

# FOREIGN AFFAIRS

DECEMBER 16, 2022

---

## Fumbling the Crystal Ball

---

Policymakers Can't Afford to Spurn the Science of  
Prediction

J. PETER SCOBLIC AND PHILIP E. TETLOCK

# Fumbling the Crystal Ball

---

## Policymakers Can't Afford to Spurn the Science of Prediction

J. PETER SCOBLIC AND PHILIP E. TETLOCK

**I**n recent months, Russian President Vladimir Putin's nuclear saber rattling has forced observers and policymakers to try to figure out how seriously to take his threats. Unfortunately, few of the countless analyses published thus far have assessed the probability that Putin will use nuclear weapons in Ukraine in specific, quantitative terms. Most merely conclude that Putin is "likely" or "unlikely" to do so without defining what those words mean in the present context.

Although some imprecision is a function of the uncertainty inherent in a unique situation, much of it stems from a resistance to quantifying the probability of rare, catastrophic events. But the qualitative descriptions commentators tend to prefer may be misleading: research has found wide variation in the likelihood people assign to probabilistic words and expressions such as "rarely" and "almost certainly."

Such terms convey both less than people want to know about the future and less than they can know about it. As we wrote in *Foreign Affairs* in 2020, interdisciplinary research teams have developed new analytic tools that can put more-accurate odds on future events. In effect, the scientific community has handed policymakers a framework for better predicting

what U.S. adversaries will do next and what their actions will (or will not) accomplish.

A CLEARER VIEW OF THE FUTURE

Better prediction tools have the potential to fundamentally change the way policymakers approach problems at all levels of policy and governance, from the most mundane questions facing local leaders—such as the likelihood that a winter storm could force schools in a particular district to delay opening by an hour or two—to complex, existential threats such as climate change and the use of nuclear weapons.

There is simply no policy situation under which we would not want better visibility into the future. National security policymakers routinely make high-consequence, difficult-to-reverse decisions, so they must use the best analytic tools available. Yet although U.S. policymakers have embraced quantitative models for threats such as pandemics and climate change, and the intelligence community recently launched a new crowd-sourced initiative to predict future threats, resistance to probabilistic forecasting remains widespread.

---

There is simply no policy situation under which we would not want better visibility into the future.

As psychologist Daniel Kahneman has noted—and as thousands of studies have amply confirmed—human beings are not natural statisticians. For example, a cognitive bias that researchers call “scope insensitivity” makes it difficult for people to adjust their beliefs to match the scale of a problem. One study found that the amount subjects were willing to donate to save wildlife affected by an oil spill varied little, regardless of whether the accident involved 2,000 or 20,000 or 200,000 animals. (The amounts were \$80, \$78, and \$88, respectively.) Even though the problem changed by orders of magnitude, the solution did not.

These quirks of innumeracy affect probability judgments, as well. Indeed, many of the cognitive biases that Kahneman and his research partner, the psychologist Amos Tversky, and their acolytes have documented are considered “biases” precisely because they lead to errors

in probabilistic reasoning. When a situation is presented as a “story,” people tend to conflate plausibility (i.e., believability) with probability (i.e., likelihood). A more detailed narrative about the future may seem more credible, but each additional detail reduces the chances that the scenario will occur.

Although human beings are not natural statisticians, they are natural storytellers. But because the forecasting community prizes measurable performance over narrative coherence, its members frequently fail to tell compelling stories that might help policymakers and ordinary citizens grasp the value of their work. Often, forecasters don’t explain their estimates, so their numbers seem to emerge from a black box. Moreover, because the best forecasters tend not to be subject-matter experts, their explanations can sound amateurish. These factors make it easier for policymakers to write off forecasting as an exercise in geopolitical dilettantism.

#### AN AFFRONT TO EXPERTISE?

Organizational change is difficult under the best of circumstances and is close to impossible when powerful insiders actively resist it. National security experts with decades of experience and access to classified information see little reason for deferring to the upstart winners of forecasting tournaments, contests that allow the public to compete at putting realistic odds on future events. Perhaps they are concerned that as forecasters get better at geopolitical analysis, they will threaten the notion of expertise and the professional identities of those who supply it. But forecasting should be seen as a complement to expert analysis, not a substitute for it.

