Good Judgment Inc

FEDERAL FORESIGHT COMMUNITY OF INTEREST

WASHINGTON DC

21 April 2017

GOOD JUDGMENT

Let's start with a forecast

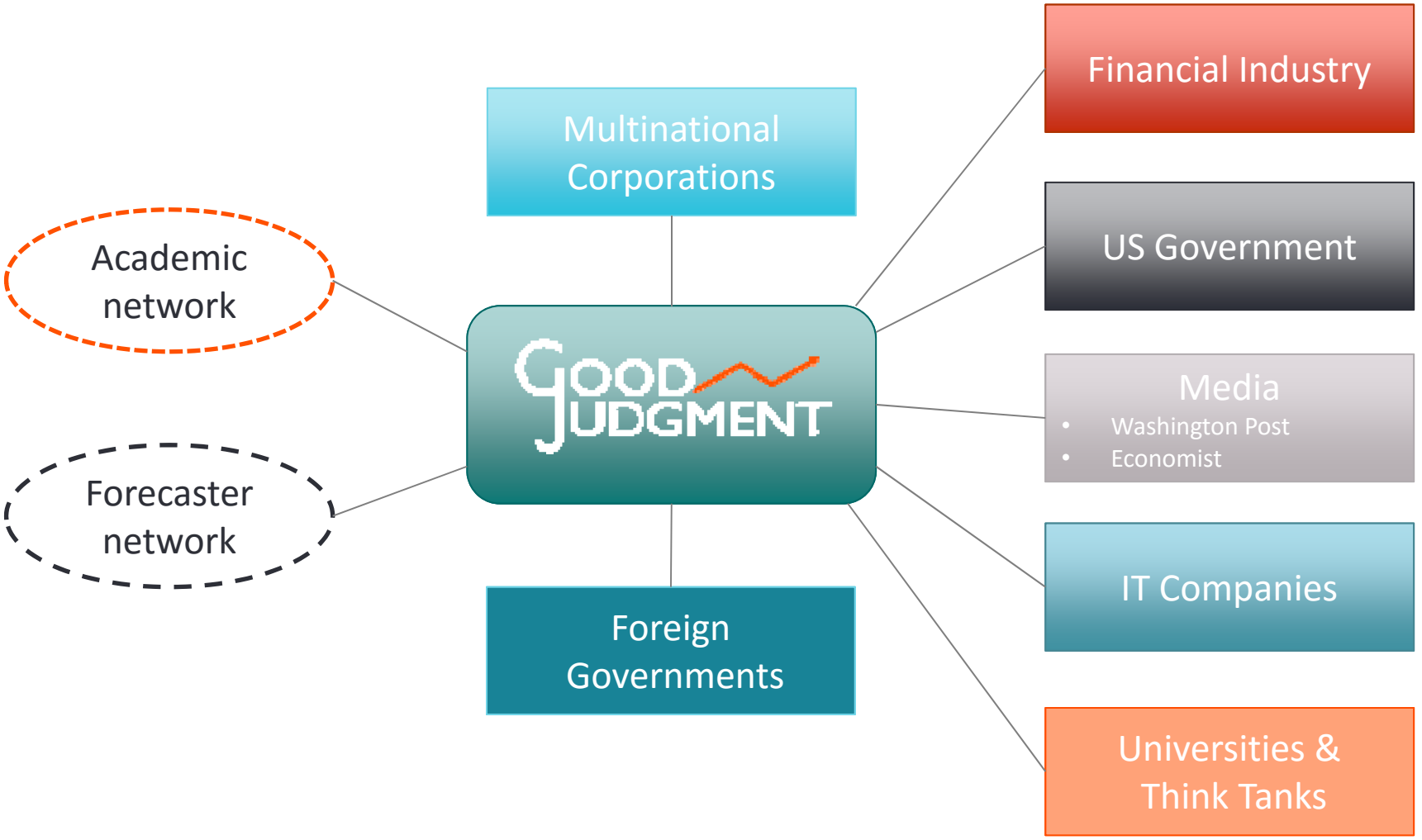
Will North Korea conduct a nuclear test before 30 June 2017?

