

HBR.ORG

# Harvard Business Review



MAY 2016  
REPRINT R1605E

**SPOTLIGHT ON MANAGING FOR AN UNPREDICTABLE FUTURE**

# Superforecasting: How to Upgrade Your Company's Judgment

*by Paul J.H. Schoemaker and Philip E. Tetlock*

# SPOTLIGHT

**ARTWORK** Sarah Morris, *Esther (São Paulo)*  
2014, household gloss paint on canvas

---





**Paul J.H. Schoemaker** is the former research director of the Wharton School's Mack Institute and a coauthor of *Peripheral Vision* (Harvard Business Review Press, 2006).  
**Philip E. Tetlock** is the

Annenberg University Professor at the University of Pennsylvania and a coauthor of *Superforecasting* (Crown, 2015). Schoemaker served as an adviser to and Tetlock co-led the Good Judgment Project.

# Superforecasting: How to Upgrade Your Company's Judgment

BY PAUL J.H. SCHOEMAKER AND PHILIP E. TETLOCK

**I**magine that you could dramatically improve your firm's forecasting ability, but to do so you'd have to expose just how unreliable its predictions—and the people making them—really are. That's exactly what the U.S. intelligence community did, with dramatic results.

Back in October 2002, the National Intelligence Council issued its official opinion that Iraq possessed chemical and biological weapons and was actively producing more weapons of mass destruction. Of course, that judgment proved colossally wrong. Shaken by its intelligence failure, the \$50 billion bureaucracy set out to determine how it could do better in the future, realizing that the process might reveal glaring organizational deficiencies.

The resulting research program included a large-scale, multiyear prediction tournament, co-led by one of us (Phil), called the Good Judgment Project. The series of contests, which pitted thousands of amateurs against seasoned intelligence analysts, generated three surprising insights: First, talented generalists often outperform specialists in making forecasts. Second, carefully crafted training can enhance predictive acumen. And third, well-run teams can outperform individuals. These findings have important implications for the way organizations and businesses forecast uncertain outcomes, such as how a competitor will respond to a new-product launch, how much revenue a promotion will generate, or whether prospective hires will perform well.

The approach we'll describe here for building an ever-improving organizational forecasting capability is not a cookbook that offers proven recipes for success. Many of the principles are fairly new and have only recently been applied in business settings. However, our research shows that they can help leaders discover and nurture their organizations' best predictive capabilities wherever they may reside.

## Find the Sweet Spot

Companies and individuals are notoriously inept at judging the likelihood of uncertain events, as studies show all too well. Getting judgments wrong, of course, can have serious consequences. Steve Ballmer's prognostication in 2007 that "there's no chance that the iPhone is going to get any significant market share" left Microsoft with no room to consider alternative scenarios. But improving a firm's forecasting competence even a little can yield a competitive advantage. A company that is right three times out of five on its judgment calls is going to have an ever-increasing edge on a competitor that gets them right only two times out of five.

Before we discuss how an organization can build a predictive edge, let's look at the types of judgments that are most amenable to improvement—and

### ABOUT THE GOOD JUDGMENT PROJECT

In 2011, Philip Tetlock teamed up with Barbara Mellers, of the Wharton School, to launch the Good Judgment Project. The goal was to determine whether some people are naturally better than others at prediction and whether prediction performance could be enhanced. The GJP was one of five academic research teams that competed in an innovative tournament funded by the Intelligence Advanced Research Projects Activity (IARPA), in which forecasters were challenged to answer the types of geopolitical and economic questions that U.S. intelligence agencies pose to their analysts.

The IARPA initiative ran from 2011 to 2015 and recruited more than 25,000 forecasters who made well over a million predictions on topics ranging from whether Greece would exit the eurozone to the likelihood of a leadership turnover in Russia to the risk of a financial panic in China. The GJP decisively won the tournament—besting even the intelligence community's own analysts.

those not worth focusing on. We can dispense with predictions that are either entirely straightforward or seemingly impossible. Consider issues that are highly predictable: You know where the hands of your clock will be five hours from now; life insurance companies can reliably set premiums on the basis of updated mortality tables. For issues that can be predicted with great accuracy using econometric and operations-research tools, there is no advantage to be gained by developing subjective judgment skills in those areas: The data speaks loud and clear.

At the other end of the spectrum, we find issues that are complex, poorly understood, and tough to quantify, such as the patterns of clouds on a given day or when the next game-changing technology will pop out of a garage in Silicon Valley. Here, too, there's little advantage in investing resources in systematically improving judgment: The problems are just too hard to crack.

