COMS E6184, LAW L7777—Introduction

Technology Issues

- The Internet
  - Watch the video listed in the reading!
- Privacy
- Machine learning

Privacy

Biases

Many of the assigned readings assume that privacy is a benefit, and should be protected to the extent possible.

"You have zero privacy anyway. Get over it".
—Scott McNealy
former CEO, Sun Microsystems

We're unabashed privacy and free speech advocates. But the course is not about *promoting* these concepts, and we don't expect you to agree with us. The course is about understanding privacy, anonymity, and free speech, and how these are affected by technology.

Definitions

anonymity The condition of an identity being unknown or concealed. (RFC 4949)

privacy The right of an entity (normally a person), acting in its own behalf, to determine the degree to which it will interact with its environment, including the degree to which the entity is willing to share its personal information with others. (RFC 4949)

pseudonym A fictitious name, especially a pen name. (The Free Dictionary)

Using versus Gathering

- The primary concern is how information is *used*
- Obtaining information is often much less of a concern
- Note, though, that a lot of personal information is considered private even from one other person

Legal Foundations of Privacy

- Common law: “[T]he house of every one is to him as his castle and fortress.” Semayne’s Case, 5 C. Rep. 91a, 77 Eng. Rep. 194 (K.B. 1603)
- Doesn’t work as well in today’s interconnected world
- Information is collected, stored, analyzed

The Role of Computers

- Computers make mass storage (more) feasible
- (Punch card storage (1880s) started the process)
- Computers allow for rapid, sophisticated matching and correlation
- Computers can make inferences and predictions, and group people into categories
- This was foreseen in the 1960s, but became reality more recently

Inferences

- Amazon, Netflix, etc., try to predict what else you might like
- These algorithms work by correlation
- Often, they're right, but sometimes, they give odd results. . .

Why Violate Privacy?

- Thoughtlessness
- Efficiency, especially for marketing
- New markets (i.e., new location-based offerings)
- Public safety and national security

How Do We Lose Privacy?

- Voluntarily
- Compulsion
- Reuse of data

☞ *This sort of secondary use is the source of most privacy violations*

Voluntary Surrender of Data

- Social networking sites
- Purchases (Netflix, Amazon)
- Warranty cards

Compulsion

- Various interactions with governments (marriage, property purchases, etc.)
- Boarding an airplane
- “Contracts”—e.g., getting a credit card in exchange for information

Secondary Use

- We may not object—or object too much—to the initial collection of certain data
- Often, we benefit from the initial collection, and hence regard it as a fair trade
- When it is used for another purpose without our knowledge or consent, trouble often results

Example: Bars and Drivers' Licenses

- Many bars use swipe readers to verify that the proferred license is genuine
- (Better-grade fakes have mag stripe data anyway. . .)
- But—the readers copy the data: name, address, gender, etc.

What are the Privacy Violations?

- Using license data to establish age
- Using license data for marketing

Data on a Driver's License

- Primary purpose: certification that you are legally allowed to drive
- Primary purpose of picture: assurance that the bearer is indeed the license holder
- Demographic data: accountability in event of violations
- *Not* intended for proof of age, *not* intended as an airplane boarding credential

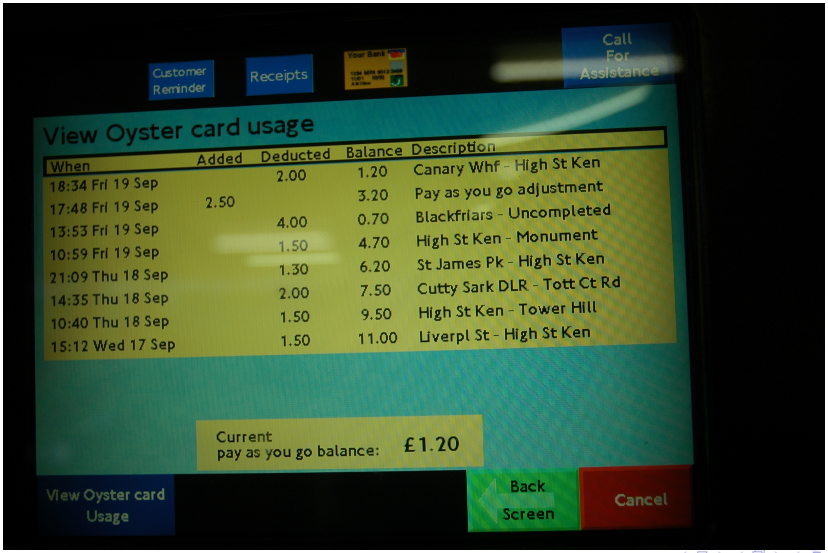
Age Verification

- Even if age verification is acceptable—and use of licenses for that is certainly accepted by the states—use of the additional data for marketing is not
- Resale of license data happens to be illegal, but not for that reason

Example: MetroCard

- Primary purpose: paying subway or bus fare
- But—the MTA retains your trip information
- This data can be and has been used for criminal and divorce cases

The London Oyster Card



Linkages

- Sometimes, items from two or more databases are linked
- Then possible to learn *much* more
- Prerequisite: common data item

Linkages: MetroCard

- How did you pay for your last MetroCard? Credit card?
- That links the MetroCard to a person
- Query: who boarded the subway at 116th and Broadway between 3:30 and 3:45 AM last Tuesday?
- In principle, at least, that question may be answerable

Deeper Linkages

- Correlate on patterns
- Example: assume a MetroCard is used infrequently, but at only two stops, Penn Station and 116th St
- Is there any one person who used a credit card to buy train (Amtrak, NJ Transit, LIRR) tickets on just those days?
- (Note: I have no idea if that has actually been done)

Identity

- Sometimes, anonymous data can be linked to a specific person
- Other times, behavior identifies you
- Linkages can be used to establish identity
- MetroCards are anonymous—but credit cards aren't

Privacy: Conclusions

- There is a vast amount of data accumulated about people
- People have little knowledge and less control over how it is used or with whom it is shared
- Per the definition, we have little ability“to determine the degree to which [we] will interact with [our] environment”
- In other words, we don’t have much privacy

Questions?

(Barred owl, Riverside Park, November 19, 2020)