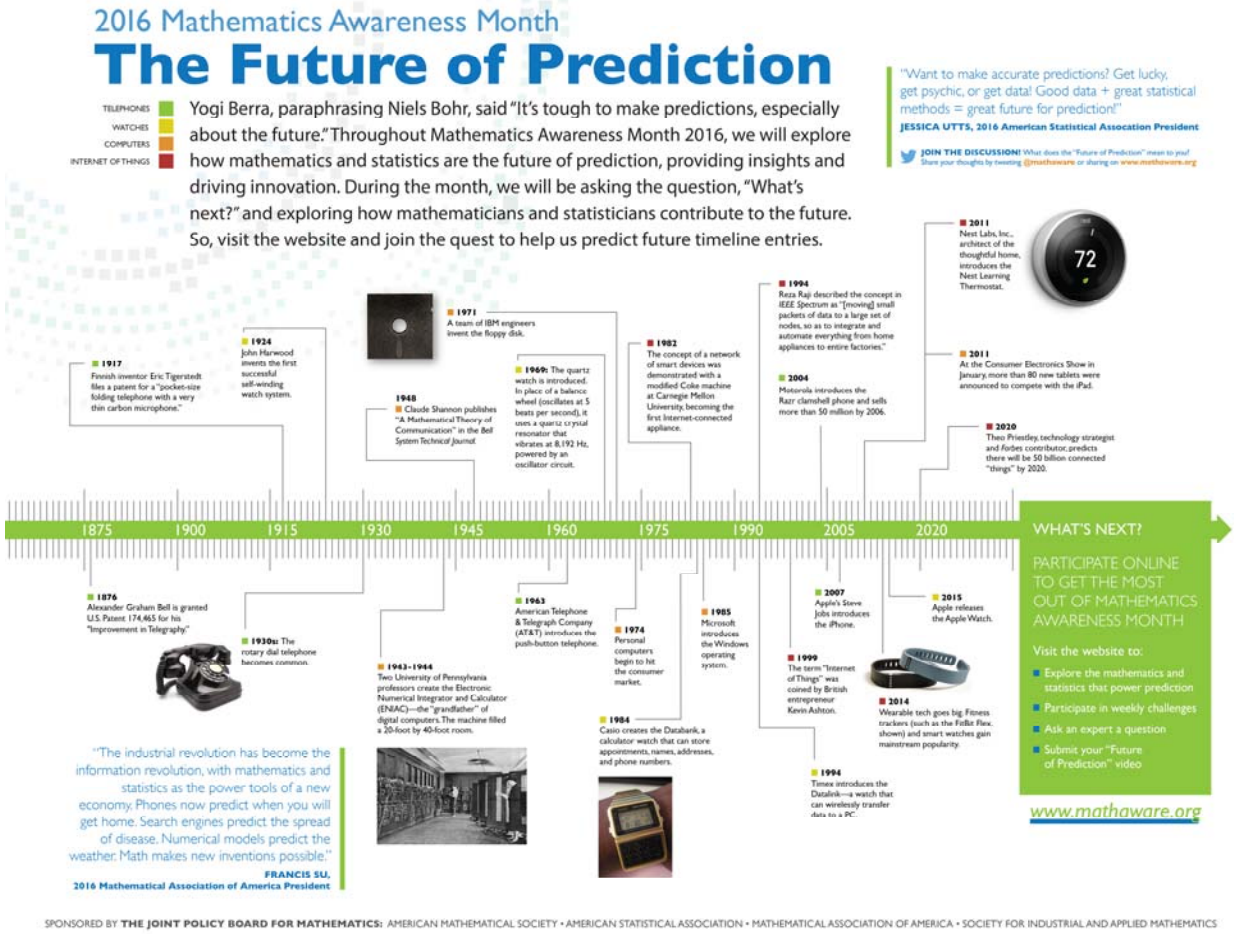
The Past, Present and Future of Prediction

Alan L. Montgomery
Associate Professor
The University of Hong Kong & Carnegie Mellon University

Inspiration
April 2016 Mathematics Awareness Month



http://www.mathaware.org/index.html

The Challenge for Prediction

Prediction underlies most decisions and range from precise numerical forecasts generated by predictive analytical models to qualitative statements made about upcoming events.

Problem: Most managers are ill equipped to predict key events or numbers, having neither the technical skills nor the right mindset.

Objective: Show how science can help us overcome biases to become better forecasters

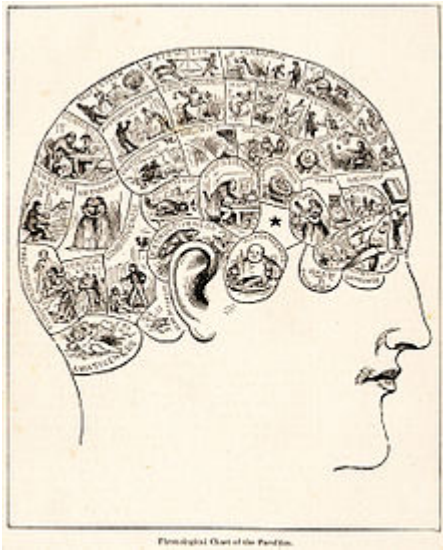
The Need for Prediction

History of Prediction

Oracle of Delphi

Advice to Philip II of Macedon
"With silver spears you
may conquer the world" (359BC)

Zhang Heng's
Seismograph (132AD)



Phrenology
19th Century

Prediction versus Forecasting

"The term 'forecasting' is used when it is a time series and we are predicting the series into the future. Hence 'business forecasts' and 'weather forecasts'.

In contrast, 'prediction' is the act of predicting in a cross-sectional setting, where the data are a snapshot in time (say, a one-time sample from a computer database).

Here you use information on a sample of records to predict the value of other records (which can be a value that will be observed in the future)."

Galit Shmueli

Distinguished Professor of Business Analytics

Taiwan's National Tsing Hua University

Prediction as the Basis for Decision Making

Mortgage loan

Your bank scored you as a low credit risk using a predictive model and approves your loan

Insurance

You volunteer to have sensors installed in your car, which transmits driving behaviors to your insurance company, which uses them in a predictive model to adjust your premium

Cellular provider

Predicts demand to locate cell towers , and enhances the system based upon those predictions

Health

Predictive models determine the medical treatments you should receive or preventative programs to keep you healthier

Movie Selection

Online movie provider uses a predictive model to compare your viewing habits with another and suggests movies that you have not viewed that you are likely to enjoy

Law Enforcement

Looks for past crime patterns to predict future crime, and deploys police accordingly to minimize crime

Climatologists

What will the average temperature be in 100 years?

