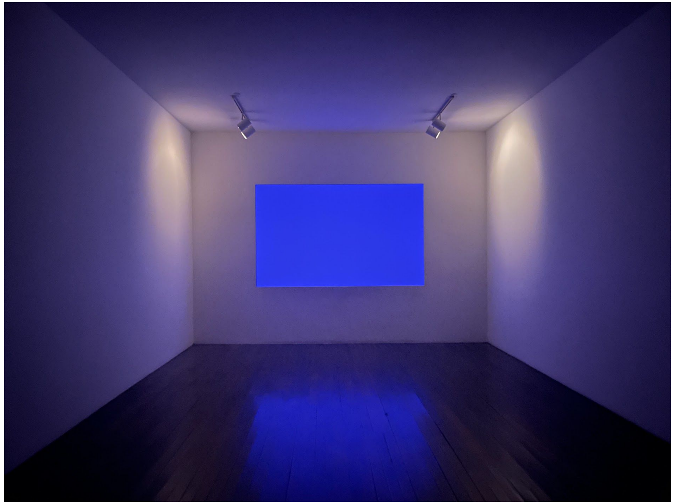


JamCoders: Week 1

Bonus Lecture: Computer Science Ethics

Material adapted from Josh Hug at UC Berkeley and Alex Krentsel at Howard.

jamcoders.org.jm

Berkeley
UNIVERSITY OF CALIFORNIA

Format

A series of Interactive Case Studies talking about some tricky ethics problems.

1. Present slides with some background
2. Ask you all for your opinion on the content
3. Discuss

Ethics

Why are we talking about this?

Computing has consequences.

CS is special.

You are powerful.

Medical Devices: Therac-25

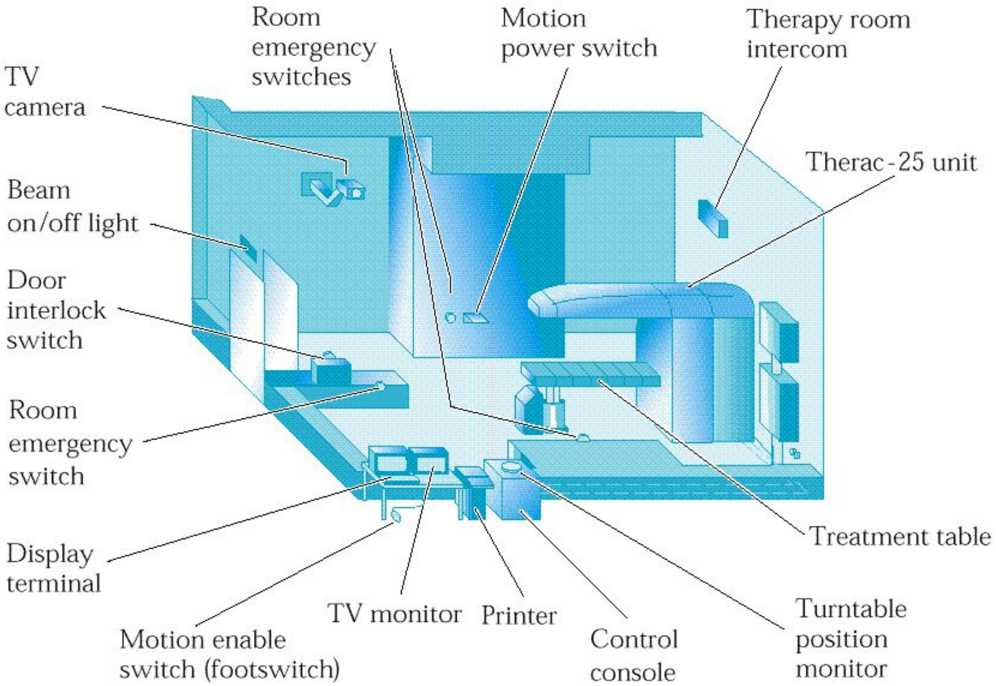
Therac-25

Radiation therapy:
Cancer is treated by damaging/killing the tissue with radiation.

- Much higher doses of radiation than imaging x-rays
- A single patient is often treated multiple times over a few weeks

Lineage:

- Therac-6
- Therac-20
- Therac-25 was first computer controlled treatment machine.

