscious memory of the triggering event. In hindsight, you will label that unease an intuition if it is followed by a bad experience. This mode of emotional learning is closely related to what happened in Pavlov’s famous conditioning experiments, in which the dogs learned to recognize the sound of the bell as a signal that food was coming. What Pavlov’s dogs learned can be described as a learned hope. Learned fears are even more easily acquired.

Fear can also be learned—quite easily, in fact—by words rather than by experience. The fireman who had the “sixth sense” of danger had certainly had many occasions to discuss and think about types of fires he was not involved in, and to rehearse in his mind what the cues might be and how he should react. As I remember from experience, a young platoon commander with no experience of combat will tense up while leading troops through a narrowing ravine, because he was taught to identify the terrain as favoring an ambush. Little repetition is needed for learning.

Emotional learning may be quick, but what we consider as “expertise”

usually takes a long time to develop. The acquisition of expertise in complex tasks such as high-level chess, professional basketball, or firefighting is intricate and slow because expertise in a domain is not a single skill but rather a large collection of miniskills. Chess is a good example. An expert player can understand a complex position at a glance, but it takes years to develop that level of ability. Studies of chess masters have shown that at least 10,000 hours of dedicated practice (about 6 years of playing chess 5 hours a day) are required to attain the highest levels of performance. During those hours of intense concentration, a serious chess player becomes familiar with thousands of configurations, each consisting of an arrangement of related pieces that can threaten or defend each other.

Learning high-level chess can be compared to learning to read. A first grader works hard at recognizing individual letters and assembling them into syllables and words, but a good adult reader perceives entire clauses. An expert reader has also acquired the ability to assemble familiar elements in a new pattern and can quickly “recognize” and correctly pronounce a word that she has never seen before. In chess, recurrent patterns of interacting pieces play the role of letters, and a chess position is a long word or a sentence.

A skilled reader who sees it for the first time will be able to read the opening stanza of Lewis Carroll’s “Jabberwocky” with perfect rhythm and intonation, as well as pleasure:

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mome raths outgrabe.

Acquiring expertise in chess is harder and slower than learning to read because there are many more letters in the “alphabet” of chess and because the “words” consist of many letters. After thousands of hours of practice, however, chess masters are able to read a chess situation at a glance. The few moves that come to their mind are almost always strong and sometimes creative. They can deal with a “word” they have never encountered, and they can find a new way to interpret a familiar one.

The Environment of Skill

Klein and I quickly found that we agreed both on the nature of intuitive skill and on how it is acquired. We still needed to agree on our key question: When can you trust a self-confident professional who claims to have an intuition?

We eventually concluded that our disagreement was due in part to the fact that we had different experts in mind. Klein had spent much time with fireground commanders, clinical nurses, and other professionals who have real expertise. I had spent more time thinking about clinicians, stock pickers, and political scientists trying to make unsupportable long-term forecasts. Not surprisingly, his default attitude was trust and respect; mine was skepticism. He was more willing to trust experts who claim an intuition because, as he told me, true experts know the limits of their knowledge. I argued that there are many pseudo-experts who have no idea that they do not know what they are doing (the illusion of validity), and that as a general proposition subjective confidence is commonly too high and often uninformative.

Earlier I traced people's confidence in a belief to two related impressions: cognitive ease and coherence. We are confident when the story we tell ourselves comes easily to mind, with no contradiction and no competing scenario. But ease and coherence do not guarantee that a belief held with confidence is true. The associative machine is set to suppress doubt and to evoke ideas and information that are compatible with the currently dominant story. A mind that follows WY SIATI will achieve high confidence much too easily by ignoring what it does not know. It is therefore not surprising that many of us are prone to have high confidence in unfounded intuitions. Klein and I eventually agreed on an important principle: the confidence that people have in their intuitions is not a reliable guide to their validity. In other words, do not trust anyone—including yourself—to tell you how much you should trust their judgment.

If subjective confidence is not to be trusted, how can we evaluate the probable validity of an intuitive judgment? When do judgments reflect true expertise? When do they display an illusion of validity? The answer comes from the two basic conditions for acquiring a skill:

- an environment that is sufficiently regular to be predictable
- an opportunity to learn these regularities through prolonged practice

When both these conditions are satisfied, intuitions are likely to be skilled. Chess is an extreme example of a regular environment, but bridge and

poker also provide robust statistical regularities that can support skill. Physicians, nurses, athletes, and firefighters also face complex but fundamentally orderly situations. The accurate intuitions that Gary Klein has described are due to highly valid cues that even the expert's System 1 has learned to use, even if System 2 has not learned to name them. In contrast, stock pickers and political scientists who make long-term forecasts operate in a zero-validity environment. Their failures reflect the basic unpredictability of the events that they try to forecast.

Some environments are worse than irregular. Robin Hogarth described "wicked" environments, in which professionals are likely to learn the wrong lessons from experience. He borrows from Lewis Thomas the example of a physician in the early twentieth century who often had intuitions about patients who were about to develop typhoid. Unfortunately, he tested his hunch by palpating the patient's tongue, without washing his hands between patients. When patient after patient became ill, the physician developed a sense of clinical infallibility. His predictions were accurate—but not because he was exercising professional intuition!

Meehl's clinicians were not inept and their failure was not due to lack of talent. They performed poorly because they were assigned tasks that did not have a simple solution. The clinicians' predicament was less extreme than the zero-validity environment of long-term political forecasting, but they operated in low-validity situations that did not allow high accuracy. We know this to be the case because the best statistical algorithms, although more accurate than human judges, were never very accurate. Indeed, the studies by Meehl and his followers never produced a "smoking gun" demonstration, a case in which clinicians completely missed a highly valid cue that the algorithm detected. An extreme failure of this kind is unlikely because human learning is normally efficient. If a strong predictive cue exists, human observers will find it, given a decent opportunity to do so. Statistical algorithms greatly outdo humans in noisy environments for two reasons: they are more likely than human judges to detect weakly valid cues and much more likely to maintain a modest level of accuracy by using such cues consistently.