Others: Companies, governments, economists, charities, universities, hospitals, ….

Some Famous Predictions

"I think there is a world market for maybe five computers." -- *Thomas Watson, chairman of IBM, 1943*

"The concept is interesting and well-formed, but in order to earn better than a 'C', the idea must be feasible." -- *A Yale University management professor in response to Fred Smith's paper proposing reliable overnight delivery service. (Smith went on to found Federal Express Corp.)*

"There is no reason anyone would want a computer in their home." -- *Ken Olson, president, chairman and founder of Digital Equipment Corp., 1977.*

"Stocks have reached what looks like a permanently high plateau." -- *Irving Fisher, Professor of Economics, Yale University, 1929.*

What is the chance of a military coup in Turkey? [in June 2016]

Jay Ulfelder, who works in the area of political forecasting, has developed a mathematical model that synthesizes this data to predict a country's level of risk.

Turkey, said Mr. Ulfelder's research, done in conjunction with the Early Warning Project, was a "very unlikely" candidate for a coup, he said in an email.

It had only a 2.5 percent probability of an attempted coup, based on 2016 data. That placed it 56th out of 160 countries, between Laos and Iran, and was within a range considered stable.

Fisher and Taub (2016), *New York Times*, July 17, 2016

http://www.nytimes.com/2016/07/17/world/europe/turkey-was-an-unlikely-victim-of-an-equally-unlikely-coup.html?_r=0

The Effects of Human Cognition in the Prediction Process

Trivia Quiz

Provide a high and low guess for each of the following questions. Make sure that you are 90% sure that the correct answer falls between the two.

1. Martin Luther King Jr.'s age at death
2. Length of the Nile River (in miles)
3. Number of counties that are members of OPEC
4. Number of books in the Old Testament
5. Diameter of the moon (in miles)
6. Weight of an empty Boeing 747 in pounds
7. Year in which Wolfgang Amadeus Mozart was born
8. Gestation period (in days) of an Asian elephant
9. Air distance from London to Tokyo
10. Deepest (known) point in the oceans (in feet)

Overconfidence

Type of People Tested	Type of Question Asked	Ideal Target	Actually Observed
Harvard MBAs	Trivia facts	2%	46%
Employees of Chemical Company	Chemical industry facts	10%	50%
	Company specific facts	50%	79%
Managers of a computer company	General business facts	5%	80%
	Company specific facts	5%	58%
Physicians	Probability that a patient has pneumonia	0-20%	82%

Source: Russo and Schoemaker, Decision Traps, 1989

Systematic Processing versus Heuristics

Answer this question:

A bat and ball together cost $1.10. The bat costs a dollar more than the ball. How much does the ball cost?

Most people answer: "ten cents"

But this is wrong, consider that if Bat is $1.00 and Ball is $0.10 then $1.00 - $0.10 = $0.90, but the question states the bat is $1.00 more than the ball.

Correct answer:

- If Bat+Ball=$1.10 and Bat=Ball+$1
- Differencing the equations gives: Ball=$.10-Ball, or Ball=$0.05

Availability Bias:
Perceived Data ≠True Picture

Choose the causes of death that you think causes more deaths in the United States each year:

Pair 1: Lung Cancer vs. Motor Vehicle Deaths
Pair 2: Tuberculosis vs. Fire and Flames

Cause of Death	People's Choice	News Reports per Year	US (in 1000s)
Pair 1	43% vs 57%	3 vs 127	140 vs 46
Pair 2	23% vs 77%	0 vs 24	4 vs 7

Anchoring Heuristic

We tend to base estimate around ‘anchors’:

- Is the population of Turkey greater or less than 30 million?
- Is the population of Turkey greater or less than 3 million?

Examples

- What are the last two digits or your phone number? What percentage of African countries are in the UN? Do you think there are more of less than your phone number?
- The median answer for people with ’10’ was 25%, while for those with ’65’ was 45%.
- The suggested retail price is $500

Hindsight Bias

Once we know the outcome, this perception skews what we thought we knew *before* the outcome

Suppose I ask you now:

- *What is the probability of the Clinton winning the election in 2016?*

Suppose next year I ask you the same question.

Consequence: “I knew it all along”

Confirmation Bias

"Economists have never predicted a recession"

Notice that the statement only considers what the forecaster got wrong, and not the fact that they were right most of the time.

Example: "If the patient is cured, it proves my treatment works; but if the Patient dies then it means nothing"

http://www.pragcap.com/the-confirmation-bias-of-the-anti-forecasters/

What is the chance to two people in this room have the same birthday?

What is the chance to two people in this room have the same birthday?

Suppose there are 60 people in the room. This gives us 1,770 pairs of individuals:

$$\frac{60 \cdot 59}{2} = 1,770$$

The chance of 2 people having *different* birthdays is:

$$1 - \frac{1}{365} = \frac{364}{365} = .997260$$

Having all 1,770 pairs be different is like getting "heads" 1,770 times in a row:

$$\left(\frac{364}{365}\right)^{1,770} = 0.007782 \approx 1\%$$

Evaluating probabilities

People are really good at understanding these two probability values:

Yes = 100%

No = 0%

They also have an intuition for a third:

Maybe = ?

Unfortunately, people like certainty so they like to round “Maybe” to either “Yes” or “No”

The Challenge of Prediction
Biases of Human Cognition

People are poor learners from experience

People have poor intuitive senses for the dollar value of information and probabilities

People are limited and biased information processor

"Declarations of high confidence mainly tell you that an individual has constructed a coherent story in his mind, not necessarily that the story is true."

Daniel Kahneman, *Thinking, Fast and Slow,* pg. 212

Overcoming our Cognitive Biases

It’s not what we don’t know that gives us trouble, it’s what we know that ain’t so.

Will Rogers

How to overcome our biases

To overcome our weaknesses we need to learn how to utilize a systematic approach to prediction.

Suggestions:

1) Level of confidence
2) Provide feedback and training to help people calibrate
3) Ask disconfirming questions
4) Expose the hidden sources of future problems
5) Limit yourself to the information you can handle

How to evaluate forecasts?

In 2007 Steve Ballmer stated:

"There's no chance that the iPhone is going to get any significant market share. No chance."

How to evaluate forecasts?

In 2007 Steve Ballmer stated:

> "There's no chance that the iPhone is going to get any significant market share. No chance. It's a $500 subsidized item. They may make a lot of money. But if you actually look at the 1.3 billion phones that get sold, I'd prefer to have our software in 60% or 70% or 80% of them, than I would to have 2% or 3%, which is what Apple might get."