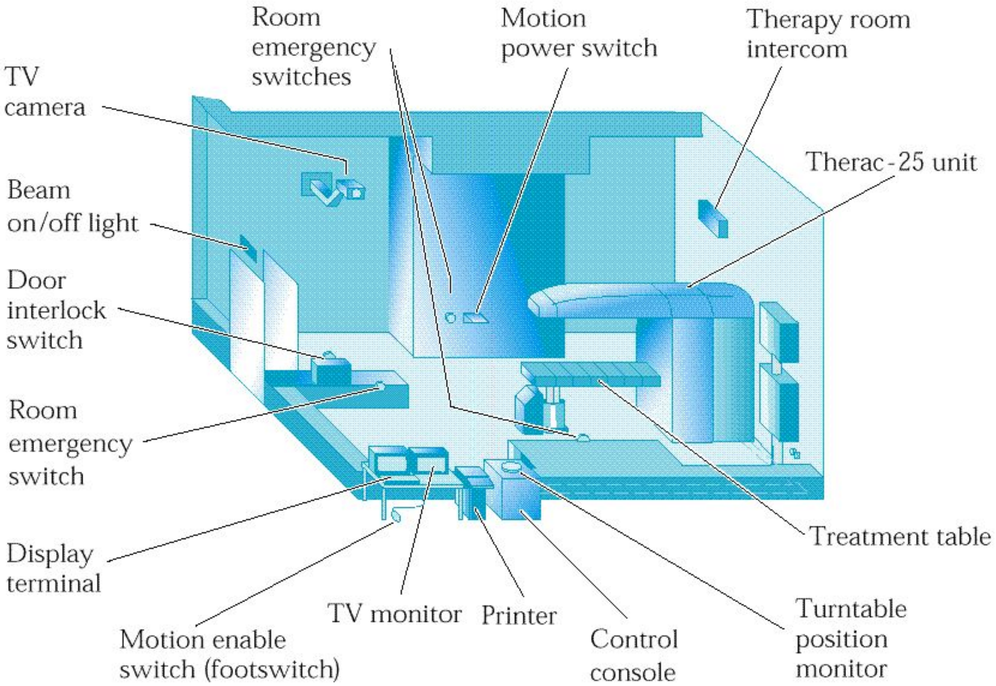


Released in 1983.

Therac-25

Impact:

- At least 100s of lives saved.
- 6 accidents, 4 deaths.



Normal dose:
80-180 rads to a small target area.

For reference, 500 rads at once to the whole body is typically fatal.

CLINAC

Therac Hardware

Therac 25 was an upgrade to Therac 20 that replaced hardware interlocks with software.

- Had two **modes**: Electron Beam and X-Ray Beam
- Interlock: Mechanism for avoiding undesirable machine states.

To pick a “mode” there were two things to coordinate (simplification):

- Beam has three states: Off, low power, high power
- Filter has three states: None, magnets, target (converts electrons to x-rays)

For electron mode: Use low power and magnets.

For x-ray mode: Use high power and target.

Therac Software

What went wrong? A few things…

- All code was written by *a single programmer*.
- During development, code was not independently reviewed by anyone

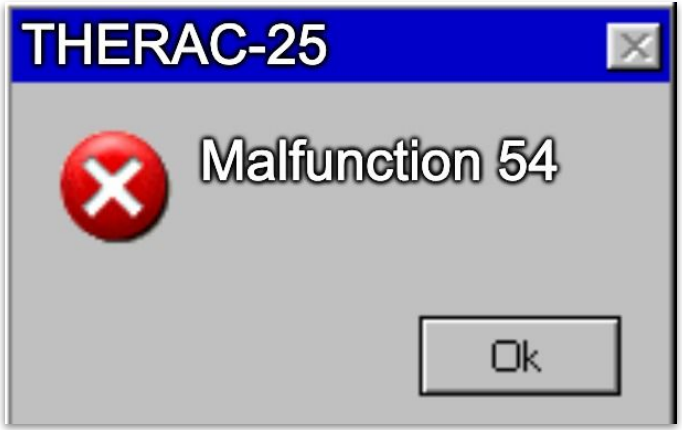
Software to run the machine was a custom operating system for PDP-11/23.

- OS scheduler had race conditions (no atomic test & set operations)
- Concurrent processes communicated through shared memory (!!)
- 8-bit counters used as signals that could overflow. 0: False, 1-255: True