GJP Set a New Standard for Forecasting Accuracy

"Team Good Judgment, led by Philip Tetlock and Barbara Mellers of the University of Pennsylvania, beat the control groups by more than 50%. This is **the largest improvement in judgmental forecasting accuracy** observed in the literature."

Steve Rieber, Program Manager, US Intelligence Advanced Research Projects Activity (IARPA), 30 June 2015

Percent improvement above benchmark by year


Financial Industry
Multinational Corporations
US Government
Academic network
Media
Washington Post
Economist
Forecaster network
IT Companies
Foreign Governments
Universities & Think Tanks

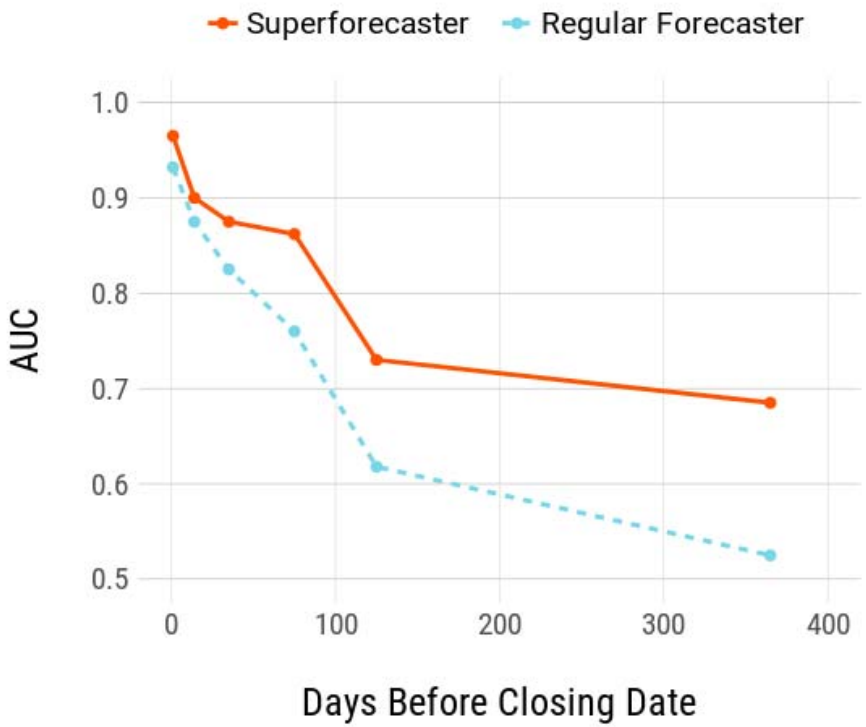
It's possible to get better … a lot better

Superforecasters saw things 400 days out as well as regular forecasters 150 days out. On the chart, a reading of 1 is perfect foresight while 0.5 is pure chance.

The AUC … corresponds to the area under the ROC curve traced by the hit and false alarm rates as the decision threshold varies. … The results clearly show that the advantage of superforecasters is especially salient on longer range forecasting problems.

– Mark Steyvers, GJP Year 4 Annual Report to IARPA

Superforecasters are more accurate, earlier



How much is accuracy worth?

Say you're going to have 10 BBQs this summer. You spend $100 for each BBQ on fresh steak tips from the butcher. You will have the BBQ on Saturday or Sunday, and it will rain on one of those days, but you need to invite your friends by Tuesday.

