

data-ppf.github.io Mar 16, 2021

lecture 9 of 14: big data, old school

chris wiggins + matt jones, Columbia

logistics

- ▶ revised HW schedule

logistics

- ▶ revised HW schedule
 - ▶ hw 3 mar 16 ->mar 25

logistics

- ▶ revised HW schedule
 - ▶ hw 3 mar 16 ->mar 25
 - ▶ hw 4 mar 25 ->apr 6

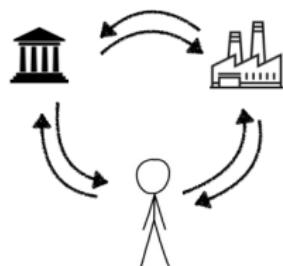
logistics

- ▶ revised HW schedule
 - ▶ hw 3 mar 16 ->mar 25
 - ▶ hw 4 mar 25 ->apr 6
 - ▶ hw 5 apr 8 ->apr 15

punchlines

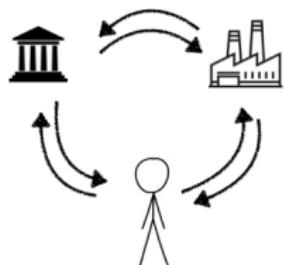
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)



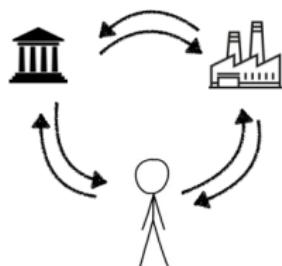
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)



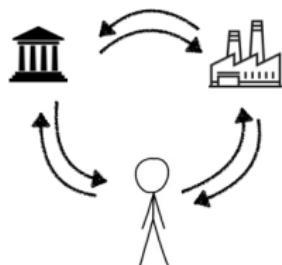
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft



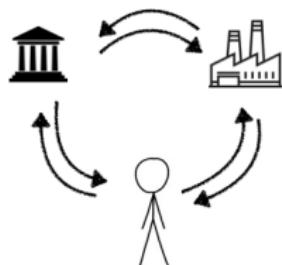
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)



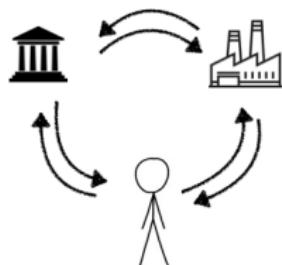
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”



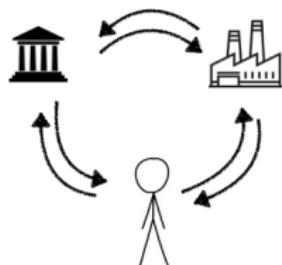
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”



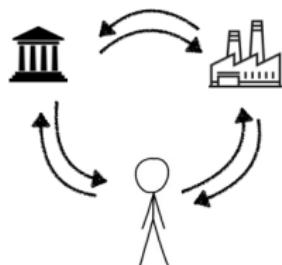
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”
- ▶ dynamics of power: people, state, corporate



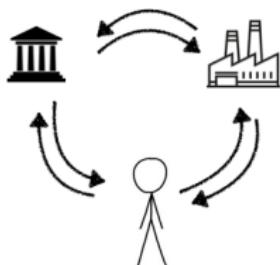
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”
- ▶ dynamics of power: people, state, corporate
 - ▶ rise in state power followed by decline in trust in state;



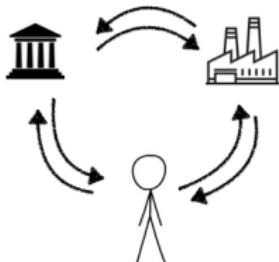
punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”
- ▶ dynamics of power: people, state, corporate
 - ▶ rise in state power followed by decline in trust in state;
 - ▶ rise in corp power followed by decline in trust in corps (now)



punchlines

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”
- ▶ dynamics of power: people, state, corporate
 - ▶ rise in state power followed by decline in trust in state;
 - ▶ rise in corp power followed by decline in trust in corps (now)
- ▶ 70s: “privacy was, as Francis Sargent, the governor of Massachusetts, put it,”good politics.” – Igo (and bipartisan!)



student reactions

87 *private/privacy*

49 surveillance

65 igo/igo's

39 government

25 companies

26 ohm/ohm's

10 1960s

9 *asimov*

6 ago

6 1970s

on Igo: “decades ago”

- ▶ many of the concerns debated about today in regard to privacy and data were being had *decades ago*.

on Igo: “decades ago”

- ▶ many of the concerns debated about today in regard to privacy and data were being had *decades ago*.
- ▶ I find it almost unfathomable how different the world was 50, 60, 70 years ago,

on Igo: “decades ago”

- ▶ many of the concerns debated about today in regard to privacy and data were being had *decades ago*.
- ▶ I find it almost unfathomable how different the world was 50, 60, 70 years ago,
- ▶ it seems that the amounts of data have continued to increase in ways unimaginable 25 years ago

on Ohm, context

like an exciting opportunity. It should have felt instead like perilous risk-taking, because it meant hurtling beyond the contextual borderlands defined by past practice.

MORE ON CUSTOMER INTELLIGENCE

[Why E-mail Marketing Is King](#)

[Turning Customer Intelligence Into Innovation](#)

[Tracking the Customer's Journey to Purchase](#)

[Pin Down Your Customer Intelligence Objectives](#)

A light blue circular icon containing the letters "PO".

Paul Ohm is an associate professor at the University of Colorado Law School. Later this year, he will join the Federal Trade Commission (FTC) as a Senior Policy Advisor.

on Ohm:

- ▶ “companies learn to say, “no” to some of the privacy-invading innovations they’re pursuing”

Ohm's solution to these potentially disastrous consequences that are coming seems to mostly lie in people just choosing to do the right thing, like industry self-regulation . . . I have a lot of skepticism about his proposed solutions; have they worked so far? I think I'm much more in favor of stronger governmental regulations like Igo described in her piece.

on Ohm:

- ▶ “companies learn to say, “no” to some of the privacy-invading innovations they’re pursuing”
- ▶ “Executives should require . . . ”

Ohm’s solution to these potentially disastrous consequences that are coming seems to mostly lie in people just choosing to do the right thing, like industry self-regulation . . . I have a lot of skepticism about his proposed solutions; have they worked so far? I think I’m much more in favor of stronger governmental regulations like Igo described in her piece.

on Ohm:

- ▶ “companies learn to say, “no” to some of the privacy-invading innovations they’re pursuing”
- ▶ “Executives should require . . . ”
- ▶ “Companies should do this . . . ”
Ohm’s solution to these potentially disastrous consequences that are coming seems to mostly lie in people just choosing to do the right thing, like industry self-regulation . . . I have a lot of skepticism about his proposed solutions; have they worked so far? I think I’m much more in favor of stronger governmental regulations like Igo described in her piece.

On Asimov:

Asimov's optimism

paid for by the DIRECT MAIL ADVERTISING ASSOCIATION...
24 page advertorial All of the ads in the section are from envelope
companies and DMAA members. labeled “Advertisement.” (pg 556)

Advertisement

The Individualism To Come

By Isaac Asimov



today: outline

- ▶ timeline (40s/50s/60s/70s/beyond)

today: outline

- ▶ timeline (40s/50s/60s/70s/beyond)
- ▶ data privacy:

today: outline

- ▶ timeline (40s/50s/60s/70s/beyond)
- ▶ data privacy:
 - ▶ laws, norms

today: outline

- ▶ timeline (40s/50s/60s/70s/beyond)
- ▶ data privacy:
 - ▶ laws, norms
 - ▶ .gov, .com (press vs PR)

today: outline

- ▶ timeline (40s/50s/60s/70s/beyond)
- ▶ data privacy:
 - ▶ laws, norms
 - ▶ .gov, .com (press vs PR)
 - ▶ definitions, principles

today: outline

- ▶ timeline (40s/50s/60s/70s/beyond)
- ▶ data privacy:
 - ▶ laws, norms
 - ▶ .gov, .com (press vs PR)
 - ▶ definitions, principles
 - ▶ .com sets norms?

today: outline

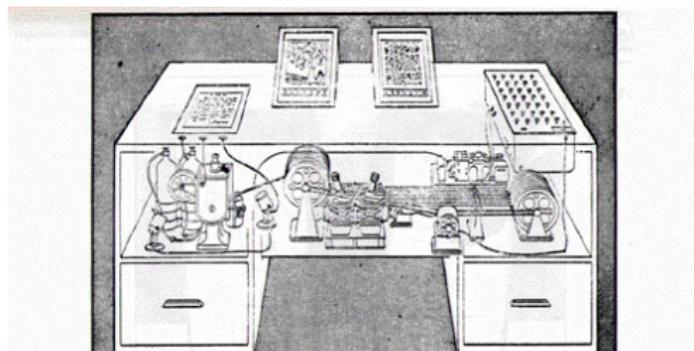
- ▶ timeline (40s/50s/60s/70s/beyond)
- ▶ data privacy:
 - ▶ laws, norms
 - ▶ .gov, .com (press vs PR)
 - ▶ definitions, principles
 - ▶ .com sets norms?
- ▶ health data (inc. COVID)

today: outline

- ▶ timeline (40s/50s/60s/70s/beyond)
- ▶ data privacy:
 - ▶ laws, norms
 - ▶ .gov, .com (press vs PR)
 - ▶ definitions, principles
 - ▶ .com sets norms?
- ▶ health data (inc. COVID)
- ▶ financial data (and learning... next week!)