Consider

What is "significant market share"? Which "market"? North American or World? Smartphones or mobile phones?

Lesson: We need to be more precise in our forecasts. Probabilities are better.

Fermi Problems

What is the weight of a 747?

I know a 747 holds about 450 people. My car holds 4 people and weighs 6,000 pounds.

6,000 pounds x (450/4) = 675,000 pounds

Systematic Approaches for Improving Predictions

Suggestions:

- Updating predictions frequently
- Using “baselines”
- Learning from mistakes
- Provide probabilistic not deterministic predictions
- Avoid ideological beliefs as basis for predictions
- Combine multiple and disparate information sources

Bayesian methods provide a systematic approach for formulating priors and updating them using new information

$$\Pr(\mathrm{H} \mid \mathrm{Data}) = \frac{\Pr(\mathrm{Data} \mid \mathrm{H})\Pr(\mathrm{H})}{\Pr(\mathrm{Data})}$$

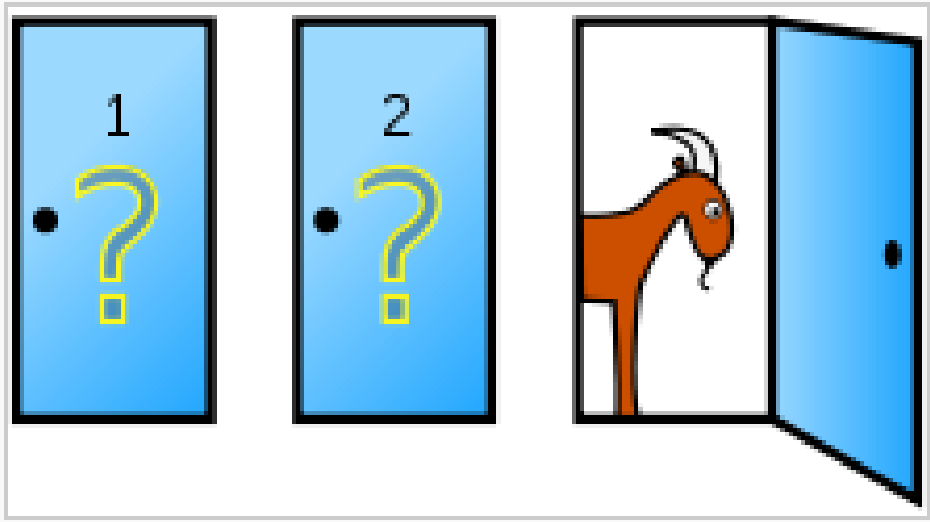
Monty Hall Problem

Conditional probabilities and priors

Source: https://en.wikipedia.org/wiki/Monty_Hall_problem

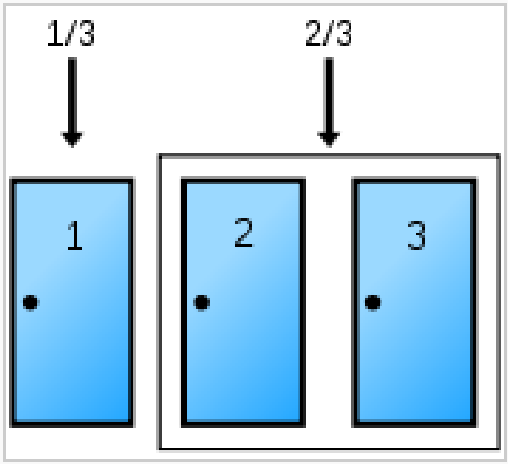
Monty Hall's
"Let's Make a Deal"

What door should you choose?

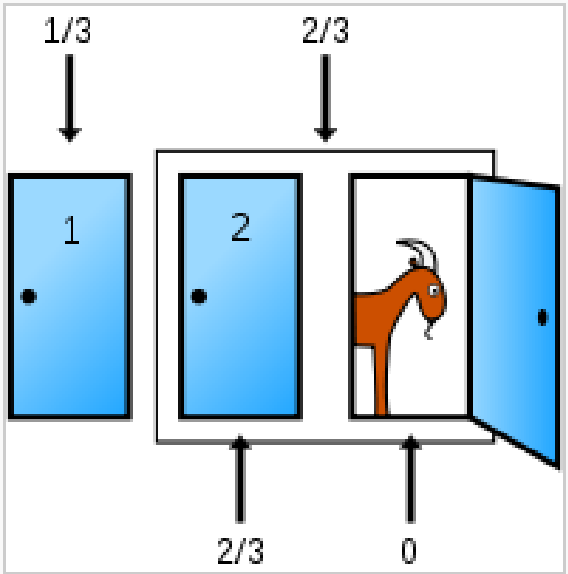


In search of a new car, the player picks a door, say 1. The game host then opens one of the other doors, say 3, to reveal a goat and offers to let the player pick door 2 instead of door 1.

What is the probability of winning the car?



Car has a 1/3 chance of being behind the player's pick and a 2/3 chance of being behind one of the other two doors



The host opens a door, the odds for the two sets don't change but the odds move to 0 for the open door and 2/3 closed door.

The Bayesian Setup

Let's start with our priors:

$$P(\text{Car}@1) = \text{P}(\text{Car}@2) = \text{P}(\text{Car}@3) = \tfrac{1}{3}$$

Suppose you choose door 1, then what is the conditional probability of Monty opening a door?

The Bayesian Setup

Suppose you choose door 1, then what is the conditional probability of Monty opening a door?

What are the chances Monty opens door 2 if the car is behind door 1?

$$P(\text{Open } 2 \mid \text{Car}@1) = \tfrac{1}{2}$$

Monty could have open either ‘door 2’ or ‘door 3’. As you have chosen ‘door 1’ he cannot open this door.

What are the chances Monty opens door 2 if the car is behind door 2?

$$P(\text{Open } 2 \mid \text{Car}@2) = 0$$

Monty will never open door 2 if it conceals the car.

What are the chances for Monty to open door 2 if the car is behind door 3?

$$P(\text{Open } 2 \mid \text{Car}@3) = 1$$

Monty has to open door 2, he has no other choice. You have already chosen ‘door 1’ and door 3 is concealing the car.

What is the posterior probability?