Therac Software

The user interface was misleading

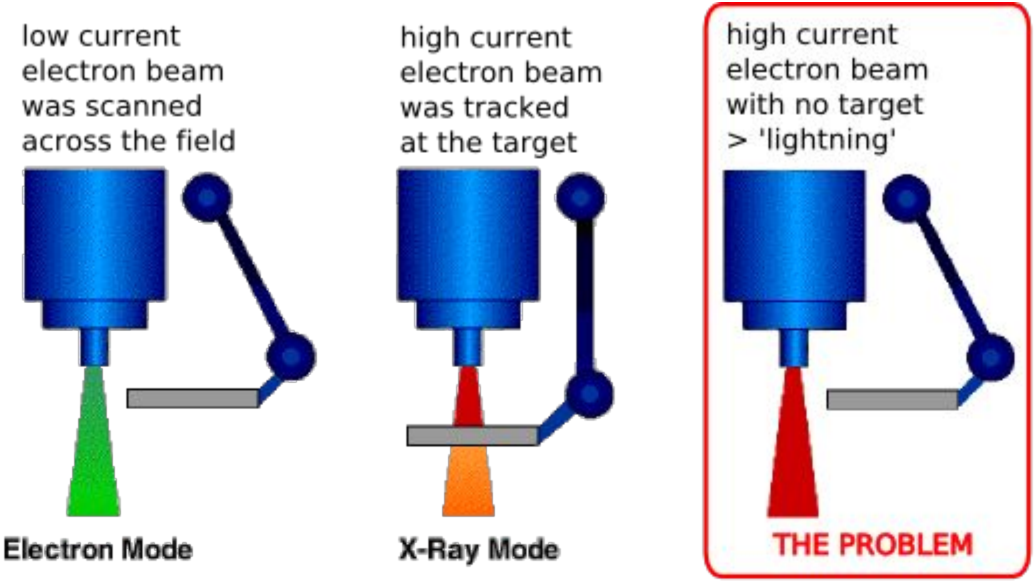
- Keyboard input accepted edit operations, but didn’t behave as expected. After editing, screen did not necessarily represent machine state.
- Default values caused undesired behavior (convenience compromised safety).
- Error messages were inscrutable.
- System paused when something went wrong, but errors were very common



THERAC-25

Malfunction 54

Therac-25 Bug


low current electron beam was scanned across the field
high current electron beam was tracked at the target
high current electron beam with no target > 'lightning'
Electron Mode
X-Ray Mode
THE PROBLEM

tray including the target, a flattening filter, the collimator jaws and an ion chamber was moved OUT for "electron" mode, and IN for "photon" mode.

Development Failures

Documentation: Development and behavior should be documented.

Concurrency: Use techniques that are easy to understand for difficult problems.

Interface: Input-output relationship should not be surprising.

Testing: A plan for how to test components and interactions should be part of design.

Auditing: Should have an independent and robust way of tracking what actually happened.

Redundancy: No single component failure should be able to cause an accident.

Discussion Question

How should we determine who is responsible for software failures that cause injury or death? Operator, Programmer, Company, Government, Patient, Doctor?

- the company – one person wrote the code, why allowing it,
- the company too
    - but also the operator?
    - company should test against a standard
- government – should have standards, prevent distribution, years of research
- the programmer themselves – you KNOW you wrote bad code?
- operator – “Malfunction 54”, there was something wrong, should have known
- hospital that implemented it -- they are the most responsible for the patient
- NEED to have timely tracking of deaths tied to the machine use
- don’t fully blame the programmer – it takes away everyone else’s responsibility, which they *do* have

Bonus Questions

- Should this depend on whether the failure was explicitly coded by a person vs. the result of a decision made by an AI algorithm trained on a dataset?

- Practically speaking, how would these cases be decided and what consequences should ensue?

Government Censorship and Surveillance

Brief History of Encryption

Modern cryptography is strong enough that it cannot be broken by anyone.

- If properly encrypted, impossible for communications to be read, altered, or modified.
- Intense debates over whether to regulate (ban!) cryptography throughout the 1990s. Ultimately, no regulation was introduced.

In 2014, as a result of Snowden disclosures, Apple updated IOS 8 so that they were physically unable to decrypt user's phones [Link].

- "There will come a day when it will matter a great deal to the lives of people . . . that we will be able to gain access" - James Comey (September 2014)

