Learning

ADEC781001: Empirical Behavioral Economics

Lawrence De Geest (lrdegeest.github.io)

BOSTON COLLEGE
Woods College of Advancing Studies

What is information?

- Data is often referred to as "information"
- But information has a very specific meaning:
  - "The reduction in uncertainty derived from learning an outcome" (McElreath 2015)
  - Information entropy: $H(p) = -E\log(p_i) = \sum_i^{n_i} p_i \log(p_i)$
    - entropy is maximized when probability is spread out as evenly as possible
    - this has to do with order: think of structure of water as gas (high entropy) versus ice (low entropy)
    - *A distribution can be thought of as a state of knowledge about something*

Why don't we always exploit new information?

- Learning alone: Confirmation bias, Law of Small Numbers
- Learning in groups: social networks a) create positive externalities for verifying info and b) create incentives to align with group ideals
  - Amplifies individual biases (e.g. confirmation bias)

Learning alone

Reminder on the importance of conditional probabilities

Lies, damn lies and statistics

Example 6

O.J.Simpson's defence attorney : "Only 0.1% of the men who abuse their wives end up murdering them. The fact that Simpson abused his wife is irrelevant to the case"

Why was this poor statistics? This ignores the fact that Simpson's wife was actually murdered. What is relevant is P(abusive husband is guilty | wife is murdered), not P(husband murders wife | husband abuses wife). Using reasonable data : P(guilty) ~ 0.8.

Bayes' Rule

- Bayes' Rule: a way to update to your beliefs about the probability of an event based on new information
- Consider two events A and B
- Recall $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  - Can be rewritten as $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
    - Why? Because $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- The power of Bayes' rule is that if you know $P(B|A)$ (the information) you can find $P(A|B)$
  - If I know something about the conditional probability of B, and I have prior knowledge of A (however good or bad), I can estimate the conditional probability of A

The canonical disease-test problem

The problem

- Suppose you get a test T for a disease D that has a 1% prevalence in the population
  - $P(D+) = 0.01$
- The test is not perfectly diagnostic:
  - It is correct 80% of the time
    - $P(T+|D+) = 0.8$
  - and it has a false positive rate of 10%
    - $P(T+|D-) = 0.1$
    - (Note the false positive rate is *not* the compliment of $P(T+|D+)$)

The canonical disease-test problem

The posterior

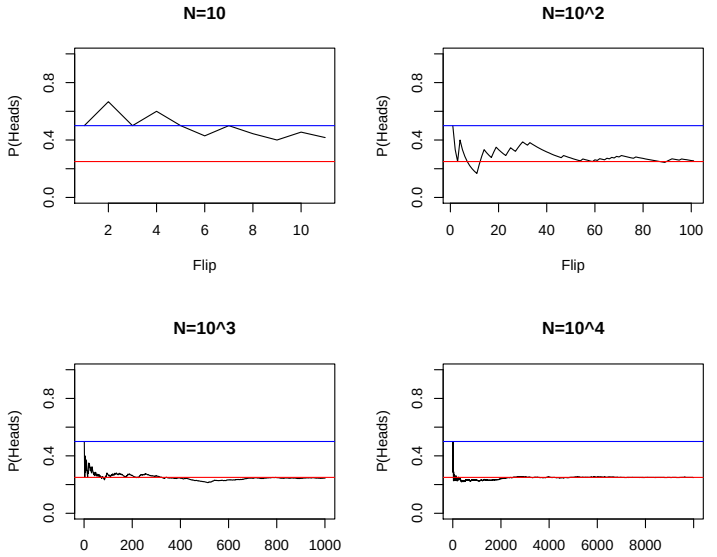
- You go in for a test. If you the test is positive, what is the probability you actually have the disease?
- What you really care about is $P(D+|T+)$, the posterior!
  - $P(D+|T+) = \frac{P(T+|D+)P(D+)}{P(T+|D+)P(D+)+P(T+|D-)P(D-)}$
  - $P(D+|T+) = \frac{(0.8)(0.01)}{(0.8)(0.01)+(0.1)(1-0.01)} = 0.07$
- Key point in the probability you have the disease *depends on its prevalence in the population* (i.e. it depends on the prior)
  - If prevalance changes, we can adjust the prior
  - If we have perfect information, we don't need the prior
    - Perfectly diagnostic test would imply we don't need a prior since $P(T+|D+) = 1$ and $P(T+|D-) = 0$