If we initial choose door 1, what is the probability given that Monty opens door 2?

$$P(\text{Car}@1 \mid \text{Open } 2) = \frac{X}{X+Y+Z} = \frac{\frac{1}{6}}{\frac{1}{6}+0+\frac{1}{3}} = \frac{1}{3}$$

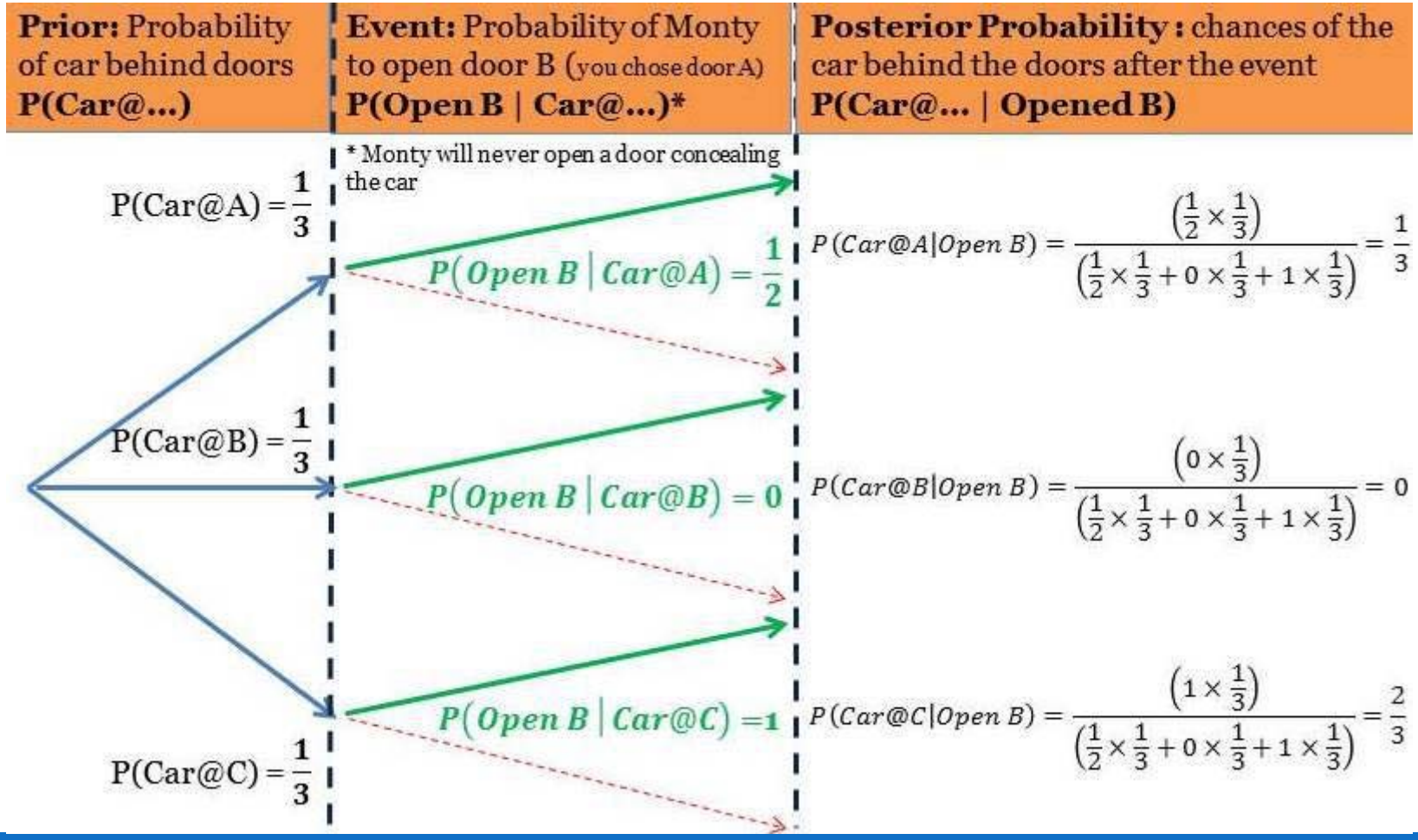
where,

$$X = P(\text{Open } 2 \mid \text{Car}@1) \times P(\text{Car}@1) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

$$Y = P(\text{Open } 2 \mid \text{Car}@2) \times P(\text{Car}@2) = 0 \times \frac{1}{3} = 0$$

$$Z = P(\text{Open } 2 \mid \text{Car}@3) \times P(\text{Car}@3) = 1 \times \frac{1}{3} = \frac{1}{3}$$

Posterior



http://ucanalytics.com/blogs/bayes-theorem-monty-hall-problem/

Conclusion

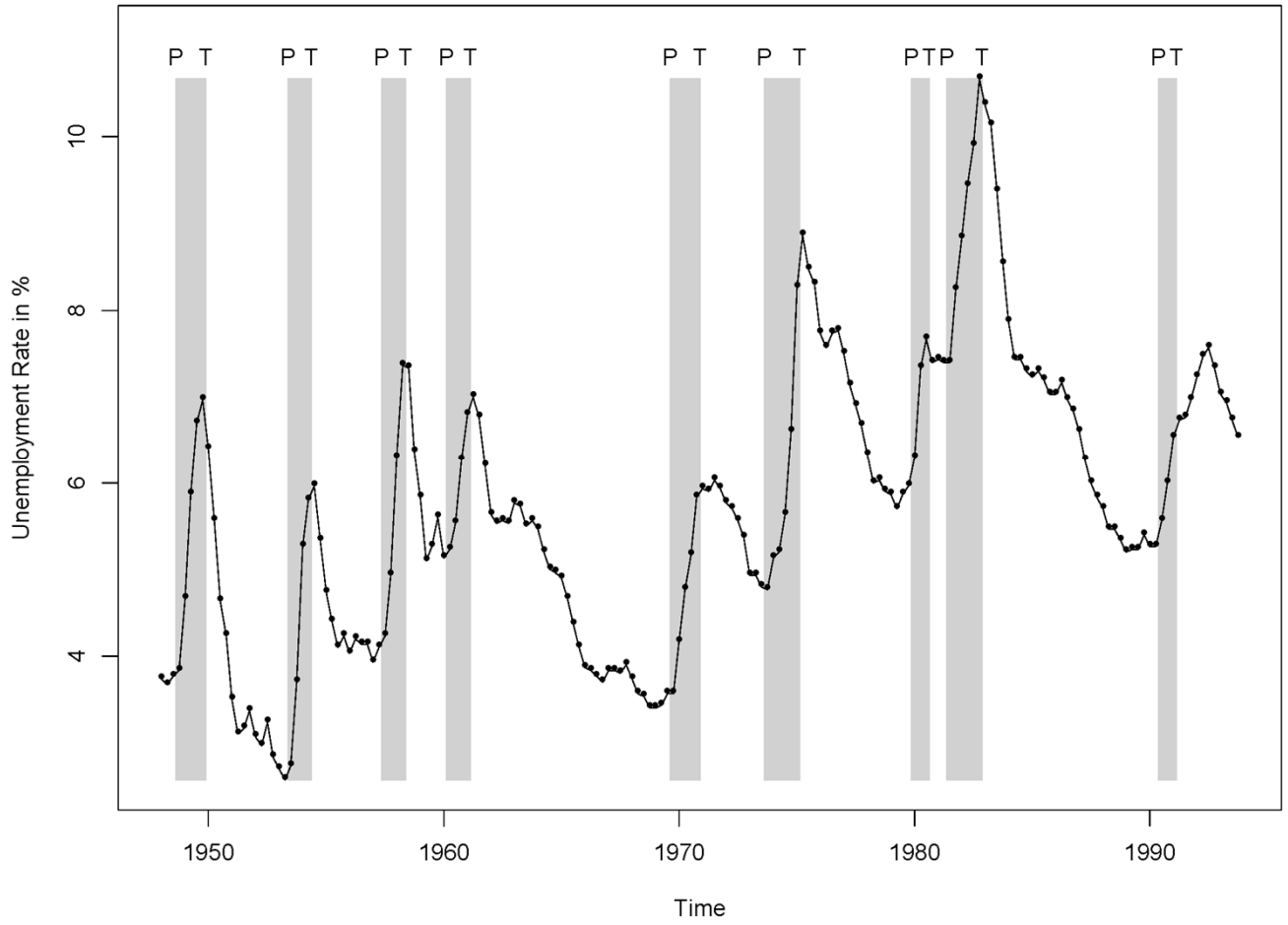
Always switch if Monty gives you the opportunity! You double the probability of winning a car.