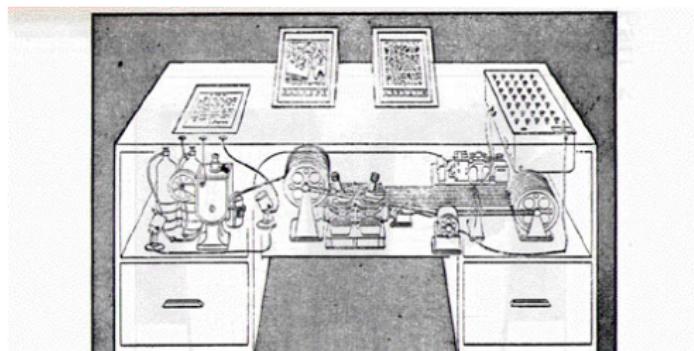
timeline: WWII demobilization

- ▶ Vannevar Bush



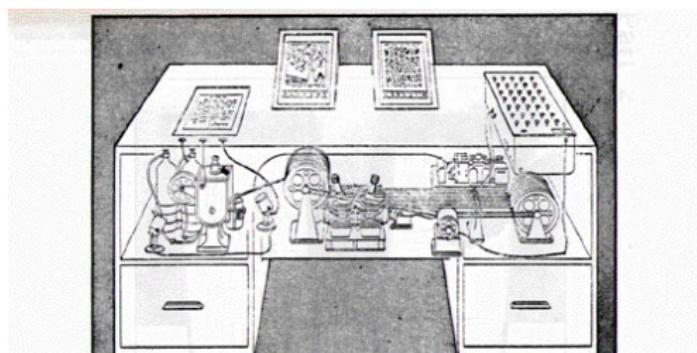
timeline: WWII demobilization

- ▶ Vannevar Bush
 - ▶ dangers of “great man” histories; see [Colin Burke](#) for more on this



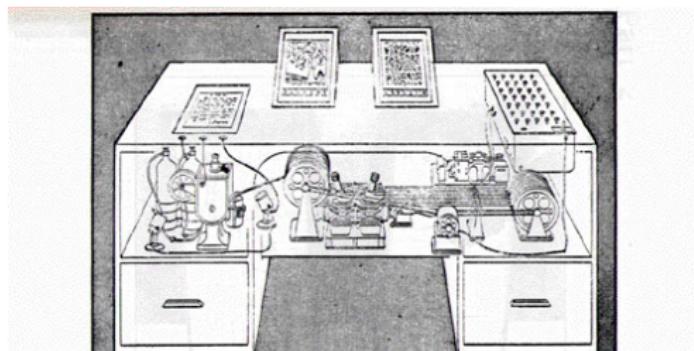
timeline: WWII demobilization

- ▶ Vannevar Bush
 - ▶ dangers of “great man” histories; see [Colin Burke](#) for more on this
 - ▶ reminder:



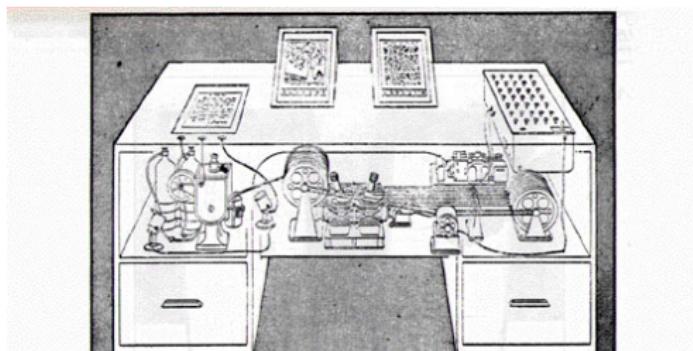
timeline: WWII demobilization

- ▶ Vannevar Bush
 - ▶ dangers of “great man” histories; see [Colin Burke](#) for more on this
 - ▶ reminder:
 - ▶ differential analyzer, funded by Weaver



timeline: WWII demobilization

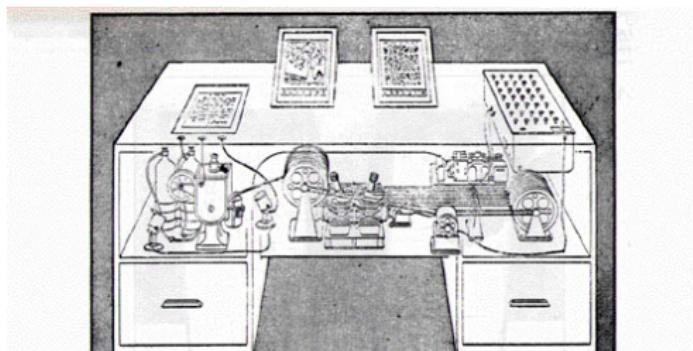
- ▶ Vannevar Bush
 - ▶ dangers of “great man” histories; see [Colin Burke](#) for more on this
 - ▶ reminder:
 - ▶ differential analyzer, funded by Weaver
 - ▶ proposed NSF



timeline: WWII demobilization

- ▶ Vannevar Bush

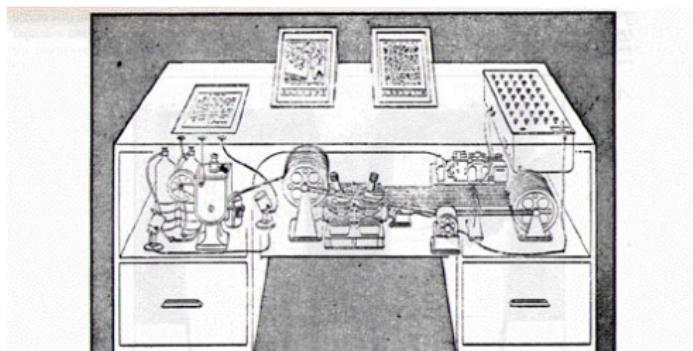
- ▶ dangers of “great man” histories; see [Colin Burke](#) for more on this
- ▶ reminder:
 - ▶ differential analyzer, funded by Weaver
 - ▶ [proposed NSF](#)
 - ▶ students included Shannon, F. Terman (son of IQ; advisor of HP)



timeline: WWII demobilization

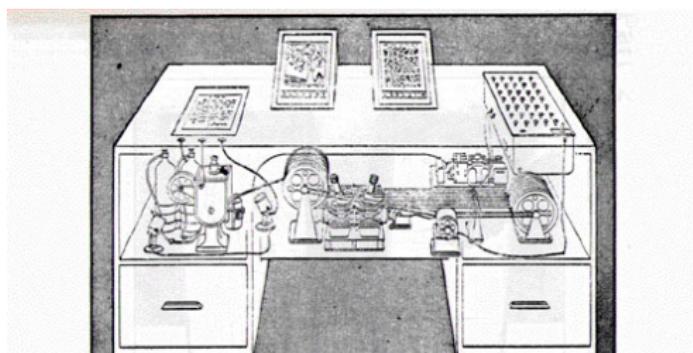
- ▶ Vannevar Bush

- ▶ dangers of “great man” histories; see [Colin Burke](#) for more on this
- ▶ reminder:
 - ▶ differential analyzer, funded by Weaver
 - ▶ [proposed NSF](#)
 - ▶ students included Shannon, F. Terman (son of IQ; advisor of HP)
 - ▶ OSRD, (3x all pre-war science funding)



