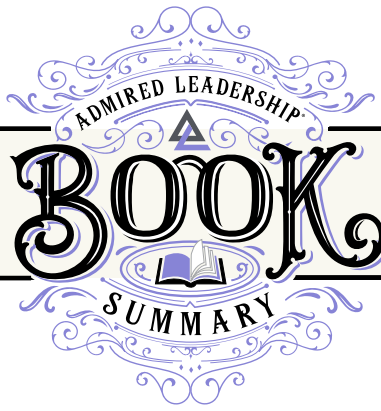




Eight Minutes, Not Eight Hours



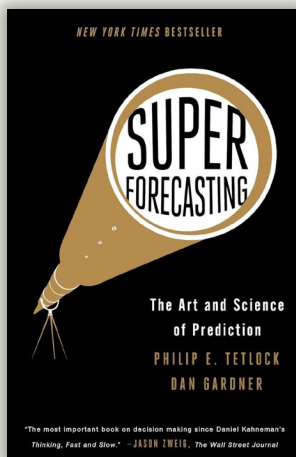
December 2024

Superforecasting

The Art and Science of Prediction

BOOK AUTHORS: PHILIP TETLOCK & DAN GARDNER

Published 2015



Wharton School psychologist Philip Tetlock famously organized a study of political prediction that showed the most highly trained political pundits and intelligence analysts were no better at predicting future events than chance, or “a dart-throwing chimp.” Following his study, he launched The Good Judgment Project, a competition among volunteers to see how well non-experts could predict political events, and whether they could improve their judgment over time by practice, or by working in groups. This book outlines Tetlock’s findings, focusing on the behaviors of the best 1% of forecasters, “superforecasters,” how the best superforecasters work in teams, and the practices that can make anyone better at prediction.

Key Quote

“There are two dangers a forecaster faces after making the initial call. One is not giving enough weight to new information. That’s under reaction. The other danger is overreacting to new information, seeing it as more meaningful than it is, and adjusting a forecast too radically” (p. 158).

KEY POINTS AND CONCEPTS

Measuring Accuracy

It is difficult to assess the accuracy of professional predictors such as political experts, intelligence analysts, and meteorologists, because predictions are seldom examined for accuracy after the fact. “The reason? It’s mostly a demand-side problem: The consumers of forecasting—governments, business, and the public—don’t demand evidence of accuracy. So there is no measurement. Which means no revision. And without revision, there can be no improvement” (p. 14).

Even in hindsight, some predictions are difficult to assess for accuracy. For example, if the weather report says there is a 90% chance of rain and it does not rain, is the prediction wrong? Not necessarily, since a 90% chance that it will rain means there is a 10% chance that it won't. For example, Steve Ballmer of Microsoft famously predicted that iPhones would never make up a significant market share. The prediction is often repeated as one of the worst business predictions of recent times. However, when you look at the context of the quotation, it becomes apparent that he was referring to the global market, and that by that metric his prediction is wrong, but much less wrong than it appears on first reading. Context and ambiguity in a prediction can greatly affect how we perceive its accuracy (pp. 46-48).


Legendary Intelligence scholar, Sherman Kent, realized that language shapes our understanding of predictions. One person's "fair chance" is not the same as someone else's "likely." **Instead, he proposed a system for the CIA to use when discussing probabilities:**

Certainty	General Area of Possibility
100%	Certain
93% (+ or - 6%)	Almost Certain
75 % (+ or -12 %)	Probable
50% (+ or -10%)	Changes About Even
30% (+ or -10%)	Probably Not
7% (+ or - 5%)	Almost Certainly Not
0%	Impossible

The CIA originally rejected his suggestion, but today the Intelligence community has adopted communicating in probabilities as standard practice (pp. 54-56).

To score his experiments, Tetlock uses the Brier score, which takes into account the accuracy of a prediction (were you correct), weighted by the certainty of a forecast (how sure you were). An omniscient score is a 0. Completely wrong is a 1. Random guessing would yield a 0.5. Tetlock's superforecasters and superforecasting teams had Brier scores in the .14 to .25 range. The Brier score makes it easier to judge the quality of predictions by isolating the element of luck to judge the quality of the prediction rather than the outcome, and compare single scores against group averages.

By judging the forecasters year-over-year, Tetlock argues that his data "suggests two key conclusions. One, we should not treat the superstars of any given year as infallible, not even [top superforecaster] Doug Lorch. Luck plays a role and it is only to be expected that the superstars will occasionally have a bad year and produce ordinary results—just as superstar athletes occasionally look less than stellar. But more basically, and more hopefully, we can conclude that the superforecasters were not just lucky. Mostly, their results reflected skill. Which raises the big question: Why are superforecasters so good?" (p. 104).



