

## The Secret Ingredients of 'Superforecasting'

**Data from a political predictions tournament are starting to yield insights into the main drivers of forecasting excellence – and how to cultivate it.**

“**Superforecasters**” walk among us – people who can predict the future with rare accuracy, outstripping even domain experts.

That was the inescapable conclusion drawn from the **Good Judgment Project** (GJP), a forecasting tournament launched by Wharton professors **Philip Tetlock** and **Barbara Mellers**. From 2011 to 2015, the US government-funded online initiative pitted the predictive powers of ordinary people against Washington, DC intelligence analysts on the most significant geopolitical questions of the day. Over successive rounds, Tetlock and Mellers identified the very best prognosticators from the 25,000-strong participant pool and shunted them into elite teams. Despite the fact that the Beltway experts had access to classified data and intelligence reports, the GJP superforecaster squads **bested them** in predictive accuracy by about 30 percent.

But there was more to the GJP’s success than merely identifying and grouping superforecasters. Along the way, Tetlock and Mellers developed three interventions – training, teaming and tracking – that improved prediction quality for superforecasters and average folks alike. This feature of the GJP may be the most appealing for companies, as even a modest increase in the overall accuracy of a firm’s predictions could unlock tremendous value.

*Training* refers specifically to probabilistic reasoning tutorials, which convey tools and techniques for testing assumptions, spotting relevant patterns in past data, avoiding common errors in judgment, etc. *Teaming*, as you might expect, involved grouping individuals together so they could share information and challenge each other prior to making a prediction. *Tracking* was the practice, mentioned above, of separating the highest performers into elite squads of superforecasters.

Four years after the close of the tournament phase of the GJP, Tetlock and Mellers continue to plumb the data for granular insights on how, exactly, these interventions improved people’s predictions. Along with INSEAD PhD graduate **Marat Salikhov**, I have been collaborating with them in this effort. Although research is still ongoing, our findings so far have introduced surprising elements to our understanding of predictions – and what makes superforecasters tick.

### **BIN: Bias, Information, Noise**

We began by positing that superforecasters excel in three areas: reducing and accounting for biases (both their own and any that may be reflected in the evidence they’re working with), efficiently extracting data from the environment to compensate

for what they don't already know, and nullifying noise in the data (i.e. errors that, unlike bias, have no pattern or system behind them).

The GJP's intervention design was based on the idea that bias would be the most easily improvable of the three. Noise, being random by definition, and information extraction, being dependent on people's curiosity and ability to hunt down useful data, were thought to be more resistant to intervention. But this was pure speculation, without numbers to confirm or refute it.

To investigate how bias, information and noise interact in predictions, we designed a statistical model (which we dubbed the BIN model, for **B**ias, **I**nformation and **N**oise) and applied it to the full 2011-2015 GJP dataset.

How does the BIN model work? Simply put, it analyses the entire "signal universe" around a given question. Signals are pieces of information that the forecasters may take into account when trying to guess whether something will happen. In formulating predictions, one can rely upon either meaningful signals (i.e. information extraction) or irrelevant signals (i.e. noise). One can also organise information along erroneous lines (i.e. bias). Comparing GJP groups that experienced one or more of the three interventions to those that did not, the BIN model was able to disaggregate the respective contributions of noise, information and bias to overall improvements in prediction accuracy.

### **A surprising result**

We found that teams of superforecasters were the least noisy, least biased and most informed. This may not be very surprising, yet it is a significant discovery. It suggests that superforecasters define the outer limit of human possibility in this area, thus giving the rest of us – and companies seeking to benefit from improved forecasting ability – an attainable goal to shoot for.

Our experiments with the BIN model have also produced results that were more unexpected. Recall that teaming, tracking and training were deployed for the express purpose of reducing bias. Yet it seems that only teaming actually did so. Two of the three – teaming and tracking – increased information. Surprisingly, all three interventions reduced noise. In light of our current study, it appears the GJP's forecasting improvements were overwhelmingly the result of noise reduction. As a rule of thumb, about 50 percent of the accuracy improvements can be attributed to noise reduction, 25 percent to tamping down bias, and 25 percent to increased information.

Visit **INSEAD Knowledge**  
<http://knowledge.insead.edu>

Again, we plan to launch further investigations into this question. For now, it seems that our initial focus on bias suppression as the key to increasing predictive power may require reinvestigation.

### **An argument for algorithms?**

While we have no definitive explanation as to why noise emerged as such an important factor, the information overload in our current media environment is one plausible cause. In a digital world swarming with fake news and sensationalist content, those who cast about widely for information are sure to reel in some strange fish. To extrapolate, superforecasters' true edge may be more about discipline – the mental rigour required to distinguish random from revealing data – than innate wisdom or intellectual objectivity.

This intuition may be supported by comparing GJP forecasts at different time horizons, from 60 to only a few days before the event. We saw that the importance of noise reduction remains rather constant over time, being responsible for about 50 percent of the accuracy improvements. Bias was more salient the further back in time a forecaster was from the moment of truth. In contrast, information became more important as the resolution date approached – perhaps in conjunction with increasing news coverage.

Whatever the reason, investing in noise reduction may not be a bad idea. One proven, if drastic, noise-reduction solution is to assign predictions to algorithms rather than humans. Bots are programmed to pay attention to patterns in data and discount random information. They are, however, ill-suited to forecast the outcome of nuanced, complex and often unique situations such as the GJP's geopolitical quandaries.

Our research implies there is hope for human forecasters seeking to improve. Indeed, if the GJP interventions, designed to reduce bias, improved predictions by shutting out the noise, presumably measures specifically targeting noise would be yet more effective.

*Ville Satopää is an Assistant Professor of Technology and Operations Management at INSEAD.*

Found this article useful? **Subscribe** to our weekly newsletter.

Follow INSEAD Knowledge on **Twitter** and **Facebook**.

**Find article at**

<https://knowledge.insead.edu/strategy/the-secret-ingredients-of-superforecasting-12721>

**Download the Knowledge app for free**



**Visit INSEAD Knowledge**  
<http://knowledge.insead.edu>