Apple vs FBI Timeline - San Bernardino Shooting

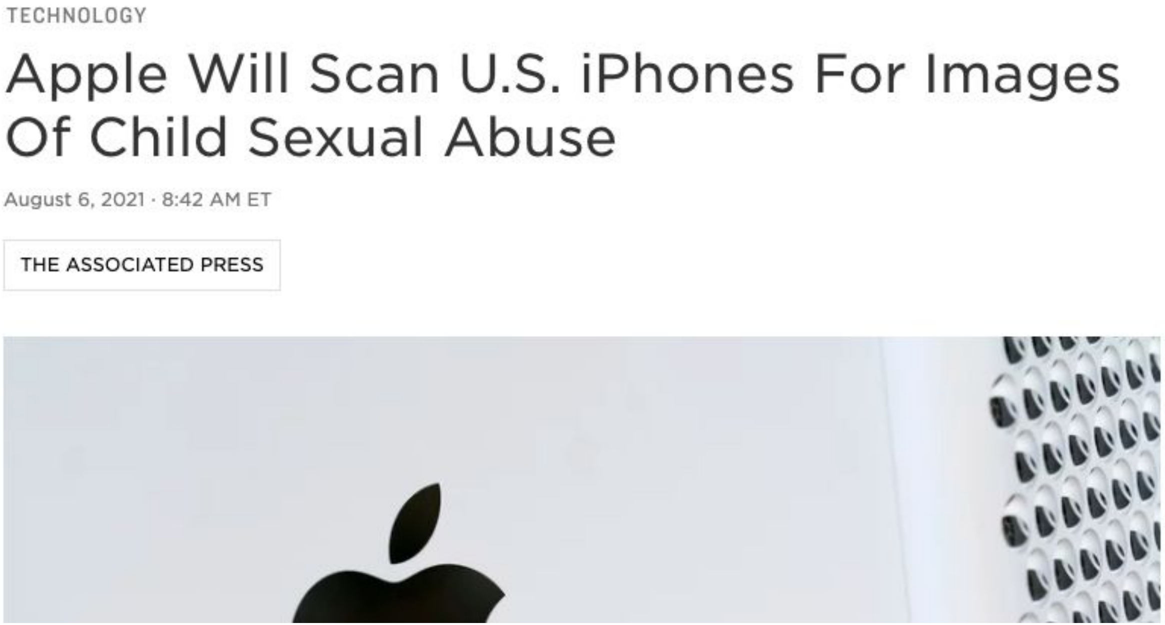
- 12/2/15: 14 people killed & 22 injured by man (phone owner) & wife.
- 2/24/16: Court orders Apple to develop a bypass of man’s phone.
- 2/25/16: Apple refuses and writes a letter to customers explaining why (Link). Court hearing scheduled for 3/22/16.
  - “Building a version of iOS that bypasses security in this way would undeniably create a backdoor. And while the government may argue that its use would be limited to this case, there is no way to guarantee such control.”
- 3/21/16: FBI requests delay in case because third party may have a way to unlock phone (Link).
- 3/28/16: FBI announced it has unlocked the phone (Link), withdraws case.
  - Cost for hacking tool: More than $1.3 million (Link (video)).
- Ultimately, the two murderers were found to have acted alone and no information of note was found on the phone.

Discussion

In 2015 and 2016, the FBI tried to force Apple to decrypt a customer's phone as part of an investigation, but Apple refused. Should Apple have complied with the court order to help the FBI unlock the iPhone?

- YES – look at bigger picture, trying to save innocent lives.
  - It could be a terrorist attack, could be beginning of whole
  - AND fix right away
  - not sure about line, but over 10 people probably enough
  - yes he’s already dead
- MAYBE – government misuse? New administration
- ABSOLUTELY not – sets the expectation that you can do this again if it’s serious enough. Authoritarian government would exploit
  - More than privacy – this is implications.
  - Unlock for “greater good” but what is that?

Apple Back in the News



TECHNOLOGY

Apple Will Scan U.S. iPhones For Images Of Child Sexual Abuse

August 6, 2021 · 8:42 AM ET

THE ASSOCIATED PRESS

link

Discussion

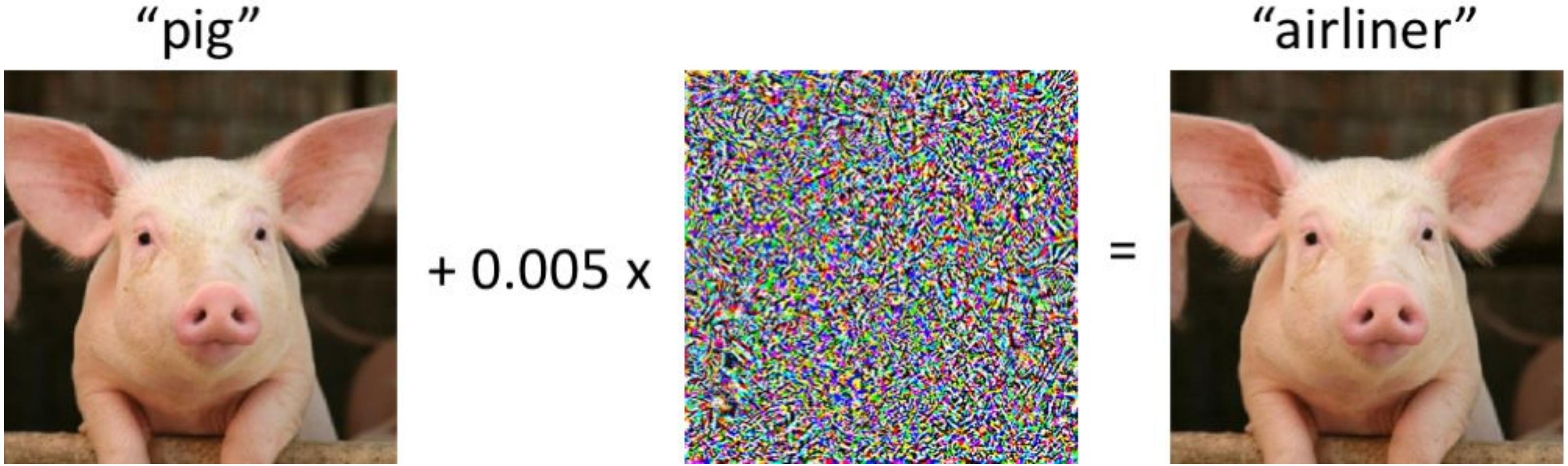
Should this be a feature built into iPhones?

