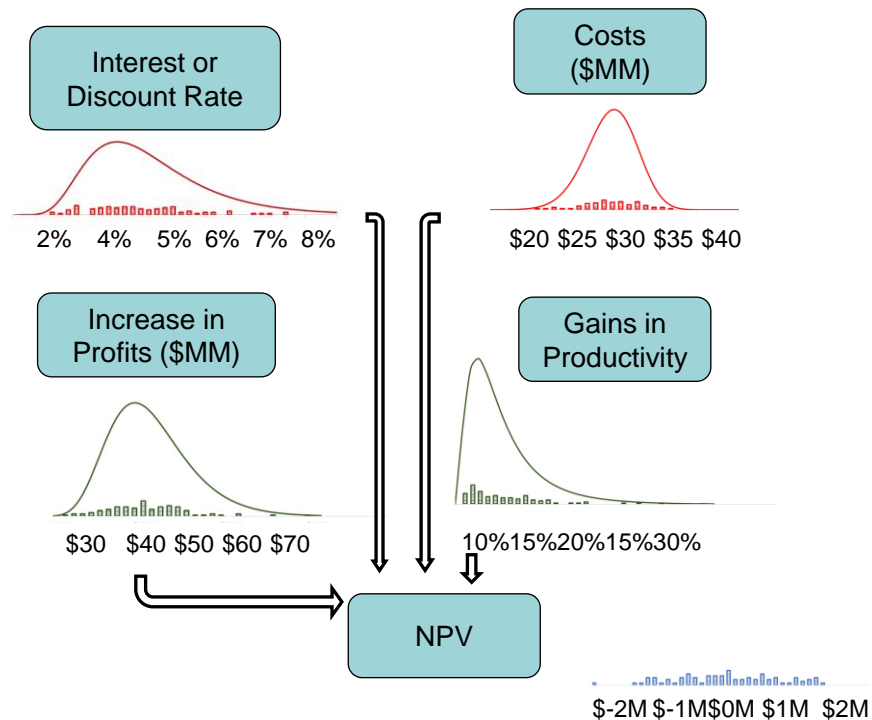# Assessing Subjective Probabilities

Using Subject Matter Experts to Estimate Uncertainty

Hubbard Decision Research
2 South 410 Canterbury Ct
Glen Ellyn, Illinois 60137
www.hubbardresearch.com

We Need Expert Estimates

The Monte Carlo Simulation

SSCAG/SCAF/EACE Joint International Conference (2008)

An Assessment of the Inherent Optimism in Early Conceptual Designs and Its Effect on Cost and Schedule Growth

D. Bearden, C. Freaner, R. Bitten, and D. Emmons

Abstract

When missions experience cost growth, cost estimators are often criticized for underestimating the cost of missions in the early conceptual design stage. The final spacecraft and instrument payload configuration at launch, however, can be significantly different as the project evolves, thereby leading to cost "growth" as compared to these lower initial estimates. In order to make a more robust initial estimate, historical mass, power, data rate, and growth

Global Risk Management Day by Copenhagen Compliance®
March 11, 2021, Online Conference

Expressing Uncertainty

© Hubbard Decision Research, 2020
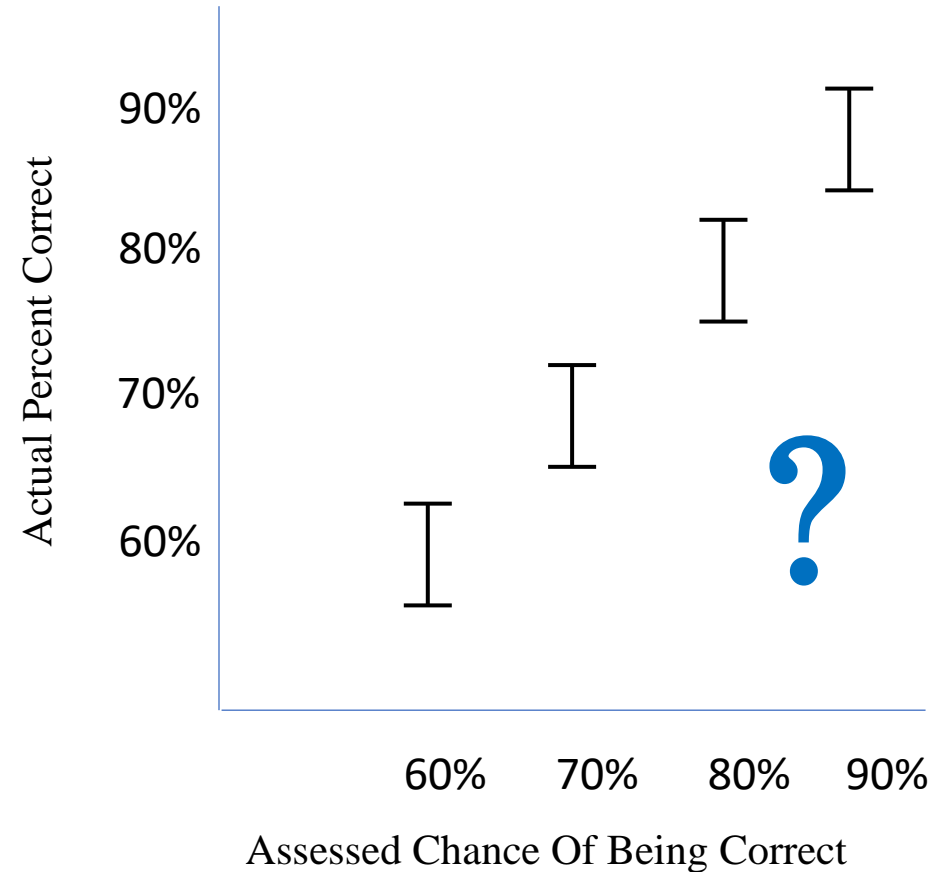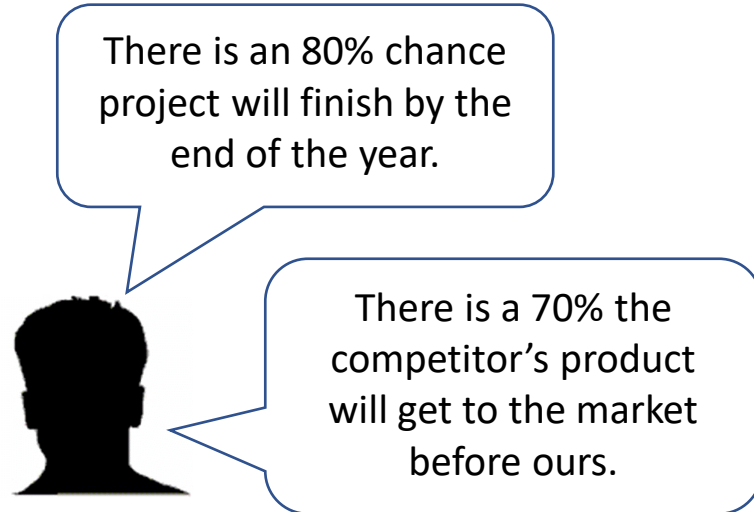
Overconfidence

