

Review of: Philip E. Tetlock. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton Univ. Press).

Adrian E. Tschoegl, Lecturer in Management, The Wharton School of The University of Pennsylvania
J. Scott Armstrong, Professor of Marketing, The Wharton School of The University of Pennsylvania*

Published in the *International Journal of Forecasting*, 2007, 339-342

Author: Philip E. Tetlock is a psychologist who is Professor of Leadership at the Haas School of Business at the University of California, Berkeley. The book combines several of his research interests such as how experts learn (or not) from experience and de-biasing judgment and choice to overcome common cognitive biases such as belief perseverance and over-confidence. In 2005 the American Political Science Association awarded *Expert Political Judgment* both the Woodrow Wilson Award for best book published on government, politics, or international affairs and the Robert E. Lane Award for best book in political psychology.

Study: Typically, researchers report new findings in scholarly journals and Tetlock (1998, 1999) has done so for some part of the findings of his study. Still, Tetlock has gone beyond journal articles, turning to a book to report on his large-scale and important study. Publishing a book has allowed him to deal with issues in detail and provide a full report, to reach a larger audience, and to have a stronger impact one hopes on experts and the consumers of their services, including political leaders and the voting public.

The book assaults common sense with evidence. In order to mount his assault on accepted wisdom, Tetlock spends more some 238 pages of text explaining his methods and findings, and considering and refuting many alternative explanations, and adds some 75 pages of technical appendices. The downside of this approach is that some readers may find the book too demanding. That would be a pity as his findings are important.

Tetlock's book reports the results of a two-decade long study of expert predictions. He recruited 284 people whose professions included "commenting or offering advice on political and economic trends." He asked them to forecast the probability that various situations would or would not occur, picking areas (geographic and substantive) within and outside their areas of expertise. In addition to eliciting forecasts, Tetlock also asked questions aimed at understanding how the forecasters came to formulate their forecasts, how they dealt with the failure of their forecasts, how they responded to contradictory information, and how they evaluated the probable accuracy of others' theories and predictions. By 2003, he had accumulated 82,361 forecasts, which provided him a database to evaluate. He then evaluated experts' predictions against outcomes, and against various alternate predictions that he derived from simple statistical procedures, from uninformed non-experts, and from well-informed non-experts.

Key Findings: The experts barely if at all outperformed informed non-experts and neither group of forecasters did well against simple rules and models. When Tetlock divided outcomes into three states, "Better", "Same", and "Worse", the forecasters frequently did worse than assuming that each state is equally likely.

* The authors would like to thank Alfred G. Cuzán, Paul Goodwin, Kesten C. Green, and Randall Jones for their helpful comments on earlier drafts of this manuscript.

That experts are no better forecasters than informed non-experts is not a new result. One of us (Armstrong 1980), based on a review of about a dozen small-scale studies, proposed the “seersucker theory” of forecasting: “No matter how much evidence exists that seers do not exist, suckers will pay for the existence of seers.” Beyond a low threshold, additional expertise did not result in greater accuracy in forecasting, and there was some evidence that those with the most expertise were less effective at using new information and thus less able to improve their accuracy.

In analyzing his respondents’ methods of arriving at their forecasts, Tetlock found it insightful to employ the metaphor that Isaiah Berlin (1953) took from the Greek poet Archilochus and used in his essay on Tolstoy: “The fox knows many things, but the hedgehog knows one big thing.” Foxes draw on many ideas and sources of information; hedgehogs interpret the world using their favorite theory or dogma. Foxes are more tolerant of ambiguity and uncertainty than hedgehogs, who tend to be confident in the rightness of their view of the world.

Tetlock found that foxes usually outperform hedgehogs; foxes produced more accurate forecasts than hedgehogs, though again neither beat simple rules. However, Tetlock reports that he was able to push more foxes than hedgehogs into forecasts that violated a fundamental axiom of probability: that the sum of a forecaster’s forecast probabilities not exceed one.

Tetlock argues that there is an inverse relationship between what works best in forecasting and what works best in the media. The fox may make the more accurate forecasts, but it is the dramatic, single-minded, combative hedgehog ideologue who makes the best TV, especially when matched against an opposite number in a point-counterpoint debate. A debate between foxes will dissolve into a miasma of nuance. It is not surprising then that Batchelor (2007) finds that “Private sector forecasters also have incentives to bias their forecasts towards optimism or pessimism,” so as to stand out from the crowd, though we believe this is not the key explanation for the existence of hedgehogs.

Understanding the Findings: One reason hedgehogs may survive despite being poorer forecasters than foxes is that consumers of the forecasts use both, but for different purposes. Users of forecasts should want to estimate risk, not just the most accurate point estimate of the outcome. The forecasts that hedgehogs provide enable the user to see the range of foreseeable outcomes. Tetlock’s results suggest that talking to both a pessimistic hedgehog and an optimistic one introduces little bias as the mean of the hedgehogs’ forecasts is little different from the mean of the foxes’ forecasts.

One of us (Tschoegl) worked for six years in Tokyo as a macroeconomist for Swiss Bank Corporation (SBC), generating economic forecasts of the Japanese economy for SBC’s fund manager clients in the UK, Europe, and elsewhere. He quickly realized that his clients talked to and read the reports of several forecasters from a variety of stockbrokers and research institutes, Japanese and foreign. The fund managers could not speak to everyone and read everything, so their problem was to find the smallest number of sources that would enable them to get both a mean and standard deviation. Although none used the following rule of thumb, it is not a bad characterization of the fund managers’ information gathering strategies:

Talk to four foxes, one pessimistic hedgehog and one optimistic hedgehog. Add the forecasts, divide by six to get the mean, and then use the mean and the six forecasts to calculate the standard deviation.

