Pre-read for Tuesday, Sept 8: Predicting geopolitical events, part 1

Matthew J. Salganik

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

Imagine:

- It is 1984

Imagine:

- It is 1984
- You've just gotten tenure at Berkeley

Imagine:

- It is 1984
- You've just gotten tenure at Berkeley
- You are sitting in a meeting at the National Research Council about the threat of nuclear war (also you are the youngest person there)

Imagine:

- It is 1984
- You've just gotten tenure at Berkeley
- You are sitting in a meeting at the National Research Council about the threat of nuclear war (also you are the youngest person there)
- What is the probability of a nuclear war?

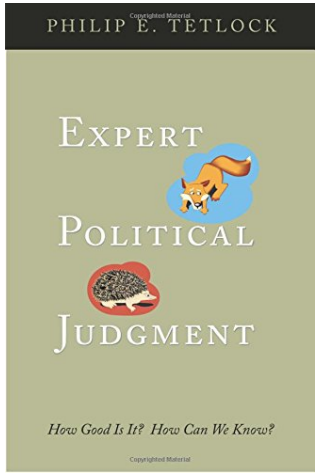
Imagine:

- It is 1984
- You've just gotten tenure at Berkeley
- You are sitting in a meeting at the National Research Council about the threat of nuclear war (also you are the youngest person there)
- What is the probability of a nuclear war?
- Everyone is making strong statements that often contradict each other, but you have no idea who is right.

Imagine:

- It is 1984
- You've just gotten tenure at Berkeley
- You are sitting in a meeting at the National Research Council about the threat of nuclear war (also you are the youngest person there)
- What is the probability of a nuclear war?
- Everyone is making strong statements that often contradict each other, but you have no idea who is right.

What if we could systematically study political judgement?



20 years in the making

“good political judgement”

"good political judgement"
Turning this idea into something measurable is hard. This problem comes up often in social science.

“good political judgement”

Turning this idea into something measurable is hard. This problem comes up often in social science.

Two focuses:

- accuracy (“do they get it right?”) [focus of this class]
- rigor (”do they think in the right way”)

Reading notes:

- Preface
- Chapter 1, Quantifying the Unquantifiable
- Chapter 2, The Ego-deflating Challenge of Radical Skepticism
- Methodological Appendix, Section I: Regional Forecasting Exercises, Chapters 2 and 3
- Technical Appendix, Part A: Correspondence Indicators of Good Judgement
- Preface to the 2017 edition

Preface

- Notice how his research questions comes from his own personal annoyances and experiences.

Preface

- Notice how his research questions comes from his own personal annoyances and experiences.
- How might this strategy be related to the projects that you will do in this class? What are the risks and benefits of this strategy for finding research topics?

Preface

- Notice how his research questions comes from his own personal annoyances and experiences.
- How might this strategy be related to the projects that you will do in this class? What are the risks and benefits of this strategy for finding research topics?
- If many academics follow this strategy (and I think they do) and if researchers do not match broader populations, what implications might there be for the kinds of questions that get studied and those that don't?

Chapter 1

- Notice that he is concern about the marketplace of ideas, not just predictive accuracy.

Chapter 1

- Notice that he is concern about the marketplace of ideas, not just predictive accuracy.
- Splits "good judgement" into getting it right and thinking the right way. We will focus on getting it right.

Chapter 1

- Notice that he is concern about the marketplace of ideas, not just predictive accuracy.
- Splits "good judgement" into getting it right and thinking the right way. We will focus on getting it right.
- Beyond chapter 2 (which we won't read) has one of the most famous findings. You might ask what separates the more accurate forecasters from the less accurate. It turns out that it is not background or political outlook, it is cognitive style. He shows "foxes" who know lots of little things are better forecasters than "hedgehogs" who know one big thing. Foxes and hedgehogs are on the cover of the book.

Chapter 2

- Notice reasons Tetlock presents for why experts might be bad at forecasting geo-political events; some are properties of the world and some of humans. Pay attention to the ones that properties of the world because we will see them again in different forms this semester.

Chapter 2

- Notice reasons Tetlock presents for why experts might be bad at forecasting geo-political events; some are properties of the world and some of humans. Pay attention to the ones that properties of the world because we will see them again in different forms this semester.
- Notice how Tetlock turns the radical skeptics ideas into 6 testable hypotheses. But also notice that the linkage is not very tight; the hypotheses don't help us distinguish between how much of unpredictability comes from punctuated equillibra or game theory.

Chapter 2

- Notice reasons Tetlock presents for why experts might be bad at forecasting geo-political events; some are properties of the world and some of humans. Pay attention to the ones that properties of the world because we will see them again in different forms this semester.
- Notice how Tetlock turns the radical skeptics ideas into 6 testable hypotheses. But also notice that the linkage is not very tight; the hypotheses don't help us distinguish between how much of unpredictability comes from punctuated equillibra or game theory.
- Notice what data Tetlock chooses to create in order to test his hypotheses.