The analysis requires us to explicitly make an assumption (or prior) on both the initial probability as well as the host's behavior. Alternatives:

- *"Monty from Hell".* The host offers the option to switch only when the player's initial choice is the winning door. Result: Switching always yields a goat. (Tierney 1991)
- "*Angelic Monty".* The host offers the option to switch only when the player has chosen incorrectly. Result: Switching always wins the car. (Granberg 1996:185)
- *"Random Monty".* The host knows what lies behind the doors, and (before the player's choice) chooses at random which goat to reveal. He offers the option to switch only when the player's choice happens to differ from his. Result: Switching wins the car half of the time.
- *"Game theory Monty".* The host opens a door and makes the offer to switch 100% of the time if the contestant initially picked the car, and 50% the time others. Result: Switching wins ½ the time at the Nash equilibrium. (Mueser and Granberg 1999)

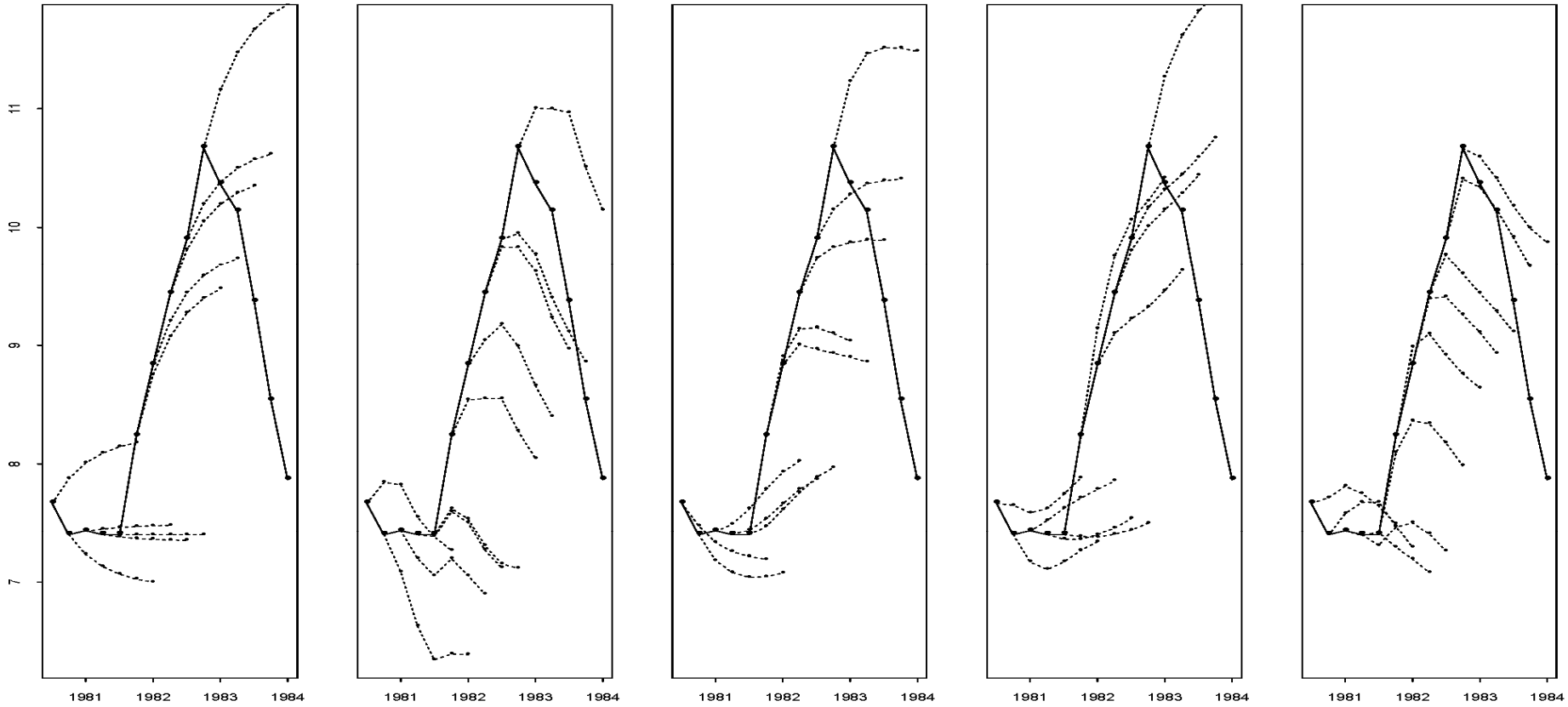
Present Research on Predictions

Forecasting the US Unemployment Rate



Montgomery, Zarnowitz, Tsay and Tiao (1998), "Forecasting the US Unemployment Rate", *Journal of the American Statistical Association*