	Tuesday Forecast	Weekend Result	Butcher Bill	Total Lost to Rainout
Regular Forecaster	50%	Half of the summer BBQs get rained out	$1,000	$500
Superforecaster	90%	One out the ten BBQs is rained out	$1,000	$100

Value added from Superforecasters = Total savings of $400

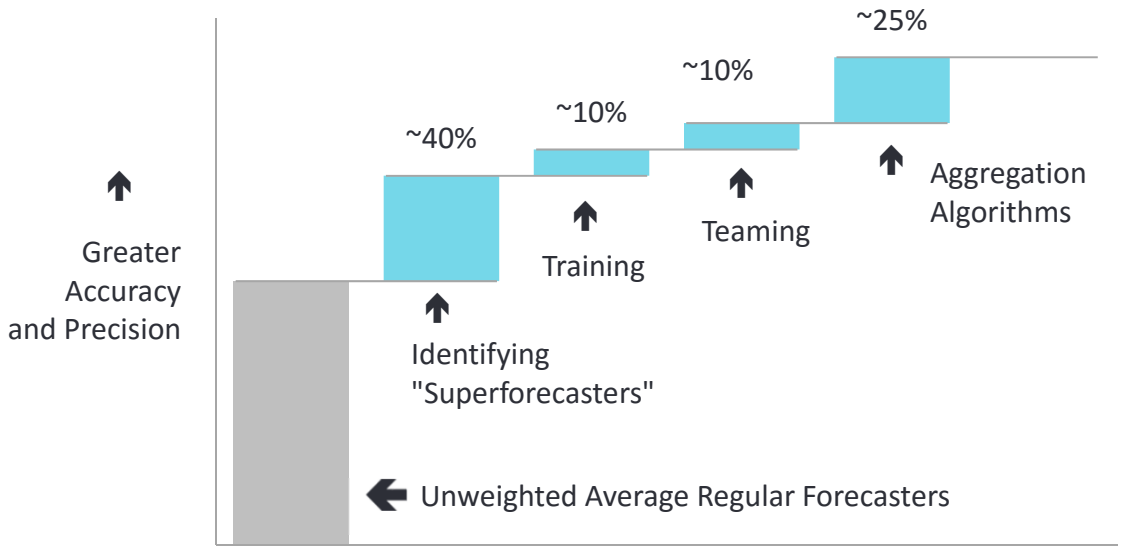
Ditto for retailers planning for Christmas, supply chain managers anticipating political disruptions, FFCOI members projecting federal budget outlays …

Five teams competed in a 4-year US government-run project. Good Judgment eliminated the other teams after the second year.

Thousands of forecasters made more than 1 million forecasts in a variety of forecasting conditions.

GJ developed advanced techniques to elicit, weight, and combine judgments.

Good Judgment's Winning Evidence-Based Four-Factor Formula to Boost Accuracy



What qualities make for a good forecaster ?

Pattern Recognition

Open to New Experiences

Numeracy

Open Mindedness

Conscientiousness

Aversion to Ambiguity

Grit/Determination

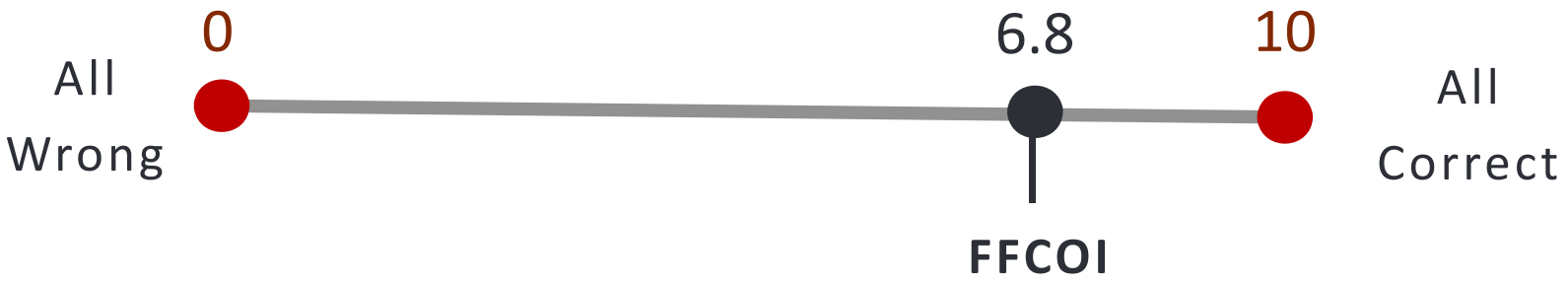
Cognitive Reflection

Knowledge Base

Superforecaster Traits

"Abstract Reasoning"

