COLLEGE OF
INFORMATION
STUDIES

INST201: Introduction to Information Science

Dr. Joel Chan

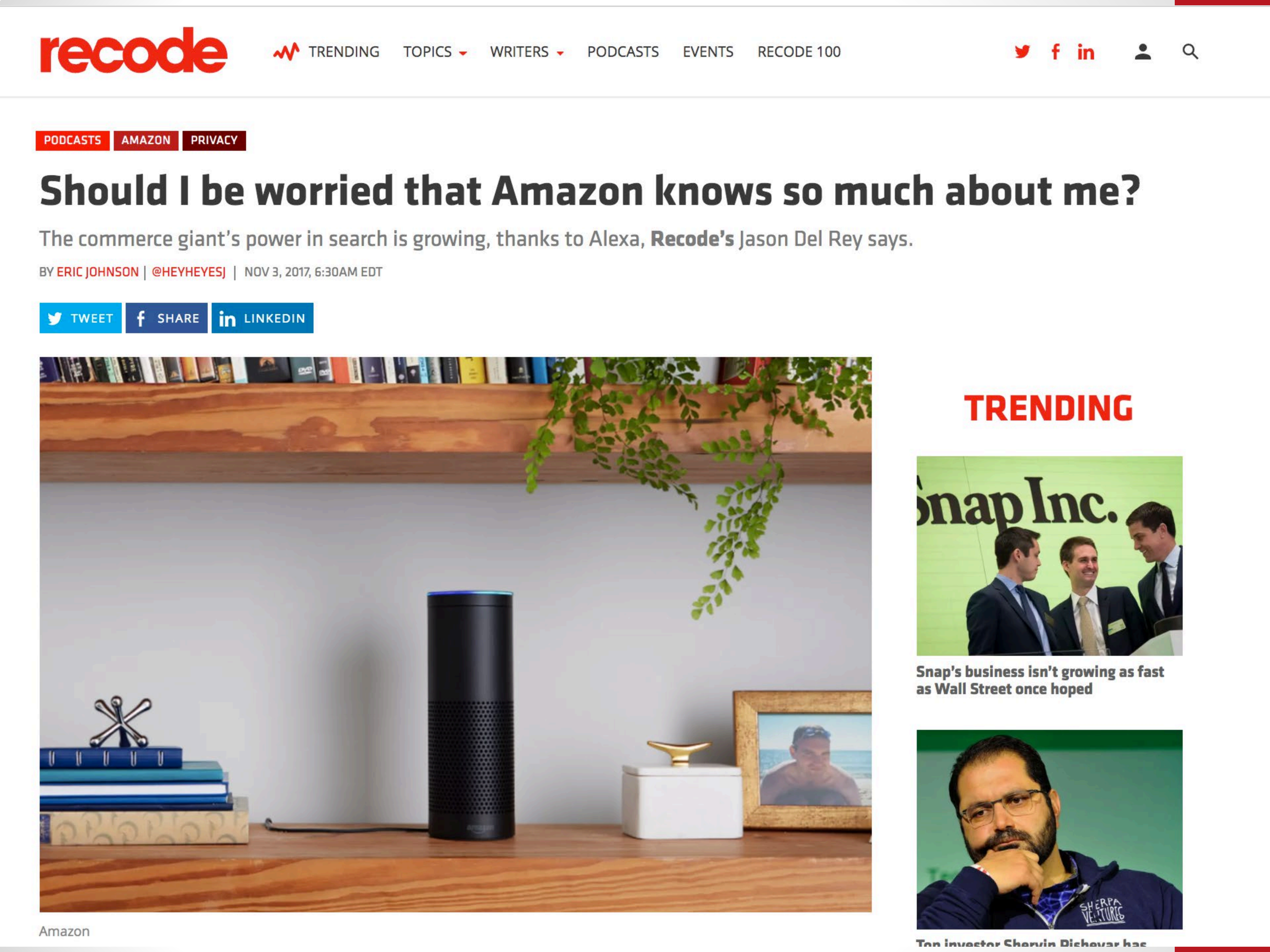
Office: 2118E Hornbake Building, South Wing

Email: joelchan@umd.edu

Week 11: Information Privacy

Learning goals

- **Define privacy** (from different major perspectives), and distinguish from related concepts
- Explain **why privacy is hard in digital spaces**
- Explain some key **privacy management mechanisms/strategies** (along with strengths/weaknesses) available to individuals and society



Should I be worried that Amazon knows so much about me?

