Inverse privacy

Yuri Gurevich, Microsoft Research

Joint work with Efim Hudis and Jeannette Wing

To my *alma mater*

In the ocean of hypocrisy, it was a rare island of decency and a temple of science.

Agenda

Part I. Privacy: Quick introduction

Part II. Inverse privacy

Pri·va·cy

Pri·va·cy, from Merriam-Webster

1*a* **:** the quality or state of being apart from company or observation **:** seclusion

1b **:** freedom from unauthorized intrusion <one's right to *privacy*>

2 (*archaic)* **:** a place of seclusion

3*a* **:** secrecy

3b **:** a private matter **:** secret

First known use*:* 15th century

Etymology of pri·va·cy, from Wikipedia

From Latin: *privatus*

"separated from the rest, deprived of something, esp. office, participation in the government",

from *privo*

"to deprive“

What is privacy?

Definitions

- Math is easy in a sense: precise definitions, deductive arguments
- Compare this to law. Try to prove something imprecise ``beyond reasonable doubt.’’
- In that sense, privacy is harder yet. “Is it a descriptive concept, a normative concept, a legal concept, or all three?” (Helen Nissenbaum 2010)

Privacy-involved disciplines

- Philosophy,
- political theory
- legal theory,
- media studies,
- information studies,
- computer science and engineering
  (where we are coming from)

Descriptive vs. normative

- School teachers tend to treat grammar as prescriptive, but the linguists agree that it is largely descriptive (and evolving).
  There is no such consensus on privacy definitions.
- There are descriptive defs: e.g.
  “a measure of the access others have to you through information, attention, and physical proximity” (Ruth Gavison 1980).
  - Moral aspects are not denied but they should build on a neutral foundation.
- But normative definitions are more common;
  they assert that privacy is valuable and deserving of protection
  - though privacy may be damaging as well; we’ll see some examples later.

Access and control in privacy defs

- The condition under which other people are deprived of access to either some information about you or some experience of you” (Jeffrey Reiman 1976). Also, Gavison’s def above.
- It is important though who has control over the degree of access.
- “Privacy as the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others” (Alan Westin 1967).

Differential privacy

- A precise quantitative definition that captures the increased risk to one’s privacy incurred by participating in a database.
- Pioneered by our MSR colleague Cynthia Dwork, it attracted attention of mathematically minded.
- A subject for a separate lecture.
- An impressive progress. But only a relatively narrow slice of the pie of privacy problems is solved by differential privacy.

Privacy is messy

- Is it a descriptive concept, a normative concept, a legal concept, or all three?
- Does it apply only to information, to actions and decisions (the so-called constitutional rights to privacy), to special seclusion, or to all three?

  (Helen Nissenbaum 2010)

Well, let’s not wait a consensus on the definition of privacy.

Some aspects of privacy

Privacy ≠ secrecy

- “If you have something that you don’t want anyone to know, maybe you shouldn’t be doing it in the first place”
  [Eric Schmidt, then Google’s CEO in a 2009 CNBC interview]
  - This was self-serving.
- “Every one of us has parents who did at least one private thing that’s not a secret, otherwise we wouldn’t be here” [Cory Doctorow]

The privacy paradox

- Judging by polls and surveys, people worry about privacy, want more control over their privacy, want laws protecting privacy.
- Yet, choosing between privacy and just about any other good, people overwhelmingly choose the other good.
  - credit cards over cash,
  - E-ZPass over traditional toll payments,
  - discount shoppers’ cards over cash,
  - traceable search engines over self-directed Web surfing.
- Few read privacy policies and even fewer complain about them.

A fallacy

- Search the internet in the *privacy* of your home.

What is privacy worth?

- There is a big gap between what people are willing to pay (WTP) for keeping an item of information private and what they are willing to accept (WTA) for giving away that same item of information.
- WTA > WTP.

Can privacy be harmful?

- Criminals use privacy laws to conduct their business.
- On many diseases, there is no big data.
- Imagine that your hard disc crashes or that you dropped your cellphone in a sea.
- Imagine you urgently need a medical specialist.