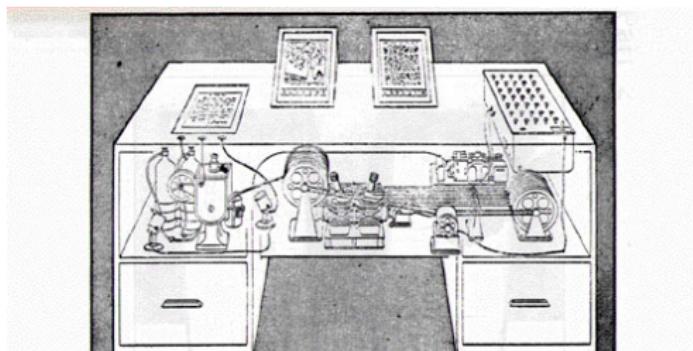
timeline: WWII demobilization

- ▶ Vannevar Bush
 - ▶ dangers of “great man” histories; see [Colin Burke](#) for more on this
 - ▶ reminder:
 - ▶ differential analyzer, funded by Weaver
 - ▶ [proposed NSF](#)
 - ▶ students included Shannon, F. Terman (son of IQ; advisor of HP)
 - ▶ OSRD, (3x all pre-war science funding)
 - ▶ HUGE document post-war effort “by end of 1947... 5M docs from Germany alone”



timeline: WWII demobilization

- ▶ Vannevar Bush
 - ▶ dangers of “great man” histories; see [Colin Burke](#) for more on this
 - ▶ reminder:
 - ▶ differential analyzer, funded by Weaver
 - ▶ [proposed NSF](#)
 - ▶ students included Shannon, F. Terman (son of IQ; advisor of HP)
 - ▶ OSRD, (3x all pre-war science funding)
 - ▶ HUGE document post-war effort “by end of 1947... 5M docs from Germany alone”
 - ▶ (related: automated MT, e.g., Soviet, Japanese, German science->English)



“As We May Think” (1943/1945) big ideas...

- ▶ futurism: “Consider a future device ... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.”

most well known for **memex** influence on hypertext

“As We May Think” (1943/1945) big ideas...

- ▶ futurism: “Consider a future device ... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.”
- ▶ information sciences: how do we organize and search info?
“When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used.”

most well known for [memex](#) influence on hypertext

“As We May Think” (1943/1945) big ideas...

- ▶ futurism: “Consider a future device ... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.”
- ▶ information sciences: how do we organize and search info?
“When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used.”
- ▶ demobilization “It has been exhilarating to work in effective partnership. Now, for many, this appears to be approaching an end. What are the scientists to do next?”

most well known for [memex](#) influence on hypertext

late 1940s: ... tech startups & Army \$\$

ENIAC (recall from “Women at the dawn...”)

- ▶ “open/high PR” entrepreneur’s journey from US Army ballistics \$ to **Census** \$



late 1940s: ... tech startups & Army \$\$

ENIAC (recall from “Women at the dawn...”)

- ▶ “open/high PR” entrepreneur’s journey from US Army ballistics \$ to **Census** \$
- ▶ 1949: Grace Hopper joins



late 1940s: ... tech startups & Army \$\$

ENIAC (recall from “Women at the dawn...”)

- ▶ “open/high PR” entrepreneur’s journey from US Army ballistics \$ to **Census** \$
- ▶ 1949: Grace Hopper joins
- ▶ 1951-03-30 ENIAC->UNIVAC “delivered” to census



late 1940s: ... tech startups & Army \$\$

ENIAC (recall from “Women at the dawn...”)

- ▶ “open/high PR” entrepreneur’s journey from US Army ballistics \$ to **Census** \$
- ▶ 1949: Grace Hopper joins
- ▶ 1951-03-30 ENIAC->UNIVAC “delivered” to census
 - ▶ **Prudential** insurance too, “study” contract \$\$



late 1940s: ... tech startups & Army \$\$

ENIAC (recall from “Women at the dawn...”)

- ▶ “open/high PR” entrepreneur’s journey from US Army ballistics \$ to **Census** \$
- ▶ 1949: Grace Hopper joins
- ▶ 1951-03-30 ENIAC->UNIVAC “delivered” to census
 - ▶ **Prudential** insurance too, “study” contract \$\$
- ▶ 1952-11-04 UNIVAC election calling



late 1940s: ... tech startups & Navy \$\$

- ▶ ERA (recall from crypto lecture)

late 1940s: . . . tech startups & Navy \$\$

- ▶ ERA (recall from crypto lecture)
 - ▶ “closed” entrepreneur’s journey from US Navy crypto to databases

late 1940s: . . . tech startups & Navy \$\$

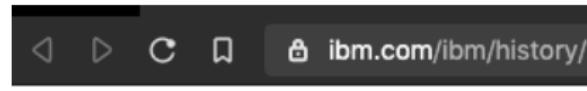
- ▶ ERA (recall from crypto lecture)
 - ▶ “closed” entrepreneur’s journey from US Navy crypto to databases
 - ▶ streams of data: surveillance presaging automated collection

late 1940s: . . . tech startups & Navy \$\$

- ▶ ERA (recall from crypto lecture)
 - ▶ “closed” entrepreneur’s journey from US Navy crypto to databases
 - ▶ streams of data: surveillance presaging automated collection
- ▶ united in Remington Rand, the (then) IBM competitor

timeline: 1950s, growth of state power (inc. compute/data)

- .mil: SAGE (picture: NORAD) drives complex info, NSA drives big compute



SAGE: A PHOTO GALLERY



07 / 11

timeline: 1950s, growth of state power (inc. compute/data)

- ▶ .mil: SAGE (picture: NORAD) drives complex info, NSA drives big compute
 - ▶ Semi-Automatic Ground Environment



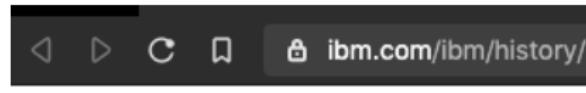
SAGE: A PHOTO GALLERY



07 / 11

timeline: 1950s, growth of state power (inc. compute/data)

- ▶ .mil: SAGE (picture: NORAD) drives complex info, NSA drives big compute
 - ▶ Semi-Automatic Ground Environment
 - ▶ .edu tie, esp MIT



SAGE: A PHOTO GALLERY



07 / 11

timeline: 1950s, growth of state power (inc. compute/data)

- ▶ .mil: SAGE (picture: NORAD) drives complex info, NSA drives big compute
 - ▶ Semi-Automatic Ground Environment
 - ▶ .edu tie, esp MIT
 - ▶ .com tie: IBM wins SAGE bid



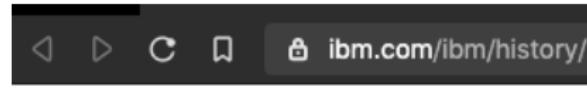
SAGE: A PHOTO GALLERY



07 / 11

timeline: 1950s, growth of state power (inc. compute/data)

- ▶ .mil: SAGE (picture: NORAD) drives complex info, NSA drives big compute
 - ▶ Semi-Automatic Ground Environment
 - ▶ .edu tie, esp MIT
 - ▶ .com tie: IBM wins SAGE bid
 - ▶ streams = automated record-keeping, presages surveillance state



SAGE: A PHOTO GALLERY



07 / 11

timeline: 1950s, growth of state power (inc. compute/data)

- ▶ .mil: SAGE (picture: NORAD) drives complex info, NSA drives big compute
 - ▶ Semi-Automatic Ground Environment
 - ▶ .edu tie, esp MIT
 - ▶ .com tie: IBM wins SAGE bid
 - ▶ streams = automated record-keeping, presages surveillance state
- ▶ .com: from records and punch cards to (digital, computational) “business intelligence”



SAGE: A PHOTO GALLERY



07 / 11

e.g., Luhn of IBM (1896-1964)

- ▶ “business”: who was his employer?



e.g., Luhn of IBM (1896-1964)

- ▶ “business”: who was his employer?
- ▶ labor: what were his patents?



e.g., Luhn of IBM (1896-1964)

- ▶ “business”: who was his employer?
- ▶ labor: what were his patents?
- ▶ impact: promoting BI to a function, centralizable, a skill



before Luhn, the deluge

who need to know it. Present methods of accomplishing this are inadequate and the general practice is to disseminate information rather broadly to be on the safe side. Since this method tends to swamp the recipients with paper, the probability of not communicating at all becomes great. The Business Intelligence System provides

Figure 3: Human problems

infosec as afterthought

There are many details which might have to be provided to adjust the general form of the system to specific applications. One such requirement might be classification, by an editor, of documents with regard to security, proprietary interests and proper utilization of information.