It is wrong to blame anyone for failing to forecast accurately in an unpredictable world. However, it seems fair to blame professionals for believing they can succeed in an impossible task. Claims for correct intuitions in an unpredictable situation are self-delusional at best, sometimes worse. In the absence of valid cues, intuitive "hits" are due either to luck or to lies. If you find this conclusion surprising, you still have a lingering belief that intuition is magic. Remember this rule: intuition cannot

be trusted in the absence of stable regularities in the environment.

Feedback and Practice

Some regularities in the environment are easier to discover and apply than others. Think of how you developed your style of using the brakes on your car. As you were mastering the skill of taking curves, you gradually learned when to let go of the accelerator and when and how hard to use the brakes. Curves differ, and the variability you experienced while learning ensures that you are now ready to brake at the right time and strength for any curve you encounter. The conditions for learning this skill are ideal, because you receive immediate and unambiguous feedback every time you go around a bend: the mild reward of a comfortable turn or the mild punishment of some difficulty in handling the car if you brake either too hard or not quite hard enough. The situations that face a harbor pilot maneuvering large ships are no less regular, but skill is much more difficult to acquire by sheer experience because of the long delay between actions and their noticeable outcomes. Whether professionals have a chance to develop intuitive expertise depends essentially on the quality and speed of feedback, as well as on sufficient opportunity to practice.

Expertise is not a single skill; it is a collection of skills, and the same professional may be highly expert in some of the tasks in her domain while remaining a novice in others. By the time chess players become experts, they have “seen everything” (or almost everything), but chess is an exception in this regard. Surgeons can be much more proficient in some operations than in others. Furthermore, some aspects of any professional’s tasks are much easier to learn than others. Psychotherapists have many opportunities to observe the immediate reactions of patients to what they say. The feedback enables them to develop the intuitive skill to find the words and the tone that will calm anger, forge confidence, or focus the patient’s attention. On the other hand, therapists do not have a chance to identify which general treatment approach is most suitable for different patients. The feedback they receive from their patients’ long-term outcomes is sparse, delayed, or (usually) nonexistent, and in any case too ambiguous to support learning from experience.

Among medical specialties, anesthesiologists benefit from good feedback, because the effects of their actions are likely to be quickly evident. In contrast, radiologists obtain little information about the accuracy of the diagnoses they make and about the pathologies they fail to detect. Anesthesiologists are therefore in a better position to develop useful

intuitive skills. If anesthesiologist says, "I have a feeling something is wrong," everyone in the operating room should be prepared for an emergency.

Here again, as in the case of subjective confidence, the experts may not know the limits of their expertise. An experienced psychotherapist knows that she is skilled in working out what is going on in her patient's mind and that she has good intuitions about what the patient will say next. It is tempting for her to conclude that she can also anticipate how well the patient will do next year, but this conclusion is not equally justified. Short-term anticipation and long-term forecasting are different tasks, and the therapist has had adequate opportunity to learn one but not the other. Similarly, a financial expert may have skills in many aspects of his trade but not in picking stocks, and an expert in the Middle East knows many things but not the future. The clinical psychologist, the stock picker, and the pundit do have intuitive skills in some of their tasks, but they have not learned to identify the situations and the tasks in which intuition will betray them. The unrecognized limits of professional skill help explain why experts are often overconfident.

Evaluating Validity

At the end of our journey, Gary Klein and I agreed on a general answer to our initial question: When can you trust an experienced professional who claims to have an intuition? Our conclusion was that for the most part it is possible to distinguish intuitions that are likely to be valid from those that are likely to be bogus. As in the judgment of whether a work of art is genuine or a fake, you will usually do better by focusing on its provenance than by looking at the piece itself. If the environment is sufficiently regular and if the judge has had a chance to learn its regularities, the associative machinery will recognize situations and generate quick and accurate predictions and decisions. You can trust someone's intuitions if these conditions are met.

Unfortunately, associativent memory also generates subjectively compelling intuitions that are false. Anyone who has watched the chess progress of a talented youngster knows well that skill does not become perfect all at once, and that on the way to near perfection some mistakes are made with great confidence. When evaluating expert intuition you should always consider whether there was an adequate opportunity to learn the cues, even in a regular environment.

In a less regular, or low-validity, environment, the heuristics of judgment are invoked. System 1 is often able to produce quick answers to difficult

questions by substitution, creating coherence where there is none. The question that is answered is not the one that was intended, but the answer is produced quickly and may be sufficiently plausible to pass the lax and lenient review of System 2. You may want to forecast the commercial future of a company, for example, and believe that this is what you are judging, while in fact your evaluation is dominated by your impressions of the energy and competence of its current executives. Because substitution occurs automatically, you often do not know the origin of a judgment that you (your System 2) endorse and adopt. If it is the only one that comes to mind, it may be subjectively undistinguishable from valid judgments that you make with expert confidence. This is why subjective confidence is not a good diagnostic of accuracy: judgments that answer the wrong question can also be made with high confidence.

You may be asking, Why didn't Gary Klein and I come up immediately with the idea of evaluating an expert's intuition by assessing the regularity of the environment and the expert's learning history—mostly setting aside the expert's confidence? And what did we think the answer could be? These are good questions because the contours of the solution were apparent from the beginning. We knew at the outset that fireground commanders and pediatric nurses would end up on one side of the boundary of valid intuitions and that the specialties studied by Meehl would be on the other, along with stock pickers and pundits.

It is difficult to reconstruct what it was that took us years, long hours of discussion, endless exchanges of drafts and hundreds of e-mails negotiating over words, and more than once almost giving up. But this is what always happens when a project ends reasonably well: once you understand the main conclusion, it seems it was always obvious.

As the title of our article suggests, Klein and I disagreed less than we had expected and accepted joint solutions of almost all the substantive issues that were raised. However, we also found that our early differences were more than an intellectual disagreement. We had different attitudes, emotions, and tastes, and those changed remarkably little over the years. This is most obvious in the facts that we find amusing and interesting. Klein still winces when the word *bias* is mentioned, and he still enjoys stories in which algorithms or formal procedures lead to obviously absurd decisions. I tend to view the occasional failures of algorithms as opportunities to improve them. On the other hand, I find more pleasure than Klein does in the come-uppance of arrogant experts who claim intuitive powers in zero-validity situations. In the long run, however, finding as much intellectual agreement as we did is surely more important than the persistent emotional differences that remained.

