

Outline: Expert Political Judgment (Tetlock)

Oct 15, 2026 · @Brant

Expert Political Judgment: How Good Is It? How Can We Know? by Philip E. Tetlock (Princeton, 2005).

The core question. Can we measure how good experts are at political judgment? To find out, Tetlock tracked 284 professional forecasters (academics, government analysts, journalists and others) over roughly two decades. They produced 82,361 probability estimates on about 27,450 questions covering leadership changes, economic performance and national security. He judges them two ways: whether their forecasts matched what happened, and whether their reasoning hung together logically.

Front matter

- **Preface:** Where the project came from. The book asks two things: how good expert judgment is, and how we could even tell.

Ch. 1 — Quantifying the Unquantifiable

- Why grading political judgment is hard ("here lurk dragons"): people disagree about values, outcomes are vague, and experts can always explain a miss after the fact.
- Two ways to measure good judgment:
 - **Getting it right:** calibration (do events you call 70% likely happen about 70% of the time?) and discrimination (do you give higher odds to things that happen than to things that don't?).
 - **Thinking the right way:** logical consistency and updating your beliefs the way Bayes' rule says you should when evidence comes in.
- Isaiah Berlin's distinction is introduced. **Hedgehogs** know one big thing and push it everywhere. **Foxes** know many little things and improvise.

Ch. 2 — The Ego-deflating Challenge of Radical Skepticism

- **Radical skepticism** holds that good judgment is mostly luck. Two kinds:

- *Ontological skeptics* say the world itself can't be predicted: history is path-dependent, small causes have big effects, and chance piles up (illustrated with a step-by-step account of a plane crash).
- *Psychological skeptics* say our minds get in the way: we prefer simple stories, dislike ambiguity, need to feel in control, and misunderstand randomness.
- Six testable hypotheses, including "debunking," "diminishing returns," "fifteen minutes of fame," "hot air" and "seduction."
- **What the data showed:** experts barely beat dilettantes. They lost to simple statistical extrapolation models and sometimes to "dart-throwing chimps" (always guessing equal odds). Knowing more paid off less and less, and famous experts did worse.
- **The compromise, "skeptical meliorism":** track records were more consistent than luck would allow, so some ways of thinking really are better.

Ch. 3 — Foxes Have Better Calibration and Discrimination Scores than Hedgehogs

- *Who* experts were made almost no difference: education, field, years of experience and access to classified information didn't predict accuracy.
- *What* they believed made no difference either: left vs. right, realist vs. institutionalist, doomster vs. boomster.
- **How they thought did matter.** Foxes beat hedgehogs, most clearly on long-range forecasts inside their own specialty.
- What foxes do differently, shown through case studies (Russia, India, Kazakhstan, Poland, Saudi Arabia, the last Communist "dominoes," causes of war):
 - They doubt sweeping laws of history and are wary of neat historical analogies.
 - They're less carried away by their own rhetoric.
 - They're less quick to judge people in the past harshly.
 - They keep political passions in check.
 - They work to reconcile conflicting views, for example on "when do leaders matter?"
- Foxes predicted fewer dramatic breaks from the status quo but rarely ruled anything out entirely.

Ch. 4 — Foxes Are Better Bayesians than Hedgehogs

- A logical-consistency test and a test of belief updating, using the experts' own "reputational bets."
- Everyone updated too little after being wrong, and hedgehogs updated least.
- **How experts defended failed forecasts:** "off on timing," "close call," "an outside shock intervened," "the other side's mistake," "it was the right mistake to make." Bad luck was blamed for misses far more often than good luck was credited for hits.
- **Hindsight bias:** experts remembered their earlier forecasts as closer to what actually happened.

Ch. 5 — Foxes Are More Willing than Hedgehogs to Entertain Self-subversive Scenarios

- Experts judged "what if" alternative histories: the USSR, the end of apartheid in South Africa, the rise of the West, World War I, the outcomes of both world wars, why the Cold War stayed cold.
- Hedgehogs held **double standards**. They welcomed "it almost happened" arguments when those saved their own forecasts, but rejected the same reasoning when it threatened their favorite lessons from history.
- **Turnabout tests** (swapping which side the evidence favors): hedgehogs judged unwelcome evidence more harshly, and were less willing than foxes to admit it.

Ch. 6 — The Hedgehogs Strike Back

- **"Not such bad forecasters after all":** Tetlock re-scores the forecasts with adjustments for which mistakes cost more, how hard the questions were, disputed outcomes, and partial credit for near misses. These narrow the gap but don't close it.
- **"Not closed-minded beyond hope":** the hedgehogs' rebuttals on double standards and hindsight bias; complaints that the questions were wrong, the wrong experts were asked, or the scoring misread what experts are actually trying to do.
- The **home-run hitter** argument: bold hedgehogs strike out a lot but land the visionary calls.
- **Verdict:** some reasonable doubt, but hedgehogs made too many errors across too many topics to be cleared.

Ch. 7 — Are We Open-minded Enough to Acknowledge the Limits of Open-mindedness?

- Scenario exercises about the futures of Canada and Japan. Breaking a possible future down into detailed scenarios inflates its probability, so the estimates add up to well over 100%.
- Debiasing how people think about the past: the Cuban Missile Crisis and "unmaking the West" experiments. Breaking down alternative histories reduces hindsight bias, but again produces probabilities that don't add up.
- **The twist:** foxes are *more* prone to this confusion. The hedgehog's failing is overconfidence; the fox's is seeing too much merit in too many stories.
- Good judgment turns out to be a skill of watching your own thinking, the "art of self-overhearing": balancing what you already know against exploring new possibilities.

Ch. 8 — Exploring the Limits on Objectivity and Accountability

- A staged debate between four philosophical positions: an unrelenting relativist, a hard-line neopositivist, a moderate neopositivist and a reasonable relativist.
- Tetlock lands on a hybrid. There will always be room to argue about who got it right, but rival views can still be held to common tests.
- **Improving policy debate** (Tetlock writing as a citizen rather than as a scientist):
 - The marketplace of ideas fails because audiences rely on cues like fame and confidence that don't track accuracy, and because they mostly want their biases confirmed.
 - He calls for public forecasting track records, beliefs stated in testable form, and honoring reputational bets. This anticipates his later Good Judgment Project and *Superforecasting*.

Appendices

- **Methodological Appendix:** who the participants were, the question types (leadership continuity, economic policy, security), scoring rules, and the procedures for each experiment in Chapters 4, 5 and 7.
- **Technical Appendix** (with Phillip Rescober): the math behind the Brier/probability score and how it breaks into calibration and discrimination; normalized discrimination; the baseline competitors; the scoring adjustments; formal tests of logical consistency.

The book in one sentence

Experts forecast only a little better than chance, and how they think (fox vs. hedgehog) predicts accuracy far better than who they are or what they believe. But too much open-mindedness has its own costs.