



## **REVOLUTIONIZING THE ASSESSMENT OF ALPHA-PUNDITS: THE SUPERFORECASTER CHALLENGE**

**Philip E. Tetlock  
Wharton School, University of Pennsylvania**

Between 2011 and 2015, the US intelligence community sponsored a series of geopolitical forecasting tournaments (Tetlock and Gardner, 2015). The winners, the so-called "superforecasters," were talented amateur/generalists who not only beat their public competitors in research universities; they even out-performed intelligence analysts making predictions on the same questions, but with access to classified information. How superforecasters pulled this off offers insights into how the intelligence community itself could do a better job of assigning realistic odds to important outcomes in the geopolitical and economic realms.

Is it possible to extend these methods to improve the quality of public debate and incentivize pundits to make accurate probability judgments--in lieu of the usual partisan blame-game incentives?

We will never know until we try—and that is why GJP is launching the alpha-pundit challenge. The goal is to assess the relative accuracy of GJP superforecasters and alpha-pundits on the high-stakes issues on which these pundits opine.

But how might this be done? We currently have no method of scoring which alpha-pundits got what right because alpha-pundits rarely make testable claims. They make vague verbiage forecasts that can later be spun as correct no matter what happens. Anything "could" happen.

Fortunately the Good Judgment Project (GJP), a public-sector entity, and its private-sector spin-off Good Judgment Inc have found a reliable scientific solution. Whenever alpha-pundits balk at making testable claims--like an 80% chance of -2% or worse global deflation in 2016—GJP's ideologically balanced panels of intelligent readers will make good-faith inferences about what the pundits meant. Using all the textual clues available, what is the most plausible interpretation of "serious possibility" of global deflation? GJP will then publish the

readers' estimates and of course invite alpha-pundits to make any corrections if they feel misinterpreted.

GJP superforecasters will also make predictions on the same issues. And the match will have begun—indeed it has already begun.

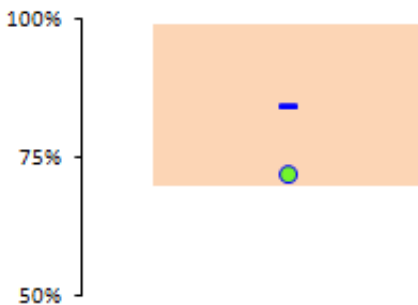
Here are reader-translated claims by four alpha-pundits on four big issues—as well as superforecaster's probability judgments. We can see that the range of possible meanings of vague verbiage forecasts is often large—sometimes as wide as 20% to 80% with a phrase like “distinct possibility”—which underscores the “heads-I-win-tails-I-don't lose” function of vague verbiage. To keep things simple, GJP uses the middle-of-the-range value.

### Superforecasting with the Pundits

■ Implied Range of Forecast ■ Probability Mid-Point ● Superforecasters

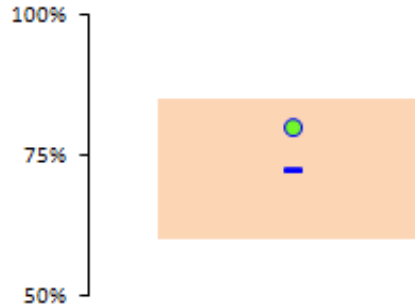
**Larry Summers:**

"The risks **tilt heavily** toward inflation rates below official targets."



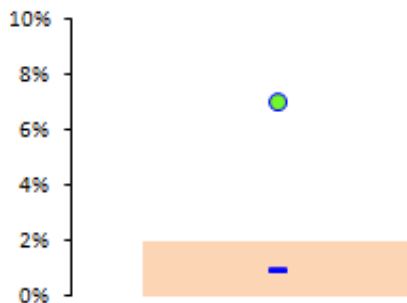
**Paul Krugman:**

"Does today's good job report mean that the Fed will raise rates next month? **Probably yes.**"



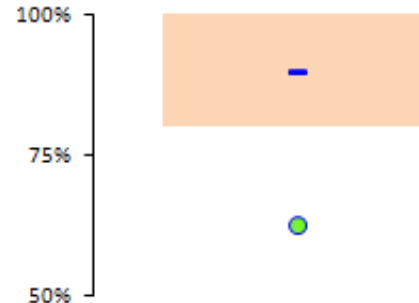
**Peggy Noonan:**

"I speak of [Jeb Bush's] candidacy in the past tense... It's just **hard** to see how this can work. By hard I mean, for me, **impossible.**"



**Tom Friedman:**

"Putin stupidly went into Syria looking for a cheap sugar high ... now he's up a tree... The **only way** Putin can get down from that tree is with our help in forging a political solution in Syria."



For instance, former Treasury Secretary Larry Summers recently published an important essay on global secular stagnation in the *Washington Post* which included a series of embedded forecasts, such as this prediction about inflation and central bank policies: "The risks *tilt heavily* toward inflation rates below official targets." It is a catchy verbal salvo, but just what it means is open to interpretation. Our panel assigned a range of 70–99% to that forecast, centering on 85%. When asked that same question, the Superforecasters give a probability of 72%. These precise forecasts can now be evaluated against reality.

As summarized in the chart, we did the same exercise with recent think pieces by Paul Krugman, Tom Friedman, and Peggy Noonan to show the value of converting vague statements into testable forecasts for important issues in public discourse.

What we propose is new, even revolutionary, and could with proper support evolve into a systemic check on hyperbolic assertions made by opinion makers in the public sphere. It is rigorous, empirical, repeatable, and backed by the widely-recognized success of the Good Judgment Project based at the University of Pennsylvania.

Of course, we will need to aggregate many judgments of many forecasters to assess the relative accuracy of pundits versus superforecasters. This project is a long-haul effort. But it offers long-term hope. We don't just have to sit idly by and bemoan the polarization of unnecessarily polarized debates. There are tangible steps that can be taken to increase the supply of an essential public good in democracies: thoughtful empirically grounded debates over consequences of policy options.

GJP and GJ Inc. are eager to collaborate with major foundations and media outlets interested in implementing these ideas with us. Those who embrace the ideas first will also advance their own long-term self-interest. They will co-own the sites that hold the punditocracy—left, center and right—accountable for the claims they make, sites to which an informed citizenry can turn to spot pundits who value polemical posturing over analytical accuracy.