If the users of forecasts consult hedgehogs to establish the range of foreseeable outcomes, then as Tetlock found, hedgehogs should resist more strongly than foxes attempts to sway their views by the introduction of an alternative possibility to the one they are already propounding. The user of the forecasts does not want the hedgehogs to move their forecasts towards the consensus when introduced to another possible outcome. The hedgehog's job is to propose a particular, non-consensus forecast. Hedgehogs who drift towards the middle simply become redundant foxes.

Hedgehogs have the problem of maintaining credibility in the face of being wrong much of the time, and consequent vulnerability to hindsight bias. Ex post, whatever has happened seems more obvious than it was ex ante. When wrong, and hedgehogs will be wrong much of the time, they will seem stupider than they actually are. Why could they not "connect the dots", after all, many others did so? Equally, when correct, they will seem less clever than they in fact were; they were just lucky, anyone could have connected the dots had they just looked carefully.

The hedgehog's first defense is that of being right enough of the time to remain in the forecast user's set of preferred sources of forecasts. As Tetlock found, the hedgehogs occasionally got things very right. The successful hedgehogs maintain an optimal distance from the consensus: their forecasts, though not consensus, are within the plausible set and worth considering, in part because the foxes have some hits to their credit. These optimally different forecasts are also usefully interesting.

Experimental studies have shown that authentic dissent, such as that of the hedgehogs, is more effective than devil's advocacy. In their experiments, Nemeth *et al.* (2001) found that authentic dissenters (that is, people arguing for what they believe) were able to get people to focus more on opposing thoughts than supporting ones, thus contributing to more opinion change. One can expect this to lead to a better assessment of risks as the experiments of Koriati *et al.* (1980) demonstrate.

The effectiveness of the hedgehogs' authentic dissent stems from the users' realization that the hedgehogs are not just making up their arguments. Furthermore, those with authentic beliefs do a better job in arguing their positions than do those who are merely assigned the role. In Tschoegl's experience, it is not unusual for a user of forecasts to ask the forecaster, "Is that the house opinion or is it yours?" If the forecaster wants the user to consider his forecast seriously, he has to maintain it seriously; a forecaster who qualifies his forecast with the words, "I don't think this is really likely but you should think about it" will find his forecasts ignored.

Having used the hedgehogs' forecasts to develop risk estimates, the users of the forecasts can then look to those items where there is the greatest risk, and put their own efforts into thinking about them. If there is widespread disagreement on forecasts that matters, then the user needs to obtain additional forecasts or engage in contingency planning.

The hedgehogs' second defense is an unshakeable conviction in their own rightness. Ex ante, hedgehogs provide a simple, consistent story to support their forecasts and this can seduce the users of forecasts. However, because the role of hedgehogs paradoxically is to be usefully wrong much of the time, it is not surprising then that we find hedgehogs caught in error employing the various ex post defenses that Tetlock documents. Producing forecasts that leave him out on a limb will generate cognitive dissonance in the forecaster; one way for hedgehogs to cope with this is to explain away their errors; as Tetlock suggests, if they missed they can say that what they predicted "almost came to pass" or "hasn't come about yet". Lastly, after an event takes place, if they came close they can then tout the fact that they made a successful forecast that everyone else missed.

Implications: Tetlock's primary finding is that political experts are poor forecasters. He demonstrates this with a large sample of forecasts and with comparison to reasonable alternatives. At the very least, then, decisions makers should invest heavily in developing contingency plans as the forecasts by experts are of such little value.

What does help? First, there is safety in numbers, as long as the user of forecasts does this in an objective manner: forecasters should draw upon foxes as well as hedgehogs and combine their forecasts. However, putting experts in unstructured groups unaided by any formal techniques only makes things worse in that they become more confident without any gain in accuracy. Thus the automatic reaction to difficult problems of "let's form a committee or study group" of experts might make people feel better in the short run, but is unlikely to produce good forecasts (or policies).

Tetlock's findings should also lead politicians to seek other ways to forecast. Experts are not the only game in town. As Tetlock shows, simple models will also do better than experts. These simple models are not substitutes for experts; mechanically combining qualitative and quantitative forecasts provides a way to increase accuracy, even when one method alone is more accurate than the other, also taken alone.

References

Armstrong, J. S. (1980) The Seer-sucker theory: The value of experts in forecasting. *Technology Review* 83 (June-July), 16-24.

Batchelor, R. A. (2007) Bias in macroeconomic forecasts. *International Journal of Forecasting*, (forthcoming).

Berlin, I. (1953) *The Hedgehog and the Fox*. New York: Simon & Schuster.

Koriat, A., S. Lichtenstein & B. Fischhoff (1980) Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory* 6, 107-118.

Nemeth, C. J., J. B. Connell, J. D. Rogers & K. S. Brown (2001) Improving decision making by means of dissent. *Journal of Applied Social Psychology* 31 (1), 48-58.

Tetlock, P. E. (1998) Close Call Counterfactuals and Belief System Defenses: I Was Not Almost Wrong But I Was Almost Right. *Journal of Personality and Social Psychology* 75 (3): 639-652.

Tetlock, P. E. (1999) Theory-Driven Reasoning about Possible Pasts and Probable Futures in World Politics: Are We Prisoners of Our Preconceptions? *American Journal of Political Science* 43 (2): 335-366.