Expert Intuition

How do I know my current uncertainty?

"Overconfident professionals sincerely believe they have expertise, act as experts and look like experts. You will have to struggle to remind yourself that they may be in the grip of an illusion."

Daniel Kahneman, Psychologist, Economics Nobel

- Studies also show that measuring *your own* uncertainty about a quantity is a general skill that can be taught with a **measurable** improvement.

- HDR has calibrated over 1,600 people in the last 22 years.

- 85% of participants reach calibration within a half-day of training.

Global Risk Management Day
by Copenhagen Compliance®
March 11, 2021, Online Conference

Before and After Calibration Training
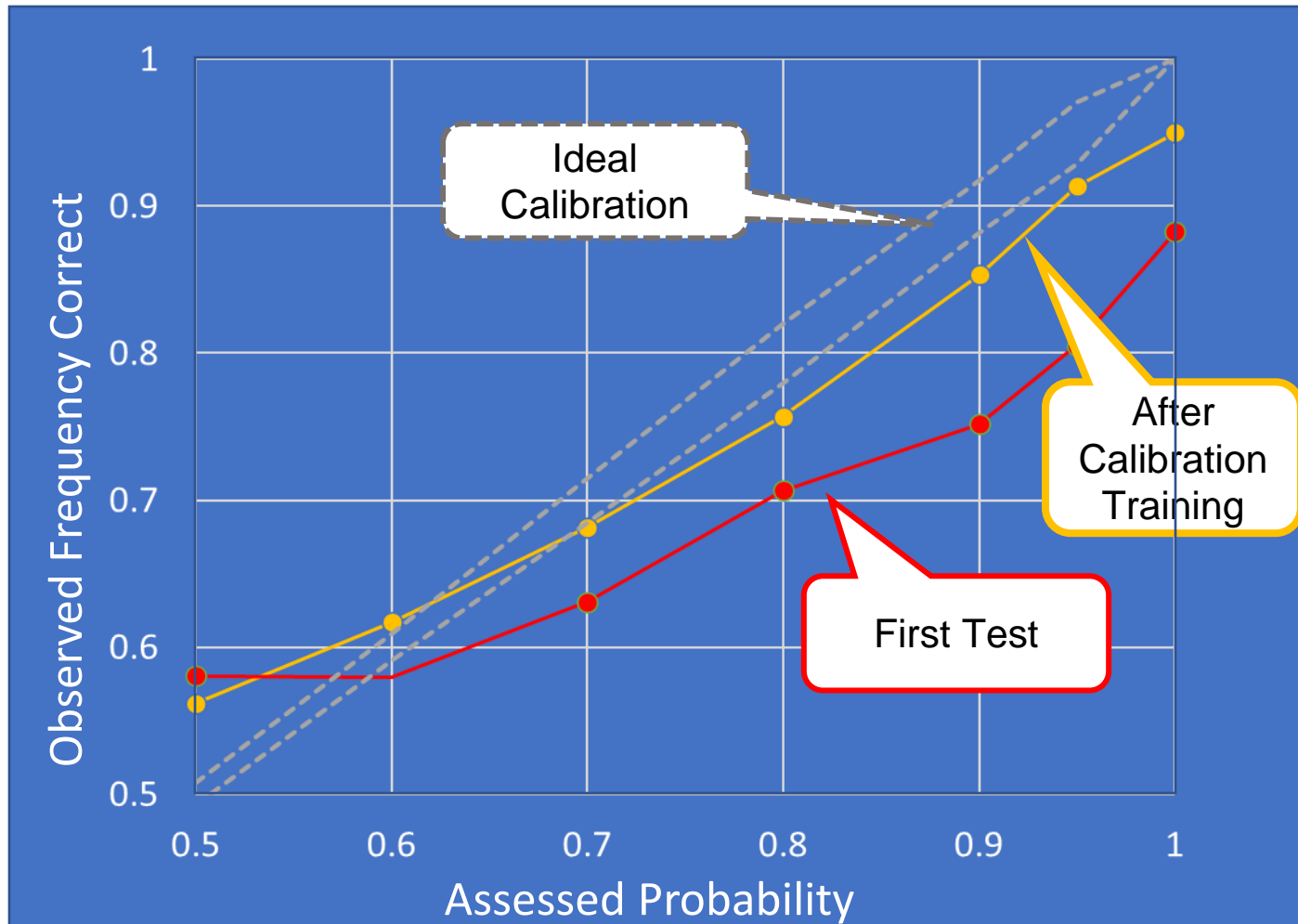
- Of the 1,600+ individuals we calibrated, we compiled the tests of 434 of the most recent training sessions totaling over 52,000 individual test responses.

- The first benchmark test showed, as expected, that participants were very overconfident.

© Hubbard Decision Research, 2020

Calibration Exercises

90% Confidence Interval Question	Lower Bound	Upper Bound
Mozart was born what year?		

True/False Question	True or False?	% Confidence
A hockey puck will fit in a golf hole.		

Global Risk Management Day
by Copenhagen Compliance®
March 11, 2021, Online Conference

Calibration Aid: "The Equivalent Bet"

For 90% Confidence Interval questions, which game would you rather play?