The sweet spot that companies should focus on is forecasts for which some data, logic, and analysis can be used but seasoned judgment and careful questioning also play key roles. Predicting the commercial potential of drugs in clinical trials requires scientific expertise as well as business judgment. Assessors of acquisition candidates draw on formal scoring models, but they must also gauge intangibles such as cultural fit, the chemistry among leaders, and the likelihood that anticipated synergies will actually materialize.

Consider the experience of a UK bank that lost a great deal of money in the early 1990s by lending to U.S. cable companies that were hot but then tanked. The chief lending officer conducted an audit of these presumed lending errors, analyzing the types of loans made, the characteristics of clients and loan officers involved, the incentives at play, and other factors. She scored the bad loans on each factor and then ran an analysis to see which ones best explained the variance in the amounts lost. In cases where the losses were substantial, she found problems in the underwriting process that resulted in loans to clients with poor financial health or no prior relationship with the bank—issues for which expertise and judgment were important. The bank was able to make targeted improvements that boosted performance and minimized losses.

On the basis of our research and consulting experience, we have identified a set of practices that leaders can apply to improve their firms' judgment

**Idea in Brief****THE PROBLEM**

Organizations and individuals are notoriously poor at judging the likelihood of uncertain events. Predictions are often colored by the forecaster's susceptibility to cognitive biases, desire to influence others, and concerns about reputation. Getting judgments wrong can of course have serious consequences.

**THE RESEARCH**

On the basis of research involving 25,000 forecasters and a million predictions, the authors identified a set of practices that can improve companies' prediction capability: training in the basics of statistics and biases; debating forecasts in teams; and tracking performance and giving rapid feedback.

**IN PRACTICE**

To improve prediction capability, companies should keep real-time accounts of how their top teams make judgments, including underlying assumptions, data sources, external events, and so on. Keys to success include requiring frequent, precise predictions and measuring accuracy for comparison.

in this middle ground. Our recommendations focus on improving individuals' forecasting ability through training; using teams to boost accuracy; and tracking prediction performance and providing rapid feedback. The general approaches we describe should of course be tailored to each organization and evolve as the firm learns what works in which circumstances.

**Train for Good Judgment**

Most predictions made in companies, whether they concern project budgets, sales forecasts, or the performance of potential hires or acquisitions, are not the result of cold calculus. They are colored by the forecaster's understanding of basic statistical arguments, susceptibility to cognitive biases, desire to influence others' thinking, and concerns about reputation. Indeed, predictions are often intentionally vague to maximize wiggle room should they prove wrong. The good news is that training in reasoning and debiasing can reliably strengthen a firm's forecasting competence. The Good Judgment Project demonstrated that as little as one hour of training improved forecasting accuracy by about 14% over the course of a year. (See the exhibit "How Training and Teams Improve Prediction.")

**Learn the basics.** Basic reasoning errors (such as believing that a coin that has landed heads three times in a row is likelier to land tails on the next flip) take a toll on prediction accuracy. So it's essential that companies lay a foundation of forecasting basics: The GJP's training in probability concepts such as regression to the mean and Bayesian revision (updating a probability estimate in light of new data), for example, boosted participants' accuracy measurably. Companies should also require that forecasts include a precise definition of what is to be predicted (say, the chance that a potential hire will meet her sales targets) and the time frame involved (one year, for

example). The prediction itself must be expressed as a numeric probability so that it can be precisely scored for accuracy later. That means asserting that one is "80% confident," rather than "fairly sure," that the prospective employee will meet her targets.

**Understand cognitive biases.** Cognitive biases are widely known to skew judgment, and some have particularly pernicious effects on forecasting. They lead people to follow the crowd, to look for information that confirms their views, and to strive to prove just how right they are. It's a tall order to debias human judgment, but the GJP has had some success in raising participants' awareness of key biases that compromise forecasting. For example, the project trained beginners to watch out for confirmation bias that can create false confidence, and to give due weight to evidence that challenges their conclusions. And it reminded trainees to not look at problems in isolation but, rather, take what Nobel laureate Daniel Kahneman calls "the outside view." For instance, in predicting how long a project will take to complete, trainees were counseled to first ask how long it typically takes to complete similar projects, to avoid underestimating the time needed.