The same situation obtains among the corps of foreign-policy columnists, think tank fellows, and former government officials who wield more influence for the confidence of their convictions than for the precision of their predictions. There is little incentive for such analysts to ask when they have been wrong and why—questions that top forecasters must constantly confront if they are to maintain their place in the accuracy hierarchy. Instead, the “thought leader” ecosystem insulates the careers of

people who would have washed out of any geopolitical forecasting tournament.

The intelligence community increasingly assigns probabilities to its assessments—albeit in words that express wide probabilistic ranges. So, for example, if a CIA analysis says something is “likely,” the agency believes there is a “55 to 80 percent” chance that it will happen. But such changes have been hard won; the battle over the utility of numerical odds goes back to the CIA’s earliest days. Only in 2015, with Intelligence Community Directive 203, did the agency begin defining probability ranges. Despite these shifts, publicly available intelligence products such as the CIA’s Annual Threat Assessment tend to rely on words, not numbers, to hedge probabilistic predictions. “The economic fallout from the pandemic is likely to continue to challenge governments,” it reads in one passage.

#### A STRUGGLE FOR LEGITIMACY

New ideas are always a hard sell, even when backed by solid, scientific data. Geopolitical forecasting has faced an uphill battle for what sociologists of science call “legitimation”—the grudging process by which an organization assimilates new norms, ideas, and practices. A group of organizations that participate in this give-and-take create a field. Once a field is established, new organizations can achieve legitimacy by mimicking other players. But what happens when no field yet exists?

The emerging management consulting industry was forced to confront this question in the years after World War II. Like many professionals, management consultants sell an intangible product, making it particularly important for them to legitimize their work by demonstrating its value and normative propriety. Put bluntly, management consultants had to prove they were not charlatans. To do this, companies took collective action, establishing a professional association that gave consultants a way to define who was a member of their community. To codify that distinction, the association adopted a code of ethics, which in turn served as the basis for hiring and training new consultants, thus reinforcing the identity of the “professional” management consultant.

Geopolitical forecasting has not yet been legitimized in this sense. It has no standardized training regimen, no degrees in geopolitical forecasting, and no formal credential establishing forecasters. The four-year tournament sponsored by the U.S. intelligence community beginning in 2010 provided a great deal of data about good forecasting traits, and it also served as a credentialing function by anointing the top two percent of competitors as “superforecasters” and by establishing a performance metric: the Brier score, named after Glenn Brier, who proposed the system in 1950 while working for the U.S. Weather Bureau. But even though Brier scores are the product of a simple mathematical formula, their meaning is not intuitive. (Brier scores, which are expressed in hundredths of a point, range from 0 to 2, with 0 representing perfect omniscience and 2 representing delusional detachment from reality. A score of, say, 0.18 would put you in superforecaster territory in periods of normal geopolitical volatility.)

Difficulties in professionalization aside, the institutional manifestations of forecasting are growing. Universities and companies now operate open forecasting platforms, and there are many prediction markets in which people bet real money on their estimates of the future. The forecasting community also feeds (and is funded by) the growing Effective Altruism movement, which aims to maximize the value of its members’ time, money, and effort—an exercise that requires measurable estimates of the future. As Effective Altruism has gained prominence, forecasting has, too. Some critics have mocked the movement for its failure to predict the collapse of FTX and the disgrace of its founder, Sam Bankman-Fried, but that can be interpreted as a teachable moment. Good intentions do not guarantee good forecasting—and good forecasters do not always have good intentions.

#### RESISTANCE TO QUANTIFYING ODDS

One obstacle to integrating forecasting in national security decision-making is that even though decision-makers certainly want to understand the chances of success, the fine-grained odds offered by the best forecasters are likely to be ignored for several reasons. For one thing,

policymakers rarely ask for quantitative odds—and when they do, they don't do so skillfully. In 2011, before U.S. President Barack Obama authorized a raid on Osama bin Laden's compound in Abbottabad, he famously went around the table in the Situation Room, asking each of his advisers to estimate the probability that the al Qaeda leader was there. After being given a wide range of probabilities—from 30 percent to 95 percent, according to Mark Bowden's reporting—Obama reportedly threw up his hands and said, "Look guys, this is a flip of the coin," by which he did not mean that the probability equaled 0.5—most people don't when they use that expression—but rather that no one truly knew. Which they didn't. Still, his aides' aggregate estimate was well above 50 percent.