Speaking of Expert Intuition

“How much expertise does she have in this particular task? How much practice has she had?”

“Does he really believe that the environment of start-ups is sufficiently regular to justify an intuition that goes against the base rates?”

“She is very confident in her decision, but subjective confidence is a poor index of the accuracy of a judgment.”

“Did he really have an opportunity to learn? How quick and how clear was the feedback he received on his judgments?”

The Outside View

A few years after my collaboration with Amos began, I convinced some officials in the Israeli Ministry of Education of the need for a curriculum to teach judgment and decision making in high schools. The team that I assembled to design the curriculum and write a textbook for it included several experienced teachers, some of my psychology students, and Seymour Fox, then dean of the Hebrew University's School of Education, who was an expert in curriculum development.

After meeting every Friday afternoon for about a year, we had constructed a detailed outline of the syllabus, had written a couple of chapters, and had run a few sample lessons in the classroom. We all felt that we had made good progress. One day, as we were discussing procedures for estimating uncertain quantities, the idea of conducting an exercise occurred to me. I asked everyone to write down an estimate of how long it would take us to submit a finished draft of the textbook to the Ministry of Education. I was following a procedure that we already planned to incorporate into our curriculum: the proper way to elicit information from a group is not by starting with a public discussion but by confidentially collecting each person's judgment. This procedure makes better use of the knowledge available to members of the group than the common practice of open discussion. I collected the estimates and jotted the results on the blackboard. They were narrowly centered around two years; the low end was one and a half, the high end two and a half years.

Then I had another idea. I turned to Seymour, our curriculum expert, and asked whether he could think of other teams similar to ours that had developed a curriculum from scratch. This was a time when several pedagogical innovations like "new math" had been introduced, and Seymour said he could think of quite a few. I then asked whether he knew the history of these teams in some detail, and it turned out that he was familiar with several. I asked him to think of these teams when they had made as much progress as we had. How long, from that point, did it take them to finish their textbook projects?

He fell silent. When he finally spoke, it seemed to me that he was blushing, embarrassed by his own answer: "You know, I never realized this before, but in fact not all the teams at a stage comparable to ours ever did complete their task. A substantial fraction of the teams ended up failing to finish the job."

This was worrisome; we had never considered the possibility that we might fail. My anxiety rising, I asked how large he estimated that fraction was. "About 40%," he answered. By now, a pall of gloom

was falling over the room. The next question was obvious: "Those who finished," I asked. "How long did it take them?" "I cannot think of any group that finished in less than seven years," he replied, "nor any that took more than ten."

I grasped at a straw: "When you compare our skills and resources to those of the other groups, how good are we? How would you rank us in comparison with these teams?" Seymour did not hesitate long this time. "We're below average," he said, "but not by much." This came as a complete surprise to all of us—including Seymour, whose prior estimate had been well within the optimistic consensus of the group. Until I prompted him, there was no connection in his mind between his knowledge of the history of other teams and his forecast of our future.

Our state of mind when we heard Seymour is not well described by stating what we "knew." Surely all of us "knew" that a minimum of seven years and a 40% chance of failure was a more plausible forecast of the fate of our project than the numbers we had written on our slips of paper a few minutes earlier. But we did not acknowledge what we knew. The new forecast still seemed unreal, because we could not imagine how it could take so long to finish a project that looked so manageable. No crystal ball was available to tell us the strange sequence of unlikely events that were in our future. All we could see was a reasonable plan that should produce a book in about two years, conflicting with statistics indicating that other teams had failed or had taken an absurdly long time to complete their mission. What we had heard was base-rate information, from which we should have inferred a causal story: if so many teams failed, and if those that succeeded took so long, writing a curriculum was surely much harder than we had thought. But such an inference would have conflicted with our direct experience of the good progress we had been making. The statistics that Seymour provided were treated as base rates normally are—noted and promptly set aside.

We should have quit that day. None of us was willing to invest six more years of work in a project with a 40% chance of failure. Although we must have sensed that persevering was not reasonable, the warning did not provide an immediately compelling reason to quit. After a few minutes of desultory debate, we gathered ourselves together and carried on as if nothing had happened. The book was eventually completed eight(!) years later. By that time I was no longer living in Israel and had long since ceased to be part of the team, which completed the task after many unpredictable vicissitudes. The initial enthusiasm for the idea in the Ministry of Education had waned by the time the text was delivered and it was never used.

This embarrassing episode remains one of the most instructive experiences of my professional life. I eventually learned three lessons from

it. The first was immediately apparent: I had stumbled onto a distinction between two profoundly different approaches to forecasting, which Amos and I later labeled the inside view and the outside view. The second lesson was that our initial forecasts of about two years for the completion of the project exhibited a planning fallacy. Our estimates were closer to a best-case scenario than to a realistic assessment. I was slower to accept the third lesson, which I call irrational perseverance: the folly we displayed that day in failing to abandon the project. Facing a choice, we gave up rationality rather than give up the enterprise.

Drawn to the Inside View

On that long-ago Friday, our curriculum expert made two judgments about the same problem and arrived at very different answers. The *inside view* is the one that all of us, including Seymour, spontaneously adopted to assess the future of our project. We focused on our specific circumstances and searched for evidence in our own experiences. We had a sketchy plan: we knew how many chapters we were going to write, and we had an idea of how long it had taken us to write the two that we had already done. The more cautious among us probably added a few months to their estimate as a margin of error.