Training can also help people understand the psychological factors that lead to biased probability estimates, such as the tendency to rely on flawed intuition in lieu of careful analysis. Statistical intuitions are notoriously susceptible to illusions and superstition. Stock market analysts may see patterns in the data that have no statistical basis, and sports fans often regard basketball free-throw streaks, or "hot hands," as evidence of extraordinary new capability when in fact they're witnessing a mirage caused by capricious variations in a small sample size.

Another technique for making people aware of the psychological biases underlying skewed estimates is to give them "confidence quizzes."

Participants are asked for range estimates about general-interest questions (such as “How old was Martin Luther King Jr. when he died?”) or company-specific ones (such as “How much federal tax did our firm pay in the past year?”). The predictors’ task is to give their best guess in the form of a range and assign a degree of confidence to it; for example, one might guess with 90% confidence that Dr. King was between 40 and 55 when he was assassinated (he was 39). The aim is to measure not participants’ domain-specific knowledge, but, rather, how well they know what they don’t know. As Will Rogers wryly noted: “It is not what we don’t know that gets us into trouble; it is what we know that ain’t so.” Participants commonly discover that half or more of their 90% confidence ranges don’t contain the true answer.

Again, there’s no one-size-fits-all remedy for avoiding these systematic errors; companies should tailor training programs to their circumstances. Susquehanna International Group, a privately held global quantitative trading firm, has its own idiosyncratic approach. Founded in 1987 by poker aficionados, the company, which transacts more than a billion dollars in trades a year, requires new hires to play lots of poker—on company time. In the process, trainees learn about cognitive traps, emotional influences such as wishful thinking, behavioral game theory, and, of course, options theory, arbitrage, and foreign exchange and trading regulations. The poker-playing exercises sensitize the trainees to the value of thinking in probability terms, focusing on information asymmetry (what the opponent might know that I don’t), learning when to fold a bad hand, and defining success not as winning each round but as making the most of the hand you are dealt.

Companies should also engage in customized training that focuses on narrower prediction domains, such as sales and R&D, or areas where past performance has been especially poor. If your sales team is prone to hubris, that bias can be systematically addressed. Such tailored programs are more challenging to develop and run than general ones, but because they are targeted, they often yield greater benefits.

### Build the Right Kind of Teams

Assembling forecasters into teams is an effective way to improve forecasts. In the Good Judgment Project, several hundred forecasters were randomly assigned to work alone and several hundred to work

collaboratively in teams. In each of the four years of the IARAP tournament, the forecasters working in teams outperformed those who worked alone. Of course, to achieve good results, teams must be deftly managed and have certain distinctive features.

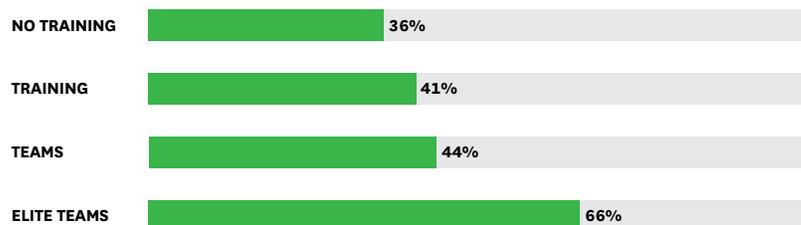
**Composition.** The forecasters who do the best in GJP tournaments are brutally honest about the source of their success, appreciating that they may have gotten a prediction right despite (not because of) their analysis. They are cautious, humble, open-minded, analytical—and good with numbers. (See the sidebar “Who Are These Superforecasters?”) In assembling teams, companies should look for natural forecasters who show an alertness to bias, a knack for sound reasoning, and a respect for data.

It’s also important that forecasting teams be intellectually diverse. At least one member should have domain expertise (a finance professional on a budget forecasting team, for example), but non-experts are essential too—particularly ones who won’t shy away from challenging the presumed experts. Don’t underestimate these generalists. In the GJP contests, nonexpert civilian forecasters often beat trained intelligence analysts at their own game.

### How Training and Teams Improve Prediction

The Good Judgment Project tracked the accuracy of participants’ forecasts about economic and geopolitical events. The control group, made up of motivated volunteers, received no training about the biases that can plague forecasters. Its members performed at about the same level as most employees in high-quality companies—perhaps even better, since they were self-selected, competitive individuals. The second group benefited from training on biases and how to overcome them. Teams of trained individuals, who debated their forecasts (usually virtually), performed even better. When the best forecasters were culled, over successive rounds, into an elite group of superforecasters, their predictions were nearly twice as accurate as those made by untrained forecasters—representing a huge opportunity for companies.

