Privacy Issues With Sensor Data Collection

A. Michael Froomkin
University of Miami School of Law
froomkin@law.miami.edu

Privacy, Quickly

- Has multiple elements including control of access to body, to thoughts, protection of private sphere, and especially intimate relations
- Today will concentrate on one aspect: ability to control data about oneself
  - Primarily an instrumental value
  - Enhances human freedom, flourishing, experimentation, innovation, self-realization; protects against discrimination, and also distant, often invisible, exercises of (often algorithmic) power

Data == Power

- People with data about you can exercise power over you
  - Market power
  - Government power
- Special case of both above: Sorting power
  - Can be invisible
  - Can be lifelong
  - Can be very empowering – or very damaging
    - Credit scores
    - Citizen safety scores
    - Issues of accurate, inaccurate, and predictive (i.e. speculative) scoring

Data Types (By Method of Collection)

- Organic data flows
  - Data streams that are currently collected either in the course of business or to comply with regulatory requirements.
  - Can be *very* detailed (e.g. NYC taxi rules require logging every pickup and drop-off)
    - And thus can be very revealing (strip club attendees; low-tipping celebrities)
- Distributed sensors
  - CCTT, License plate readers, biometrics
- Special-purpose sensors
  - NY skyscraper cams & heat sensors
- Self-surveillance
  - Instagram metadata, many cell phone apps
- [Information derived from above:] Correlations (i.e.. Big Data products)

Big Data, Quickly

- US government definition: the growing technological ability to capture, aggregate, and process data
- EU: “the massive and rapid processing of data (through modern data analytics) in the search for information (including unforeseen information) The practice of data mining poses a significant challenge due to the degree of opacity characterising many contemporary data processing activities. ...
  - Data mining practices may result in “behavioural targeting” and further encourage a “datafication” of society that poses significant challenges for privacy and digital rights in general. Due to such risks as statistical discrimination...
- Bottom line: Big Data is *both a technology and a process* with many (reinforcing) parts, each of which can be enabled or, in theory, regulated. It is also, arguably, an ideology.

Big Data as a Process

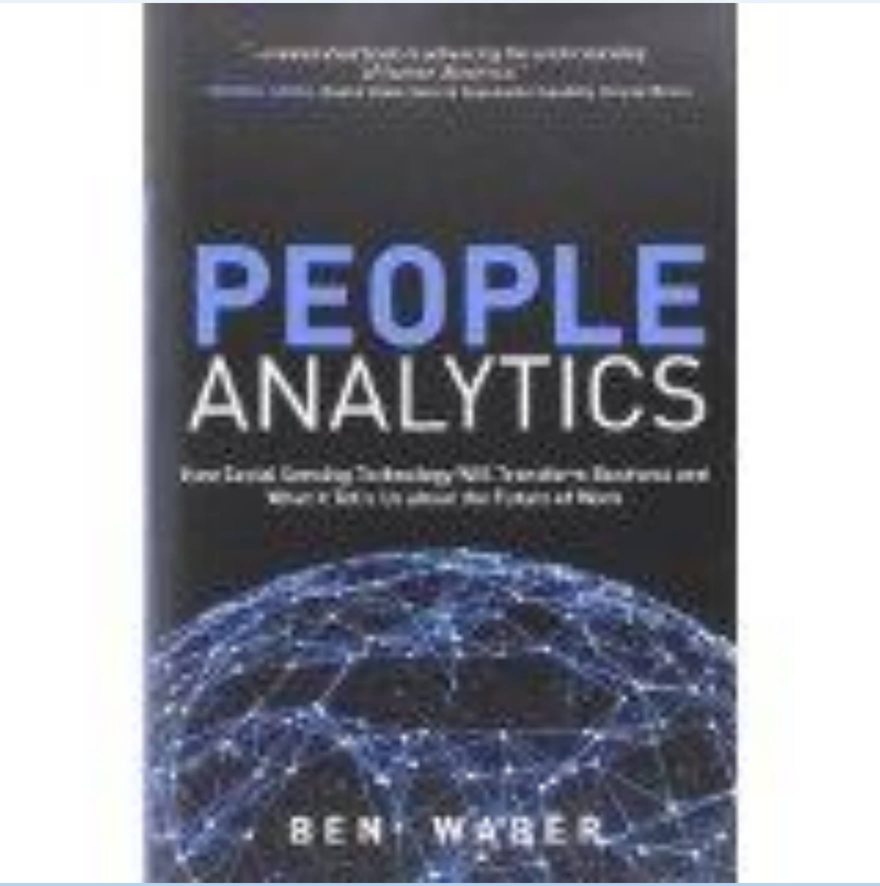
- Data collection
  - Sensors: Smart city, Cell Phones
  - Self-surveillance
  - Transactional data: Communications, Shopping, Medical
- Storage
- Processing
  - Correlation / Data Mining
    - Issues of identification/ de-identification/ re-identification
- Re-Use
  - Nature of Big Data projects is to seek unexpected connections

The Trend is to Ever-More Data Collection, Hence More (and Bigger) Data

- More apps creating occasions to create and collect data
- More monitoring
  - VACCINE (ccTV nets)
  - So called ‘self-surveillance’ e.g. Fitbits, Nests
- Ubiquitous hardware/sensors
- Bigger aggregations of data
  - Smart cities
  - Fusion centers, DHS, NSA
- Cost/benefit motivates private sector: costs are dropping, perceived benefits growing

The Lure of Unexpected Insights

"What if I told you that ...one of the biggest decisions a company makes revolves around the size of its lunch tables?"