Small versus Large Data

- Bayes' Rule can be simplified to $P(A|B) \propto P(B|A)P(A)$
  - "The posterior is proportional to the likelihood times the prior"
- When modeling you have to choose distributions for $P(B|A)$ and $P(A)$
  - Life is simple when these are chosen so that the posterior is closed form (we don't have to simulate the integrals)
  - These are known as conjugate priors
  - Classic example: likelihood is Binomial distribution, prior is Beta distribution $Beta(a, b)$
  - Suppose you are interetested in frequency of outcome x (e.g. heads in a coin flip) and n is the number of flips
  - Implies posterior is $Beta(a + x, b + n - x)$
- Posterior mean will lie in between sample mean $\frac{\sum x}{n}$ and Beta mean $\frac{x+a}{n+a+b}$
  - as $n \to \infty$ the effect of the prior diminishes and the posterior converges to the sample mean

Small versus Large Data

Demo



Do people Bayesian update?

- Bayes' Rule is nice!
  - more information means we update our beliefs about the world and rely less on our prior beliefs
  - many papers begin with null hypothesis that people Bayesian update
- But evidence suggests people don't always update according to Bayes' Rule
  - Charness and Levin (2005): reinforcement learning more common than Bayesian updating
- Deryugina (2013): test for Bayesian updating using survey data on beliefs about the occurrence of the effects of global warming

Do people Bayesian update?

Deryugina (2013)

- estimate how local temperature fluctuations influence what individuals believe about these effects
- both short (1 day–2 weeks) and prolonged (1 month–12 months) periods of abnormal temperatures
- Bayesian updater should acknowledge the following:
  - Longer periods of abnormal temperatures will have a greater effect than shorter periods do
  - The more extreme are temperatures the larger are changes in beliefs
  - Within a relatively short period of time, such as a year, whether extreme temperatures occurred more or less recently should not matter
- Findings mildly support Bayesian updating
  - short-run fluctuations in temperatures over 1 day–2 week periods prior to the survey do not significantly affect beliefs
  - longer periods of abnormally warm or cold temperatures (1 month–1 year) do change the probability that respondents believe that the effects of global warming have already begun to happen
  - However, updating also consistent with other forms of updating (e.g. representativeness: beyond a 180-day period, longer periods of abnormal weather do not have a larger effect)

Confirmation Bias

Prison Study

Table 5.3 Evidence of confirmatory bias in attitudes to capital punishment. Mean evaluations on a −8 to +8 scale where +8 meant convincing and well done and −8 not convincing and not well done

Question	*Report*	*Evaluation*	
		Proponents	*Opponents*
Brief description changes my attitude to capital punishment	Pro-deterrence	1.3	0.4
	Anti-deterrence	−0.7	−0.9
How convincing the research was	Pro-deterrence	1.4	−2.1
	Anti-deterrence	−1.8	0.1
How well the research was conducted	Pro-deterrence	1.5	−2.1
	Anti-deterrence	−1.6	−0.3
Detailed description changes my attitude to capital punishment	Pro-deterrence	0.8	−0.9
	Anti-deterrence	0.7	−0.8

Source: Lord, Ross and Lepper (1979).

Confirmation Bias

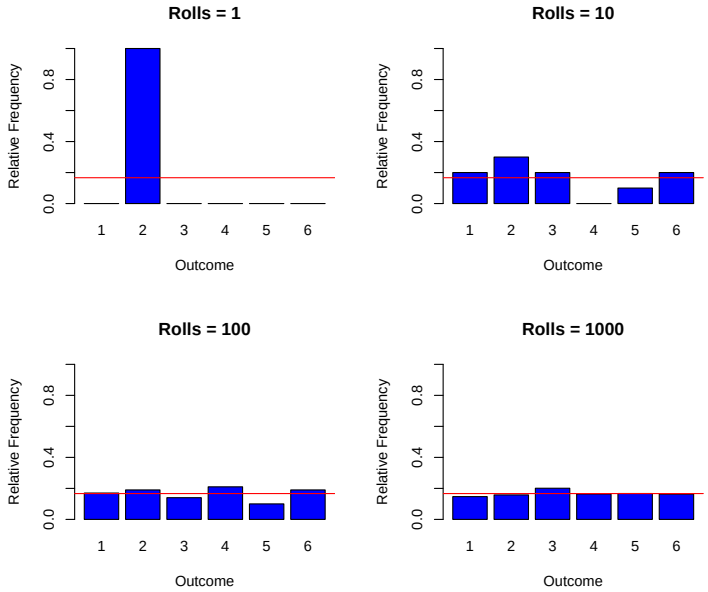
Video Study

Table 5.4 Evidence of confirmatory bias in assessing the ability of a child. Predicted grade placement goes down after watching a video if the child was perceived to be from a low income background, and goes up if she was perceived to be from a middle-class background

Watched the video or not?	*Information provided*	*Grade level placement in*		
		Mathematics	*Reading*	*Liberal arts*
Did not	Low income	3.98	3.90	3.85
	Middle class	4.30	4.29	4.03
Did	Low income	3.79	3.71	3.04
	Middle class	4.83	4.67	4.10

Source: Darley and Gross (1983).