Forecasting the US Unemployment Rate



ARIMA(1,1,0) SARIMA(1,1,0) TAR model MSA model Consensus

Montgomery, Zarnowitz, Tsay and Tiao (1998), "Forecasting the US Unemployment Rate", *Journal of the American Statistical Association*

Forecasting the US Unemployment Rate

- Unemployment does not have a consistent trend, but it does have asymmetrical cycles that go up quickly and decline slowly.
- Nonlinear time series models work better than linear time series models.
- Forecasting unemployment is much more difficult during recessions when it is rising rapidly.
- Expert forecasters are better in the near term, because they use more information, but one year ahead forecasts are no better than models.
- No methods (that we know of) can forecast when the turning point of unemployment occurs

Montgomery, Zarnowitz, Tsay and Tiao (1998), "Forecasting the US Unemployment Rate", *Journal of the American Statistical Association*

Aggregative Contingent Estimation

In 2011 the Intelligence Advanced Research Projects Activity (IARPA) announced a forecasting tournament that would run from 2011 to 2015. Five teams would compete by answering 500 questions on world affairs

- Will protests in Russian spread?
- Will the price of gold would plummet?
- Will the Nikkei close above 9500?
- Will war would erupt on the Korean peninsula?

The Good Judgment Project

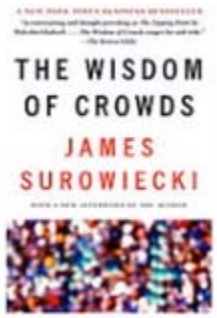
One team, the “Good Judgment Project” (GJP), created by Barbara Mellers and Philip Tetlock, was comprised of 20,000 forecasters in the first year

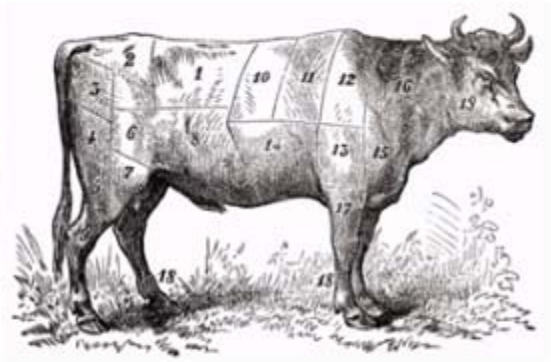
- Assigned its participants a rating called a “Brier Score”, which measures the accuracy of predictions on a scale from 0 to 2 (0=perfection).
- Identified the best forecasters (called superforecasters) and then put them into 12-person “superteams” so that they could share information

Won the tournament by 60% in its first year, and 78% in the second year. IARPA decide to drop the others leaving only the GJP

“Wisdom of the Crowds”

Sir Francis Galton (1907) proposed that the collective knowledge of a group of people could be more accurate than individual predictions





average of 800 guesses = 1,197
actual weight of the ox = 1,198

“Wisdom of the Crowds”

In Williams and Reade (2015, “Forecasting Elections”) compared prediction markets to traditional methods.

- Using data from InTrade and Betfair, plus statistical modeling, expert opinion and every opinion poll. We compared hundreds of different elections.
- Found that Prediction markets significantly outperformed the other methodologies. (The closer you get to the event the better the forecasts)

Berg et al (2008) compared Iowa Electronic Market predictions to 964 polls over five US presidential elections and found that the market is closer to the eventual outcome 74% of the time, and that the market significantly outperforms the pools in every election when the forecasts are more than 100 days in advance.

GJP “superteams” competed against regular forecasters. Tetlock found “Teams of ordinary forecasters beat the wisdom of the crowd by about 10%. Prediction markets beat ordinary teams by about 20%. And superteams beat prediction markets by 15% to 30%”.

Future of Prediction

When will predictive analytics work?

In the 1980s “People realised that if you took all of the ideas that people have developed in different fields, and you thought of it as a collection of techniques and overlaid that with analytical and scientific thinking, then forecasting itself could be considered a scientific discipline”

Predictability of an event boils down to three factors:

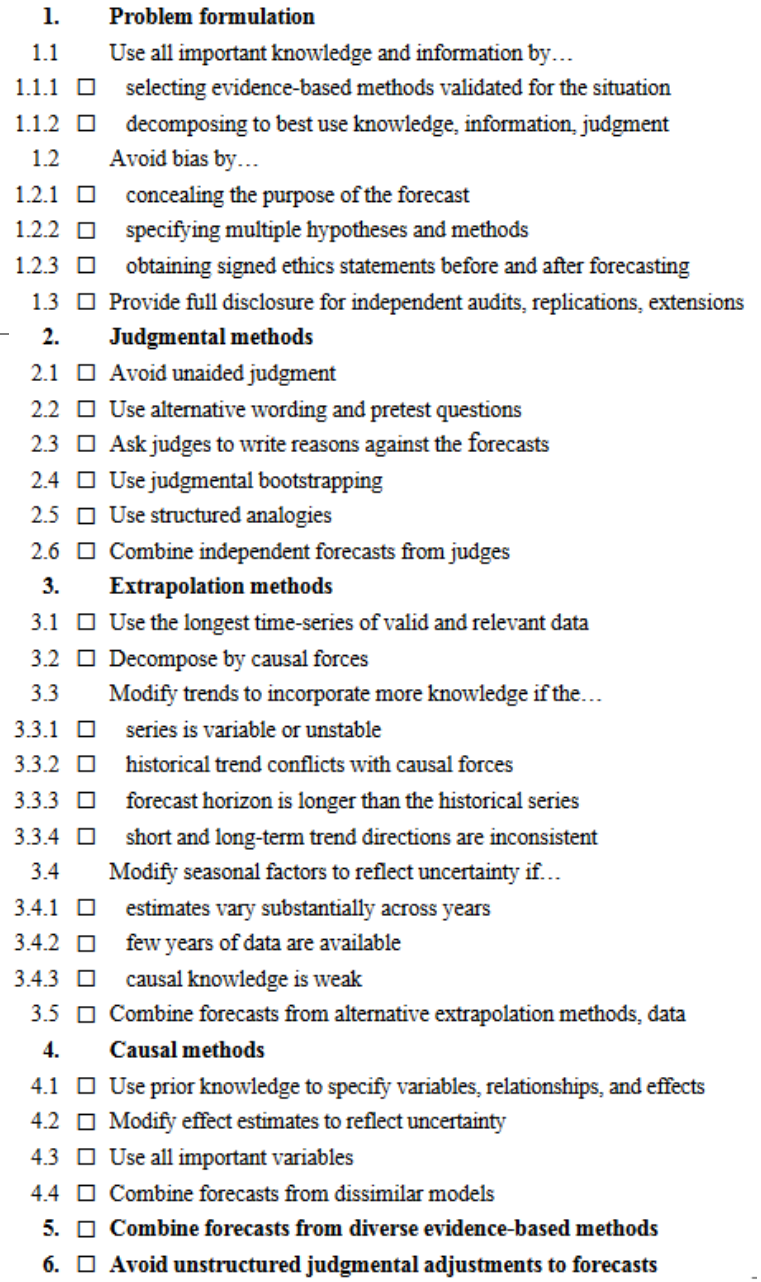
- *Causation:* Do you have an idea of what’s driving it
- *Data:* Do you have past examples to leverage?
- *Endogeneity:* Will forecasts themselves affect what you are trying to predict?