Contextual integrity (Nissenbaum 2010)

- We interact as coworkers, professionals, clients, teachers, students, citizens, family members, club members, congregants, and neighbors.
- Once these characteristics are factored into an account of privacy expectations …, the law-like character of these privacy expectations, or norms, is much more evident.

Things are more complicated yet

“We investigate individual privacy valuations in a series of experiments ... [R]esults paint a more nuanced and detailed picture of privacy valuations than the one currently in the literature. They suggest that the value of privacy, while not entirely arbitrary, is highly malleable and sensitive to non-normative factors” (Alessandro Acquisti et al. 2010).

That concludes the introductory part.
We turn attention to inverse privacy.

What is inverse privacy?

Personal information

- We do not define term. It suffices to say that we understand the term broadly, without the restriction to personally identifiable information.
- The whole set of personal information about a person P will be called the P *infoset*.

Classification of personal information

The P infoset splits into four buckets.

1. The information about P that P has and nobody else does.
   To contrast this bucket with the next one, we call it *directly private*.
2. The *inversely private* information about P, information that some party has but P doesn't.
3. The *partially private* information about P, information that P has and a limited number of other parties do as well.
4. The *public information* about P.

Remarks on the classification

- Note that our classification is solely in terms of who possesses information. In particular, the definition of the direct privacy bucket abstracts from how private various items are.
- The entity P in our classification does not have to be a person. It could be a family for example. For simplicity we stick here to the case where P is a person.

The virtual directory of P infoset

- In addition to the P infoset itself, there is also information about the infoset, the metadata.
- Think of it as a virtual directory on P. For each item of information about P, it species the parties that possess the item. For each party, it species the items of information about P that the party possesses.
- One may say that the P infoset comprises first-order information about P while the virtual directory on P comprises second-order information about P.

The dynamics of inverse privacy

- New information arises. A person P can create new info-items on her own. She may write down a diary record or make a selfie.
- Some information may be lost, intentionally or unintentionally. People forget things. They lose or destroy some of their records.

The provenance of personal information

- The overwhelming majority of new info-items result from interactions of P with other parties.
  - These could be people – relatives, neighbors, coworkers, clerks, waiters, medical personnel, etc.
  - They could be institutions – employers, banks, internet providers, brick and mortar shops, online shops, government organizations, etc.
- Institutions also may destroy or lose information but in general they are better in keeping records.

Direct privacy to inverse privacy

Direct-privacy dominated world

- Until recently the capacity of a person to take and keep records was comparable to that of institutions.
- Among the four buckets of the P infoset, the direct privacy bucket was the largest, and the inverse privacy bucket was the smallest.

Welcome to inverse-privacy dominated world

- A radical change was brought recently by technology. The capacity of institutions to take and keep records became superior to that of a person.
- Typically the new partially private items quickly decay into inversely private because the institutions remember it all while the person often hardly remembers that the interaction took place.
- For a regular citizen P of an advanced society today, the volume – and value! – of the inverse privacy bucket vastly exceeds that of the direct privacy bucket.
- Furthermore, the value of the inverse privacy bucket grows much faster.
- We need to understand legal, political, sociological, technological implications of inverse privacy.

Inverse privacy to partial privacy

Social norms

- Directly private, partially private and public kinds of information are familiar. There are more or less accepted norms.
- But the ascent to dominance of inverse privacy is new.
- The dominance if inverse privacy, is it here to stay?

What all this info for?

- The collection of personal data is problematic.
- It creates opportunities as well.
  - To serve customers better.
  - To diagnose personal and societal problems and to facilitate solutions.

Share back personal info

- But, even if we accept that the collection of personal data by commercial institutions may be highly beneficial,
  it is hard to see why to keep this data inversely rather than partially private.
- There are numerous benefits to converting all that inversely private information into partially private.
  - Correct possible errors in the information.
  - Get a better picture of your health status, your credit status, etc.