(% more accurate than random guesses)



## Who Are These Superforecasters?

The Good Judgment Project identified the traits shared by the best-performing forecasters in the Intelligence Advanced Research Projects Activity tournament. A public tournament is ongoing at [gjoin.com](http://gjoin.com); join to see if you have what it takes.

### PHILOSOPHICAL APPROACH AND OUTLOOK

**CAUTIOUS** They understand that few things are certain

**HUMBLE** They appreciate their limits

**NONDETERMINISTIC** They don't assume that what happens is meant to be

### ABILITIES AND THINKING STYLE

**OPEN-MINDED** They see beliefs as hypotheses to be tested

**INQUIRING** They are intellectually curious and enjoy mental challenges

**REFLECTIVE** They are introspective and self-critical

**NUMERATE** They are comfortable with numbers

### METHODS OF FORECASTING

**PRAGMATIC** They are not wedded to any one idea or agenda

**ANALYTICAL** They consider other views

**SYNTHESIZING** They blend diverse views into their own

**PROBABILITY-FOCUSED** They judge the probability of events not as certain or uncertain but as more or less likely

**THOUGHTFUL UPDATERS** They change their minds when new facts warrant it

**INTUITIVE SHRINKS** They are aware of their cognitive and emotional biases

### WORK ETHIC

**IMPROVEMENT-MINDED** They strive to get better

**TENACIOUS** They stick with a problem for as long as needed

### Diverging, evaluating, and converging.

Whether a team is making a forecast about a single event (such as the likelihood of a U.S. recession two years from now) or making recurring predictions (such as the risk each year of recession in an array of countries), a successful team needs to manage three phases well: a diverging phase, in which the issue, assumptions, and approaches to finding an answer are explored from multiple angles; an evaluating phase, which includes time for productive disagreement; and a converging phase, when the team settles on a prediction. In each of these phases, learning and progress are fastest when questions are focused and feedback is frequent.

The diverging and evaluating phases are essential; if they are cursory or ignored, the team develops tunnel vision—focusing too narrowly and quickly locking into a wrong answer—and prediction quality suffers. The right norms can help prevent this, including a focus on gathering new information and testing assumptions relevant to the forecasts. Teams must also focus on neutralizing a common prediction error called anchoring, wherein an early—and possibly ill-advised—estimate skews subsequent

opinions far too long. This often happens unconsciously because easily available numbers serve as convenient starting points. (Even random numbers, when used in an initial estimate, have been shown to anchor people's final judgments.)

One of us (Paul) ran an experiment with University of Chicago MBA subjects that demonstrated the impact of divergent exploration on the path to a final prediction. In one test, subjects in the control group were asked to estimate how many gold medals the U.S. would win relative to another top country in the next summer Olympics and to provide their 90% confidence ranges around these estimates. The other group was asked to first sketch out various reasons why the ratio of medals might be lower or higher than in years past and then make an estimate. This group naturally thought back to terrorist attacks and boycotts, and considered other factors that might influence the outcome, from illness to improved training to performance-enhancing drugs. As a consequence of this divergent thinking, this group's ranges were significantly wider than the control group's, often by more than half. In general, wider ranges reflect more carefully weighed predictions; narrow ranges commonly indicate overconfident—and often less accurate—forecasts.

**Trust.** Finally, trust among members of any team is required for good outcomes. It is particularly critical for prediction teams because of the nature of the work. Teams that are predicting the success or failure of a new acquisition, or handicapping the odds of successfully divesting a part of the business, may reach conclusions that raise turf issues or threaten egos and reputations. They are also likely to expose areas of the firm, and perhaps individuals, with poor forecasting abilities. To ensure that forecasters share their best thinking, members must trust one another and trust that leadership will defend their work and protect their jobs and reputations. Few things chill a forecasting team faster than a sense that its conclusions could threaten the team itself.

### Track Performance and Give Feedback

Our work on the Good Judgment Project and with a range of companies shows that tracking prediction outcomes and providing timely feedback is essential to improving forecasting performance.

Consider U.S. weather forecasters, who, though much maligned, excel at what they do. When they

## BRIER SCORES REVEAL YOUR BEST—AND WORST—PREDICTORS

It's important that forecasters make precise estimates of probability—for example, pegging at 80% the likelihood that their firm will sell between 9,000 and 11,000 units of a new product in the first quarter. That way, the predictions can be analyzed and compared using a method called Brier scoring, allowing managers to reliably rank forecasters on the basis of skill.