Extrapolating was a mistake. We were forecasting based on the information in front of us—WYSIATI—but the chapters we wrote first were probably easier than others, and our commitment to the project was probably then at its peak. But the main problem was that we failed to allow for what Donald Rumsfeld famously called the “unknown unknowns.” There was no way for us to foresee, that day, the succession of events that would cause the project to drag out for so long. The divorces, the illnesses, the crises of coordination with bureaucracies that delayed the work could not be anticipated. Such events not only cause the writing of chapters to slow down, they also produce long periods during which little or no progress is made at all. The same must have been true, of course, for the other teams that Seymour knew about. The members of those teams were also unable to imagine the events that would cause them to spend seven years to finish, or ultimately fail to finish, a project that they evidently had thought was very feasible. Like us, they did not know the odds they were facing. There are many ways for any plan to fail, and although most of them are too improbable to be anticipated, the likelihood that *something* will go wrong in a big project is high.

The second question I asked Seymour directed his attention away from us and toward a class of similar cases. Seymour estimated the base rate of success in that reference class: 40% failure and seven to ten years for

completion. His informal survey was surely not up to scientific standards of evidence, but it provided a reasonable basis for a baseline prediction: the prediction you make about a case if you know nothing except the category to which it belongs. As we saw earlier, the *baseline prediction* should be the anchor for further adjustments. If you are asked to guess the height of a woman about whom you know only that she lives in New York City, your baseline prediction is your best guess of the average height of women in the city. If you are now given case-specific information, for example that the woman's son is the starting center of his high school basketball team, you will adjust your estimate away from the mean in the appropriate direction. Seymour's comparison of our team to others suggested that the forecast of our outcome was slightly worse than the baseline prediction, which was already grim.

The spectacular accuracy of the outside-view forecast in our problem was surely a fluke and should not count as evidence for the validity of the *outside view*. The argument for the outside view should be made on general grounds: if the reference class is properly chosen, the outside view will give an indication of where the ballpark is, and it may suggest, as it did in our case, that the inside-view forecasts are not even close to it.

For a psychologist, the discrepancy between Seymour's two judgments is striking. He had in his head all the knowledge required to estimate the statistics of an appropriate reference class, but he reached his initial estimate without ever using that knowledge. Seymour's forecast from his inside view was not an adjustment from the baseline prediction, which had not come to his mind. It was based on the particular circumstances of our efforts. Like the participants in the Tom W experiment, Seymour knew the relevant base rate but did not think of applying it.

Unlike Seymour, the rest of us did not have access to the outside view and could not have produced a reasonable baseline prediction. It is noteworthy, however, that we did not feel we needed information about other teams to make our guesses. My request for the outside view surprised all of us, including me! This is a common pattern: people who have information about an individual case rarely feel the need to know the statistics of the class to which the case belongs.

When we were eventually exposed to the outside view, we collectively ignored it. We can recognize what happened to us; it is similar to the experiment that suggested the futility of teaching psychology. When they made predictions about individual cases about which they had a little information (a brief and bland interview), Nisbett and Borgida's students completely neglected the global results they had just learned. "Pallid" statistical information is routinely discarded when it is incompatible with

one's personal impressions of a case. In the competition with the inside view, the outside view doesn't stand a chance.

The preference for the inside view sometimes carries moral overtones. I once asked my cousin, a distinguished lawyer, a question about a reference class: "What is the probability of the defendant winning in cases like this one?" His sharp answer that "every case is unique" was accompanied by a look that made it clear he found my question inappropriate and superficial. A proud emphasis on the uniqueness of cases is also common in medicine, in spite of recent advances in evidence-based medicine that point the other way. Medical statistics and baseline predictions come up with increasing frequency in conversations between patients and physicians. However, the remaining ambivalence about the outside view in the medical profession is expressed in concerns about the impersonality of procedures that are guided by statistics and checklists.

The Planning Fallacy

In light of both the outside-view forecast and the eventual outcome, the original estimates we made that Friday afternoon appear almost delusional. This should not come as a surprise: overly optimistic forecasts of the outcome of projects are found everywhere. Amos and I coined the term *planning fallacy* to describe plans and forecasts that

- are unrealistically close to best-case scenarios
- could be improved by consulting the statistics of similar cases

Examples of the planning fallacy abound in the experiences of individuals, governments, and businesses. The list of horror stories is endless.

- In July 1997, the proposed new Scottish Parliament building in Edinburgh was estimated to cost up to £40 million. By June 1999, the budget for the building was £109 million. In April 2000, legislators imposed a £195 million "cap on costs." By November 2001, they demanded an estimate of "final cost," which was set at £241 million. That estimated final cost rose twice in 2002, ending the year at

£294.6 million. It rose three times more in 2003, reaching £375.8 million by June. The building was finally completed in 2004 at an ultimate cost of roughly £431 million.

- A 2005 study examined rail projects undertaken worldwide between 1969 and 1998. In more than 90% of the cases, the number of passengers projected to use the system was overestimated. Even though these passenger shortfalls were widely publicized, forecasts did not improve over those thirty years; on average, planners overestimated how many people would use the new rail projects by 106%, and the average cost overrun was 45%. As more evidence accumulated, the experts did not become more reliant on it.
- In 2002, a survey of American homeowners who had remodeled their kitchens found that, on average, they had expected the job to cost \$18,658; in fact, they ended up paying an average of \$38,769.

The optimism of planners and decision makers is not the only cause of overruns. Contractors of kitchen renovations and of weapon systems readily admit (though not to their clients) that they routinely make most of their profit on additions to the original plan. The failures of forecasting in these cases reflect the customers' inability to imagine how much their wishes will escalate over time. They end up paying much more than they would if they had made a realistic plan and stuck to it.

Errors in the initial budget are not always innocent. The authors of unrealistic plans are often driven by the desire to get the plan approved—whether by their superiors or by a client—supported by the knowledge that projects are rarely abandoned unfinished merely because of overruns in costs or completion times. In such cases, the greatest responsibility for avoiding the planning fallacy lies with the decision makers who approve the plan. If they do not recognize the need for an outside view, they commit a planning fallacy.

Mitigating the Planning Fallacy

The diagnosis of and the remedy for the planning fallacy have not changed since that Friday afternoon, but the implementation of the idea has come a long way. The renowned Danish planning expert Bent Flyvbjerg, now at Oxford University, offered a forceful summary:

The prevalent tendency to underweight or ignore distributional information is perhaps the major source of error in forecasting. Planners should therefore make every effort to frame the

forecasting problem so as to facilitate utilizing all the distributional information that is available.