The challenge

- We challenge the inverse privacy dominance.
- It may be diminished. It must be diminished.
- Governments have legitimate security concerns, but by and large people should be entitled to know
  - who knows what about them
  - and what is there to know.

The role of technology

- In the inverse-privacy case, technology created the problem, and technology can solve it.
- How? Here are some approaches.

Personal technology

- Develop tools that enhance people's capacity to take and keep records.
- For example, your smartphone may become eventually a trusted and universal recorder of many things you do.

Legal means

- Make institutions legally responsible to share back information.
- It should become available to you routinely,
  not only if and when you submit a legal request
  but as matter of course.

Incentives

- Technological, business and social incentives should be created to entice institutions to share information back.

A new common norm

- Encourage technological innovation and social campaigning, the creation of a new social norm according to which, by default, person-to-institution interactions produce partially private information.

Remark

- Enabling people to access information makes it also easier for invaders to find information.
- Any technology invented to allow inverse privacy information to be shared back has to be made secure.
  Communication channels have to be secure, encryption has to be secure, etc.
- Note, however, that today hackers are in much better position to find information about you that you are.
  Sharing information back should improve the situation.

Engineering

Toward Biggish

- Suppose that a company is willing to share back personal info with its clients. Two questions arise.
  - How practical is it for the company?
  - Will you bother to look up your personal info?
- This brings us to our engineering proposal that we call Biggish.
- Why “Biggish”? Your personal data is huge and growing.
  It isn’t quite *big data* but it is sort of *biggish*
  [великоватый, assez grand, ziemlich groß].

