

Pre-read for Thursday Sept 10:  
Predicting geopolitical events, part 2

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction  
Fall 2020, Princeton University

Copyrighted Material  
PHILIP E. TETLOCK

EXPERT  
  
POLITICAL  
  
JUDGMENT

*How Good Is It? How Can We Know?*

Copyrighted Material

20 years in the making

BOOKS DECEMBER 5, 2005 ISSUE

# EVERYBODY'S AN EXPERT

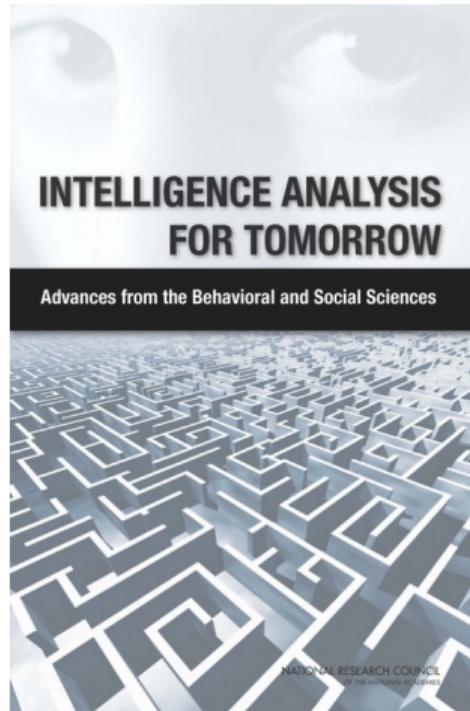
*Putting predictions to the test.*



By Louis Menand

November 28, 2005

<https://www.newyorker.com/magazine/2005/12/05/everybodys-an-expert>



<https://doi.org/10.17226/13040>



Intelligence Assessment Secretariat, part of the Canadian Security Intelligence Service

# Accuracy of forecasts in strategic intelligence

David R. Mandel<sup>a,1</sup> and Alan Barnes<sup>b,2</sup>

<sup>a</sup>Socio-Cognitive Systems Section, Defence Research and Development Canada, Toronto, ON, Canada M3K 2C9; and <sup>b</sup>Formerly of the Intelligence Assessment Secretariat, Privy Council Office, Ottawa, ON, Canada K1A 0A3

<https://doi.org/10.1073/pnas.1406138111>

Predictions embedded in reports:

- ▶ “The intense distrust that exists between Country X and Country Y is almost certain [9/10] to prevent the current relationship of convenience from evolving into a stronger alliance.”
- ▶ “It is very unlikely [1/10] that either of these countries will make a strategic decision to launch an offensive in the coming six months.”

Reading notes:

- ▶ How does the accuracy of these predictions compare to those in Tetlock?

Reading notes:

- ▶ How does the accuracy of these predictions compare to those in Tetlock?
- ▶ What are the similarities and differences between the tasks in the two studies?

Reading notes:

- ▶ How does the accuracy of these predictions compare to those in Tetlock?
- ▶ What are the similarities and differences between the tasks in the two studies?
- ▶ Note: Tetlock (2005): roughly 15% of cases had ambiguous outcomes (p 296),  
Mandel and Barnes (2014): 20% of cases had ambiguous outcomes (p 10987)

# Judging political judgment

**Philip Tetlock<sup>1</sup> and Barbara Mellers**

*Department of Psychology and The Wharton School, University of Pennsylvania, Philadelphia, PA 19104*

<https://doi.org/10.1073/pnas.1412524111>

Reading notes:

- ▶ How do they explain any difference in findings between Mandel and Barnes and Tellock (1995) and Mellers and Tetlock (comes after Tetlock 1995)?

Reading notes:

- ▶ How do they explain any difference in findings between Mandel and Barnes and Tetlock (1995) and Mellers and Tetlock (comes after Tetlock 1995)?
- ▶ What does this mean for our ability to learn anything from an empirical study like this?

LETTERS

<https://doi.org/10.1038/s41562-019-0620-8>

nature  
human behaviour

# Predicting history

Joseph Risi<sup>1,4</sup>, Amit Sharma<sup>2,4</sup>, Rohan Shah<sup>1</sup>, Matthew Connelly<sup>1</sup> and Duncan J. Watts<sup>1</sup>

<https://doi.org/10.1038/s41562-019-0620-8>

ARTHUR C. DANTO

Analytical  
Philosophy  
of History



CAMBRIDGE UNIVERSITY PRESS 17s. 6d. net \$2.95  
IN U.K. IN U.S.A.

