THE NEW YORKER
THE CRITICS
BOOKS

EVERYBODY'S AN EXPERT

by LOUIS MENAND

Putting predictions to the test.

Issue of 2005-12-05
Posted 2005-11-28

Prediction is one of the pleasures of life. Conversation would wither without it. "It won't last. She'll dump him in a month." If you're wrong, no one will call you on it, because being right or wrong isn't really the point. The point is that you think he's not worthy of her, and the prediction is just a way of enhancing your judgment with a pleasant prevision of doom. Unless you're putting money on it, nothing is at stake except your reputation for wisdom in matters of the heart. If a month goes by and they're still together, the deadline can be extended without penalty. "She'll leave him, trust me. It's only a matter of time." They get married: "Funny things happen. You never know." You still weren't *wrong*. Either the marriage is a bad one—you erred in the right direction—or you got beaten by a low-probability outcome.

It is the somewhat gratifying lesson of Philip Tetlock's new book, "Expert Political Judgment: How Good Is It? How Can We Know?" (Princeton; $35), that people who make prediction their business—people who appear as experts on television, get quoted in newspaper articles, advise governments and businesses, and participate in punditry roundtables—are no better than the rest of us. When they're wrong, they're rarely held accountable, and they rarely admit it, either. They insist that they were just off on timing, or blindsided by an improbable event, or almost right, or wrong for the right reasons. They have the same repertoire of self-justifications that everyone has, and are no more inclined than anyone else to revise their beliefs about the way the world works, or ought to work, just because they made a mistake. No one is paying you for your gratuitous opinions about other people, but the experts are being paid, and Tetlock claims that the better known and more frequently quoted they are, the less reliable their guesses about the future are likely to be. The accuracy of an expert's predictions actually has an inverse relationship to his or her self-confidence, renown, and, beyond a certain point, depth of knowledge. People who follow current events by reading the papers and newsmagazines regularly can guess what is likely to happen about as accurately as the specialists whom the papers quote. Our system of expertise is completely inside out: it rewards bad judgments over good ones.

"Expert Political Judgment" is not a work of media criticism. Tetlock is a psychologist—he teaches at Berkeley—and his conclusions are based on a long-term study that he began twenty years ago. He picked two hundred and eighty-four people who made their living "commenting or offering advice on political and economic trends," and he started asking them to assess the probability that various things would or would not come to pass, both in the areas of the world in which they specialized and in areas about which they were not expert. Would there be a nonviolent end to

apartheid in South Africa? Would Gorbachev be ousted in a coup? Would the United States go to war in the Persian Gulf? Would Canada disintegrate? (Many experts believed that it would, on the ground that Quebec would succeed in seceding.) And so on. By the end of the study, in 2003, the experts had made 82,361 forecasts. Tetlock also asked questions designed to determine how they reached their judgments, how they reacted when their predictions proved to be wrong, how they evaluated new information that did not support their views, and how they assessed the probability that rival theories and predictions were accurate.

Tetlock got a statistical handle on his task by putting most of the forecasting questions into a "three possible futures" form. The respondents were asked to rate the probability of three alternative outcomes: the persistence of the status quo, more of something (political freedom, economic growth), or less of something (repression, recession). And he measured his experts on two dimensions: how good they were at guessing probabilities (did all the things they said had an *x* per cent chance of happening happen *x* per cent of the time?), and how accurate they were at predicting specific outcomes. The results were unimpressive. On the first scale, the experts performed worse than they would have if they had simply assigned an equal probability to all three outcomes—if they had given each possible future a thirty-three-per-cent chance of occurring. Human beings who spend their lives studying the state of the world, in other words, are poorer forecasters than dart-throwing monkeys, who would have distributed their picks evenly over the three choices.

Tetlock also found that specialists are not significantly more reliable than non-specialists in guessing what is going to happen in the region they study. Knowing a little might make someone a more reliable forecaster, but Tetlock found that knowing a lot can actually make a person less reliable. "We reach the point of diminishing marginal predictive returns for knowledge disconcertingly quickly," he reports. "In this age of academic hyperspecialization, there is no reason for supposing that contributors to top journals—distinguished political scientists, area study specialists, economists, and so on—are any better than journalists or attentive readers of the New York *Times* in 'reading' emerging situations." And the more famous the forecaster the more overblown the forecasts. "Experts in demand," Tetlock says, "were more overconfident than their colleagues who eked out existences far from the limelight."

People who are not experts in the psychology of expertise are likely (I predict) to find Tetlock's results a surprise and a matter for concern. For psychologists, though, nothing could be less surprising. "Expert Political Judgment" is just one of more than a hundred studies that have pitted experts against statistical or actuarial formulas, and in almost all of those studies the people either do no better than the formulas or do worse. In one study, college counsellors were given information about a group of high-school students and asked to predict their freshman grades in college. The counsellors had access to test scores, grades, the results of personality and vocational tests, and personal statements from the students, whom they were also permitted to interview. Predictions that were produced by a formula using just test scores and grades were more accurate. There are also many studies showing that expertise and experience do not make someone a better reader of the evidence. In one, data from a test used to diagnose brain damage were given to a group of clinical psychologists and their secretaries. The psychologists' diagnoses were no better than the secretaries'.