- no, don’t trust apple – it’s a company
- yes, it should be allowed
  - give a little bit of privacy, get safety for population

Update - Adversarial ML

Someone reverse-engineered the ML model Apple had put out onto phones for testing.

- Able to make arbitrary images get flagged



Internet Freedom

2024 Freedom on the Net (Link)

Freedom House (a U.S. funded think tank) publishes a report every year evaluating the freedom of various countries.

- Various quantitative measures (e.g. Freedom on the Net score).
- Useful anecdotal data (e.g. specific abuses of power in various countries).

Let’s look at the Freedom on the Net map from 2024:
https://freedomhouse.org/explore-the-map?type=fotn&year=2024

Long Term Trend, from Freedom House

- Internet freedom around the world declined in 2024 **for the fourteenth consecutive year**.
- “Social media platforms have enabled the collection and analysis of vast amounts of data on entire populations.”
- “Sophisticated mass surveillance that was once feasible only for the world’s leading intelligence agencies is now affordable for a much broader range of states.”
- “ The Chinese firm Semptian has touted its Aegis surveillance system as providing “a full view to the virtual world” with the capacity to “store and analyze unlimited data.”
  - “The company even markets a “national firewall” product, mimicking the so-called Great Firewall that controls internet traffic in China.”

Chinese Internet Censorship

Internet access in China is heavily censored [Link].

- Nearly all popular western websites are blocked: https://en.wikipedia.org/wiki/List_of_websites_blocked_in_mainland_China
- Content in SMS and social media is aggressively filtered.
- Even images of text are automatically scanned by OCR software and blocked if deemed inappropriate.
  - Analysis of WeChat OCR filtering tool: https://citizenlab.ca/2018/08/cant-picture-this-an-analysis-of-image-filtering-on-wechat-moments/
  - Includes filtering of relatively innocuous ideas like the fact that Xi Jinping vaguely resembles Winnie the Pooh. Messages with images containing the words 维尼 and 领导人 were filtered.

Chinese Internet Censorship

Users face real world penalties for violations.

- An internet user sent a private message on WeChat insulting Xi Jinpeng and wishing for the end of the communist party, and was detained for 7 days [Link, including the message if you read Chinese and are curious].
- Dr. Li Wenliang “warned colleagues on social media in late December [2019] about a mysterious virus that would become the coronavirus epidemic and was detained by police in Wuhan on 3 January for “spreading false rumours”. He was forced to sign a police document to admit he had breached the law and had “seriously disrupted social order.””

Open Question

What are the negatives of China’s surveillance? What are some of the positives of China’s surveillance? Does it seem worth it to you?

-

Algorithmic Bias

What is Bias in an Algorithmic Decision?

In stats/DS/ML class, *bias* of an estimation procedure is

E[prediction] *(AKA the average predicted value)* - **truth**

What’s the problem here?

- “Truth” - for an ML model, “truth” means whatever data was shown to the model when it was being created.

A more socially relevant form of bias is due to the (often erroneous) assumption that training examples are representative of the population to which prediction is applied and that the labels of those examples are correct.

Algorithm Bias in Hiring Algorithms

Amazon had 1,335,000 employees by end of 2021. They hired ~400,000 in 2021.

To process all of their applications, back in 2014-2015, they used a machine learning recruiting tool to take a first pass through all of the resumes.

“But by 2015, the company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way.

That is because Amazon’s computer models were trained to vet applicants by observing patterns in résumés submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry.” (link)

Discussion

Should Machine Learning algorithms be restricted from making some decisions? Are they more fair or less fair than humans?