- **Game A**: Win $1,000 if your interval contains the correct answer
- **Game B**: Spin a dial with a 90% chance to win $1,000

For the Binary Confidence questions, which game would you rather play?

- **Game A**: Win $1,000 if your answer is correct
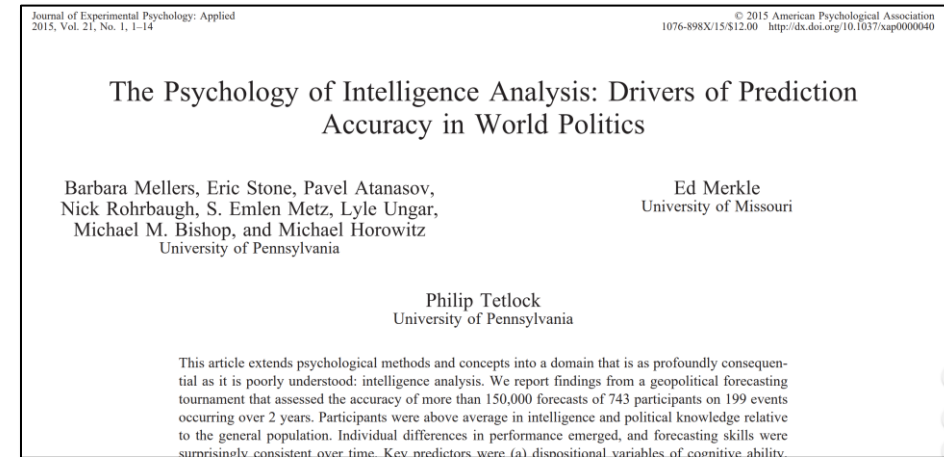- **Game B**: Spin a dial with a chance to win $1,000 equal to your stated confidence

Game B:

Win $0

Win $1,000

Spin the Dial!

Global Risk Management Day
by Copenhagen Compliance®
March 11, 2021, Online Conference

Three Ways to Make Better Forecasts

Philip Tetlock tracked over 150,000 forecasts from 743 experts in world affairs. He determined factors that made the biggest difference in the performance of forecasting.

The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics

Barbara Mellers, Eric Stone, Pavel Atanasov, Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar, Michael M. Bishop, and Michael Horowitz
University of Pennsylvania

Ed Merkle
University of Missouri

Philip Tetlock
University of Pennsylvania

This article extends psychological methods and concepts into a domain that is as profoundly consequential as it is poorly understood: intelligence analysis. We report findings from a geopolitical forecasting tournament that assessed the accuracy of more than 150,000 forecasts of 743 participants on 199 events occurring over 2 years. Participants were above average in intelligence and political knowledge relative to the general population. Individual differences in performance emerged, and forecasting skills were surprisingly consistent over time. Key predictors were (a) dispositional variables of cognitive ability,

1. Training: Calibrated probabilities basic probabilistic thinking

2. Aptitude: Measured by tests and past performance

3. …

Global Risk Management Day by Copenhagen Compliance®

March 11, 2021, Online Conference

Before and After Calibration Training

- Of the 1600+ individuals calibrated, we compiled the tests of 434 of the most recent training sessions totaling over 20,000 individual test responses.

- The first benchmark test showed, as expected, that participants were very overconfident.

© Hubbard Decision Research, 2020
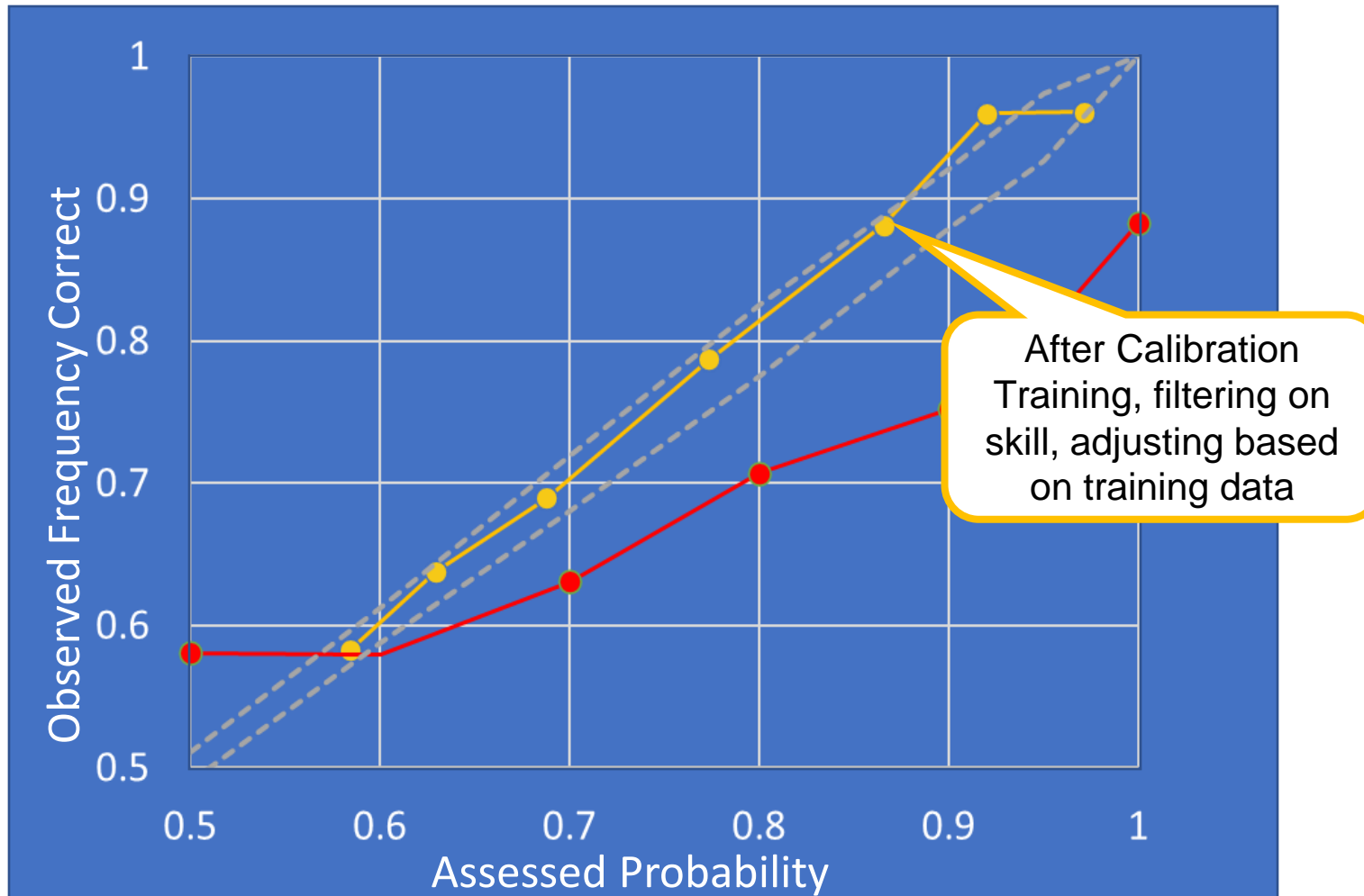
Adjustments