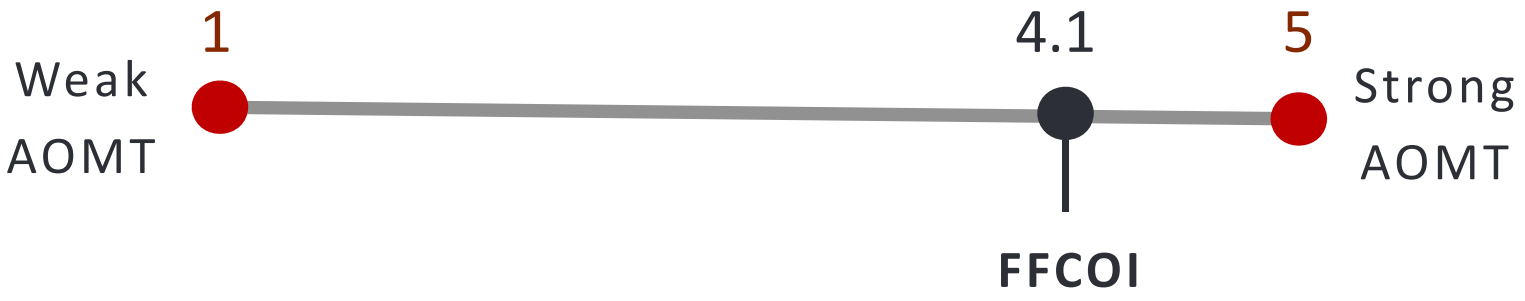
Superforecasters were identified as having superior abstract reasoning when compared to our control group. We used 10 puzzles with shapes to identify how people apply logic. The puzzles got progressively more difficult.



Superforecaster Traits

"Actively Open-Minded Thinking"

Research on forecasting by Good Judgment found that people who consider evidence that contradicts their initial beliefs are more accurate. These AOMT individuals tend to engage in more critical thinking and are less likely to hunker down on a losing position.



Superforecasting Checklist

1 Base Rates: Start with the Outside View and ***then*** add the Inside View

2 Confidence Check: Think in terms of 90% confidence bands around a forecast, with wider bands where you have less certainty

3 Scope: Consider the probability distribution across different outcomes than posed by the question, such as longer/shorter timeframes

4 Diagnostics: Evaluate the strength of new evidence to avoid over- and under-reacting. Watch for conjunction fallacies, sample size, etc

5 Comments: Explain the reasons for a forecast to head off later hindsight bias. Rigorous post-mortems turbo charge the feedback that improves calibration

1 Base Rates: Reference Classes

One of Good Judgment's most important lessons, and generally a good starting point for making an initial forecast, is taking the inside and outside views on a question.

People often neglect the outside view.

Inside View

- Most people ask themselves: How happy is this couple?
- Then they translate that feeling into a probability estimate.

Outside View

- Smart forecasters first ask: How often do couples in this demographic divorce?
- If 40% of marriages end in divorce, the smart forecaster starts there and adjusts appropriately for idiosyncratic factors.

Will North Korea conduct a nuclear test before 30 June 2017?

- **BASE RATES**
- CONFIDENCE
- SCOPE
- DIAGNOSTICS
- COMMENTS

How many tests has North Korea conducted?

How often?