-

Political Campaigns and Polarization

Effect of Technology on Who Succeeds at Elections

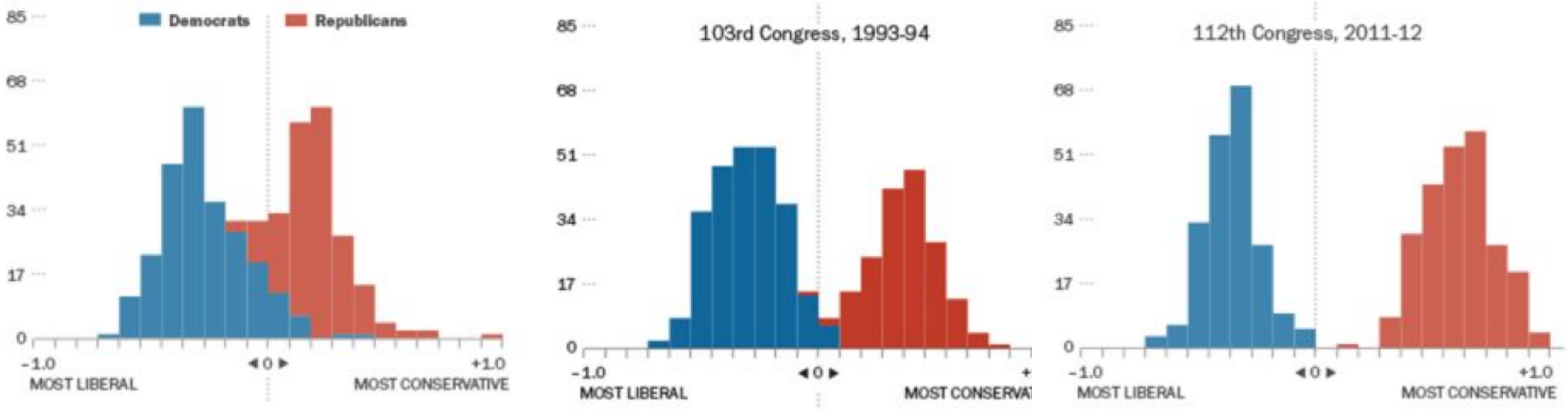
Author Nicholas Carr has a model that technology has induced 3 phase shifts in politics:

- **Radio**: Disembodied candidates, reducing them to voices. "**The blustery rhetoric that stirred big, partisan crowds came off as shrill and off-putting when piped into a living room or a kitchen**."
- **TV**: Gave candidates their bodies back. "With its jumpy cuts and pitiless close-ups, TV placed a stress on sound bites, good teeth and an easy manner. **Image became everything, as the line between politician and celebrity blurred**."
- **Social Media**: "The more visceral the message, the more quickly it circulates and the longer it holds the darting public eye. In something of a return to the pre-radio days, **the fiery populist now seems more desirable**, more worthy of attention, than the cool wonk."

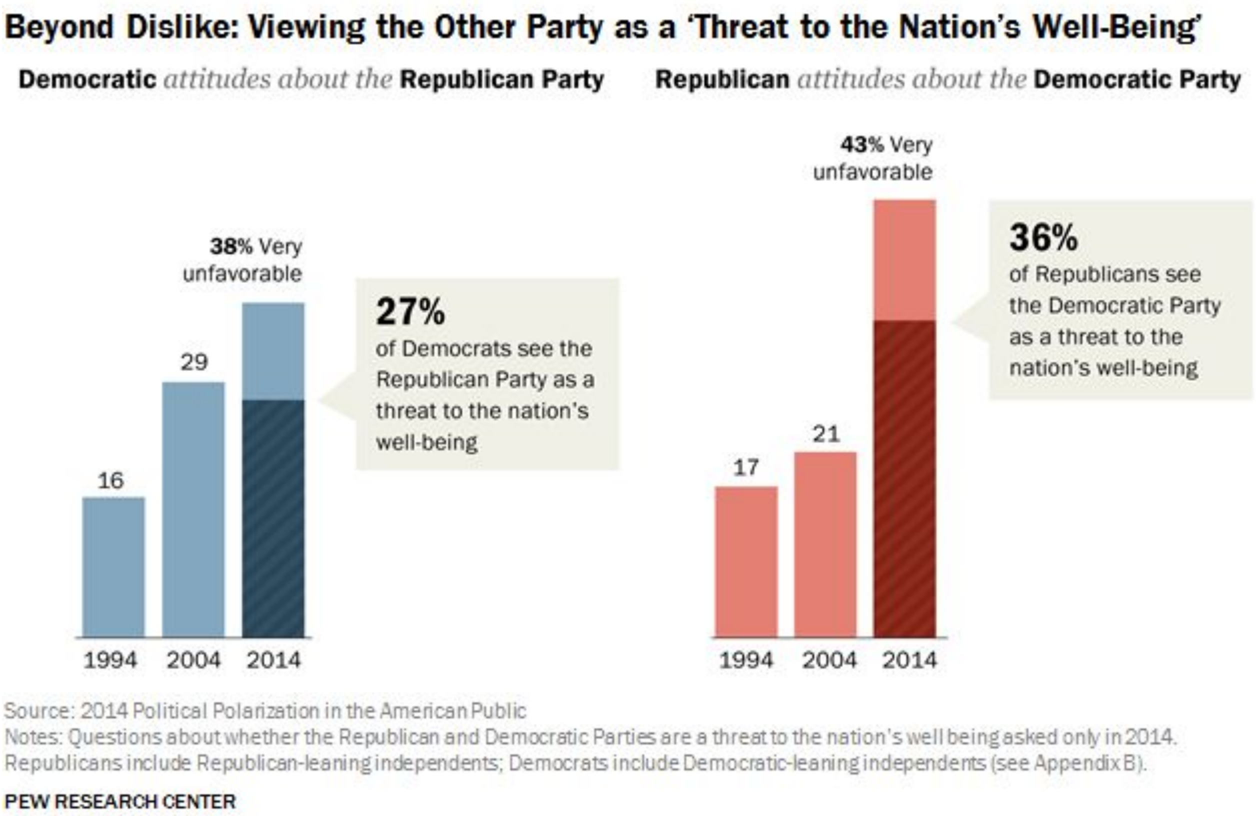
Polarization

America is highly polarized. Pew claims that American populace is more politically polarized than at any time in the last two decades.

