Probabilistic Thinking

Idea In Short

Replace verbal hedges with numbers. Starting this quarter, require every forecast in strategy reviews to carry a probability, a resolution date, and a named owner. Record the numbers, score them when events resolve, and share the results inside the team. Words such as likely or possible mean different things to different readers, and that gap has misled senior decision makers before. Teams that state odds argue about evidence instead of vocabulary, spot overconfidence early, and learn which of their people forecast well. Begin with ten live questions, such as whether a launch will ship on schedule or a competitor will cut prices, and review the first scores in ninety days.

Most executives can recall a meeting where everyone agreed that an outcome was likely, and where each person meant something different by the word. Nobody noticed until the outcome arrived. Probabilistic thinking removes that gap by asking for a number, a date, and an owner. It does not require advanced mathematics. It requires the habit of saying how sure you are, and the discipline to check afterward. This article describes why verbal forecasts fail, what forecasting tournaments taught researchers, and how a leadership team can build a simple practice around stated odds.

Why words fail leaders

Sherman Kent, an analyst at the Central Intelligence Agency, made the case for consistent odds language in a 1964 essay titled Words of Estimative Probability.1 He saw that analysts used terms such as probable and serious possibility with no shared meaning. Readers filled in their own numbers. Kent proposed that probable should mean roughly 75 percent, give or take 12 points, and that other terms should map to similar ranges. His proposal never took hold across the intelligence community, but the problem he named has not gone away.

A well-known story shows the cost. Mellers and her co-authors retell an account in which an evaluator described the odds of the Bay of Pigs operation in a phrase he meant as not too

good, while President Kennedy allegedly heard something more favorable.2 The evaluator later believed that his vague wording contributed to a serious strategic error. Business settings repeat the pattern at lower stakes. A product lead says a launch is probably on track, a finance lead hears 90 percent, and the budget gets built around it.

What stating odds changes

A stated probability turns an opinion into a claim that reality can confirm or refute. If you say an event has a 70 percent chance and it happens, you were neither right nor wrong on that one case. Over fifty such statements, however, about 35 of the events should occur, and the record will tell you whether your sense of 70 percent deserves trust. Verbal forecasts allow no such test. A leader who says something is likely can always claim vindication, whatever happens.

Numbers also make disagreement useful. When one person says 30 percent and another says 80, the team has found something worth discussing. They can ask what evidence each person weighs, and which of them knows something the other lacks. When both say likely, the conversation ends early and the difference stays hidden.

Choosing the granularity

Precision should match knowledge. For most business questions, steps of five percentage points are enough. A forecast of 62.5 percent claims a refinement that no one in the room possesses. At the other extreme, labels such as high, medium, and low blur real differences, because 55 percent and 85 percent both land in high. Reserve finer steps for the ends of the scale, where the gap between 1 percent and 5 percent changes the decision. Pick one convention and apply it to every forecast the team makes, so that results can be compared across quarters.

Lessons from the forecasting tournaments

The most useful evidence comes from forecasting contests. Philip Tetlock's work with Barbara Mellers and Don Moore recruited thousands of volunteers to predict world events, in a program funded by the US intelligence research agency IARPA, the Intelligence Advanced Research Projects Activity.3 The team launched its first contest in 2011 as the Good Judgment Project. Tetlock summarized the broader finding in an interview: most

people forecast poorly, and even well-known experts lack a solid track record.4

Some participants, though, did consistently well. Researchers labeled the best of them superforecasters. The Good Judgment Project also built a short training module in probabilistic reasoning, and it won the tournaments by wide margins in each year of the competition. Training mattered, which means the skill is teachable. One framework from that work uses the acronym CHAMP: Comparisons, Historical trends, Average opinions, Mathematical models, and Predictable biases.5 Each element is a prompt that a team can run through before it commits to a number.

Starting with base rates

Good forecasters begin with a reference class. They ask how often projects like this one finish on time, how often a competitor in this position cuts prices, or how often a pilot of this type converts into a contract. Suppose a company has completed ten large software migrations and seven overran their schedules. A fresh migration therefore starts near a 70 percent chance of overrun, whatever the sponsor feels. The team then adjusts for specifics, such as a stronger vendor or a narrower scope, and moves the number a few points in either direction.

Most teams reverse the order. They start from the plan, tell a persuasive story about why this case is different, and treat the base rate as an afterthought. Fixing the order requires no extra data, only a decision to ask the historical question first. If the company has no records, the exercise still helps, because it reveals how little history the team has to rely on.

Keeping score

Forecasts improve only when someone checks them. Keep a log with the question, the date, each person's probability, and the eventual outcome. After enough questions have resolved, group the forecasts by stated probability and compare each group with how often the events happened. If your 90 percent calls come true only 70 percent of the time, you are overconfident, and you can discount your own certainty accordingly. Many forecasters also use the Brier score, a measure that rewards accurate and confident calls and penalizes confident misses. Scores are for learning, not blame. Teams that publicize individual rankings too early tend to see people retreat to safe 50 percent answers.

Combining views

Independent estimates, once averaged, usually outperform most of the individuals who produced them. The condition is independence. If the most senior person speaks first, others anchor on that number, and the group loses the benefit of its diversity. Ask for written probabilities before discussion, then reveal them together. Look at the spread. A narrow cluster suggests shared information, while a wide one points to a gap in knowledge that someone can close with a phone call. Superforecasting research puts weight on the same idea, listing average opinions as one of the five habits in the CHAMP scheme described above.

A forecasting routine for leadership teams

A simple cycle fits inside existing meetings. It adds little time, and it builds a record that gets more valuable each quarter. The steps are:

1. Convert each key judgment in the plan into a question with a clear resolution date
2. Collect independent probabilities in writing
3. Compare the numbers and discuss the largest gaps
4. Record the agreed forecast and the reasons behind it
5. Score the results when each question resolves and review them quarterly

After two or three cycles, patterns emerge. Some people prove consistently well calibrated, some topics produce persistent overconfidence, and some data sources prove less reliable than expected. Use those findings to assign questions, to commission better information, and to adjust how much weight each input receives.

Where the approach strains

The method has limits. Rare events resist calibration because a team may wait years to see enough cases. Questions framed vaguely produce numbers that cannot be scored, so each question needs a crisp resolution rule. Numbers can also create false confidence if leaders treat a 72 percent estimate as a measurement rather than a judgment. Finally, incentives matter: if forecasts feed into bonuses, people will shade them. Keep the practice separate from performance pay, and present results as a learning tool. Within those boundaries, stating odds remains the most direct way to expose the difference between what a team

knows and what it merely feels.

- 1Words of estimative probability
- 2Improving the accuracy of geopolitical risk assessments
- 3Superforecasters: the art of accurate predictions
- 4Why an open mind is key to making better predictions
- 5Superforecaster

Summary

Probabilistic thinking asks leaders to give up the comfort of confident language. A forecast that carries a number can be tested, compared, and improved, while a forecast that carries an adjective can only be defended. Start from base rates, write down your odds, gather independent views before discussion, and keep score. The practice feels awkward for the first few weeks because it exposes how often confidence outruns evidence. After a few quarters, the record of resolved forecasts becomes one of the most useful management assets a team owns, since it shows who sees clearly, which questions mislead, and where the organization needs better information.