- By dividing the questions into two sets we could use one set to provide adjustments to answers on a different set of questions.

- If we used the adjusted estimates instead of their stated estimates, the results are almost perfectly calibrated.

© Hubbard Decision Research, 2020

The Lens Method: Reducing Inconsistency

Methods that statistically "smooth" estimates of experts show reduced error in several studies for many different kinds of problems.
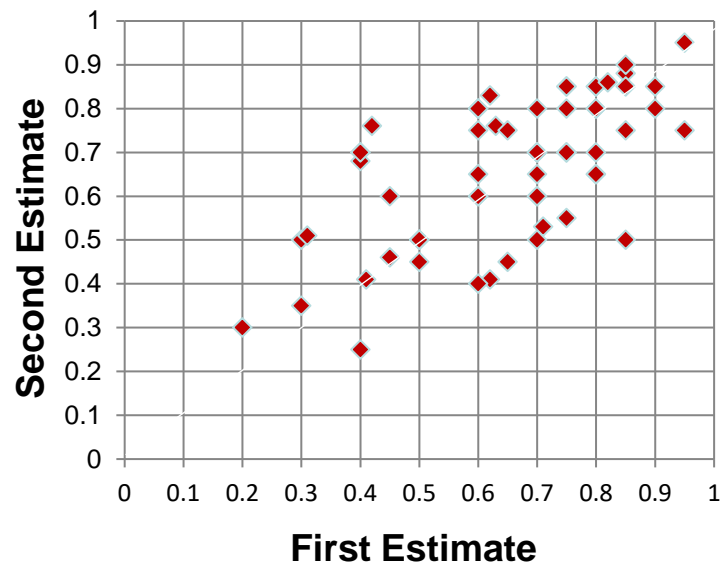
In Cybersecurity, SME inconsistency accounts for 21% of variation in estimates.

Three Ways to Make Better Forecasts

Philip Tetlock tracked over 150,000 forecasts from 743 experts in world affairs. He determined factors that made the biggest difference in the performance of forecasting.

The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics

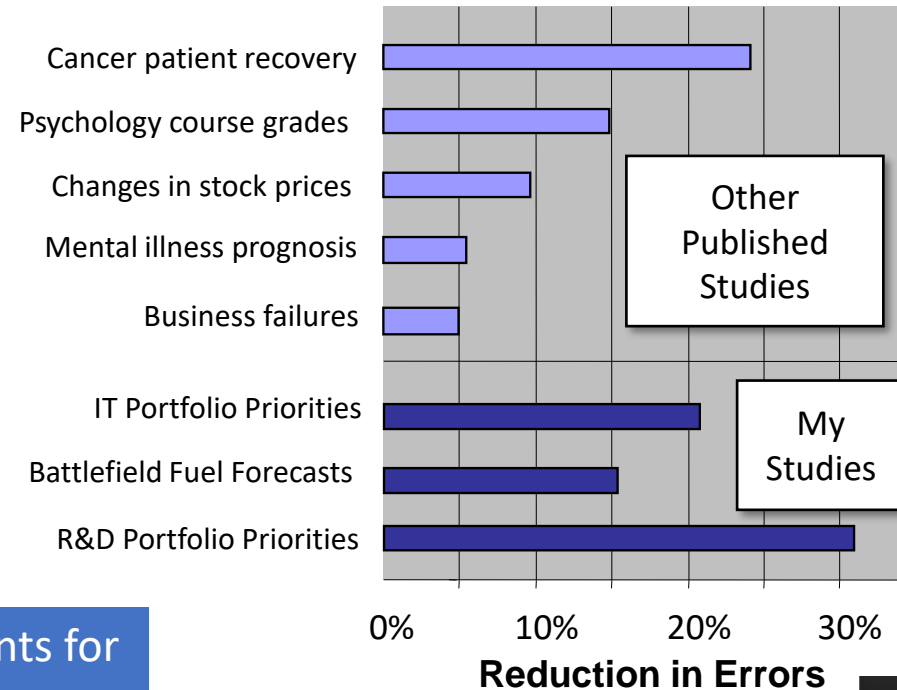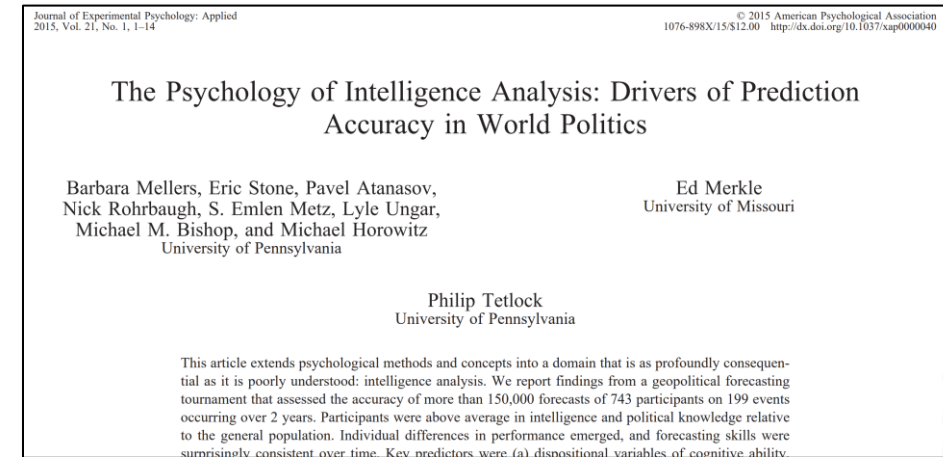Barbara Mellers, Eric Stone, Pavel Atanasov, Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar, Michael M. Bishop, and Michael Horowitz
University of Pennsylvania