House of representatives is similarly polarized: 1973, 1993, and 2011 (Pew)



Polarization



Beyond Dislike: Viewing the Other Party as a 'Threat to the Nation's Well-Being'

27% of Democrats see the Republican Party as a threat to the nation's well-being

36% of Republicans see the Democratic Party as a threat to the nation's well-being

Source: 2014 Political Polarization in the American Public
Notes: Questions about whether the Republican and Democratic Parties are a threat to the nation's well being asked only in 2014. Republicans include Republican-leaning independents; Democrats include Democratic-leaning independents (see Appendix B).

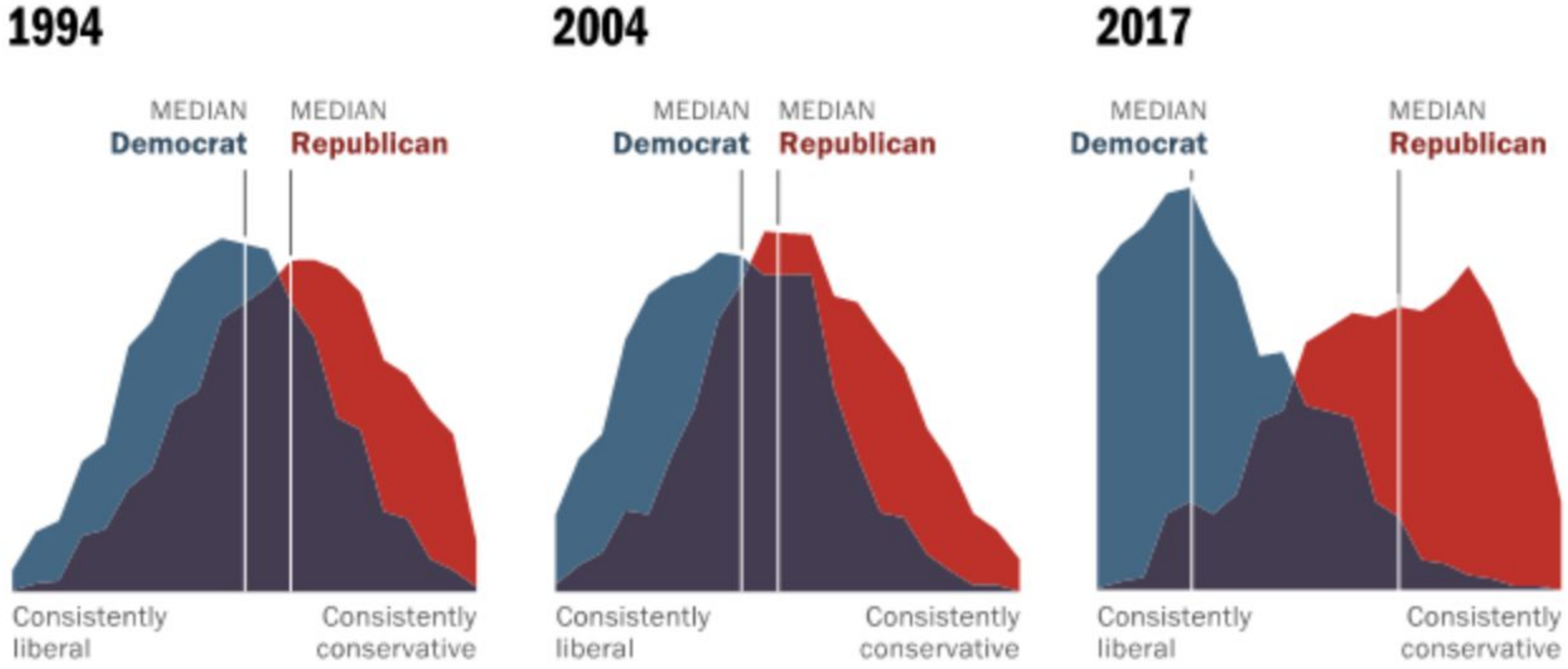
PEW RESEARCH CENTER

Source: http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/

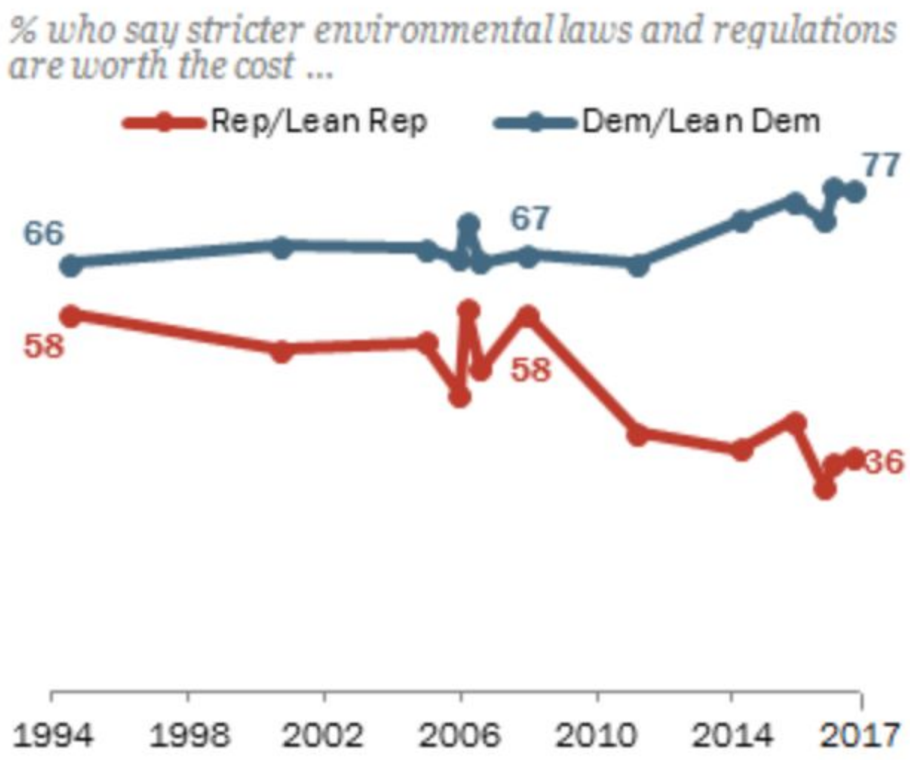
Issue Polarization (Link)

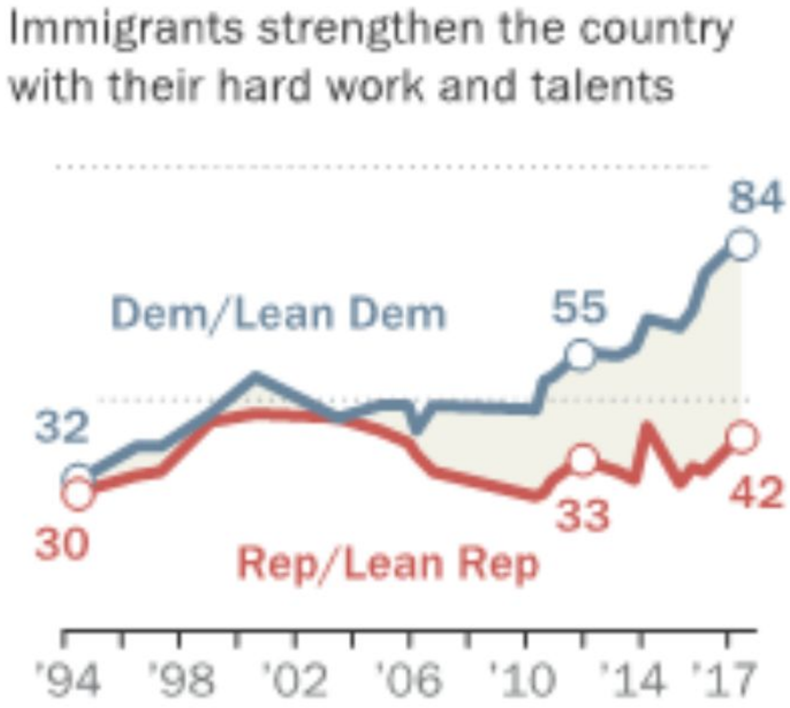
Democrats and Republicans more ideologically divided than in the past

Distribution of Democrats and Republicans on a 10-item scale of political values



Issue Polarization ([Link](), [Link]())





jamcoders.org.jm

Berkeley
UNIVERSITY OF CALIFORNIA

Information and Attention

Herbert Simon (1971): “Information … consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention effectively among the overabundance of information sources.” [Designing Organizations in an Information Rich World]

Vanessa Otero (2016): “We are living in a time where we have more information available to each of us than ever before in history. However, we are not all proficient at distinguishing between good information and bad information… I submit that these two circumstances are highly related to why our country is so politically polarized at the moment.”

Discussion

Do modern technologies reduce or increase exposure to differing opinions? Does this have a positive, negative, or insignificant effect on people's personal beliefs and on society at large?

-

Discussion

Bonus question: Are these technologies driving polarization?

-