By judging the forecasters year-over-year, Tetlock argues that his data "suggests two key conclusions. One, we should not treat the superstars of any given year as infallible, not even [top superforecaster] Doug Lorch.



Prediction is a Discipline

Tetlock draws two major conclusions from his research. First, foresight is real and some people have it. “They aren’t gurus or oracles with the power to peer decades into the future, but they do have a real measurable skill at judging how high-stakes events are likely to unfold [...] The other conclusion is what makes these forecasters so good. It’s not really who they are. It’s what they do [emphasis added]. Foresight isn’t a particular gift bestowed at birth. It is a product of particular ways of thinking, of gathering information, of updating beliefs. These habits of thought can be learned and cultivated by any intelligent, thoughtful, determined person” (p. 18).

The volunteers in Tetlock’s study scored higher than 70% of the population on general intelligence tests. Superforecasters scored better than about 80%, but not many were genius-level. Tetlock concludes that intelligence helps, but only to a certain threshold, after which there is a diminishing return on higher intelligence (p. 109).

The best forecasters are committed to staying abreast of the best information and not building attachment to their level of certainty. To avoid overreacting or underreacting to new information, “the key is commitment—in this case [stock trading] the absence of commitment. Traders who constantly buy and sell are not cognitively or emotionally connected with their stocks. They expect some to fall and they sell off those losers with a shrug [...] they are no more committed to these stocks than a gin rummy player is to cards in his hand” (pp. 164-165).


Prediction is Both Science and Art

While some dynamics at work in prediction—such as understanding regression to the mean—are mathematically based, prediction also requires a healthy dose of luck. The danger is to take the wrong lessons away when a solid prediction turns out wrong—or a terrible prediction turns out correct—because of a freak occurrence that no one could or should have seen coming (p. 99).


The Best Predictions are Highly Granular

Barbara Mellers has shown that granularity predicts accuracy. Not only were the best predictions the most precise ones, but when she rounded the Brier scores to make forecasts less granular, she discovered that superforecasters lost accuracy in response to even the smallest-scale rounding, to the nearest 0.05, whereas regular forecasters lost little even from rounding four times as large, to the nearest 0.20 (pp. 145-144).

Top superforecasters revised their predictions by tiny increments whenever they learned data that changed their certainty level. However, “there are two dangers a forecaster faces after making the initial call. One is not giving enough weight to new information. That’s under reaction. The other danger is overreacting to new information, seeing it as more meaningful than it is, and adjusting a forecast too radically” (p. 158).



“While some dynamics at work in prediction—such as understanding regression to the mean—are mathematically based, prediction also requires a healthy dose of luck. The danger is to take the wrong lessons away when a solid prediction turns out wrong—or a terrible prediction turns out correct—because of a freak occurrence that no one could or should have seen coming.”



Because predictions are made up of many smaller predictions and educated guesses, the margin of uncertainty gets larger as more sub-predictions are made. Therefore, the more granular and precise individual sub-predictions are, the more accurate the final answer will be.

Physicist Enrico Fermi teaches an approach to prediction that breaks down each question into sub-questions. For example, “How many piano tuners are there in Chicago?” can be broken down into: “How many pianos are there in the city? How often are pianos tuned? How long does it take to tune a piano? How many hours a year does a piano tuner work?” Since some of those questions can be estimated accurately, it helps to anchor the guesses at a number that is closer to reality. Tetlock calls this process “Fermi-izing” (pp. 110-113).

The decision to invade Iraq in 2003 based on allegations Hussein was hiding Weapons of Mass Destruction is one of the most egregiously poor predictions in modern political history. Counter intuitively, the Intelligence Community’s investigation of the analysis did not find that the forecasts of 100% certainty were that off the mark. However, they did find that the certainty based on available evidence should have been closer to 60-70%, and that makes a huge difference.

Common Errors in Prediction

People commonly make bad predictions because they unconsciously answer an easier question than the one they are asked. Tetlock calls this the “bait and switch.” For example, instead of answering, “Will this politician do x?” they tend to search for evidence to answer the question, “Would I do x?” The oversimplification of the question leads to poor thinking and research strategy.

Poor judgment commonly occurs because predictors don’t counteract their confirmation bias by asking themselves how they might be wrong. “It is wise to take admissions of uncertainty seriously,” Daniel Kahneman noted, “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” (p. 39).