All US State Department Cables (1973-1979). About 2 million cables of which about 2 thousand ( $\sim 1$  in 1,000) were judged to be historically significant.

All US State Department Cables (1973-1979). About 2 million cables of which about 2 thousand (~1 in 1,000) were judged to be historically significant.

(U) NAMIBIA: MARCH 19 MEETING OF CONTACT GROUP - MINISTERS WITH NNF AND SWAPO(D)

- MINISTERS WITH NNF AND SWAPO(D)  
1. (C-ENTIRE TEXT).  
2. REPRESENTATIVES OF THE NAMIBIA NATIONAL FRONT (NNF)  
AND SWAPO(D) MET WITH THE WESTERN MINISTERS AT 3:45 PM  
MARCH 19. BRIAN O'LINN AND R. RUKORO SPOKE FOR NNF;  
ANDREAS SHIPANGA FOR SWAPO(D).  
3. VANCE - THE CONTACT GROUP(CG) STANDS BEHIND ITS PROPOSAL.  
IT IS FULLY COMMITTED TO DEMOCRATIC ELECTIONS. IT WILL FOL-  
LOW THE WHOLE TRANSITION PROCESS WITH THIS END IN MIND.  
THERE HAS BEEN A MISUNDERSTANDING FOLLOWING THE FEB 26  
REPORT OF THE SYG. WE BELIEVE THE REPORT IS CONSISTENT  
WITH OUR PROPOSAL. THE SYG IS WORKING WITH US TO ENSURE  
THAT SWAPO BASES OUTSIDE NAMIBIA AND SWAPO'S ARMED PER-  
SONNEL INSIDE CANNOT INTERFERE WITH THE POLITICAL PROCESS.  
THE SYG IS WORKING ON WHO WOULD QUALIFY FOR CONFINEMENT  
INSIDE NAMIBIA AND UNDER WHAT CONDITIONS. THEREFORE,  
WE BELIEVE SWAPO ARMED PERSONNEL WOULD HAVE NO IMPACT.  
WE HOPE THE TRANSITION PROCESS CAN START AS SOON AS POSSI-  
BLE AND THAT REMAINING ISSUES CAN BE RESOLVED HERE IN NEW  
YORK.

Reading notes:

- ▶ We see new measures of performance: precision, recall, and F1 score. If you are not familiar with these read the Methods section carefully because we will see these again.

Reading notes:

- ▶ We see new measures of performance: precision, recall, and F1 score. If you are not familiar with these read the Methods section carefully because we will see these again.
- ▶ How accurately can the perceived contemporaneous impression of importance judgements predict subsequent judgement by historians?

Reading notes:

- ▶ We see new measures of performance: precision, recall, and F1 score. If you are not familiar with these read the Methods section carefully because we will see these again.
- ▶ How accurately can the perceived contemporaneous impression of importance judgements predict subsequent judgement by historians?
- ▶ How well can an ML model trained on the full cables predict subsequent judgement by historians?

Reading notes:

- ▶ We see new measures of performance: precision, recall, and F1 score. If you are not familiar with these read the Methods section carefully because we will see these again.
- ▶ How accurately can the perceived contemporaneous impression of importance judgements predict subsequent judgement by historians?
- ▶ How well can an ML model trained on the full cables predict subsequent judgement by historians?
- ▶ What do the errors of both approaches teach us?

Reading notes:

- ▶ We see new measures of performance: precision, recall, and F1 score. If you are not familiar with these read the Methods section carefully because we will see these again.
- ▶ How accurately can the perceived contemporaneous impression of importance judgements predict subsequent judgement by historians?
- ▶ How well can an ML model trained on the full cables predict subsequent judgement by historians?
- ▶ What do the errors of both approaches teach us?
- ▶ Overall, what do these results suggest for Danto's argument about the ideal chronicler?



### Accuracy of forecasts in strategic intelligence

David R. Mandel<sup>1,\*</sup> and Alan Barroso<sup>2,3</sup>

\*Socio-Cognitive Systems Section, Defence Research and Development Canada, Toronto, ON, Canada M3K 2J6; and <sup>1</sup>Formerly of the Intelligence Assessment Secretariat, Privy Council Office, Ottawa, ON, Canada K1A 0L2 (D.R.M.)