The experts' trouble in Tetlock's study is exactly the trouble that all human beings have: we fall in love with our hunches, and we really, really hate to be wrong. Tetlock describes an experiment that he witnessed thirty years ago in a Yale classroom. A rat was put in a T-shaped maze. Food was placed in either the right or the left transept of the T in a random sequence such that, over the long run, the food was on the left sixty per cent of the time and on the right forty per cent. Neither the students nor (needless to say) the rat was told these frequencies. The students were asked to predict on which side of the T the food would appear each time. The rat eventually figured out that the food was on the left side more often than the right, and it therefore nearly always went to the left, scoring roughly sixty per cent—D, but a passing grade. The students looked for patterns of left-right placement, and ended up scoring only fifty-two per cent, an F. The rat, having no reputation to begin with, was not embarrassed about being wrong two out of every five tries. But Yale students, who do have reputations, searched for a hidden order in the sequence. They couldn't deal with forty-per-cent error, so they ended up with almost fifty-per-cent error.

The expert-prediction game is not much different. When television pundits make predictions, the more ingenious their forecasts the greater their cachet. An arresting new prediction means that the expert has discovered a set of interlocking causes that no one else has spotted, and that could lead to an outcome that the conventional wisdom is ignoring. On shows like "The McLaughlin Group," these experts never lose their reputations, or their jobs, because long shots are their business. More serious commentators differ from the pundits only in the degree of showmanship. These serious experts—the think tankers and area-studies professors—are not entirely out to entertain, but they are a little out to entertain, and both their status as experts and their appeal as performers require them to predict futures that are not obvious to the viewer. The producer of the show does not want you and me to sit there listening to an expert and thinking, I could have said that. The expert also suffers from knowing too much: the more facts an expert has, the more information is available to be enlisted in support of his or her pet theories, and the more chains of causation he or she can find beguiling. This helps explain why specialists fail to outguess non-specialists. The odds tend to be with the obvious.

Tetlock's experts were also no different from the rest of us when it came to learning from their mistakes. Most people tend to dismiss new information that doesn't fit with what they already believe. Tetlock found that his experts used a double standard: they were much tougher in assessing the validity of information that undercut their theory than they were in crediting information that supported it. The same deficiency leads liberals to read only *The Nation* and conservatives to read only *National Review*. We are not natural falsificationists: we would rather find more reasons for believing what we already believe than look for reasons that we might be wrong. In the terms of Karl Popper's famous example, to verify our intuition that all swans are white we look for lots more white swans, when what we should really be looking for is one black swan.

Also, people tend to see the future as indeterminate and the past as inevitable. If you look backward, the dots that lead up to Hitler or the fall of the Soviet Union or the attacks on September 11th all connect. If you look forward, it's just a random scatter of dots, many potential chains of causation leading to many possible outcomes. We have no idea today how tomorrow's invasion of a foreign land is going to go; after the invasion, we can actually persuade ourselves that we knew all along.

The result seems inevitable, and therefore predictable. Tetlock found that, consistent with this asymmetry, experts routinely misremembered the degree of probability they had assigned to an event after it came to pass. They claimed to have predicted what happened with a higher degree of certainty than, according to the record, they really did. When this was pointed out to them, by Tetlock's researchers, they sometimes became defensive.

And, like most of us, experts violate a fundamental rule of probabilities by tending to find scenarios with more variables more likely. If a prediction needs two independent things to happen in order for it to be true, its probability is the product of the probability of each of the things it depends on. If there is a one-in-three chance of *x* and a one-in-four chance of *y*, the probability of both *x* and *y* occurring is one in twelve. But we often feel instinctively that if the two events "fit together" in some scenario the chance of both is greater, not less. The classic "Linda problem" is an analogous case. In this experiment, subjects are told, "Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations." They are then asked to rank the probability of several possible descriptions of Linda today. Two of them are "bank teller" and "bank teller and active in the feminist movement." People rank the second description higher than the first, even though, logically, its likelihood is smaller, because it requires two things to be true—that Linda is a bank teller and that Linda is an active feminist—rather than one.

Plausible detail makes us believers. When subjects were given a choice between an insurance policy that covered hospitalization for any reason and a policy that covered hospitalization for all accidents and diseases, they were willing to pay a higher premium for the second policy, because the added detail gave them a more vivid picture of the circumstances in which it might be needed. In 1982, an experiment was done with professional forecasters and planners. One group was asked to assess the probability of "a complete suspension of diplomatic relations between the U.S. and the Soviet Union, sometime in 1983," and another group was asked to assess the probability of "a Russian invasion of Poland, and a complete suspension of diplomatic relations between the U.S. and the Soviet Union, sometime in 1983." The experts judged the second scenario more likely than the first, even though it required two separate events to occur. They were seduced by the detail.