This may be considered the single most important piece of advice regarding how to increase accuracy in forecasting through improved methods. Using such distributional information from other ventures similar to that being forecasted is called taking an “outside view” and is the cure to the planning fallacy.

The treatment for the planning fallacy has now acquired a technical name, *reference class forecasting*, and Flyvbjerg has applied it to transportation projects in several countries. The outside view is implemented by using a large database, which provides information on both plans and outcomes for hundreds of projects all over the world, and can be used to provide statistical information about the likely overruns of cost and time, and about the likely underperformance of projects of different types.

The forecasting method that Flyvbjerg applies is similar to the practices recommended for overcoming base-rate neglect:

1. Identify an appropriate reference class (kitchen renovations, large railway projects, etc.).
2. Obtain the statistics of the reference class (in terms of cost per mile of railway, or of the percentage by which expenditures exceeded budget). Use the statistics to generate a baseline prediction.
3. Use specific information about the case to adjust the baseline prediction, if there are particular reasons to expect the optimistic bias to be more or less pronounced in this project than in others of the same type.

Flyvbjerg’s analyses are intended to guide the authorities that commission public projects, by providing the statistics of overruns in similar projects. Decision makers need a realistic assessment of the costs and benefits of a proposal before making the final decision to approve it. They may also wish to estimate the budget reserve that they need in anticipation of overruns, although such precautions often become self-fulfilling prophecies. As one official told Flyvbjerg, “A budget reserve is to contractors as red meat is to lions, and they will devour it.”

Organizations face the challenge of controlling the tendency of executives competing for resources to present overly optimistic plans. A well-run organization will reward planners for precise execution and

penalize them for failing to anticipate difficulties, and for failing to allow for difficulties that they could not have anticipated—the unknown unknowns.

Decisions and Errors

That Friday afternoon occurred more than thirty years ago. I often thought about it and mentioned it in lectures several times each year. Some of my friends got bored with the story, but I kept drawing new lessons from it. Almost fifteen years after I first reported on the planning fallacy with Amos, I returned to the topic with Dan Lovallo. Together we sketched a theory of decision making in which the optimistic bias is a significant source of risk taking. In the standard rational model of economics, people take risks because the odds are favorable—they accept some probability of a costly failure because the probability of success is sufficient. We proposed an alternative idea.

When forecasting the outcomes of risky projects, executives too easily fall victim to the planning fallacy. In its grip, they make decisions based on delusional optimism rather than on a rational weighting of gains, losses, and probabilities. They overestimate benefits and underestimate costs. They spin scenarios of success while overlooking the potential for mistakes and miscalculations. As a result, they pursue initiatives that are unlikely to come in on budget or on time or to deliver the expected returns—or even to be completed.

In this view, people often (but not always) take on risky projects because they are overly optimistic about the odds they face. I will return to this idea several times in this book—it probably contributes to an explanation of why people litigate, why they start wars, and why they open small businesses.

Failing a Test

For many years, I thought that the main point of the curriculum story was what I had learned about my friend Seymour: that his best guess about the future of our project was not informed by what he knew about similar projects. I came off quite well in my telling of the story, in which I had the role of clever questioner and astute psychologist. I only recently realized that I had actually played the roles of chief dunce and inept leader.

The project was my initiative, and it was therefore my responsibility to ensure that it made sense and that major problems were properly discussed by the team, but I failed that test. My problem was no longer the planning fallacy. I was cured of that fallacy as soon as I heard Seymour's statistical summary. If pressed, I would have said that our earlier estimates

had been absurdly optimistic. If pressed further, I would have admitted that we had started the project on faulty premises and that we should at least consider seriously the option of declaring defeat and going home. But nobody pressed me and there was no discussion; we tacitly agreed to go on without an explicit forecast of how long the effort would last. This was easy to do because we had not made such a forecast to begin with. If we had had a reasonable baseline prediction when we started, we would not have gone into it, but we had already invested a great deal of effort—an instance of the sunk-cost fallacy, which we will look at more closely in the next part of the book. It would have been embarrassing for us—especially for me—to give up at that point, and there seemed to be no immediate reason to do so. It is easier to change directions in a crisis, but this was not a crisis, only some new facts about people we did not know. The outside view was much easier to ignore than bad news in our own effort. I can best describe our state as a form of lethargy—an unwillingness to think about what had happened. So we carried on. There was no further attempt at rational planning for the rest of the time I spent as a member of the team—a particularly troubling omission for a team dedicated to teaching rationality. I hope I am wiser today, and I have acquired a habit of looking for the outside view. But it will never be the natural thing to do.

Speaking of the Outside View

“He’s taking an inside view. He should forget about his own case and look for what happened in other cases.”

“She is the victim of a planning fallacy. She’s assuming a best-case scenario, but there are too many different ways for the plan to fail, and she cannot foresee them all.”

“Suppose you did not know a thing about this particular legal case, only that it involves a malpractice claim by an individual against a surgeon. What would be your baseline prediction? How many of these cases succeed in court? How many settle? What are the amounts? Is the case we are discussing stronger or weaker than similar claims?”

“We are making an additional investment because we do not

want to admit failure. This is an instance of the sunk-cost fallacy.”

The Engine of Capitalism

The planning fallacy is only one of the manifestations of a pervasive optimistic bias. Most of us view the world as more benign than it really is, our own attributes as more favorable than they truly are, and the goals we adopt as more achievable than they are likely to be. We also tend to exaggerate our ability to forecast the future, which fosters optimistic overconfidence. In terms of its consequences for decisions, the optimistic bias may well be the most significant of the cognitive biases. Because optimistic bias can be both a blessing and a risk, you should be both happy and wary if you are temperamentally optimistic.