Chapter 2

- Notice reasons Tetlock presents for why experts might be bad at forecasting geo-political events; some are properties of the world and some of humans. Pay attention to the ones that properties of the world because we will see them again in different forms this semester.
- Notice how Tetlock turns the radical skeptics ideas into 6 testable hypotheses. But also notice that the linkage is not very tight; the hypotheses don't help us distinguish between how much of unpredictability comes from punctuated equillibra or game theory.
- Notice what data Tetlock chooses to create in order to test his hypotheses.
- Tetlock pays a lot of attention to which types of errors occur (eg, Fig 2.6 over-predict low probability events and under-predict high probability events). Think a bit about how this teaches us something important beyond just accuracy. This is a theme we will see again.

Chapter 2

- Notice reasons Tetlock presents for why experts might be bad at forecasting geo-political events; some are properties of the world and some of humans. Pay attention to the ones that properties of the world because we will see them again in different forms this semester.
- Notice how Tetlock turns the radical skeptics ideas into 6 testable hypotheses. But also notice that the linkage is not very tight; the hypotheses don't help us distinguish between how much of unpredictability comes from punctuated equillibra or game theory.
- Notice what data Tetlock chooses to create in order to test his hypotheses.
- Tetlock pays a lot of attention to which types of errors occur (eg, Fig 2.6 over-predict low probability events and under-predict high probability events). Think a bit about how this teaches us something important beyond just accuracy. This is a theme we will see again.
- I recommend you flip back and forth between Methodological and Technical appendix. Otherwise the empirical results are hard to understand.

Technical appendix

- For our class, key part is page 273 - 283

Technical appendix

- For our class, key part is page 273 - 283
- If I give you some predictions you should be able to calculate what Tetlock calls "probability score" and "variability, calibration, and discrimination".

Technical appendix

- For our class, key part is page 273 - 283
- If I give you some predictions you should be able to calculate what Tetlock calls "probability score" and "variability, calibration, and discrimination".
- A theme we will see often is expert vs. simple algorithm vs. complex algorithm. Note how Tetlock creates simple and complex algorithms.

Technical appendix

- For our class, key part is page 273 - 283
- If I give you some predictions you should be able to calculate what Tetlock calls "probability score" and "variability, calibration, and discrimination".
- A theme we will see often is expert vs. simple algorithm vs. complex algorithm. Note how Tetlock creates simple and complex algorithms.
- Pay less attention (unless you are interested) to section "Adjustments of probability scores to address conceptual objections" except section "Controversy-adjusted probability scores" where he points out that 15% of the outcomes are unclear.

Preface to the 2017 edition

- It is neat to see how people reflect on their work after it has been out in the world

Further context

Possible Futures Study

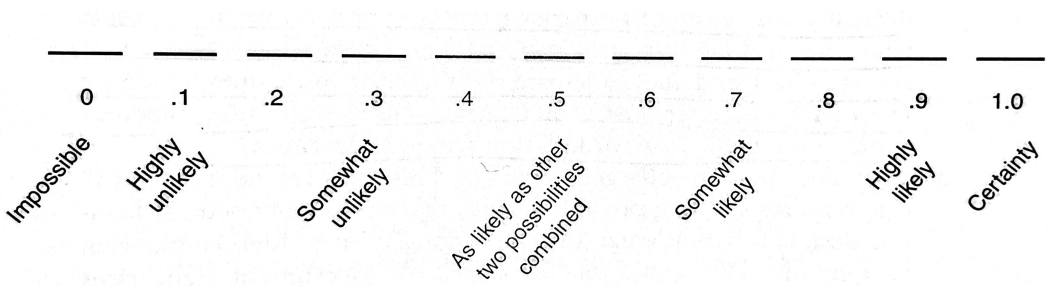
"Although political forecasting is obviously an inexact science, educated guesswork is still critical for setting priorities and making contingency plans. Your answers to the forecasting questions posed here will not be traceable either to you personally or to any institution with which you may be affiliated. Our goal is not to proclaim 'winners' and 'losers' in a forecasting contest but rather to study how highly trained professionals reason about complex real-world processes under conditions of uncertainty."

Part 1: Questions about your professional background, preferred style of thinking, and ideological and theoretical commitments.