Ed Merkle
University of Missouri

Philip Tetlock
University of Pennsylvania

This article extends psychological methods and concepts into a domain that is as profoundly consequential as it is poorly understood: intelligence analysis. We report findings from a geopolitical forecasting tournament that assessed the accuracy of more than 150,000 forecasts of 743 participants on 199 events occurring over 2 years. Participants were above average in intelligence and political knowledge relative to the general population. Individual differences in performance emerged, and forecasting skills were surprisingly consistent over time. Key predictors were (a) dispositional variables of cognitive ability,

1. Training: Calibrated probabilities basic probabilistic thinking

2. Aptitude: Measured by tests and past performance

3. …

Global Risk Management Day
by Copenhagen Compliance®

March 11, 2021, Online Conference

Three Ways to Make Better Forecasts

Philip Tetlock tracked over 150,000 forecasts from 743 experts in world affairs. He determined factors that made the biggest difference in the performance of forecasting.
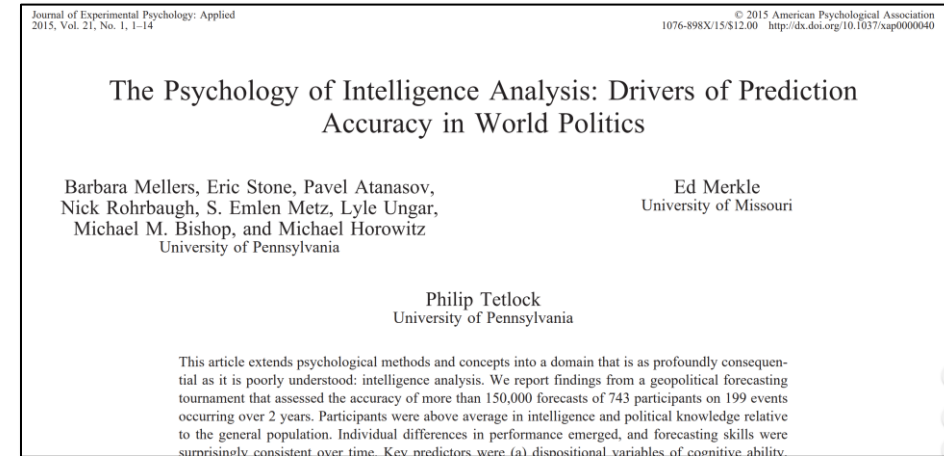
The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics

Barbara Mellers, Eric Stone, Pavel Atanasov, Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar, Michael M. Bishop, and Michael Horowitz
University of Pennsylvania

Ed Merkle
University of Missouri

Philip Tetlock
University of Pennsylvania

This article extends psychological methods and concepts into a domain that is as profoundly consequential as it is poorly understood: intelligence analysis. We report findings from a geopolitical forecasting tournament that assessed the accuracy of more than 150,000 forecasts of 743 participants on 199 events occurring over 2 years. Participants were above average in intelligence and political knowledge relative to the general population. Individual differences in performance emerged, and forecasting skills were surprisingly consistent over time. Key predictors were (a) dispositional variables of cognitive ability,

1. Training: Calibrated probabilities basic probabilistic thinking

2. Aptitude: Measured by tests and past performance

3. Teams!

Global Risk Management Day by Copenhagen Compliance®
March 11, 2021, Online Conference

Aggregating Experts

Accuracy, consistency, calibration, etc.

A team can be selected, trained and mathematically aggregated in a way that outperforms any single individual.

Expert Elicitation: Using the Classical Model to Validate Experts' Judgments

Abigail R. Colson* and Roger M. Cooke[†]

Automatica, Vol. 24, No. 1, pp. 87–94, 1988
Printed in Great Britain.

0005-1098/88 $3.00 + 0.00
Pergamon Journals Ltd.
© 1988 International Federation of Automatic Control

Brief Paper

Calibration and Information in Expert Resolution; a Classical Approach*

...GER COOKE†, MAX MENDEL‡ and WIM THIJS§

...rds—Expert resolution; expert opinion; subjective probability; calibration.

Risk Analysis, Vol. 19, No. 2, 1999

Combining Probability Distributions From Experts in Risk Analysis

Robert T. Clemen[1,2] and Robert L. Winkler[1]

Tentative Name: "FrankenSME"

practice. The role of experts is important because their judgments can provide valuable information, particularly in view of the limited availability of "hard data" regarding many important uncertainties in risk analysis. Because uncertainties are represented in terms of probability distributions in probabilistic risk analysis (PRA), we consider expert information in terms of probability distributions. The motivation for the use of multiple experts is simply the desire to obtain as much information as possible. Combining experts' probability distributions summarizes the accumulated information for risk analysts and decision-makers. Procedures for combining probability distributions are often compartmentalized as mathematical aggregation methods or behavioral approaches, and we discuss both categories.

Global Risk Management Day
by Copenhagen Compliance®
March 11, 2021, Online Conference

The Effect of Combining SMEs

- A special case of aggregation is when two people give the same answer (they agree and both state 80% confidence).

- What is the chance they are correct?