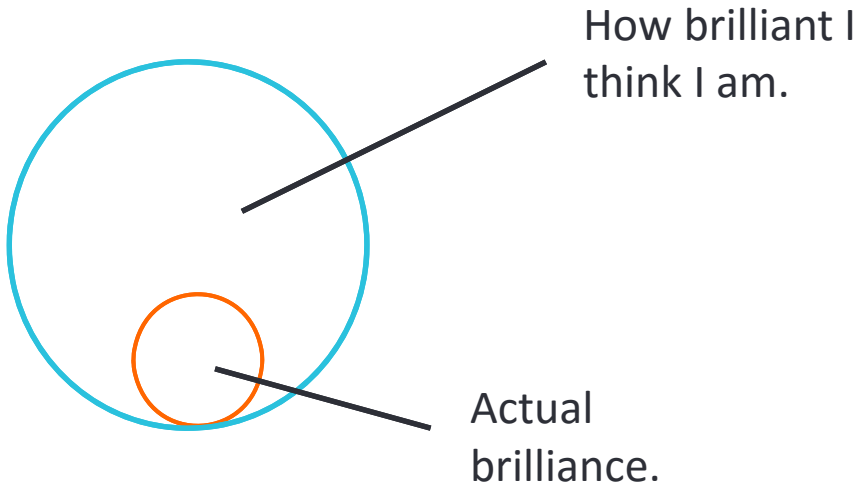
Is the frequency rising/falling/immaterial?

2 Calibrating Confidence

The overconfidence effect measures how well people really know what they think they know.

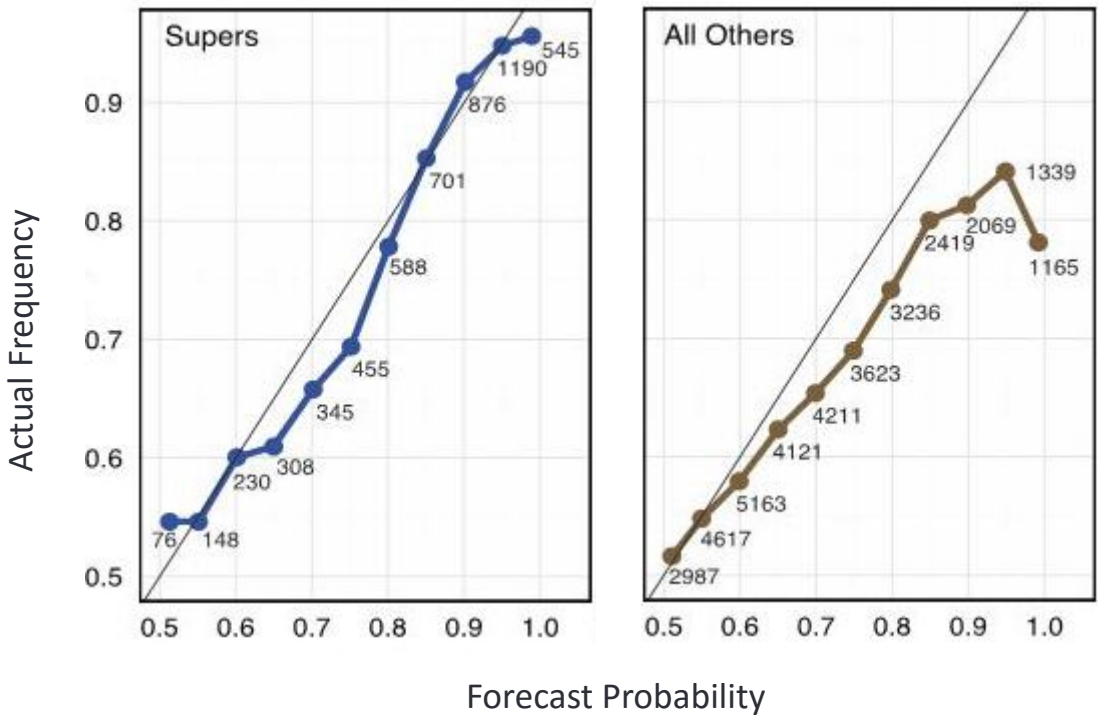
Most people tend to be over-confident about their abilities, their strengths relative to others, and in the precision of their knowledge.

Setting a confidence interval can help strike a better balance, with wider bands to offset over-confidence.



The Good Judgment Process Strives for Accuracy

Superforecasters deliver **consistently accurate** probability estimates.



Will North Korea conduct a nuclear test before 30 June 2017?

BASE RATES

CONFIDENCE

SCOPE

DIAGNOSTICS

COMMENTS

How much do you know about North Korea? About nuclear testing?

Are you sure?

What's the strongest counterargument?