Law of Large Numbers



Law of Small Numbers

- LLN tells us that sample size matters
  - Large samples are more representative than small samples
- But people often expect small samples to be more representative than they really are
  - this is the Law of Small Numbers: over-inferring from small samples
  - e.g. letting first impression dictate final impression
  - e.g. “housing prices will only go up!”
  - e.g. “4/5 people love the De Geest Juicer Deluxe, so out of any sample of 5 five people, 4 will love it!”
  - Kahnemann Nobel lecture: “As I understood clearly only when I taught statistics some years later, the idea that predictions should be less extreme than the information on which they are based is deeply counterintuitive.”

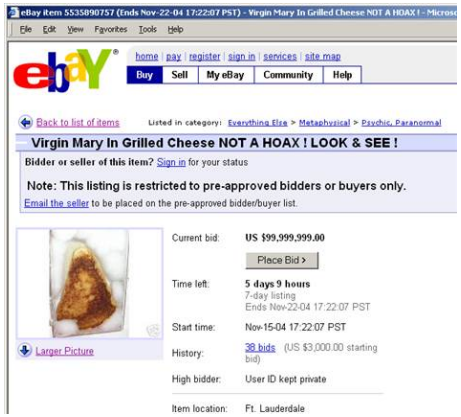
Consequences of Law of Small Numbers

- **Gambler's Fallacy**: "five heads means the next must be tails"
  - probability of outcome is known (e.g. fair coin)
  - Clotfelter and Cook (1991), Terrell (1994): Maryland vs New Jersey Lottery
    - pick three numbers and win if they are correct
    - fallacy: "the same three numbers can't win the next time"
- **Hot Hand Fallacy**: "five heads means the next will be heads"
  - probability of outcome is unknown (e.g. probability of making a three-pointer)
  - what people think: previous sequence is good predictor for next outcome
- Broader point: people can draw unmerited conclusions from small samples
  - e.g. "Mozart effect" (study with $n = 20$ subjects) or "vaccinnes cause autism" ("study" with $n = 12$ patients)
    - Hviid et al. (2019): no link between vaccines and autism (study with $n = 657,461$ patients)

What drives these biases?

Pattern (mis)recognition

- Humans are pattern-seekers and look for causal relationships
  - do not expect to observe patterns in randomness
  - but when they do, they overstate it



What drives these biases?

First impressions

- Ross et al. (1975): "...once formed, impressions are remarkably perseverent."
  - Subjects presented with pairs of suicide notes, told that one was real (they were; came from LA coroner's office), the other fake, asked to choose which was real
    - First phase: in treatment group, subjects told they were very good at picking real one (this was a lie)
    - Second phase: same subjects told their feedback was in fact a lie and learned true feedback (they weren't better than average)
    - Third phase: same subjects asked to rate how good they were compared to average (Most said they were better than average – despite just being informed the opposite)
  - "initial impressions structure and distort the attributional processes through which subsequently considered evidence is interpreted."

What drives these biases?

Heuristics

- Thinking fast and slow
  - Erikson and Mattson (1981): "How many animals of each kind did Moses take on the Ark?" "Margaret Thatcher was the president of what country?"
    - Many subjects forget that a) Noah built the Ark and b) Thatcher was prime minister, not president
    - People go with what "feels right" and miss details
    - Scharwz et al. (2016) on cognitive fluency: "when thoughts flow smoothly, people nod along"
- People overestimate their own knowledge (Sloman and Fernbach)
  - "the illusion of explanatory depth" driven by specialization
    - e.g. the toilet was well-designed so we can use it easily
    - study: ask people to rate their understanding of how a toilet works (most people rate themselves highly)
    - then: ask same people to provide step-by-step explanation of how toilet works
    - conclude: ask same people again to rate their understanding (most people reduce their rating)

What drives these biases?
Ambiguity aversion

- Two forms of uncertainty:
  - Risk: known probabilities (e.g. rolling a 6-sided die)
  - Ambiguity: unknown probabilities (e.g. rolling a ?-sided die)
- Evidence that individuals are ambiguity averse (e.g. compound lotteries)
- If new info is perceived as ambiguous, more likely to discount (and thus stick with prior)