Another challenge is that policymakers are rarely playing iterated games in which incremental improvements in judgment make big differences over time. They are not equities traders. The decision threshold for a government official is unlikely to vary because of a ten percent shift one way or another. A poker player's career would be made (or broken) on the ability to differentiate 45:55 odds from 55:45 odds, but to a decision-maker faced with a high-stakes one-off decision, both sound like a tossup. And even after the fact, there is no way to say that a probability estimate was "right," so no decision-maker can defend a failed policy on the basis that success had been deemed ten percent more likely.

Yet another challenge to forecasting is that, in national security, decision quality tends to be judged based on outcome, not process. Decision scientists recommend the precise opposite, however, because, over time, good decision-making hygiene will lead to better results. But given a single event, even an impeccable decision process can lead to a poor outcome simply because the randomness of the universe interferes. Moreover, the public rarely has access to policymakers' deliberative processes until well after their decisions are made, and judging decisions by procedural quality is counterintuitive.

Finally, forecasting requires specific questions with answers that one can ultimately score as true or false based on whether an event happened. So one might ask: "Will Russian or Belarusian troops cross the land border



between Belarus and either the Volyn or Rivne oblasts before July 2023?” But the answer may not address the question policymakers are most interested in: “Will Russia prevail in Ukraine?” As we argued in our 2020 article, one way to deal with this problem is to take broader, longer-term questions and develop sets of uncorrelated, shorter-term questions that suggest the answer. So one might ask not only the probability of specific Russian military operations this year but also the probabilities that Russia will meet certain economic targets or lose support in specific international forums—developments that could influence the odds of military victory.

#### WARY POLICYMAKERS

At the most abstract level, policymakers have not flocked to forecasting because they consider future geopolitical events unique and believe that it is thus impossible to assign them meaningful odds. Those who hold this view tend to regard scientific analysis demonstrating the validity of forecasting as epistemological sleight of hand.

This is an understandable objection but also conveniently exculpatory, allowing policymakers to elide responsibility for their failure to predict the predictable. When COVID-19 struck, everyone from *New York Times* writers to Republican supporters of President Donald Trump cried, “Black swan!”—inappropriately borrowing Nassim Taleb’s memorable term for events that are inherently unforeseeable. Humanity has experienced many pandemics, and for decades health experts warned the public that a deadly bug could spread globally and kill millions. They ran dozens of crisis simulations with ominous names such as “Dark Winter,” “Crimson Contagion,” and “Event 201.” If anything, the COVID-19 pandemic was more of a “gray rhino”—to use the writer Michele Wucker’s term for a foreseeable but often overlooked threat—than a black swan.

In reality, most events fall roughly into a comparison set that provides a base rate—that is, a measure of how often such events occur. Just as each person is both a specific individual and a human being, so are many geopolitical events both unique and common. Consider coups, for example. Even in a situation that would be truly novel, such as a strategic



nuclear exchange—an event with no precedent—one can look at the odds of conventional war and estimate the likelihood of escalation.

---

New ideas are always a hard sell, even when backed by solid, scientific data.

Thoughtful skeptics often complain about the specificity of forecasts. They argue that forecasters cannot meaningfully distinguish between a 35 percent probability and a 45 percent probability. (They can.) They worry that policymakers who are provided precise estimates will become overconfident. (They do not.) And they object that any forecast containing three significant digits cannot be valid. (Fair, but such forecasts—e.g., there is a 0.037 percent chance of dying in a nuclear strike in the next month—are just a function of multiplying out component forecasts. And they can highlight significant differences of opinion regarding very low-probability events.)

All this suggests that to make forecasting a resource that policymakers use, the quality of both supply and demand needs to improve. The former requires giving subject-matter experts a role in producing forecasts—in formulating questions (because they know which indicators are most germane) and in vetting the rationales that inform forecasts (because they can gut-check causal claims and fact-check evidence). The latter requires making the national security establishment more numerate or at least more open to quantitative appraisals of the future.

These are challenging tasks, but forecasting scholars are already testing methods for not only measuring the best forecasts but also judging the most persuasive rationales for those forecasts. For example: What story best conveys that there is a 10–15 percent chance of between one and three million people dying in the Ukraine war by the end of 2024? Where forecasters provide probability, subject-matter experts can provide plausibility, making well-calibrated quantitative future estimates more convincing and palatable to policymakers—and therefore making their decisions a little less wrong. And in national security, being a little less wrong can be a lot less dangerous.