Part 2: Questions inside your area of expertise

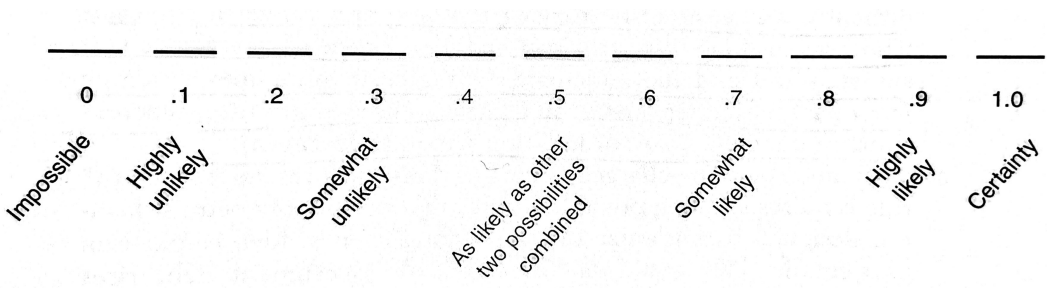
For Vietnam:

- Should we expect—over the next 2 years—increases, decreases, or essentially no changes in the marginal tax rate?



For Vietnam:

- Should we expect—over the next 2 years—increases, decreases, or essentially no changes in the marginal tax rate?



- Think of spreading probability over three buckets, not making a point prediction
- Option to select "maximum uncertainty", which put 0.33 weight in each category (dart throwing chimp)

For Vietnam:

- Should we expect—over the next 2 years—increases, decreases, or essentially no changes in the marginal tax rate?
- Should we expect—over the next 5 years—increases, decreases, or essentially no changes in the marginal tax rate?
- Should we expect—over the next 5 years—defense spending as a percentage of central government expenditure to rise, fall, or stay the same?
- Should we expect—over the next 10 years—defense spending as a percentage of central government expenditure to rise, fall, or stay the same?

Part 3: Questions outside your area of expertise

Four content categories:

- continuity of domestic political leadership (eg, In the next election [Two elections from now], will the party with the most seats in the legislature keep that status, lose that status, or strengthen its position?)

Four content categories:

- continuity of domestic political leadership (eg, In the next election [Two elections from now], will the party with the most seats in the legislature keep that status, lose that status, or strengthen its position?)
- domestic policy and economic performance (eg, Should we expect—over the next two years [or five years]—increases, decreases, or essentially no changes in the marginal tax rate?)

Four content categories:

- continuity of domestic political leadership (eg, In the next election [Two elections from now], will the party with the most seats in the legislature keep that status, lose that status, or strengthen its position?)
- domestic policy and economic performance (eg, Should we expect—over the next two years [or five years]—increases, decreases, or essentially no changes in the marginal tax rate?)
- national security and defense policy (eg, Should we expect—over the next 5 years [or 10 years]—defense spending as a percentage of central government expenditure to rise, fall, or stay the same?)

Four content categories:

- continuity of domestic political leadership (eg, In the next election [Two elections from now], will the party with the most seats in the legislature keep that status, lose that status, or strengthen its position?)
- domestic policy and economic performance (eg, Should we expect—over the next two years [or five years]—increases, decreases, or essentially no changes in the marginal tax rate?)
- national security and defense policy (eg, Should we expect—over the next 5 years [or 10 years]—defense spending as a percentage of central government expenditure to rise, fall, or stay the same?)
- special-purpose exercises (eg, weapons of mass destruction, Persian Gulf War I, etc.)

Four content categories:

- continuity of domestic political leadership (eg, In the next election [Two elections from now], will the party with the most seats in the legislature keep that status, lose that status, or strengthen its position?)
- domestic policy and economic performance (eg, Should we expect—over the next two years [or five years]—increases, decreases, or essentially no changes in the marginal tax rate?)
- national security and defense policy (eg, Should we expect—over the next 5 years [or 10 years]—defense spending as a percentage of central government expenditure to rise, fall, or stay the same?)
- special-purpose exercises (eg, weapons of mass destruction, Persian Gulf War I, etc.)

These questions were given at many different times to many different people so I don't think there is a single questionnaire.

Notes about data:

- Probabilities are categorical (eg, 0, 0.1, 0.2) not continuous (eg, 0.2369), which impacts scoring

Notes about data:

- Probabilities are categorical (eg, 0, 0.1, 0.2) not continuous (eg, 0.2369), which impacts scoring
- Outcomes generally split into three groups: increase, stay the same ($\pm$ 1 standard deviation), or down. I'm not sure why it was set up that way.

Notes about data:

- Probabilities are categorical (eg, 0, 0.1, 0.2) not continuous (eg, 0.2369), which impacts scoring
- Outcomes generally split into three groups: increase, stay the same ($\pm$ 1 standard deviation), or down. I'm not sure why it was set up that way.
- There were many different exercises at different times so it is probably better to think of this as a collection of related studies rather than a single study

What to read next?

- Readings for class on Thursday
- More work by Tetlock and colleagues: https://www.sas.upenn.edu/tetlock/
- Watch Tetlock's book talk at Google
- Murphy (1973) A New Vector Partition of the Probability Score
- Blattenberg and Lad (1985) Separating the Brier Score into Calibration and Refinement Components: A Graphical Exposition
- Murphy and Winkler (1987) A General Framework for Forecast Verification