Figure 4: move fast, break things

5. The system is not to impose conditions on its user which require special training to obtain its services. Instead the system is to be operated by experienced library workers. Thus, in the case of an inquiry, the user will be required only to call the librarian, who will accept the query and will ask for any amplification which, in accordance with his experience, will be most helpful in securing the desired information.

Figure 5: “librarian” but “not to impose... special training”

Luhn the inventor + tinkerer

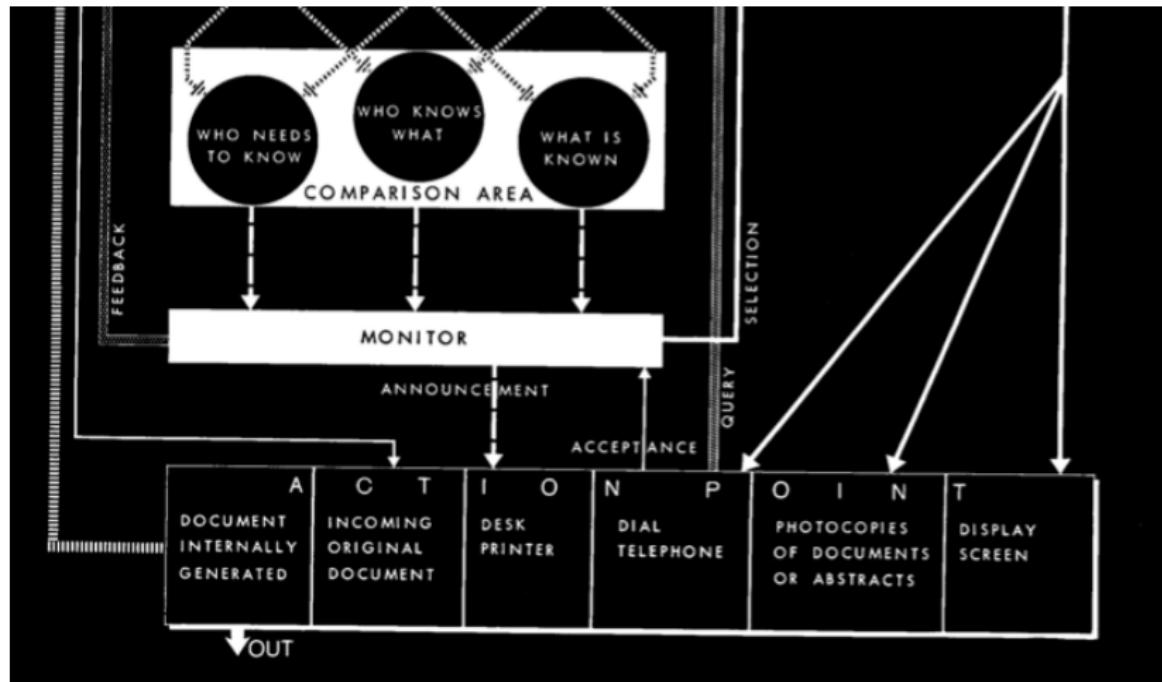


Figure 6: people are part of the information machine

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"
- ▶ expansion of business record-keeping, including

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"
- ▶ expansion of business record-keeping, including
 - ▶ earlier: payroll

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"
- ▶ expansion of business record-keeping, including
 - ▶ earlier: payroll
 - ▶ consumer data

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"
- ▶ expansion of business record-keeping, including
 - ▶ earlier: payroll
 - ▶ consumer data
 - ▶ purchasing

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"
- ▶ expansion of business record-keeping, including
 - ▶ earlier: payroll
 - ▶ consumer data
 - ▶ purchasing
 - ▶ marketing

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"
- ▶ expansion of business record-keeping, including
 - ▶ earlier: payroll
 - ▶ consumer data
 - ▶ purchasing
 - ▶ marketing
 - ▶ from SAGE to SABRE (1957/1960, via IBM)

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"
- ▶ expansion of business record-keeping, including
 - ▶ earlier: payroll
 - ▶ consumer data
 - ▶ purchasing
 - ▶ marketing
 - ▶ from SAGE to SABRE (1957/1960, via IBM)
 - ▶ Semi Automatic Ground Environment (1954, was: Whirlwind 1951)

timeline: 1960s (backdrop: Civil rights, Vietnam)

- ▶ people notice (e.g., National Data Center backlash)
- ▶ (for later: *political campaigns* notice)
 - ▶ cf. Jill Lepore's "If Then", cited in #discussion: "much of the animus against the National Data Center was plainly partisan: conservatives hated Johnson's Great Society and the rising authority of the state"
- ▶ expansion of business record-keeping, including
 - ▶ earlier: payroll
 - ▶ consumer data
 - ▶ purchasing
 - ▶ marketing
 - ▶ from SAGE to SABRE (1957/1960, via IBM)
 - ▶ Semi Automatic Ground Environment (1954, was: Whirlwind 1951)
 - ▶ Semi-Automated Business Research Environment,

timeline: 1970s (backdrop: Nixon, inc. tapping+tapes)

- ▶ people notice (from Miller 1971)...

When the Secret Service's unregulated computerized system containing dossiers on "activists," "malcontents," and "potential presidential assassins" is considered, as well as the recent disclosure that the Justice Department's civil disturbance group is maintaining an intelligence data bank, along with the Army's activities, no one should be surprised that there is concern lest the government's surveillance efforts be the genesis of a police state or a return to McCarthyism. The senator is right—a thorough ventilation of the subject is obviously needed.

Figure 7: Dossier... McCarthyism (1950s)... senatorial investigation "is needed"

sinister surveys and cybernetics (Miller 1971)

To some degree, the widespread use of computers to determine consumer appetites and voter attitudes adds new dimensions to the study of human dynamics. Unfortunately, the name of the game is not necessarily to give the citizenry what it wants; often these surveys are intended to divine a method of making palatable what industry or government already has decided to offer the public. One New Jersey firm is developing a data bank on doctors in order to enable drug companies to promote their products in a way that is suited to the habits and personality of individual doctors.³² As this illustrates, the line between the use of cybernetics to understand an individual and its use to control or affect his conduct and beliefs is shadowy at best and one that is likely to be transgressed with some frequency.

Figure 8: “understand an individual” vs “control..conduct and beliefs”

timeline: 1970s (backdrop: Nixon, inc. tapping+tapes)

- ▶ voters notice...

timeline: 1970s (backdrop: Nixon, inc. tapping+tapes)

- ▶ voters notice...
- ▶ ...and government notices

timeline: 1970s (backdrop: Nixon, inc. tapping+tapes)

- ▶ voters notice...
- ▶ ... and government notices
- ▶ (for later: *advertisers* notice)

timeline: 1970s (backdrop: Nixon, inc. tapping+tapes)

- ▶ voters notice...
- ▶ ... and government notices
- ▶ (for later: *advertisers* notice)
- ▶ privacy revolution

timeline: 1970s (backdrop: Nixon, inc. tapping+tapes)

- ▶ voters notice...
- ▶ ... and government notices
- ▶ (for later: *advertisers* notice)
- ▶ privacy revolution
 - ▶ but what exactly is privacy?

timeline: 1970s (backdrop: Nixon, inc. tapping+tapes)

- ▶ voters notice...
- ▶ ... and government notices
- ▶ (for later: *advertisers* notice)
- ▶ privacy revolution
 - ▶ but what exactly is privacy?
 - ▶ (we can't legislate what we can't define)

timeline: 1970s (backdrop: Nixon, inc. tapping+tapes)

- ▶ voters notice...
- ▶ ... and government notices
- ▶ (for later: *advertisers* notice)
- ▶ privacy revolution
 - ▶ but what exactly is privacy?
 - ▶ (we can't legislate what we can't define)
 - ▶ and what *technically* is privacy (later)

Igo: by 1974, data-woke was in

- ▶ Feb 23, 1974

But until the day comes when science finds a way of installing a conscience in every computer, we must develop human, personal safeguards that prevent computers from becoming huge, mechanical, impersonal robots that deprive us of our essential liberties.