What drives these biases?
Social learning

- If reasoning is supposed to provide sound judgements about states of the world, hard to understand why confirmation bias would not be selected out by evolution
- Alternative hypothesis: reasoning evolved for social learning (Mericer and Sperber 2017)[1]
  - we are very good at pointing out other's peoples mistakes but not our own
  - humans evolved in groups where most important things were influence and social standing
    - more important to not get screwed over than to be right
    - more important to win arguments than to be right
- Evidence of neurological benefits to confirmation bias
  - dopamine
  - neuroatypical individuals (e.g. schizophreniacs) less sucsceptible to confirmation bias

[1] Mercier, Hugo, and Dan Sperber. The enigma of reason. Harvard University Press, 2017.

Learning in groups

Example: health care

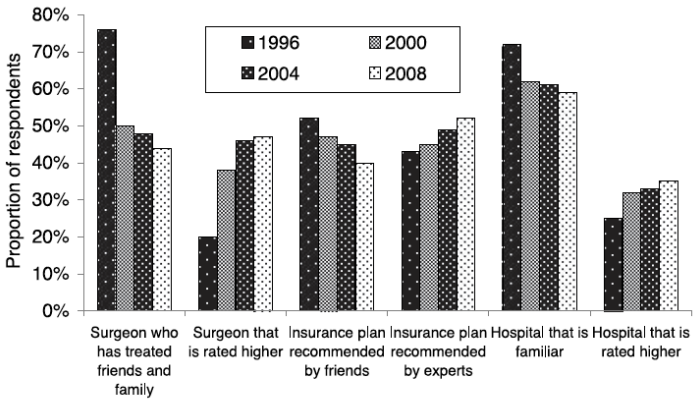


Figure 5.5 Preferences of respondents in surveys of patients in 1996, 2000, 2004 and 2008. Most people prefer things that are familiar or recommended by friends and family, rather than things recommended by experts.

Source: Kaiser Family Foundation (2008).

Fake news

- Lazer et al (2018): "Homogeneous social networks, in turn, reduce tolerance for alternative views, amplify attitudinal polarization, boost the likelihood of accepting ideologically compatible news, and increase closure to new information"
  - fact checking not always helpful
  - people tend not to question information unless it violates preconceived notions or they are incentivized
  - people align beliefs with community
  - people prefer and seek information that confirms beliefs
  - "people tend to remember information, or how they feel about it, while forgetting the context within which they encountered it."
- Vosughi (2018): lies spread faster than truth
  - fake news more likely to be novel
  - "Contrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it."

Allcott and Gentzkow (2016)

Background

- Longheld concerns about tech and information quality (cheaper presses leading to worse newspapers, television and radio, etc)
- Norms about journalism objectivity upheld by high barriers to entry that later fell with rise of internet
- Fake news arises in equilibrium because:
  - it is cheap to supply and monetize
    - setting up website is easy
    - social networks provide reach and advertising
    - firms don't care about a) truth and b) reputation
  - when feedback about true state is limited, people prefer information that aligns with priors
- Social costs to fake news
  - less accurate beliefs impose negative externalities that undermine process of electing high-quality candidates
    - High probability Trump would be not be president without fake news influence (Parkinson 2016; Read 2016; Dewey 2016)
  - increased skepticism of high-quality news
  - smaller incentive to produce high-quality news

Allcott and Gentzkow (2016)

Social media and fake news

Figure 2
Trends Related to Fake News

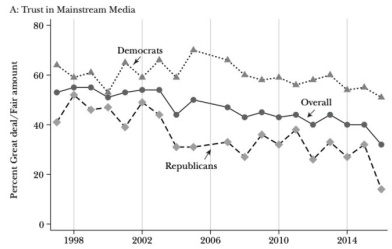
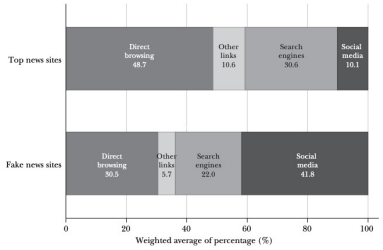


Figure 3
Share of Visits to US News Websites by Source



Allcott and Gentzkow (2016)