It was no news to Tetlock, therefore, that experts got beaten by formulas. But he does believe that he discovered something about why some people make better forecasters than other people. It has to do not with what the experts believe but with the way they think. Tetlock uses Isaiah Berlin's metaphor from Archilochus, from his essay on Tolstoy, "The Hedgehog and the Fox," to illustrate the difference. He says:

> Low scorers look like hedgehogs: thinkers who "know one big thing," aggressively extend the explanatory reach of that one big thing into new domains, display bristly impatience with those who "do not get it," and express considerable confidence that they are already pretty proficient forecasters, at least in the long term. High scorers look like foxes: thinkers who know many small things (tricks of their trade), are skeptical of grand schemes, see explanation and prediction not as deductive exercises but rather as exercises in flexible "ad hocery" that require stitching together diverse sources of information, and are rather diffident about their own forecasting prowess.

A hedgehog is a person who sees international affairs to be ultimately determined by a single bottom-line force: balance-of-power considerations, or the clash of civilizations, or globalization and the spread of free markets. A hedgehog is the kind of person who holds a great-man theory of history, according to which the Cold War does not end if there is no Ronald Reagan. Or he or she might adhere to the "actor-dispensability thesis," according to which Soviet Communism was doomed no matter what. Whatever it is, the big idea, and that idea alone, dictates the probable outcome of events. For the hedgehog, therefore, predictions that fail are only "off on timing," or are "almost right," derailed by an unforeseeable accident. There are always little swerves in the short run, but the long run irons them out.

Foxes, on the other hand, don't see a single determining explanation in history. They tend, Tetlock says, "to see the world as a shifting mixture of self-fulfilling and self-negating prophecies: self-fulfilling ones in which success breeds success, and failure, failure but only up to a point, and then self-negating prophecies kick in as people recognize that things have gone too far."

Tetlock did not find, in his sample, any significant correlation between how experts think and what their politics are. His hedgehogs were liberal as well as conservative, and the same with his foxes. (Hedgehogs were, of course, more likely to be extreme politically, whether rightist or leftist.) He also did not find that his foxes scored higher because they were more cautious—that their appreciation of complexity made them less likely to offer firm predictions. Unlike hedgehogs, who actually performed worse in areas in which they specialized, foxes enjoyed a modest benefit from expertise. Hedgehogs routinely over-predicted: twenty per cent of the outcomes that hedgehogs claimed were impossible or nearly impossible came to pass, versus ten per cent for the foxes. More than thirty per cent of the outcomes that hedgehogs thought were sure or near-sure did not, against twenty per cent for foxes.

The upside of being a hedgehog, though, is that when you're right you can be really and spectacularly right. Great scientists, for example, are often hedgehogs. They value parsimony, the simpler solution over the more complex. In world affairs, parsimony may be a liability—but, even there, there can be traps in the kind of highly integrative thinking that is characteristic of foxes. Elsewhere, Tetlock has published an analysis of the political reasoning of Winston Churchill. Churchill was not a man who let contradictory information interfere with his idées fixes. This led him to make the wrong prediction about Indian independence, which he opposed. But it led him to be right about Hitler. He was never distracted by the contingencies that might combine to make the elimination of Hitler unnecessary.

Tetlock also has an unscientific point to make, which is that "we as a society would be better off if participants in policy debates stated their beliefs in testable forms"—that is, as probabilities—"monitored their forecasting performance, and honored their reputational bets." He thinks that we're suffering from our primitive attraction to deterministic, overconfident hedgehogs. It's true that the only thing the electronic media like better than a hedgehog is two hedgehogs who don't agree. Tetlock notes, sadly, a point that Richard Posner has made about these kinds of public intellectuals, which is that most of them are dealing in "solidarity" goods, not "credence" goods.

Their analyses and predictions are tailored to make their ideological brethren feel good—more white swans for the white-swan camp. A prediction, in this context, is just an exclamation point added to an analysis. Liberals want to hear that whatever conservatives are up to is bound to go badly; when the argument gets more nuanced, they change the channel. On radio and television and the editorial page, the line between expertise and advocacy is very blurry, and pundits behave exactly the way Tetlock says they will. Bush Administration loyalists say that their predictions about postwar Iraq were correct, just a little off on timing; pro-invasion liberals who are now trying to dissociate themselves from an adventure gone bad insist that though they may have sounded a false alarm, they erred "in the right direction"—not really a mistake at all.

The same blurring characterizes professional forecasters as well. The predictions on cable news commentary shows do not have life-and-death side effects, but the predictions of people in the C.I.A. and the Pentagon plainly do. It's possible that the psychologists have something to teach those people, and, no doubt, psychologists are consulted. Still, the suggestion that we can improve expert judgment by applying the lessons of cognitive science and probability theory belongs to the abiding modern American faith in expertise. As a professional, Tetlock is, after all, an expert, and he would like to believe in expertise. So he is distressed that political forecasters turn out to be as unreliable as the psychological literature predicted, but heartened to think that there might be a way of raising the standard. The hope for a little more accountability is hard to dissent from. It would be nice if there were fewer partisans on television disguised as "analysts" and "experts" (and who would not want to see more foxes?). But the best lesson of Tetlock's book may be the one that he seems most reluctant to draw: Think for yourself. ✦