Igo: by 1974, data-woke was in

- ▶ Feb 23, 1974
- ▶ Pres. Richard Nixon “Radio Address About the American Right of Privacy.”

But until the day comes when science finds a way of installing a conscience in every computer, we must develop human, personal safeguards that prevent computers from becoming huge, mechanical, impersonal robots that deprive us of our essential liberties.

Balancing of privacy clear

Privacy must be weighed today against the value gained from the collection and availability of information at central points or data banks . . . How much personal information is worth the convenience of a credit card?

- ▶ testimony in United States Congress Senate Committee on Government Operations Ad Hoc Subcommittee on Privacy and Information, Privacy. 1:651, June 18, 19, and 20, 1974.

1973 “Code of Fair Information Practices”

1. There must be no personal data system whose very existence is secret.

U.S. Department of Health, Education and Welfare, Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computer, and the Rights of Citizens (1973) — a gov “spec” for privacy!

1973 “Code of Fair Information Practices”

1. There must be no personal data system whose very existence is secret.
2. There must be a way for an individual to find out that information about him is in a record and how that information is to be used.

U.S. Department of Health, Education and Welfare, Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computer, and the Rights of Citizens (1973) — a gov “spec” for privacy!

1973 “Code of Fair Information Practices”

1. There must be no personal data system whose very existence is secret.
2. There must be a way for an individual to find out that information about him is in a record and how that information is to be used.
3. There must be a way for an individual to correct information about him, if it is erroneous.

U.S. Department of Health, Education and Welfare, Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computer, and the Rights of Citizens (1973) — a gov “spec” for privacy!

1973 “Code of Fair Information Practices”

1. There must be no personal data system whose very existence is secret.
2. There must be a way for an individual to find out that information about him is in a record and how that information is to be used.
3. There must be a way for an individual to correct information about him, if it is erroneous.
4. There must be a record of every significant access to any personal data in the system, including the identity of all persons and organizations to whom access has been given.

U.S. Department of Health, Education and Welfare, Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computer, and the Rights of Citizens (1973) — a gov “spec” for privacy!

1973 “Code of Fair Information Practices”

1. There must be no personal data system whose very existence is secret.
2. There must be a way for an individual to find out that information about him is in a record and how that information is to be used.
3. There must be a way for an individual to correct information about him, if it is erroneous.
4. There must be a record of every significant access to any personal data in the system, including the identity of all persons and organizations to whom access has been given.
5. There must be a way for an individual to prevent information about him collected for one purpose from being used for other purposes, without his consent.

U.S. Department of Health, Education and Welfare, Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computer, and the Rights of Citizens (1973) — a gov “spec” for privacy!

Privacy bill limited to 1) Fed 2) sectoral

- ▶ the US Federal Government affirmed *no general principle* of the protection of personal data.

Privacy bill limited to 1) Fed 2) sectoral

- ▶ the US Federal Government affirmed *no general principle* of the protection of personal data.
- ▶ provided no generalized form of *accounting* for the collection, exchange, and sale of that data.

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975
- ▶ Social Security Number Task Force of 1970

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975
- ▶ Social Security Number Task Force of 1970
- ▶ Fair Credit Reporting Act of 1970

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975
- ▶ Social Security Number Task Force of 1970
- ▶ Fair Credit Reporting Act of 1970
- ▶ Privacy Act of 1974

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975
- ▶ Social Security Number Task Force of 1970
- ▶ Fair Credit Reporting Act of 1970
- ▶ Privacy Act of 1974
- ▶ The Family Educational Rights and Privacy Act of 1974

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975
- ▶ Social Security Number Task Force of 1970
- ▶ Fair Credit Reporting Act of 1970
- ▶ Privacy Act of 1974
- ▶ The Family Educational Rights and Privacy Act of 1974
- ▶ Electronic Communications Privacy Act 1986

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975
- ▶ Social Security Number Task Force of 1970
- ▶ Fair Credit Reporting Act of 1970
- ▶ Privacy Act of 1974
- ▶ The Family Educational Rights and Privacy Act of 1974
- ▶ Electronic Communications Privacy Act 1986
- ▶ Video Privacy Protection Act of 1987

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975
- ▶ Social Security Number Task Force of 1970
- ▶ Fair Credit Reporting Act of 1970
- ▶ Privacy Act of 1974
- ▶ The Family Educational Rights and Privacy Act of 1974
- ▶ Electronic Communications Privacy Act 1986
- ▶ Video Privacy Protection Act of 1987
- ▶ Computer Matching and Privacy Protection Act of 1988

Igo: state as reformer

"sectoral approach" via acts and commissions:

- ▶ FOIA 1966/1967
- ▶ "Church" (Select Committee to Study Governmental Operations with Respect to Intelligence Activities of the United States Senate) 1975 (NSA/FBI/CIA/IRS)
- ▶ Watergate hearings
- ▶ Rockefeller Commission 1975 (CIA)
- ▶ Pike Committee (NSA/FBI/CIA) 1975
- ▶ Social Security Number Task Force of 1970
- ▶ Fair Credit Reporting Act of 1970
- ▶ Privacy Act of 1974
- ▶ The Family Educational Rights and Privacy Act of 1974
- ▶ Electronic Communications Privacy Act 1986
- ▶ Video Privacy Protection Act of 1987
- ▶ Computer Matching and Privacy Protection Act of 1988
- ▶ Drivers Privacy Protection Act of 1994

Business PR's back (1973)

imagined the writer, might beget a whole new form of individualism. Basing his discussion on the targeted mailing lists then beginning to make an appearance, Asimov reasoned, “What [the consumer] receives will be so likely to be of interest to him and to be slanted to his particular needs that, even if he does not buy, he will feel that someone has gone to the trouble of knowing what he might want.”⁷⁷ Indeed, it was perhaps only in a thoroughly computerized world that a person *could* be fully individualized. Until that day, “his wants and needs are unknown to anyone but himself and his immediate associates,” making him a “faceless nothing.”⁷⁸ The known citizen, in his futuristic scenario, was a well-cared-for citizen.

Figure 9: quote, igo

Asimov (paid to) love persuasion architectures (DMAA advert)
Asimov.. see “page” 556

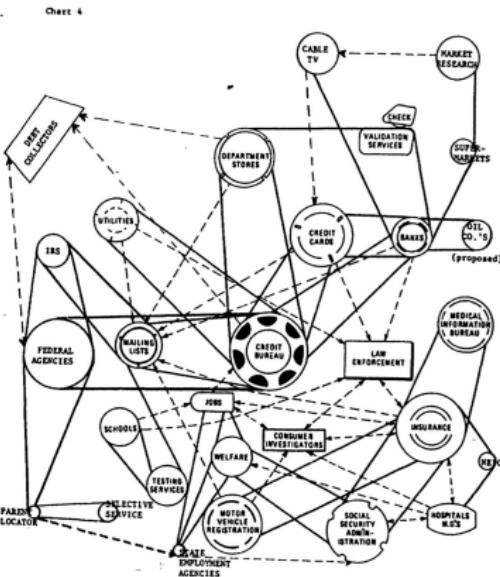
Business pushes back (1974)

Modern technology permits credit grantors to respond to consumers efficiently and rapidly partially by virtue of accessing credit information through on-line terminal facilities or alternatively by telephone Inquiries. If the free flow of information is impeded by law, the resulting inefficiencies will necessarily be translated into higher costs to industry and consumer.