Rob Hyndman, Professor of Statistics
Monash University in Australia

Principles of Forecasting

Scott Armstrong (2001), *Principles of Forecasting: A Handbook for Researchers and Practitioners*

Based upon a group of forecasting experts developed a series of 139 principles about forecasting.



1.		**Problem formulation**
1.1		Use all important knowledge and information by…
1.1.1	☐	selecting evidence-based methods validated for the situation
1.1.2	☐	decomposing to best use knowledge, information, judgment
1.2		Avoid bias by…
1.2.1	☐	concealing the purpose of the forecast
1.2.2	☐	specifying multiple hypotheses and methods
1.2.3	☐	obtaining signed ethics statements before and after forecasting
1.3	☐	Provide full disclosure for independent audits, replications, extensions
2.		**Judgmental methods**
2.1	☐	Avoid unaided judgment
2.2	☐	Use alternative wording and pretest questions
2.3	☐	Ask judges to write reasons against the forecasts
2.4	☐	Use judgmental bootstrapping
2.5	☐	Use structured analogies
2.6	☐	Combine independent forecasts from judges
3.		**Extrapolation methods**
3.1	☐	Use the longest time-series of valid and relevant data
3.2	☐	Decompose by causal forces
3.3		Modify trends to incorporate more knowledge if the…
3.3.1	☐	series is variable or unstable
3.3.2	☐	historical trend conflicts with causal forces
3.3.3	☐	forecast horizon is longer than the historical series
3.3.4	☐	short and long-term trend directions are inconsistent
3.4		Modify seasonal factors to reflect uncertainty if…
3.4.1	☐	estimates vary substantially across years
3.4.2	☐	few years of data are available
3.4.3	☐	causal knowledge is weak
3.5	☐	Combine forecasts from alternative extrapolation methods, data
4.		**Causal methods**
4.1	☐	Use prior knowledge to specify variables, relationships, and effects
4.2	☐	Modify effect estimates to reflect uncertainty
4.3	☐	Use all important variables
4.4	☐	Combine forecasts from dissimilar models
5.	☐	**Combine forecasts from diverse evidence-based methods**
6.	☐	**Avoid unstructured judgmental adjustments to forecasts**

http://www.forecastingprinciples.com/index.php

Forecasting Global Warming

Green and Armstrong (2007)

"The IPCC predicted dramatic increases in average world temperatures. Using forecasting principles as our guide we asked: Were proper forecasting methods used?

We audited the forecasting processes described by the IPCC to assess the extent to which they complied with scientific forecasting principles. We found enough information to make judgments on 89 out of a total of 140 forecasting principles. 72 of the 89 principles were violated.

The IPCC "forecasts" are, in effect, the opinions of scientists transformed by mathematics and obscured by complex writing. "

Green and Armstrong (2014)

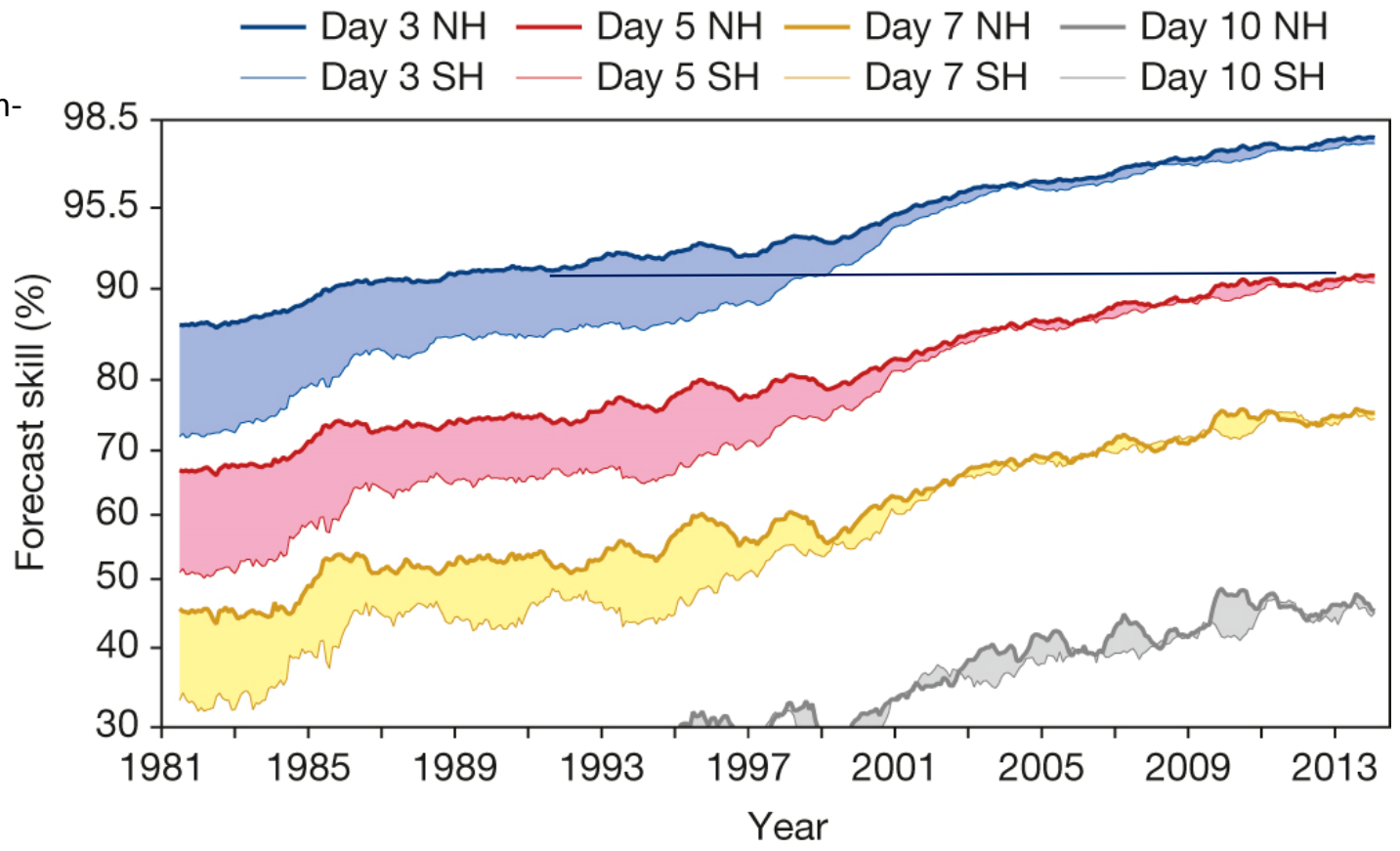
"We tested the predictive validity of the global warming hypothesis (+0.03°C per year with increasing CO2) against a relatively conservative global cooling hypothesis of -0.01°C per year, and against the even more conservative simple no-change hypothesis (0.0°C per year). The errors of forecasts from the global warming hypothesis for horizons 11 to 100 years ahead over the period 1851 to 1975 were nearly four times larger than those from the global cooling hypothesis. Findings from our tests using the latest data and other data covering a period of nearly 2,000 years support the predictive validity of the no-change hypothesis for horizons from one year to centuries ahead. "