3 Scope Sensitivity

Humans tend be inconsistent when it comes to scope and scale.

In this study, people were randomly assigned to see what they would pay to rescue migrating birds affected by an oil spill.

When they were asked how much they were willing to pay, the total amount remained roughly the same regardless of how many birds would be saved.

"How much would you pay to protect the lives of … "

"2,000 migrating birds?"	$80
"20,000 migrating birds?"	$70
"200,000 migrating birds?"	$88

Source: Desvousges et al

Will North Korea conduct a nuclear test before 30 June 2017?

- BASE RATES
- CONFIDENCE
- **SCOPE**
- DIAGNOSTICS
- COMMENTS

What is your forecast for different time-frames? Does it make a difference at 6 months? At 2 years?

Is it more likely sooner?

Or later?

Diagnostics: Conjunction Fallacy

Most people give higher odds in response to the second question because they've fallen for the more vivid narrative in the "conjunction fallacy."

1. What are the odds of a 30% selloff in the S&P 500 in the next year?

2. What are the odds of a 30% selloff in the S&P 500 following a wave of defaults in high yield energy bonds in the next year?

Diagnostics: Sample Size

Our brains look for patterns and sometimes see patterns without sufficient evidence.

In this example, most people answer a. because there is more variance in the order and frequency of 0 and 1.

The answer is both because the small sample size doesn't allow for proper judgment of pattern.

A random number generator produced a sequence of 0s and 1s with the same likelihood.

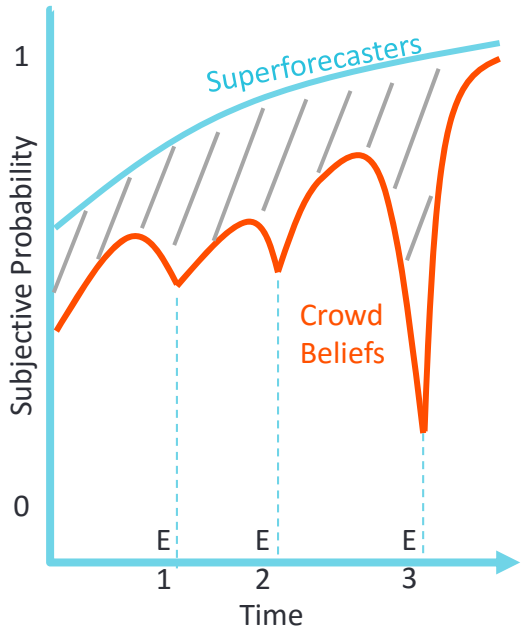
Which one of these sequences is more likely to come from the generator:

a. 0-0-1-1-0-1-0-1-0-1

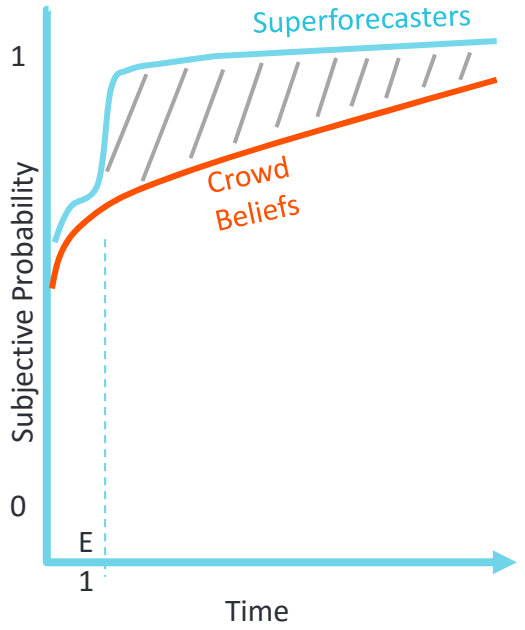
b. 1-1-1-1-1-0-0-0-0-1

Superforecasters **avoid over- and under-reacting** to new information.

- Discounting pseudo-diagnostic news to which crowd over-reacts



Spotting subtly-diagnostic news to which the crowd under-reacts



Will North Korea conduct a nuclear test before 30 June 2017?