© Hubbard Decision Research, 2020

The Effect of Combining SMEs

After Calibration Training, filtering on skill, adjusting based on training data

Observed Frequency Correct

Assessed Probability
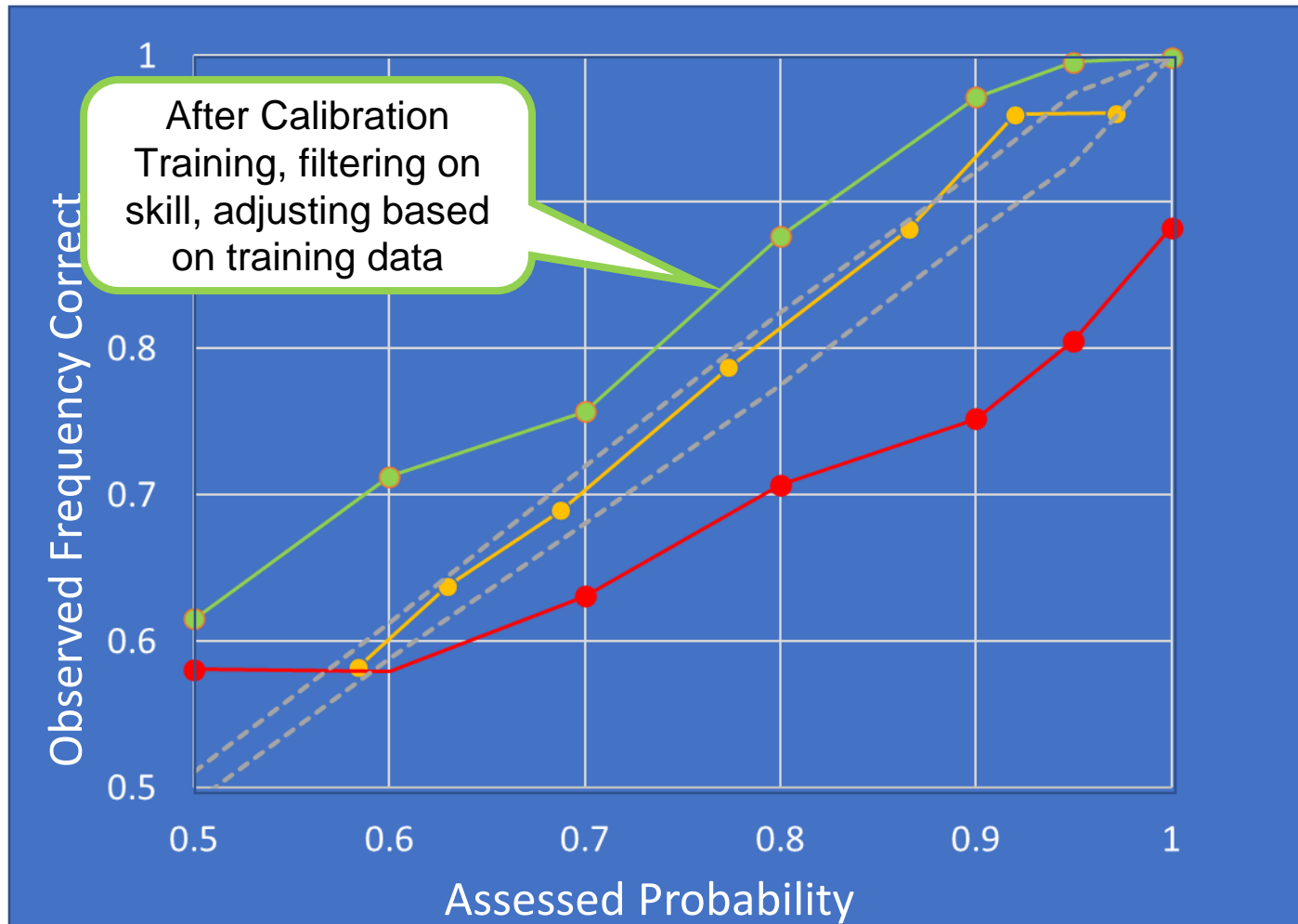
- A special case of aggregation is when two people give the same answer (they agree and both state 80% confidence).

- What is the chance they are correct?

- **Surprise!**

Global Risk Management Day *by Copenhagen Compliance*

March 11, 2021, Online Conference

It's all about Bayes

Bayes Theorem: $P(X|Y) = \dfrac{P(X)P(Y|X)}{P(Y)} = \dfrac{P(X)P(Y|X)}{\sum\limits_{i} P(Y|X_i)\,P(X_i)}$

P(X) = the probability of X

P(X|Y) = the probability of X given the condition Y

$\sum$ P(Y | X$_i$) P(X$_i$) = the sum of the probability of Y under each possible condition

$$\frac{P(X|C_1 \ldots C_n)}{1 - P(X|C_1 \ldots C_n)} = \left(\frac{1 - P(X)}{P(X)}\right)^{n-1} \prod_{i=1}^{n} \frac{P(X|C_i)}{1 - P(X|C_i)}$$
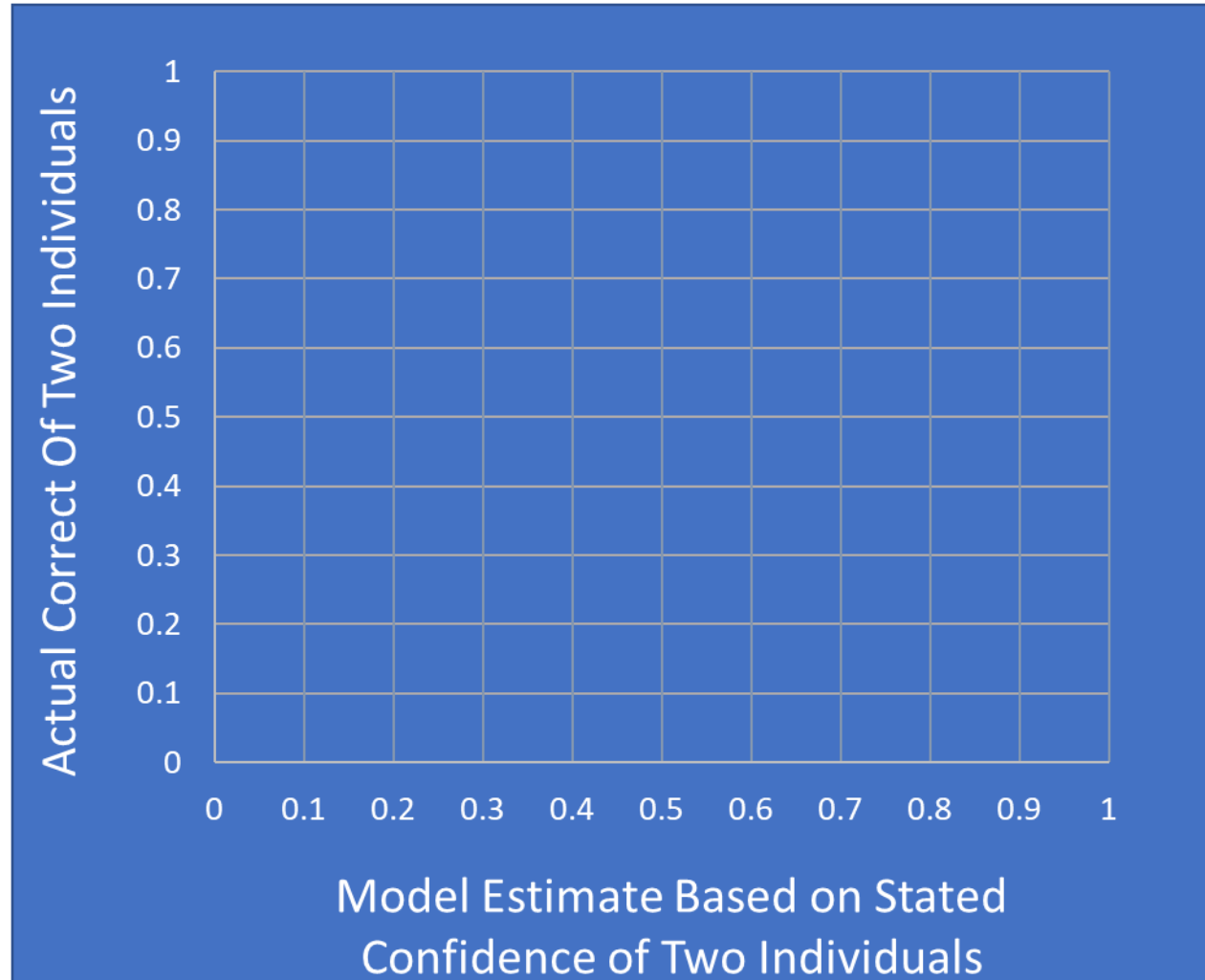
Global
Risk Management
Day   by Copenhagen Compliance®