Brier scores are calculated by squaring the difference between a probability prediction and the actual outcome, scored as 1 if the event happened and 0 if not. For example, if a forecaster assigns a 0.9 probability (a 90% confidence level) that the firm will exceed a sales target and the firm then does, her Brier score for that forecast is:

$$(0.9 - 1)^2, \text{ or } 0.01.$$

If the firm misses the target, her score is:

$$(0.9 - 0)^2, \text{ or } 0.81.$$

The closer to zero the score is, the smaller the forecast error and the better the prediction.

Brier scoring makes it readily apparent who's good at forecasting and who isn't. By enabling direct comparison among forecasters, the tool encourages thoughtful analysis while exposing "shooting from the hip" and biased prognostications.

say there's a 30% chance of rain, 30% of the time it rains on those days, on average. Key to their superior performance is that they receive timely, continual, and unambiguous feedback about their accuracy, which is often tied to their performance reviews. Bridge players, internal auditors, and oil geologists also shine at prediction thanks in part to robust feedback and incentives for improvement.

The purest measure for the accuracy of predictions and tracking them over time is the Brier score. It allows companies to make direct, statistically reliable comparisons among forecasters across a series of predictions. Over time, the scores reveal those who excel, be they individuals, members of a team, or entire teams competing with others. (See the sidebar "Brier Scores Reveal Your Best—and Worst—Predictors.")

But simply knowing a team's score does little to improve performance; you have to track the process it used as well. It's important to audit why outcomes were achieved—good or bad—so that you can learn from them. Some audits may reveal that certain process steps led to a good or a bad prediction. Others may show that a forecast was correct despite a faulty rationale (that is, it was lucky), or that a forecast was wrong because of unusual circumstances rather than a flawed analysis. For example, a retailer may make very accurate forecasts of how many customers will visit a store on a given day, but if a black-swan event—say, a bomb threat—closes the store, its forecast for that day will be badly off. Its Brier score would indicate poor performance, but a process audit would show that bad luck, not bad process, accounted for the outlying score.

Gauging group dynamics is also a critical part of the process audit. No amount of good data and by-the-book forecasting can overcome flawed team dynamics. Consider the discussions that took place between NASA and engineering contractor Morton Thiokol before the doomed launch of the space shuttle *Challenger* in 1986. At first, Thiokol engineers advised against the launch, concerned that cold temperatures could compromise the O-rings that sealed the rocket boosters' joints. They predicted a much higher than usual chance of failure because of the temperature. Ultimately, and tragically, Thiokol reversed its stance.

The engineers' analysis was good; the organizational process was flawed. A reconstruction of the events that day, based on congressional hearings, revealed the interwoven conditions that compromised

the forecast: time pressure, directive leadership, failure to fully explore alternate views, silencing of dissenters, and a sense of infallibility (after all, 24 previous flights had gone well).

To avoid such catastrophes—and to replicate successes—companies should systematically collect real-time accounts of how their top teams make judgments, keeping records of assumptions made, data used, experts consulted, external events, and so on. Videos or transcripts of meetings can be used to analyze process; asking forecasters to record their own process may also offer important insights. Recall Susquehanna International Group, which trains its traders to play poker. Those traders are required to document their rationale for entering or exiting a trade before making a transaction. They are asked to consider key questions: What information might others have that you don't that might affect the trade? What cognitive traps might skew your judgment on this transaction? Why do you believe the firm has an edge on this trade? Susquehanna further emphasizes the importance of process by pegging traders' bonuses not just to the outcome of individual trades but also to whether the underlying analytic process was sound.

Well-run audits can reveal post facto whether forecasters coalesced around a bad anchor, framed the problem poorly, overlooked an important insight, or failed to engage (or even muzzled) team members with dissenting views. Likewise, they can highlight the process steps that led to good forecasts and thereby provide other teams with best practices for improving predictions.

**EACH OF** the methods we've described—training, team building, tracking, and talent spotting—is essential to good forecasting. The approach must be customized across businesses, and no firm, to our knowledge, has yet mastered them all to create a fully integrated program. This presents a great opportunity for companies that take the lead—particularly those with a culture of organizational innovation and those who embrace the kind of experimentation the intelligence community did.

But companies will capture this advantage only if respected leaders champion the effort, by broadcasting an openness to trial and error, a willingness to ruffle feathers, and a readiness to expose "what we know that ain't so" in order to hone the firm's predictive edge. ♥

HBR Reprint R1605E