BASE RATES

CONFIDENCE

SCOPE

DIAGNOSTICS

COMMENTS

How reliable are media reports? The experts? Other sources?

Is there a variety of sources with different perspectives pointing to a similar conclusion?

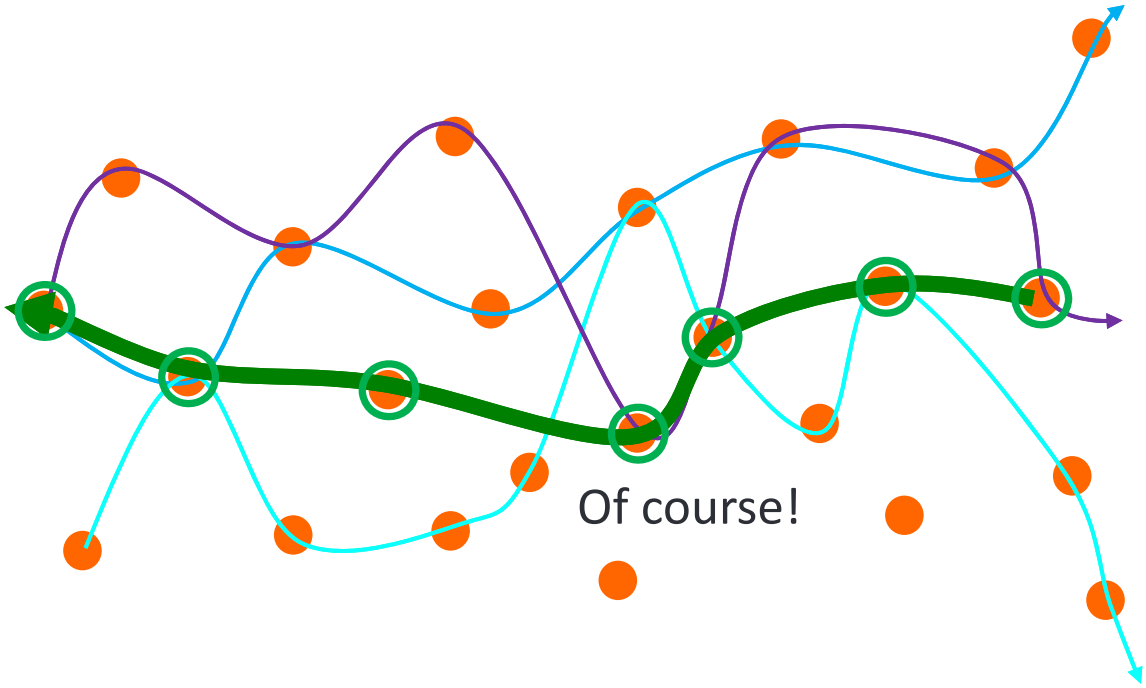
What would get you to change your mind?

#5 Comments to Prevent Hindsight Bias

Once we learn of an outcome, x, we mobilize cognitive resources to explain why x had to occur when and how it did.

We lose contact with how indeterminate the world once felt (and exaggerate degree to which we saw it coming all along).

So it is critical to write comments as you update forecasts to compare your reasoning to the outcome in post-mortem reviews.



Will North Korea conduct a nuclear test before 30 June 2017?

Make regular comments and read others' comments

Do a Pre-post-mortem: If you are proved "wrong," why might that be?

Then do a post-mortem when the question is resolved and take the time to review your comments. Were you:

- Right for the right reasons?
- Right for the wrong reasons?
- Wrong for the right reasons? ...

Will North Korea conduct a nuclear test before 30 June 2017?

BASE RATES

CONFIDENCE

SCOPE

DIAGNOSTICS

COMMENTS

UPDATE !!

Has your forecast changed?

Good Judgment Open: 57%

Superforecasters: 59%

The Future: Human/Machine Collaboration

Outside View:

Big Data

Inside View:

Subject Matter Experts

Synthesis:

Superforecasters

GOOD JUDGMENT

For additional information:

Michael Story
story@goodjudgment.com