- ▶ Statement Of The National Bank Americard, United States Congress Senate Committee on Government Operations Ad Hoc Subcommittee on Privacy and Information, I: 606

by 1984, concern over .com<->.gov data web

275



©1984 Robert Ellis Smith

-13-

Figure 10: 1984 hearings on "CIVIL LIBERTIES AND THE NATIONAL SECURITY STATE"

timeline: 1990s (Lesk '96)

- ▶ optimism for WWW

timeline: 1990s (Lesk '96)

- ▶ optimism for WWW
 - ▶ NB: WWW vs Internet (oh hi! it's our friend ARPA)

timeline: 1990s (Lesk '96)

- ▶ optimism for WWW
 - ▶ NB: WWW vs Internet (oh hi! it's our friend ARPA)
- ▶ (oh hi! it's our friend AI)

timeline: 1990s (Lesk '96)

- ▶ optimism for WWW
 - ▶ NB: WWW vs Internet (oh hi! it's our friend ARPA)
- ▶ (oh hi! it's our friend AI)
 - ▶ expert systems

timeline: 1990s (Lesk '96)

- ▶ optimism for WWW
 - ▶ NB: WWW vs Internet (oh hi! it's our friend ARPA)
- ▶ (oh hi! it's our friend AI
 - ▶ expert systems
 - ▶ speech recognition)

timeline: 1990s (Lesk '96)

- ▶ optimism for WWW
 - ▶ NB: WWW vs Internet (oh hi! it's our friend ARPA)
- ▶ (oh hi! it's our friend AI
 - ▶ expert systems
 - ▶ speech recognition)
- ▶ EFF and 1st amendment values: www will free us from big brother

2000s:

- ▶ surveillance capitalism

Technology

Massive camera hack exposes the growing reach and intimacy of American surveillance

A breach of the camera start-up Verkada ‘should be a wake-up call to the dangers of self-surveillance,’ one expert said: ‘Our desire for some fake sense of security is its own security threat’

Figure 11: which some people are OK with, but

2000s:

- ▶ surveillance capitalism
- ▶ surveillance as a service

Technology

Massive camera hack exposes the growing reach and intimacy of American surveillance

A breach of the camera start-up Verkada ‘should be a wake-up call to the dangers of self-surveillance,’ one expert said: ‘Our desire for some fake sense of security is its own security threat’

Figure 11: which some people are OK with, but

2000s:

- ▶ surveillance capitalism
- ▶ surveillance as a service
- ▶ always-listening IoT ('the "S" stands for security')

Technology

Massive camera hack exposes the growing reach and intimacy of American surveillance

A breach of the camera start-up Verkada 'should be a wake-up call to the dangers of self-surveillance,' one expert said: 'Our desire for some fake sense of security is its own security threat'

Figure 11: which some people are OK with, but

data privacy: from laws to norms and principles

data rhetoric: Miller on truth in tech, '71

date, and seek out more or better data. True, most information users insist that they understand that the computer's utility and a data base's reliability necessarily are limited by the quality of the input, typically emphasizing their alleged awareness by reciting the maxim "garbage in, garbage out" (GIGO). Nonetheless, the hypnotic effect of being able to manipulate enormous data bases is likely to encourage people to use the computer as an electronic security blanket and to view it as a device for quantifying the unquantifiable.

Figure 12: Miller on GIGO

Nor should we ignore the real possibility that prejudice to the individual will not end with the decision to seek an indictment. Commenting on this particular computer application, a lawyer observed: "[T]he . . . computer can tell you where the stars are going to be a million years from now. Do you think a jury is not going to believe that it can tell you where a bookie is in the Bronx?"²⁴

Figure 13: astronomy ergo predictive policing

Igo: principle of data *transparency* (distinct from privacy)

FERPA

- ▶ then: parent access to student records to correct errors

Igo: principle of data *transparency* (distinct from privacy)

FERPA

- ▶ then: parent access to student records to correct errors
 - ▶ "Unlocking School Files: The Buckley Amendment,"
Washington Post, November 17, 1974"

Igo: principle of data *transparency* (distinct from privacy)

FERPA

- ▶ then: parent access to student records to correct errors
 - ▶ "Unlocking School Files: The Buckley Amendment,"
Washington Post, November 17, 1974"
- ▶ now: non disclosure of student records to non-parents

Norms

Miller (and Igo)

- ▶ 'dossier' implies criminal

the word “dossier” conjured up a suspect citizen who had through his own activities earned the attention of authorities. Merely having a dossier implied having something to hide; in- deed, possessing “a record” was one feature distinguishing the bad citizen, or the non-citizen, from the good. For this reason, keeping tabs on law- abiding Americans, no matter that it was a routine practice of credit card companies and the IRS alike, could generate considerable unease. The headline of one negative report on the National Data Center, “There’s a Dossier on You,” tapped into just this sense of disquiet.”?

Data Privacy: who defines it?

Miller 1971: what is privacy anyway?

- ▶ independence?

Privacy is a special kind of independence, which can be understood as an attempt to secure autonomy in at least a few personal and spiritual concerns, if necessary in defiance of all the pressures of modern society. . [I]t seeks to erect an unbreachable wall of dignity and reserve against the entire world. The free man is the private man, the man who still keeps some of his thoughts and judgments entirely to himself, who feels no over-riding compulsion to share everything of value with others, not even those he loves and trusts.

Miller 1971: independence or... control?

lawyers and social scientists have been reaching the conclusion that the basic attribute of an effective right of privacy is the individual's ability to control the circulation of information relating to him—a power that often is essential to maintaining social relationships and personal freedom.

Helen Nissenbaum: the alternative “contextual integrity” 2010

- ▶ Privacy is provided by appropriate flows of information.

in facets:

Helen Nissenbaum: the alternative “contextual integrity”

2010

- ▶ Privacy is provided by appropriate flows of information.
- ▶ Appropriate information flows are those that conform with contextual information norms

in facets:

Helen Nissenbaum: the alternative “contextual integrity”

2010

- ▶ Privacy is provided by appropriate flows of information.
- ▶ Appropriate information flows are those that conform with contextual information norms
- ▶ Contextual informational norms refer to five independent parameters: data subject, sender, recipient, information type, and transmission principle

in facets:

Helen Nissenbaum: the alternative “contextual integrity”

2010

- ▶ Privacy is provided by appropriate flows of information.
- ▶ Appropriate information flows are those that conform with contextual information norms
- ▶ Contextual informational norms refer to five independent parameters: data subject, sender, recipient, information type, and transmission principle
- ▶ Conceptions of privacy are based on ethical concerns that evolve over time

in facets:

Helen Nissenbaum: the alternative “contextual integrity”

2010

- ▶ Privacy is provided by appropriate flows of information.
- ▶ Appropriate information flows are those that conform with contextual information norms
- ▶ Contextual informational norms refer to five independent parameters: data subject, sender, recipient, information type, and transmission principle
- ▶ Conceptions of privacy are based on ethical concerns that evolve over time

in facets:

1. The data subject

Helen Nissenbaum: the alternative “contextual integrity”

2010

- ▶ Privacy is provided by appropriate flows of information.
- ▶ Appropriate information flows are those that conform with contextual information norms
- ▶ Contextual informational norms refer to five independent parameters: data subject, sender, recipient, information type, and transmission principle
- ▶ Conceptions of privacy are based on ethical concerns that evolve over time

in facets:

1. The data subject
2. The sender of the data

Helen Nissenbaum: the alternative “contextual integrity”

2010

- ▶ Privacy is provided by appropriate flows of information.
- ▶ Appropriate information flows are those that conform with contextual information norms
- ▶ Contextual informational norms refer to five independent parameters: data subject, sender, recipient, information type, and transmission principle
- ▶ Conceptions of privacy are based on ethical concerns that evolve over time

in facets:

1. The data subject
2. The sender of the data
3. The recipient of the data

Helen Nissenbaum: the alternative “contextual integrity”

2010

- ▶ Privacy is provided by appropriate flows of information.
- ▶ Appropriate information flows are those that conform with contextual information norms
- ▶ Contextual informational norms refer to five independent parameters: data subject, sender, recipient, information type, and transmission principle
- ▶ Conceptions of privacy are based on ethical concerns that evolve over time

in facets:

1. The data subject
2. The sender of the data
3. The recipient of the data
4. The information type

Helen Nissenbaum: the alternative “contextual integrity”

2010

- ▶ Privacy is provided by appropriate flows of information.
- ▶ Appropriate information flows are those that conform with contextual information norms
- ▶ Contextual informational norms refer to five independent parameters: data subject, sender, recipient, information type, and transmission principle
- ▶ Conceptions of privacy are based on ethical concerns that evolve over time

in facets:

1. The data subject
2. The sender of the data
3. The recipient of the data
4. The information type
5. The transmission principle.