Optimists

Optimism is normal, but some fortunate people are more optimistic than the rest of us. If you are genetically endowed with an optimistic bias, you hardly need to be told that you are a lucky person—you already feel fortunate. An optimistic attitude is largely inherited, and it is part of a general disposition for well-being, which may also include a preference for seeing the bright side of everything. If you were allowed one wish for your child, seriously consider wishing him or her optimism. Optimists are normally cheerful and happy, and therefore popular; they are resilient in adapting to failures and hardships, their chances of clinical depression are reduced, their immune system is stronger, they take better care of their health, they feel healthier than others and are in fact likely to live longer. A study of people who exaggerate their expected life span beyond actuarial predictions showed that they work longer hours, are more optimistic about their future income, are more likely to remarry after divorce (the classic “triumph of hope over experience”), and are more prone to bet on individual stocks. Of course, the blessings of optimism are offered only to individuals who are only mildly biased and who are able to “accentuate the positive” without losing track of reality.

Optimistic individuals play a disproportionate role in shaping our lives. Their decisions make a difference; they are the inventors, the entrepreneurs, the political and military leaders—not average people. They got to where they are by seeking challenges and taking risks. They are talented and they have been lucky, almost certainly luckier than they acknowledge. They are probably optimistic by temperament; a survey of founders of small businesses concluded that entrepreneurs are more sanguine than midlevel managers about life in general. Their experiences of success have confirmed their faith in their judgment and in their ability to

control events. Their self-confidence is reinforced by the admiration of others. This reasoning leads to a hypothesis: the people who have the greatest influence on the lives of others are likely to be optimistic and overconfident, and to take more risks than they realize.

The evidence suggests that an optimistic bias plays a role—sometimes the dominant role—whenever individuals or institutions voluntarily take on significant risks. More often than not, risk takers underestimate the odds they face, and do not invest sufficient effort to find out what the odds are. Because they misread the risks, optimistic entrepreneurs often believe they are prudent, even when they are not. Their confidence in their future success sustains a positive mood that helps them obtain resources from others, raise the morale of their employees, and enhance their prospects of prevailing. When action is needed, optimism, even of the mildly delusional variety, may be a good thing.

Entrepreneurial Delusions

The chances that a small business will survive for five years in the United States are about 35%. But the individuals who open such businesses do not believe that the statistics apply to them. A survey found that American entrepreneurs tend to believe they are in a promising line of business: their average estimate of the chances of success for “any business like yours” was 60%—almost double the true value. The bias was more glaring when people assessed the odds of their own venture. Fully 81% of the entrepreneurs put their personal odds of success at 7 out of 10 or higher, and 33% said their chance of failing was zero.

The direction of the bias is not surprising. If you interviewed someone who recently opened an Italian restaurant, you would not expect her to have underestimated her prospects for success or to have a poor view of her ability as a restaurateur. But you must wonder: Would she still have invested money and time if she had made a reasonable effort to learn the odds—or, if she did learn the odds (60% of new restaurants are out of business after three years), paid attention to them? The idea of adopting the outside view probably didn't occur to her.

One of the benefits of an optimistic temperament is that it encourages persistence in the face of obstacles. But persistence can be costly. An impressive series of studies by Thomas Åstebro sheds light on what happens when optimists receive bad news. He drew his data from a Canadian organization—the Inventor's Assistance Program—which

collects a small fee to provide inventors with an objective assessment of the commercial prospects of their idea. The evaluations rely on careful ratings of each invention on 37 criteria, including need for the product, cost of production, and estimated trend of demand. The analysts summarize their ratings by a letter grade, where D and E predict failure—a prediction made for over 70% of the inventions they review. The forecasts of failure are remarkably accurate: only 5 of 411 projects that were given the lowest grade reached commercialization, and none was successful.

Discouraging news led about half of the inventors to quit after receiving a grade that unequivocally predicted failure. However, 47% of them continued development efforts even after being told that their project was hopeless, and on average these persistent (or obstinate) individuals doubled their initial losses before giving up. Significantly, persistence after discouraging advice was relatively common among inventors who had a high score on a personality measure of optimism—on which inventors generally scored higher than the general population. Overall, the return on private invention was small, “lower than the return on private equity and on high-risk securities.” More generally, the financial benefits of self-employment are mediocre: given the same qualifications, people achieve higher average returns by selling their skills to employers than by setting out on their own. The evidence suggests that optimism is widespread, stubborn, and costly.

Psychologists have confirmed that most people genuinely believe that they are superior to most others on most desirable traits—they are willing to bet small amounts of money on these beliefs in the laboratory. In the market, of course, beliefs in one’s superiority have significant consequences. Leaders of large businesses sometimes make huge bets in expensive mergers and acquisitions, acting on the mistaken belief that they can manage the assets of another company better than its current owners do. The stock market commonly responds by downgrading the value of the acquiring firm, because experience has shown that efforts to integrate large firms fail more often than they succeed. The misguided acquisitions have been explained by a “hubris hypothesis”: the executives of the acquiring firm are simply less competent than they think they are.

The economists Ulrike Malmendier and Geoffrey Tate identified optimistic CEOs by the amount of company stock that they owned personally and observed that highly optimistic leaders took excessive risks. They assumed debt rather than issue equity and were more likely than others to “overpay for target companies and undertake value-destroying mergers.” Remarkably, the stock of the acquiring company suffered substantially more in mergers if the CEO was overly optimistic by

the authors' measure. The stock market is apparently able to identify overconfident CEOs. This observation exonerates the CEOs from one accusation even as it convicts them of another: the leaders of enterprises who make unsound bets do not do so because they are betting with other people's money. On the contrary, they take greater risks when they personally have more at stake. The damage caused by overconfident CEOs is compounded when the business press anoints them as celebrities; the evidence indicates that prestigious press awards to the CEO are costly to stockholders. The authors write, "We find that firms with award-winning CEOs subsequently underperform, in terms both of stock and of operating performance. At the same time, CEO compensation increases, CEOs spend more time on activities outside the company such as writing books and sitting on outside boards, and they are more likely to engage in earnings management."

Many years ago, my wife and I were on vacation on Vancouver Island, looking for a place to stay. We found an attractive but deserted motel on a little-traveled road in the middle of a forest. The owners were a charming young couple who needed little prompting to tell us their story. They had been schoolteachers in the province of Alberta; they had decided to change their life and used their life savings to buy this motel, which had been built a dozen years earlier. They told us without irony or self-consciousness that they had been able to buy it cheap, "because six or seven previous owners had failed to make a go of it." They also told us about plans to seek a loan to make the establishment more attractive by building a restaurant next to it. They felt no need to explain why they expected to succeed where six or seven others had failed. A common thread of boldness and optimism links businesspeople, from motel owners to superstar CEOs.