### LETTERS

human  
human behaviour

### Predicting history

Joseph Rusconi<sup>1</sup>, Amit Sharma<sup>1,2</sup>, Rohan Shah<sup>3,4</sup>, Matthew Connolly<sup>3,4\*</sup> and Duncan J. Watts<sup>3,4</sup>

## Judging political judgment

Philip Tetlock<sup>1</sup> and Barbara Mellers

Department of Psychology and The Wharton School, University of Pennsylvania, Philadelphia, PA 19104

3 additional thoughts

1) Underconfidence without confidence intervals

## 1) Underconfidence without confidence intervals

- ▶ Normally, when we produce an estimate we need to produce a confidence interval.  
In this case underconfidence, is having confidence intervals that are too big.

## 1) Underconfidence without confidence intervals

- ▶ Normally, when we produce an estimate we need to produce a confidence interval.  
In this case underconfidence, is having confidence intervals that are too big.
- ▶ In Mandel and Barnes, analysts don't produce confidence intervals but they still talk about underconfidence.

## 1) Underconfidence without confidence intervals

- ▶ Normally, when we produce an estimate we need to produce a confidence interval. In this case underconfidence, is having confidence intervals that are too big.
- ▶ In Mandel and Barnes, analysts don't produce confidence intervals but they still talk about underconfidence.
- ▶ Is there a role for confidence intervals (sometimes called predictive intervals) here? Would making accurate predictive intervals require understanding the mechanisms that lead to uncertainty . . . and hence the limits of prediction?

2) What is hard and what is easy?

2) What is hard and what is easy?

- ▶ Tetlock and Mellers: “We should, however, focus on the core problem that neither past nor current work has yet solved: how best to measure the deceptively simple concept of accuracy.

## 2) What is hard and what is easy?

- ▶ Tetlock and Mellers: “We should, however, focus on the core problem that neither past nor current work has yet solved: how best to measure the deceptively simple concept of accuracy. One challenge is that the standardization of difficulty. Getting a good Brier score by predicting weather in a low-variance world (eg, Phoenix) is a lot easier than it is in a high-variance world (eg, St. Louis).”

## 2) What is hard and what is easy?

- ▶ Tetlock and Mellers: “We should, however, focus on the core problem that neither past nor current work has yet solved: how best to measure the deceptively simple concept of accuracy. One challenge is that the standardization of difficulty. Getting a good Brier score by predicting weather in a low-variance world (eg, Phoenix) is a lot easier than it is in a high-variance world (eg, St. Louis).”
- ▶ Mandel and Barnes have a non-statistical way of thinking about difficulty (based on information base, number of factor, role for irrationality, time frame), see SI

## 2) What is hard and what is easy?

- ▶ Tetlock and Mellers: “We should, however, focus on the core problem that neither past nor current work has yet solved: how best to measure the deceptively simple concept of accuracy. One challenge is that the standardization of difficulty. Getting a good Brier score by predicting weather in a low-variance world (eg, Phoenix) is a lot easier than it is in a high-variance world (eg, St. Louis).”
- ▶ Mandel and Barnes have a non-statistical way of thinking about difficulty (based on information base, number of factor, role for irrationality, time frame), see SI
- ▶ Mandel and Barnes did not provide a measure of inter-coder reliability

## 2) What is hard and what is easy?

- ▶ Tetlock and Mellers: "We should, however, focus on the core problem that neither past nor current work has yet solved: how best to measure the deceptively simple concept of accuracy. One challenge is that the standardization of difficulty. Getting a good Brier score by predicting weather in a low-variance world (eg, Phoenix) is a lot easier than it is in a high-variance world (eg, St. Louis)."
- ▶ Mandel and Barnes have a non-statistical way of thinking about difficulty (based on information base, number of factor, role for irrationality, time frame), see SI
- ▶ Mandel and Barnes did not provide a measure of inter-coder reliability
- ▶ Is there some way to use these expert created measures of difficulty to learn about the limits of prediction?

### 3) Readymades vs Custommades





readymades



readymades





readymades



custommades



R. Mutt  
1917

3 additional thoughts:

- ▶ Underconfidence without confidence intervals
- ▶ What is hard and what is easy?
- ▶ Readymades vs Custommades

What to read next:

- ▶ “Bobby W.” (2019) [The Limits of Prediction—or How I Learned to Stop Worrying About Black Swans and Love Analysis](#) “The key struggle for intelligence analysts is that they are made to produce and what their consumers think they can produce are often two different things.”
- ▶ What to read next from last class

Pre-read for Thursday Sept 10:  
Predicting geopolitical events, part 2

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction  
Fall 2020, Princeton University