Data Privacy: who sets norms?

counter-claim: privacy not a norm

"You have zero privacy anyway. Get over it." - Scott McNealy, the CEO of Sun Microsystems, 1999



SCOTT McNEALY
Sun Microsystems, Inc.
Chairman & Chief Executive Officer

Eric Schmidt (Google)

- ▶ “If you have something that you don’t want anyone to know, maybe you shouldn’t be doing it in the first place.” (2009)
(‘Nothing to hide argument’)

Eric Schmidt (Google)

- ▶ “If you have something that you don’t want anyone to know, maybe you shouldn’t be doing it in the first place.” (2009)
(‘Nothing to hide argument’)
- ▶ “It is possible that [Google’s] information could be made available to the authorities.” (2009)

Eric Schmidt (Google)

- ▶ “If you have something that you don’t want anyone to know, maybe you shouldn’t be doing it in the first place.” (2009)
(‘Nothing to hide argument’)
- ▶ “It is possible that [Google’s] information could be made available to the authorities.” (2009)
- ▶ “much greater transparency and no anonymity... true anonymity is too dangerous.” (2010)

Privacy not a norm, 2010

Privacy no longer a social norm, says Facebook founder

Bobbie Johnson, Las Vegas

Mon 11 Jan 2010 01.58 GMT



▲ People have become more comfortable sharing private information online, says Facebook founder Mark Zuckerberg. Photograph: Eric Risberg/AP

The rise of social networking online means that people no longer have an expectation of privacy, according to Facebook founder [Mark Zuckerberg](#).

Talking at the Crunchie awards in San Francisco this weekend, the 25-year-old chief executive of the world's most popular social network said that privacy was no longer a "social norm".

Figure 14: zuck

data and health: Radin

- ▶ health data and digital colonialism / “Bio-colonialism”

researchers who donated the data were primarily concerned with attribution — making certain that their work and funders were credited properly — not about compensating patients, whose data had been made anonymous and was therefore regarded as protected.

data and health: Radin

- ▶ health data and digital colonialism / “Bio-colonialism”
- ▶ “invisible labor” of big data, hidden even the word ‘data’!
researchers who donated the data were primarily concerned with attribution — making certain that their work and funders were credited properly — not about compensating patients, whose data had been made anonymous and was therefore regarded as protected.

data and health: COVID

- ▶ insightful infographics abound

data and health: COVID

- ▶ insightful infographics abound
 - ▶ [Our world in data has a gallery](#)

data and health: COVID

- ▶ insightful infographics abound
 - ▶ Our world in data has a gallery
 - ▶ NYT Graphics, dynamic dashboard

data and health: COVID

- ▶ insightful infographics abound
 - ▶ Our world in data has a gallery
 - ▶ NYT Graphics, dynamic dashboard
- ▶ Self-reporting apps are launching, e.g.,

data and health: COVID

- ▶ insightful infographics abound
 - ▶ Our world in data has a gallery
 - ▶ NYT Graphics, dynamic dashboard
- ▶ Self-reporting apps are launching, e.g.,
 - ▶ Kings College London

data and health: COVID

- ▶ insightful infographics abound
 - ▶ Our world in data has a gallery
 - ▶ NYT Graphics, dynamic dashboard
- ▶ Self-reporting apps are launching, e.g.,
 - ▶ Kings College London
- ▶ Academics join in as well, e.g.,

data and health: COVID

- ▶ insightful infographics abound
 - ▶ Our world in data has a gallery
 - ▶ NYT Graphics, dynamic dashboard
- ▶ Self-reporting apps are launching, e.g.,
 - ▶ Kings College London
- ▶ Academics join in as well, e.g.,
 - ▶ 19 February 2020 An interactive web-based dashboard to track COVID-19 in real time

privacy and COVID

privacy and COVID

- ▶ JHU dashboard ; these data now hosted by google, too

privacy and COVID

- ▶ JHU dashboard ; these data now hosted by google, too
- ▶ sharing data and code on GitHub, UK, US, CN, including “social” data, hosting apps

privacy and COVID

- ▶ JHU dashboard ; these data now hosted by google, too
- ▶ sharing data and code on GitHub, UK, US, CN, including “social” data, hosting apps
- ▶ Citizen scientists too

options differ; opinions surprise

- ▶ Pinboard guy opens [We Need A Massive Surveillance Program:](#) (Mar 23)

I am a privacy activist who has been riding a variety of high horses about the dangers of permanent, ubiquitous data collection since 2012. But warning people about these dangers today is like being concerned about black mold growing in the basement when the house is on fire.

options differ; opinions surprise

- ▶ Pinboard guy opens [We Need A Massive Surveillance Program](#): (Mar 23)
I am a privacy activist who has been riding a variety of high horses about the dangers of permanent, ubiquitous data collection since 2012. But warning people about these dangers today is like being concerned about black mold growing in the basement when the house is on fire.
- ▶ Former philosophy/ethics prof: "[If I Were the Government or Big Tech Looking to Destroy Privacy, I'd Invent the Coronavirus](#)" (Mar 24)

options differ; opinions surprise

- ▶ Pinboard guy opens [We Need A Massive Surveillance Program](#): (Mar 23)
I am a privacy activist who has been riding a variety of high horses about the dangers of permanent, ubiquitous data collection since 2012. But warning people about these dangers today is like being concerned about black mold growing in the basement when the house is on fire.
- ▶ Former philosophy/ethics prof: "[If I Were the Government or Big Tech Looking to Destroy Privacy, I'd Invent the Coronavirus](#)" (Mar 24)
- ▶ et tu, Zoom?: [Access Now urges transparency from Zoom on privacy and security](#) (Human rights NGO; Mar 19)

data, epidemics and control/policy: a long tradition

- ▶ E.g., [A decision-support framework to optimize border control for global outbreak mitigation](#), 2019

data, epidemics and control/policy: a long tradition

- ▶ E.g., [A decision-support framework to optimize border control for global outbreak mitigation](#), 2019
- ▶ E.g., from [SARS 2003](#)

data, epidemics and control/policy: a long tradition

- ▶ E.g., [A decision-support framework to optimize border control for global outbreak mitigation](#), 2019
- ▶ E.g., from [SARS 2003](#)
- ▶ E.g., R_0 , the "[Basic reproduction number](#)", traces to models of the early 1900s

Finance data, granularity, and models

Poon on FICO (why saved for last...)

- ▶ end with FICO because it's a living fossil

- ▶ credit became statistical (learned coefficients)

They compare the USA [and] Russia... [where] credit cards are distributed on the basis of subjective evaluations, on social networks and on 'trust' (i.e., demonstrable friendships, kinship ties and employment), while in the USA, a mature market, institutions such as credit bureaus make calculative credit scoring practices feasible.

- ▶ credit became statistical (learned coefficients)
- ▶ was led by biz not academia

They compare the USA [and] Russia... [where] credit cards are distributed on the basis of subjective evaluations, on social networks and on 'trust' (i.e., demonstrable friendships, kinship ties and employment), while in the USA, a mature market, institutions such as credit bureaus make calculative credit scoring practices feasible.

- ▶ credit became statistical (learned coefficients)
- ▶ was led by biz not academia
 - ▶ cf. death tables, Gaunt, etc

They compare the USA [and] Russia... [where] credit cards are distributed on the basis of subjective evaluations, on social networks and on 'trust' (i.e., demonstrable friendships, kinship ties and employment), while in the USA, a mature market, institutions such as credit bureaus make calculative credit scoring practices feasible.

- ▶ credit became statistical (learned coefficients)
- ▶ was led by biz not academia
 - ▶ cf. death tables, Gaunt, etc
- ▶ was not immediately adopted relative to craft

They compare the USA [and] Russia... [where] credit cards are distributed on the basis of subjective evaluations, on social networks and on 'trust' (i.e., demonstrable friendships, kinship ties and employment), while in the USA, a mature market, institutions such as credit bureaus make calculative credit scoring practices feasible.

- ▶ credit became statistical (learned coefficients)
- ▶ was led by biz not academia
 - ▶ cf. death tables, Gaunt, etc
- ▶ was not immediately adopted relative to craft
 - ▶ cf. death tables, Gaunt, etc

They compare the USA [and] Russia... [where] credit cards are distributed on the basis of subjective evaluations, on social networks and on 'trust' (i.e., demonstrable friendships, kinship ties and employment), while in the USA, a mature market, institutions such as credit bureaus make calculative credit scoring practices feasible.