Personal experience can cause us to overemphasize small details to view problems from an “inside view” (our perspective), rather than from an “outside view” (our perspective in relation to the mathematical base rate). For example, if you are trying to determine whether a family owns a dog, small pieces of information may convince you of the inside view that they certainly don’t, while the outside view tells you that the base rate of dog ownership is 62%. Since the family is mathematically more likely than not to own a dog, the outside view suggests that it may be lower than a 62% chance, but it is certainly not zero (p. 118).

Prediction Requires a Growth Mindset

Tetlock cites the work of Carol Dweck, who argues that individuals have one of two mindsets. People with a “fixed” mindset think of themselves as a finished product with innate talents, while those with “growth” mindsets understand that their intelligence and performance is malleable. Those with growth mindsets are less likely to take negative feedback personally. Tetlock argues that the growth mindset is critical for improving forecasting accuracy.

In his original study, Tetlock found that there were two groups of forecasters: dogmatic thinkers who were committed to their conclusions and were reluctant to change their minds. “They would tell us ‘just wait.’” The other group consisted of more pragmatic experts who were more willing to change their mind based on new evidence (p. 69).

The more famous an expert, the less likely they were to have a high accuracy rate. Tetlock believes that the high level of confidence that makes someone appealing as a TV expert makes them more likely to overlook evidence that counters their view (p. 72).

Feedback is Critical

Repetition and practice of making forecasts does not necessarily lead to better accuracy without disciplined feedback. Many police officers believe they are unusually accurate at figuring out who is lying and who is telling the truth. However, the research shows that they aren't nearly as good as they think. They seldom get clear feedback if they were correct (p. 180).

The best forecasters look for evidence that their predictions could be wrong, by asking teammates for feedback, or by engaging in rigorous self-critique. Psychologist Jonathan Baron calls this characteristic “active open-mindedness.”

Superforecasters performed even better in teams, as long as they maintained their growth mindset and appealed to others for feedback. Normal predictors were prone to groupthink and performed worse in teams. “Teams of ordinary forecasters beat the wisdom of the crowd by about 10%. Prediction markets [aggregate scores of individuals] beat the ordinary teams by 20%. And super-teams beat prediction markets by 15% to 30%” (p. 207).

Profile of a Superforecaster (pp. 191, 210)

Philosophic outlook

- Cautious: Nothing is certain
- Humble: Reality is infinitely complex
- Nondeterministic: What happens is not meant to be, and does not have to happen

Abilities and thinking style

- Actively open-minded: Beliefs are hypotheses to be tested, not treasures to be protected
- Intelligent and knowledgeable with a “need for cognition”
- Reflective: Introspective and self-critical
- Numerate: Comfortable with numbers

Forecasting methodology

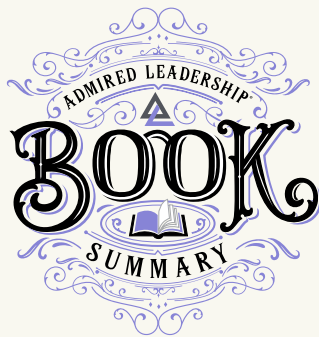
- **Pragmatic:** Not wedded to any idea or agenda
- **Analytical:** Capable of stepping back from the tip-of-your-nose perspective and considering other's views

- **Dragonfly-eyed:** Value diverse views and synthesize them into their own
- **Probabilistic:** Judge using many grades of maybe
- **Thoughtful updates:** When facts change, they change their minds

Superteams

- **Psychologically safe:** Encourage feedback
- Diverse views with a high average aptitude

Tetlock, P. & Gardner, D. (2015) **Superforecasting: The Art and Science of Prediction**. New York: Crown Publishers.



Tetlock cites the work of Carol Dweck, who argues that individuals have one of two mindsets. People with a “fixed” mindset think of themselves as a finished product with innate talents, while those with “growth” mindsets understand that their intelligence and performance is malleable.

The Latest and Greatest Books for Leaders

We work hard to stay abreast of the current writings on leadership, especially those books our clients are reading or have been recommended to read. As a benefit to our clients and to facilitate our own learning, the Admired Leadership® team has long maintained a tradition of summarizing the newest books of interest to leaders. Better to read a summary for eight minutes before investing 8 hours in the entire book. After reading a good summary, we believe leaders are able to make better choices as to what to ignore, what to peruse and what to make the time to read closely.