March 11, 2021, Online Conference

The Model vs. Reality

- I generated over 380,000 random pairs of individuals who responded to the same question.

- When we look at all the combinations of probabilities that two people put on a claim being true, the Bayesian model which estimates team performance based on individual performance is a good predictor of actual team performance.

Global Risk Management Day
by Copenhagen Compliance®

March 11, 2021, Online Conference

The Model vs. Reality

$R^2 = 0.9885$

Actual Correct Of Two Individuals

Model Estimate Based on Stated Confidence of Two Individuals

- I generated over 380,000 random pairs of individuals who responded to the same question.

- When we look at all the combinations of probabilities that two people put on a claim being true, the Bayesian model which estimates team performance based on individual performance is a good predictor of actual team performance.

Global Risk Management Day
by Copenhagen Compliance®
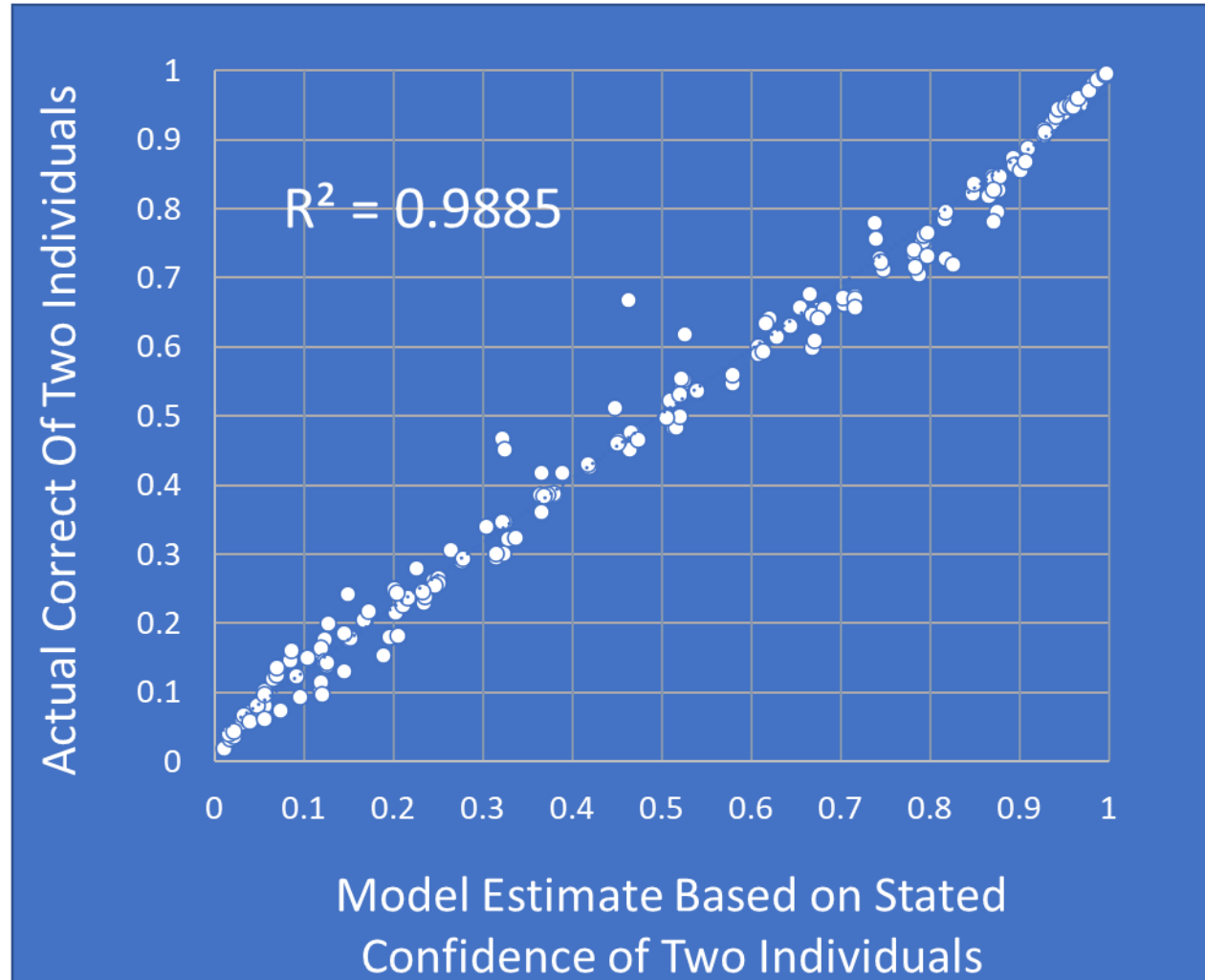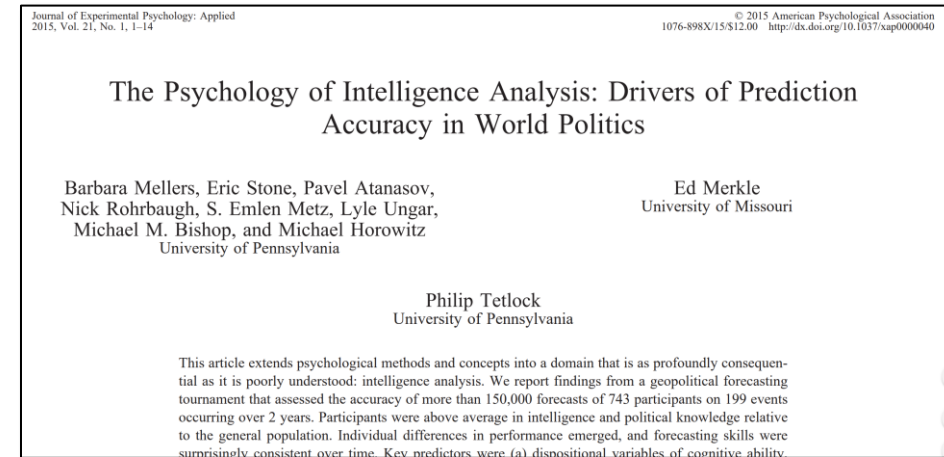March 11, 2021, Online Conference