Data

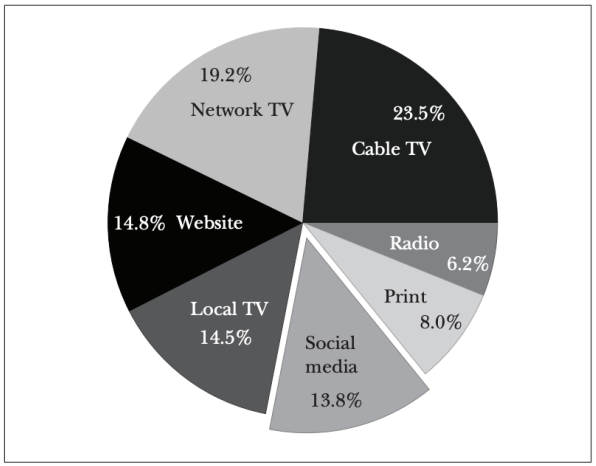
- Fake news database: 156 fake news articles circulating 3 months before 2016 election
  - scraped stories from fact-checking sites Snopes and Polifact
  - Buzzfeed-compiled fake news articles
  - 41 pro-Clinton/anti-Trump articles shared 7.6 million times on Facebook
  - 115 pro-Trump/anti-Clinton articles shared 30.3 million times on Facebook
- Post-election survey
  - online survey of 1208 adults during week of November 28th, 2016
  - collected demographics and media consumption
  - then showed each respondent 15 randomly-selected news headlines from pool about election and asked:
    - do you recall seeing this article?
    - at the time of election, would your best guess be that it is true

Allcott and Gentzkow (2016)

Results: media consumption

Figure 4

Most Important Source of 2016 Election News



Notes: Our post-election survey asked, "Which of these sources was your most important source of news and information about the 2016 election?" This figure plots responses. Observations are

Allcott and Gentzkow (2016)

Results: who believes fake news?

Table 1

What Predicts Correct Beliefs about News Headlines?

	(1)	(2)	(3)	(4)	(5)
Democrat	0.029	–0.004	0.028	–0.010	0.015
	(0.020)	(0.023)	(0.019)	(0.021)	(0.013)
Republican	–0.024	0.040	–0.037*	0.021	–0.018
	(0.024)	(0.027)	(0.020)	(0.023)	(0.014)
ln(Daily media time)			–0.002	0.042***	0.013***
			(0.007)	(0.008)	(0.004)
Social media most important			–0.066***	0.065***	–0.023
			(0.025)	(0.024)	(0.016)
Use social media			0.014	–0.023	0.002
			(0.030)	(0.038)	(0.019)
Social media ideological segregation			–0.027	0.028	–0.008
			(0.036)	(0.046)	(0.024)
Education			0.014***	0.004	0.011***
			(0.004)	(0.004)	(0.003)
Undecided			–0.011	0.006	–0.005
			(0.017)	(0.022)	(0.013)
Age			0.002***	0.000	0.002***
			(0.000)	(0.001)	(0.000)
N	12,080	6,040	12,080	6,040	18,120
p-value (Democrat = Republican)	0.029	0.124	0.004	0.207	0.035

Allcott and Gentzkow (2016)

Results: confirmation bias?

Table 2

Ideological Alignment and Belief of News Headlines

	(1)	(2)	(3)
Democrat × Pro-Clinton	0.172***		
	(0.021)		
Republican × Pro-Trump	0.147***		
	(0.023)		
Aligned		0.161***	0.096
		(0.016)	(0.140)
Aligned × Republican			0.000
			(0.027)
Aligned × ln(Daily media time)			0.024***
			(0.009)
Aligned × Social media most important			–0.031
			(0.037)
Aligned × Use social media			–0.068
			(0.050)
Aligned × Social media ideological segregation			0.147***
			(0.046)
Aligned × Education			–0.004
			(0.007)
Aligned × Undecided			–0.099***
			(0.030)
Aligned × Age			0.001
			(0.001)

Solutions?

Overcoming biases in information processing

- Follow up to 1979 capital punishment study: motivational vs cognition focused subjects
  - motivational: ask subjects to be “as objective and unbiased as possible”
  - cognition or “consider the opposite”: “Ask yourself at each step whether you would have made the same high or low evaluations had exactly the same study produced results on the other side of the issue.”
  - motivational frame preserved biases but cognition frame did not
  - takeaway: wanting to make unbiased decisions not enough
- Sloman and Fernbach: push people to explain viewpoints
  - subjects asked if there should be a single-payer health-care system and then rated their positions
  - subjects then asked to explain their position and the impacts of implementing one policy or another
  - then asked to rate position again
  - most subjects reduced the intensity of their views