The optimistic risk taking of entrepreneurs surely contributes to the economic dynamism of a capitalistic society, even if most risk takers end up disappointed. However, Marta Coelho of the London School of Economics has pointed out the difficult policy issues that arise when founders of small businesses ask the government to support them in decisions that are most likely to end badly. Should the government provide loans to would-be entrepreneurs who probably will bankrupt themselves in a few years? Many behavioral economists are comfortable with the "libertarian paternalistic" procedures that help people increase their savings rate beyond what they would do on their own. The question of whether and how government should support small business does not have

an equally satisfying answer.

Competition Neglect

It is tempting to explain entrepreneurial optimism by wishful thinking, but emotion is only part of the story. Cognitive biases play an important role, notably the System 1 feature WYSIATI.

- We focus on our goal, anchor on our plan, and neglect relevant base rates, exposing ourselves to these planning fallacy.
- We focus on what we want to do and can do, neglecting the plans and skills of others.
- Both in explaining the past and in predicting the future, we focus on the causal role of skill and neglect the role of luck. We are therefore prone to an *illusion of control*.
- We focus on what we know and neglect what we do not know, which makes us overly confident in our beliefs.

The observation that “90% of drivers believe they are better than average” is a well-established psychological finding that has become part of the culture, and it often comes up as a prime example of a more general above-average effect. However, the interpretation of the finding has changed in recent years, from self-aggrandizement to a cognitive bias. Consider these two questions:

Are you a good driver?

Are you better than average as a driver?

The first question is easy and the answer comes quickly: most drivers say yes. The second question is much harder and for most respondents almost impossible to answer seriously and correctly, because it requires an assessment of the average quality of drivers. At this point in the book it comes as no surprise that people respond to a difficult question by answering an easier one. They compare themselves to the average without ever thinking about the average. The evidence for the cognitive interpretation of the above-average effect is that when people are asked about a task they find difficult (for many of us this could be “Are you better than average in starting conversations with strangers?”), they readily rate themselves as below average. The upshot is that people tend to be overly

optimistic about their relative standing on any activity in which they do moderately well.

I have had several occasions to ask founders and participants in innovative start-ups a question: To what extent will the outcome of your effort depend on what you do in your firm? This is evidently an easy question; the answer comes quickly and in my small sample it has never been less than 80%. Even when they are not sure they will succeed, these bold people think their fate is almost entirely in their own hands. They are surely wrong: the outcome of a start-up depends as much on the achievements of its competitors and on changes in the market as on its own efforts. However, WY SIATI plays its part, and entrepreneurs naturally focus on what they know best—their plans and actions and the most immediate threats and opportunities, such as the availability of funding. They know less about their competitors and therefore find it natural to imagine a future in which the competition plays little part.

Colin Camerer and Dan Lovallo, who coined the concept of competition neglect, illustrated it with a quote from the then chairman of Disney Studios. Asked why so many expensive big-budget movies are released on the same days (such as Memorial Day and Independence Day), he replied:

Hubris. Hubris. If you only think about your own business, you think, "I've got a good story department, I've got a good marketing department, we're going to go out and do this." And you don't think that everybody else is thinking the same way. In a given weekend in a year you'll have five movies open, and there's certainly not enough people to go around. re

The candid answer refers to hubris, but it displays no arrogance, no conceit of superiority to competing studios. The competition is simply not part of the decision, in which a difficult question has again been replaced by an easier one. The question that needs an answer is this: Considering what others will do, how many people will see our film? The question the studio executives considered is simpler and refers to knowledge that is most easily available to them: Do we have a good film and a good organization to market it? The familiar System 1 processes of WY SIATI and substitution produce both competition neglect and the above-average effect. The consequence of competition neglect is excess entry: more competitors enter the market than the market can profitably sustain, so their average outcome is a loss. The outcome is disappointing for the typical entrant in the market, but the effect on the economy as a whole could well be positive. In fact, Giovanni Dosi and Dan Lovallo call

entrepreneurial firms that fail to signal new markets to more qualified competitors “optimistic martyrs”—good for the economy but bad for their investors.

Overconfidence

For a number of years, professors at Duke University conducted a survey in which the chief financial officers of large corporations estimated the returns of the Standard & Poor’s index over the following year. The Duke scholars collected 11,600 such forecasts and examined their accuracy. The conclusion was straightforward: financial officers of large corporations had no clue about the short-term future of the stock market; the correlation between their estimates and the true value was slightly less than zero! When they said the market would go down, it was slightly more likely than not that it would go up. These findings are not surprising. The truly bad news is that the CFOs did not appear to know that their forecasts were worthless.

In addition to their best guess about S&P returns, the participants provided two other estimates: a value that they were 90% sure would be too high, and one that they were 90% sure would be too low. The range between the two values is called an “80% confidence interval” and outcomes that fall outside the interval are labeled “surprises.” An individual who sets confidence intervals on multiple occasions expects about 20% of the outcomes to be surprises. As frequently happens in such exercises, there were far too many surprises; their incidence was 67%, more than 3 times higher than expected. This shows that CFOs were grossly overconfident about their ability to forecast the market. *Overconfidence* is another manifestation of WYSIATI: when we estimate a quantity, we rely on information that comes to mind and construct a coherent story in which the estimate makes sense. Allowing for the information that does not come to mind—perhaps because one never knew it—is impossible.

The authors calculated the confidence intervals that would have reduced the incidence of surprises to 20%. The results were striking. To maintain the rate of surprises at the desired level, the CFOs should have said, year after year, “There is an 80% chance that the S&P return next year will be between -10% and +30%.” The confidence interval that properly reflects the CFOs’ knowledge (more precisely, their ignorance) is more than 4 times wider than the intervals they actually stated.