PEOPLE
ANALYTICS
How Social Sensing Technology Will Transform Business and What It Tells Us about the Future of Work
BEN WABER

And the Data Streams Keep Coming

- Smart Meters
- Smart Cars (or self-driving & gridded)
- Mobile & personal care robots
- ‘Always on’ voice recognition devices
    - Amazon Echo, Google Home, Apple TBA
- IoT devices more generally

NowTheEndBegins.com
The New Normal

Data Can Hurt People; So Can Data Collection

- Discrimination
  - Redlining
  - Unintentional, invidious algorithms
- State Power
  - Investigation of Occupy Wall St. protestors
- Being watched is intimidating, silencing

Are You Trusted? Or ‘Dangerous’?

Because There are no Routine Calls

Intrado Beware®

Alert call takers, dispatchers and responders to potentially dangerous situations when and where it matters most with Intrado Beware®.

"Beware," a program made by West Safety Services is capable of quickly sorting through billions of publicly available commercial records to alert first responders to potentially dangerous situations, according to the product website. Beware calculates ‶threat scores″ by assigning people and addresses green, yellow, or red scores with red being highest threat and green being the safest.

Mixed Blessings of Big Data

Benefits – Numerous, Usually Concentrated

- Medical, Epidemiology
- Security
- Finance
- Marketing
- Urban planning
- “Quantified Self”

Costs – Widespread, Diffuse, Harder to Quantify

- Privacy reduced or eliminated
    - Increases risk of reprisal for acts
        - Less experimentation
    - Dossiers are bigger
- Increased opacity in decision-making
- Polity
    - Self-censorship
        - Unpredictable consequences of speech, civic choices
    - Less whistleblowing

More & ‘Better’ Algorithms

As they are fed more data

- Can infer results about you based on people who share observed traits about you, another way in which there is no occasion for consent on your part, if other have consented to deeper dive into their data
- Thus, big data is a power shift - away from data subject towards those who control and can access the data.

Common Privacy Protection Techniques Fail

- Consent to data collection loses what little meaning it has
  - “Consent” is meaningless for remote sensing
  - In US ‘consent’ not required for most transactional collection unless “sensitive”
  - “Informed consent” is *impossible if one cannot predict what effect the data use might have*
- Processing limits are inadequate
  - US law bans ‘discrimination’ against certain classes e.g. race, religion
  - But no law against discriminating against people who pay for pizza with credit
  - Battle over no-fly lists is going very slowly
- Data anonymization easily undermined
  - Re-identification is easy and likely profitable

Metaproblem: Attitudes

- As Julie Cohen says, Big Data repackages surveillance as innovation. Plus, “Big Data ... equates information with truth and more information with more truth, and that denies the possibility that information processing designed simply to identify a ‘pattern’ might be systematically infused with a particular ideology.”
- “[P]rivacy is increasingly cast as the spoiler in this tale, the obstacle to the triumphant march of predictive rationalism. ... [I]f information processing is rational, then anything that disrupts information processing, including privacy protection, is presumptively irrational.”
  - If Cohen is correct that Big Data is the result of an ideology then we should be wary of the argument that the value of insights drawn from huge datasets justify the creation of centralized (OR federated!) repositories and their use.

This is Not Just a Data Collection But Also a Systems Issue

- Privatizing, distributing, or ‘federalizing’ the data makes little difference.
- US Law (1st Am) makes ‘unseeing’ and RTBF unlikely
    - But there is a very limited ability to prohibit invidious uses in regulated industry
- Rules for sharing and re-use are as critical as rules for collection
- Also need algorithmic transparency (but must be comprehensible)
- Data moves between public and private sectors unless legislation forbids sale or disclosure.
    - This is rare, but does happen:
        - Video Privacy Protection Act (no sharing Video Rentals)
        - Driver Privacy Protection Act (no selling drivers license data)

What Can Be Done

- Limit sensors/data collection
  - Publicize fact of collection; publicize use/re-use of data
  - Impose costs on collection to cure market failure
- Regulate storage or flows (sharing)
  - Age out data
  - RTBF
- Limit use/re-use once holding data
  - Forbid sales of public data to private parties without **stringent** conditions on re-use, re-sale
- Focus on discrimination: ban ‘bad’ algorithms (but enforcement hard, and unconscious biases in algorithmic design or in data set construction)
- Create a new due process in data processing
  - Right to contest distant, invisible algorithmic decision making
  - Give individualized (expensive!) notice of when, how and why your data is being processed
- Invent new ways to opt-in and opt-out
  - But opt-out options need to be realistic not formalistic

Takeaways

- Key point1: No one can predict the implications and uses of ANY data stream, given the power of correlations
  - Effects of data streams will change
  - Transparency is essential – but alone is not nearly enough, because…
- Key point 2: INFORMED CONSENT IS NOT POSSIBLE (Long range + #1 above)
- If system design capable of info collection OR correlation then must design strong technical AND legal barriers against misuse:
  - Pre-collection/use review (IRB on steroids)
  - Regular **audit** of compliance and also of possible new uses/ harms of data
  - Mechanism for external review/enforcement (overcome standing, valuation issues that block legal process)

00
• WASHINGTON DC •
I SEE U
TAXATION WITHOUT REPRESENTATION