http://www.kestencgreen.com/GR-Climate-Flyer.pdf

The Quiet Revolution of Numerical Weather Prediction

A measure of forecast skill at three-, five-, seven- and ten-day ranges, computed over the extra-tropical northern and southern hemispheres.

The accuracy of 5 day forecasts today is as good as 3 day forecasts 20 years ago



P Bauer *et al. Nature* **525**, 47-55 (2015)

http://www.nature.com/nature/journal/v525/n7567/full/nature14956.html

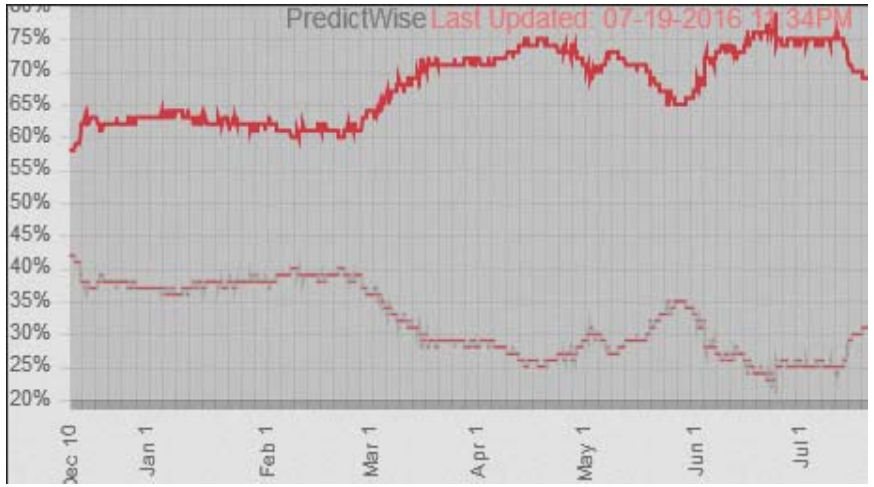
New types of Prediction Markets

Almanis

Incentivizes forecasters with points and makes money by charges those who post questions.

PredictWise

Meta-models using statistical methods to combine forecasts from prediction markets, opinion polls and bookmakers' odds



Ensemble Learning

Combining multiple learning algorithms yields better predictive performance than using any one of the individual learning algorithms.

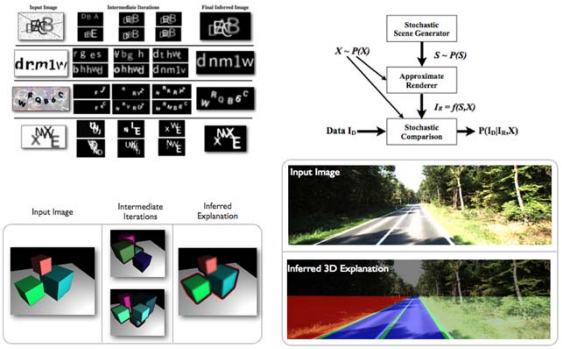
Some common methods:

- Bagging (or bootstrap aggregating) (eg, random forest)
- Boosting
- Bayesian Model Averaging
- Stacking

Probabilistic Programming

Probabilistic programs are programming languages that integrates general purpose programming with probabilistic models

- Write a program to simulate a problem of interest
- Present the inference engine with real-world data that has information about your problem
- Let the inference engine learn which random choices generate data like yours



Concluding Thoughts

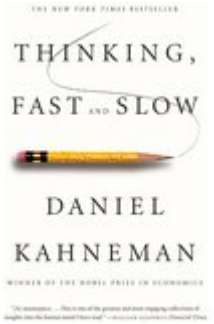
Prediction is incredibly important to decision making.

Great advances in predictive modeling and forecasting are being made.

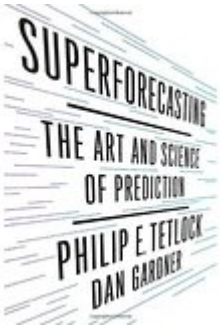
Since predictions are used by people to make decisions, we need to evaluate the potential for bias that enters because of cognitive biases

Systematic methods can help overcome them

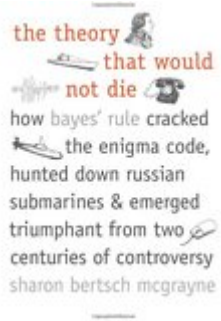
Further Reading



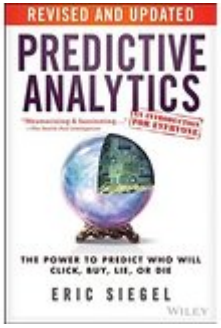
Daniel Kahneman (2013)
Thinking, Fast and Slow



Philip Tetlock and Dan Gardner (2015)
Superforecasting: The Art and Science of Prediction



Sharon Bertsch McGrayne (2012)
The Theory That Would Not Die



Eric Siegel (2016)
Predictive Analytics