Three Ways to Make Better Forecasts

Philip Tetlock tracked over 150,000 forecasts from 743 experts in world affairs. He determined factors that made the biggest difference in the performance of forecasting.

The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics

Barbara Mellers, Eric Stone, Pavel Atanasov,
Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar,
Michael M. Bishop, and Michael Horowitz
University of Pennsylvania

Ed Merkle
University of Missouri

Philip Tetlock
University of Pennsylvania

This article extends psychological methods and concepts into a domain that is as profoundly consequential as it is poorly understood: intelligence analysis. We report findings from a geopolitical forecasting tournament that assessed the accuracy of more than 150,000 forecasts of 743 participants on 199 events occurring over 2 years. Participants were above average in intelligence and political knowledge relative to the general population. Individual differences in performance emerged, and forecasting skills were surprisingly consistent over time. Key predictors were (a) dispositional variables of cognitive ability,

1. Training: Calibrated probabilities basic probabilistic thinking

2. Aptitude: Measured by tests and past performance

3. Teams!

Global Risk Management Day
by Copenhagen Compliance

March 11, 2021, Online Conference

Three Ways to Make Better Forecasts

Philip Tetlock tracked over 150,000 forecasts from 743 experts in world affairs. He determined factors that made the biggest difference in the performance of forecasting.
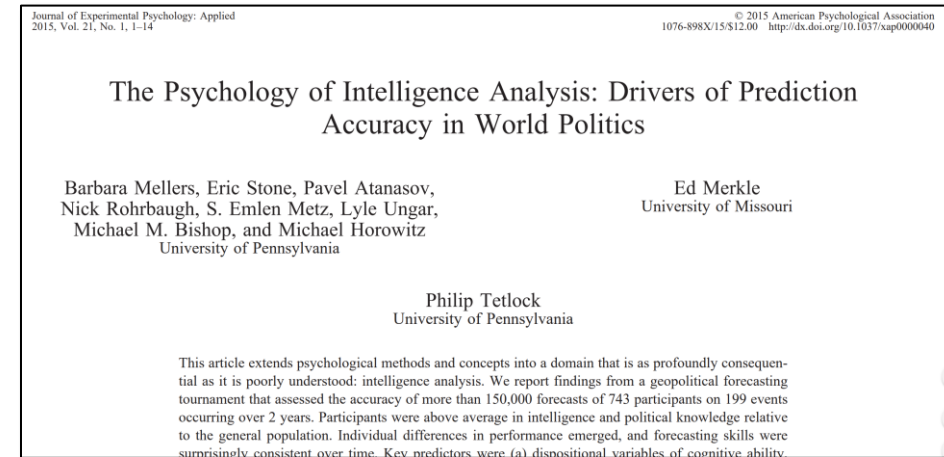
The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics

Barbara Mellers, Eric Stone, Pavel Atanasov,
Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar,
Michael M. Bishop, and Michael Horowitz
University of Pennsylvania

Ed Merkle
University of Missouri

Philip Tetlock
University of Pennsylvania

This article extends psychological methods and concepts into a domain that is as profoundly consequential as it is poorly understood: intelligence analysis. We report findings from a geopolitical forecasting tournament that assessed the accuracy of more than 150,000 forecasts of 743 participants on 199 events occurring over 2 years. Participants were above average in intelligence and political knowledge relative to the general population. Individual differences in performance emerged, and forecasting skills were surprisingly consistent over time. Key predictors were (a) dispositional variables of cognitive ability,

1. Training: Calibrated probabilities basic probabilistic thinking

2. Aptitude: Measured by tests and past performance

3. Teams! – but only if they consist of "belief updaters"

Global Risk Management Day
by Copenhagen Compliance®
March 11, 2021, Online Conference

Summary

- Even in quantitative models, subjective estimates are useful and often unavoidable.

- Calibrating SMEs is critical!

- A team is more than the sum of its parts.

- Some aggregation methods are *MUCH BETTER* than others based on:
  - How SMEs are selected
  - How SMEs are trained
  - How SME input is elicited
  - How SME estimates are combined mathematically

- A team of SMEs could be optimized for any type of application.

  Questions?: dwhubbard@hubbardresearch.com