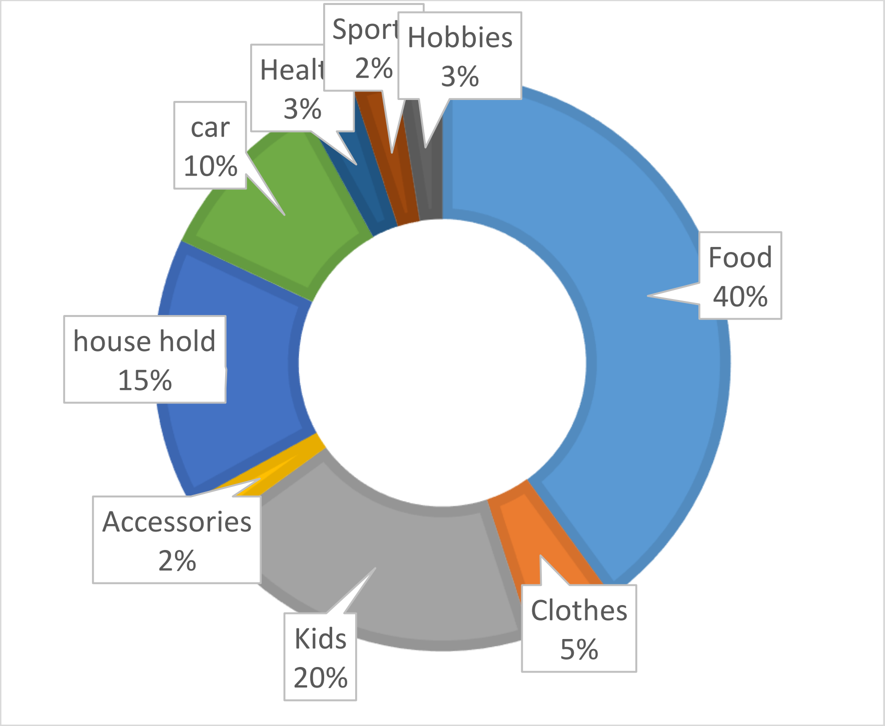
Biggish

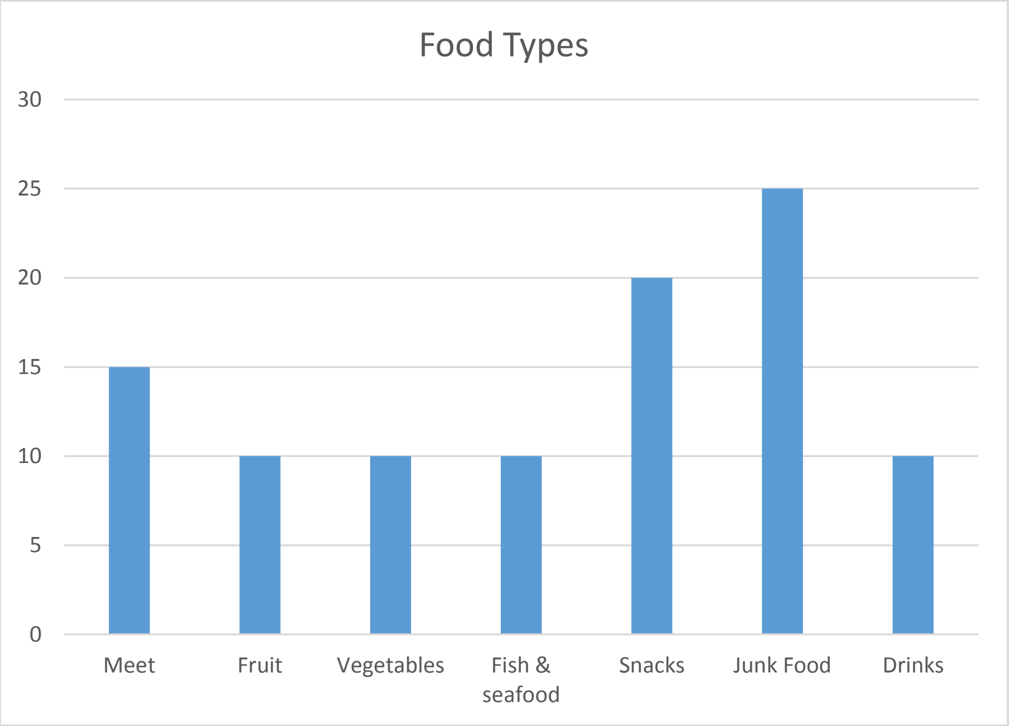
- Biggish provides a platform that allows companies to conveniently share your personal info with you.
- Biggish gives you access to your personal data in such a way
  that you can mine it with much benefit to you.

Health

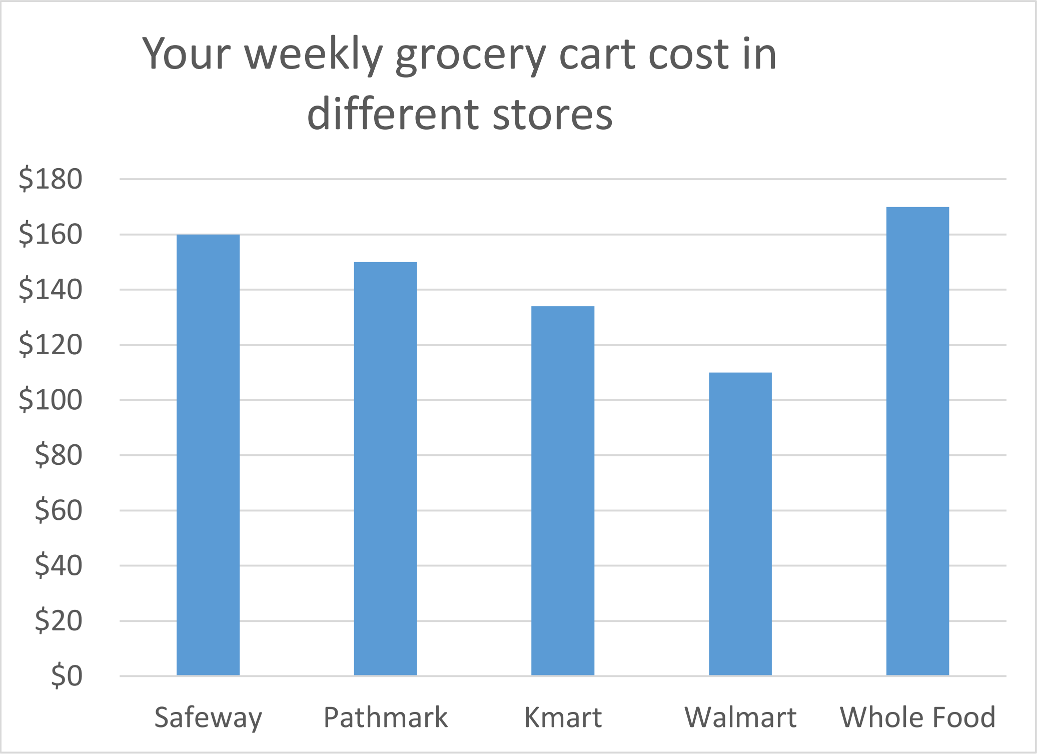
- Run useful algorithms on
  - the history of all your medical encounters,
  - your medications and vitamins,
  - your fitbit data
  - ...
- Compare your data with
  - the averages,
  - the recommended
  - ...
- Get alerts when ...