Poon on Labor

This was a two-stage process performed by housewives working at piece rates of a few cents per sample application out of their homes. The meticulous work of 'the home-coders' was the backbone of the scorecard since it was their job to interpret the writing on the ledger cards and reliably convert it into the standardized numerical codes demanded by the analytic process. Codes were transferred to paper, reviewed for accuracy by a woman assigned as a 'checker', and subsequently transferred to punch cards so that the data could be read by machine. As one of the women who headed coding described it, a punch card machine is 'like a typewriter, you put your IBM cards in — they're about five by seven — and you have to sort them. If we punched a certain digit that would mean [occupation]: housewife' (Senior coder A).

Poon on Labor, cont.

Because of its repetitive and mechanical nature, coding was considered a mundane task in the company. Yet upon scrutiny it is clear that the work that was done involved its own form of skilled decision-making that was far from obvious. A former coder made clear to me that '[t]here was some interpretation on all of this. You couldn't just copy it. That was the hard part, coding it. [. . .] They didn't just say he's been three times thirty days late in nice English' (Senior coder B). Another drew attention to the fact that, "We had to read these logs of payments and every company didn't do the same thing, and we'd get so confused' (Senior coder C).

The Wells system involved teams of women working with boxes of punch cards and 42 pound Fridan [sic] calculators.

Poon: role of state

[if you] wonder why such flawed techniques continue[d] to prevail in the lending industry, the author points to the role of the state in sanctioning these methods,

- ▶ cf., e.g., p-values

each time statistical practices have been introduced to a problem in a substantive domain, expertise, networks of associations, technical objects and even new interpretations of probability must be formed to accommodate this extension (Desrosières, 2000; Gigerenzer et al., 1989; Hackинг, 1975; Porter, 1988). If statistical theories change as they travel, and if the places they go must be rebuilt and rearranged to fit to accommodate them in practice, then it is from the details of this mutual refitting that novel calculative effects must emerge.

- ▶ where have we seen this before?

Poon: subjective design choices

*the very distinction between 'good' and 'bad' was flexible.
... [e.g.,] one missed payment, two missed payments in
12 months, not paying at all for three months...) and
policies on what was considered an account in default varied
between firms.*

*just how many files, from how narrow and recent a time
period in a firm's history were necessary to build a repre-
sentative model.*

Poon, summary

- ▶ finance data, e.g., Poon on FICO (Fair, Issacs. . .)

Poon, summary

- ▶ finance data, e.g., Poon on FICO (Fair, Issacs. . .)
 - ▶ founded 1956

Poon, summary

- ▶ finance data, e.g., Poon on FICO (Fair, Issacs. . .)
 - ▶ founded 1956
 - ▶ FCRA 1970

Poon, summary

- ▶ finance data, e.g., Poon on FICO (Fair, Issacs. . .)
 - ▶ founded 1956
 - ▶ FCRA 1970
 - ▶ key idea: in the data vs in the “model”

Poon, summary

- ▶ finance data, e.g., Poon on FICO (Fair, Issacs. . .)
 - ▶ founded 1956
 - ▶ FCRA 1970
 - ▶ key idea: in the data vs in the “model”
 - ▶ (next week!)

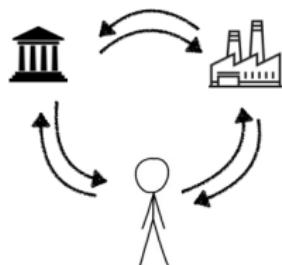
power and principles

how did this capability rearrange power? who can now do what, from what, to whom?

role of rights, harms, justice?

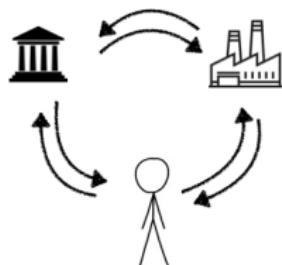
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)



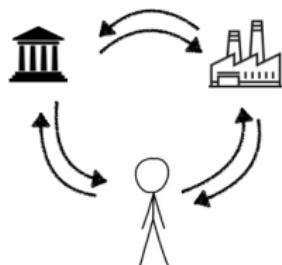
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)



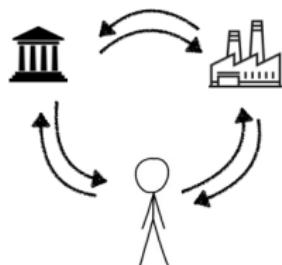
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft



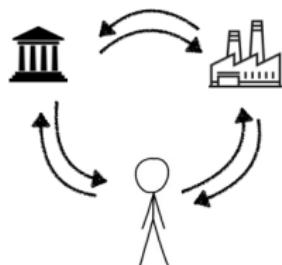
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)



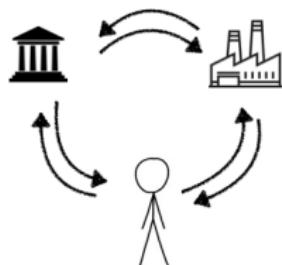
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”



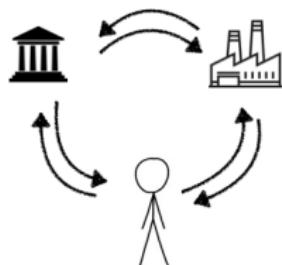
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”



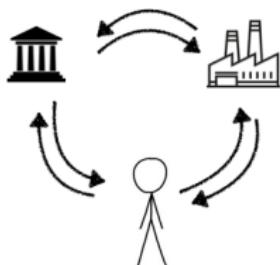
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”
- ▶ dynamics of power: people, state, corporate



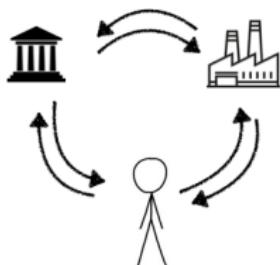
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”
- ▶ dynamics of power: people, state, corporate
 - ▶ rise in state power followed by decline in trust in state;



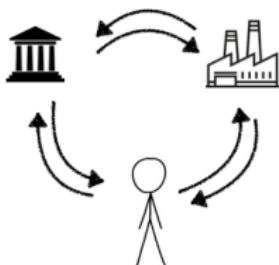
reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”
- ▶ dynamics of power: people, state, corporate
 - ▶ rise in state power followed by decline in trust in state;
 - ▶ rise in corp power followed by decline in trust in corps (now)



reminder: themes for today

- ▶ computers “scale up” both “compute” and “data” (and more is different)
 - ▶ .mil: crypto streams, logistics streams (e.g., SAGE)
 - ▶ .gov: vulgar quantitative statecraft
 - ▶ .com: efficiency, then *learning* (e.g., FICO)
- ▶ database is itself something with “consequences”
 - ▶ technologies have capabilities, but a database can be “ruinous”
- ▶ dynamics of power: people, state, corporate
 - ▶ rise in state power followed by decline in trust in state;
 - ▶ rise in corp power followed by decline in trust in corps (now)
- ▶ 70s: “privacy was, as Francis Sargent, the governor of Massachusetts, put it,”good politics.” – Igo (and bipartisan!)



Appendix

Appendix

- ▶ 2021-01-12: intro to course

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-30: AI2.0 = ML, 1959-2015

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-30: AI2.0 = ML, 1959-2015
- ▶ 2021-03-23: data science, 1962-2017

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-30: AI2.0 = ML, 1959-2015
- ▶ 2021-03-23: data science, 1962-2017
- ▶ 2021-04-06: ethics

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-30: AI2.0 = ML, 1959-2015
- ▶ 2021-03-23: data science, 1962-2017
- ▶ 2021-04-06: ethics
- ▶ 2021-04-13: present problems: attention
economy+VC=dumpsterfire

Appendix

- ▶ 2021-01-12: intro to course
- ▶ 2021-01-19: setting the stakes
- ▶ 2021-01-26: risk and social physics
- ▶ 2021-02-02: statecraft and quantitative racism
- ▶ 2021-02-09: intelligence, causality, and policy
- ▶ 2021-02-16: data gets real: mathematical baptism
- ▶ 2021-02-23: WWII, dawn of digital computation
- ▶ 2021-03-09: birth and death of AI
- ▶ 2021-03-16: big data, old school (1958-1980)
- ▶ 2021-03-30: AI2.0 = ML, 1959-2015
- ▶ 2021-03-23: data science, 1962-2017
- ▶ 2021-04-06: ethics
- ▶ 2021-04-13: present problems: attention
economy+VC=dumpsterfire
- ▶ 2021-04-15: future solutions