

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 itsignл 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 more download 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*: 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 scoop 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 have 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. 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."