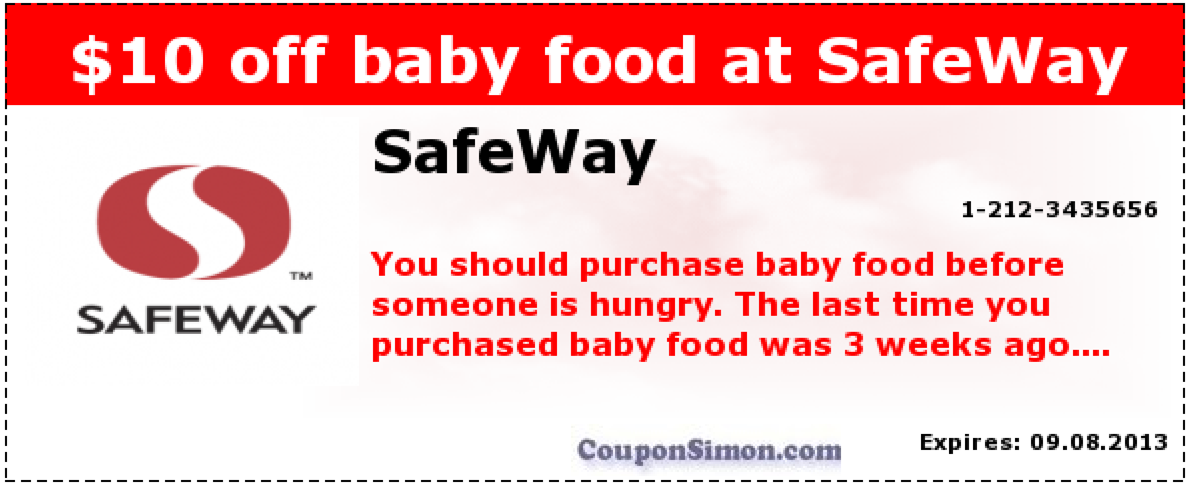
Learn your own shopping practice. Manage your budget better.





Shop wiser





$10 off baby food at SafeWay

SafeWay

1-212-3435656

You should purchase baby food before someone is hungry. The last time you purchased baby food was 3 weeks ago....

CouponSimon.com

Expires: 09.08.2013

If you shop at Walmart you will save $50 weekly

An academic thinks of Biggish

- Match the faces with the names of people I should know, match the names with the phone numbers.
- All the countries you visited in the last 10 years.
- If I return sick from X, I’d like to find out why. What I have eaten? What did I visit? …
- What lectures I’ve given? What courses I taught, who were my students? …

Epilog

Summary

We haven’t solved THE privacy problem; this may take eternity. Instead, we did the following.

- Pointed out a troubling phenomenon and named it *inverse privacy*.
- Analyzed the phenomenon, providing a simple language for the purpose.
- Argued that the trouble can be mended.
- Propose an engineering solution for the purpose.

Civilization is the progress toward a society of privacy. The savage's whole existence is public, ruled by the laws of his tribe. Civilization is the process of setting man free from men.

Ayn Rand 1961