Social psychology comes into the picture here, because the answer that a truthful CFO would offer is plainly ridiculous. A CFO who informs his colleagues that “there is a good chance that the S&P returns will be

between -10% and +30%" can expect to be laughed out of the room." The wide confidence interval is a confession of ignorance, which is not socially acceptable for someone who is paid to be knowledgeable in financial matters. Even if they knew how little they know, the executives would be penalized for admitting it. President Truman famously asked for a "one-armed economist" who would take a clear stand; he was sick and tired of economists who kept saying, "On the other hand..."

Organizations that take the word of overconfident experts can expect costly consequences. The study of CFOs showed that those who were most confident and optimistic about the S&P index were also overconfident and optimistic about the prospects of their own firm, which went on to take more risk than others. As Nassim Taleb has argued, inadequate appreciation of the uncertainty of the environment inevitably leads economic agents to take risks they should avoid. However, optimism is highly valued, socially and in the market; people and firms reward the providers of dangerously misleading information more than they reward truth tellers. One of the lessons of the financial crisis that led to the Great Recession is that there are periods in which competition, among experts and among organizations, creates powerful forces that favor a collective blindness to risk and uncertainty.

The social and economic pressures that favor overconfidence are not restricted to financial forecasting. Other professionals must deal with the fact that an expert worthy of the name is expected to display high confidence. Philip Tetlock observed that the most overconfident experts were the most likely to be invited to strut their stuff in news shows. Overconfidence also appears to be endemic in medicine. A study of patients who died in the ICU compared autopsy results with the diagnosis that physicians had provided while the patients were still alive. Physicians also reported their confidence. The result: "clinicians who were 'completely certain' of the diagnosis antemortem were wrong 40% of the time." Here again, expert overconfidence is encouraged by their clients: "Generally, it is considered a weakness and a sign of vulnerability for clinicians to appear unsure. Confidence is valued over uncertainty and there is a prevailing censure against disclosing uncertainty to patients." Experts who acknowledge the full extent of their ignorance may expect to be replaced by more confident competitors, who are better able to gain the trust of clients. An unbiased appreciation of uncertainty is a cornerstone of rationality—but it is not what people and organizations want. Extreme uncertainty is paralyzing under dangerous circumstances, and the admission that one is merely guessing is especially unacceptable when the stakes are high. Acting on pretended knowledge is often the preferred solution.

When they come together, the emotional, cognitive, and social factors that support exaggerated optimism are a heady brew, which sometimes leads people to take risks that they would avoid if they knew the odds. There is no evidence that risk takers in the economic domain have an unusual appetite for gambles on high stakes; they are merely less aware of risks than more timid people are. Dan Lovallo and I coined the phrase “bold forecasts and timid decisions” to describe the background of risk taking.

The effects of high optimism on decision making are, at best, a mixed blessing, but the contribution of optimism to good implementation is certainly positive. The main benefit of optimism is resilience in the face of setbacks. According to Martin Seligman, the founder of positive psychology, an “optimistic explanation style” contributes to resilience by defending one’s self-image. In essence, the optimistic style involves taking credit for successes but little blame for failures. This style can be taught, at least to some extent, and Seligman has documented the effects of training on various occupations that are characterized by a high rate of failures, such as cold-call sales of insurance (a common pursuit in pre-Internet days). When one has just had a door slammed in one’s face by an angry homemaker, the thought that “she was an awful woman” is clearly superior to “I am an inept salesperson.” I have always believed that scientific research is another domain where a form of optimism is essential to success: I have yet to meet a successful scientist who lacks the ability to exaggerate the importance of what he or she is doing, and I believe that someone who lacks a delusional sense of significance will wilt in the face of repeated experiences of multiple small failures and rare successes, the fate of most researchers.

The Premortem: A Partial Remedy

Can overconfident optimism be overcome by training? I am not optimistic. There have been numerous attempts to train people to state confidence intervals that reflect the imprecision of their judgments, with only a few reports of modest success. An often cited example is that geologists at Royal Dutch Shell became less overconfident in their assessments of possible drilling sites after training with multiple past cases for which the outcome was known. In other situations, overconfidence was mitigated (but not eliminated) when judges were encouraged to consider competing hypotheses. However, overconfidence is a direct consequence of features

of System 1 that can be tamed—but not vanquished. The main obstacle is that subjective confidence is determined by the coherence of the story one has constructed, not by the quality and amount of the information that supports it.

Organizations may be better able to tame optimism and individuals than individuals are. The best idea for doing so was contributed by Gary Klein, my “adversarial collaborator” who generally defends intuitive decision making against claims of bias and is typically hostile to algorithms. He labels his proposal the *premortem*. The procedure is simple: when the organization has almost come to an important decision but has not formally committed itself, Klein proposes gathering for a brief session a group of individuals who are knowledgeable about the decision. The premise of the session is a short speech: “Imagine that we are a year into the future. We implemented the plan as it now exists. The outcome was a disaster. Please take 5 to 10 minutes to write a brief history of that disaster.”

Gary Klein’s idea of the *premortem* usually evokes immediate enthusiasm. After I described it casually at a session in Davos, someone behind me muttered, “It was worth coming to Davos just for this!” (I later noticed that the speaker was the CEO of a major international corporation.) The *premortem* has two main advantages: it overcomes the groupthink that affects many teams once a decision appears to have been made, and it unleashes the imagination of knowledgeable individuals in a much-needed direction.

As a team converges on a decision—and especially when the leader tips her hand—public doubts about the wisdom of the planned move are gradually suppressed and eventually come to be treated as evidence of flawed loyalty to the team and its leaders. The suppression of doubt contributes to overconfidence in a group where only supporters of the decision have a v filepos-id="filepos726557"> nacea and does not provide complete protection against nasty surprises, but it goes some way toward reducing the damage of plans that are subject to the biases of WY SIATI and uncritical optimism.

Speaking of Optimism

“They have an illusion of control. They seriously underestimate the obstacles.”

“They seem to suffer from an acute case of competitor neglect.”

“This is a case of overconfidence. They seem to believe they know more than they actually do know.”

“We should conduct a premortem session. Someone may come up with a threat we have neglected.”