The commerce giant’s power in search is growing, thanks to Alexa, **Recode’s** Jason Del Rey says.

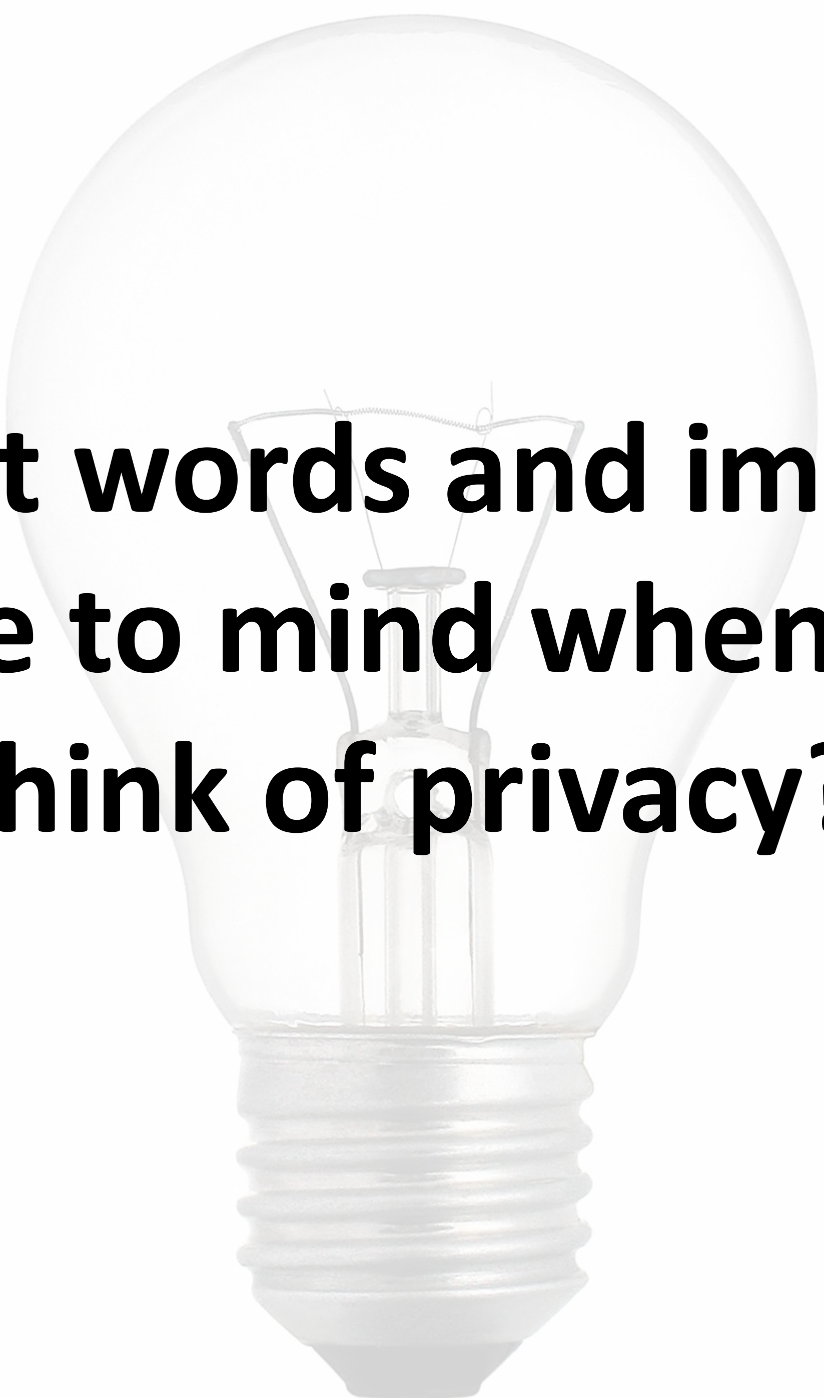
BY ERIC JOHNSON | @HEYHEYESJ | NOV 3, 2017, 6:30AM EDT

Amazon

TRENDING

Snap’s business isn’t growing as fast as Wall Street once hoped

Top investor Shervin Pishevar has



What words and images come to mind when you think of privacy?

We often treat privacy like it’s a BLACK OR WHITE issue

Upon closer examination, we realize most privacy issues are shades of gray.

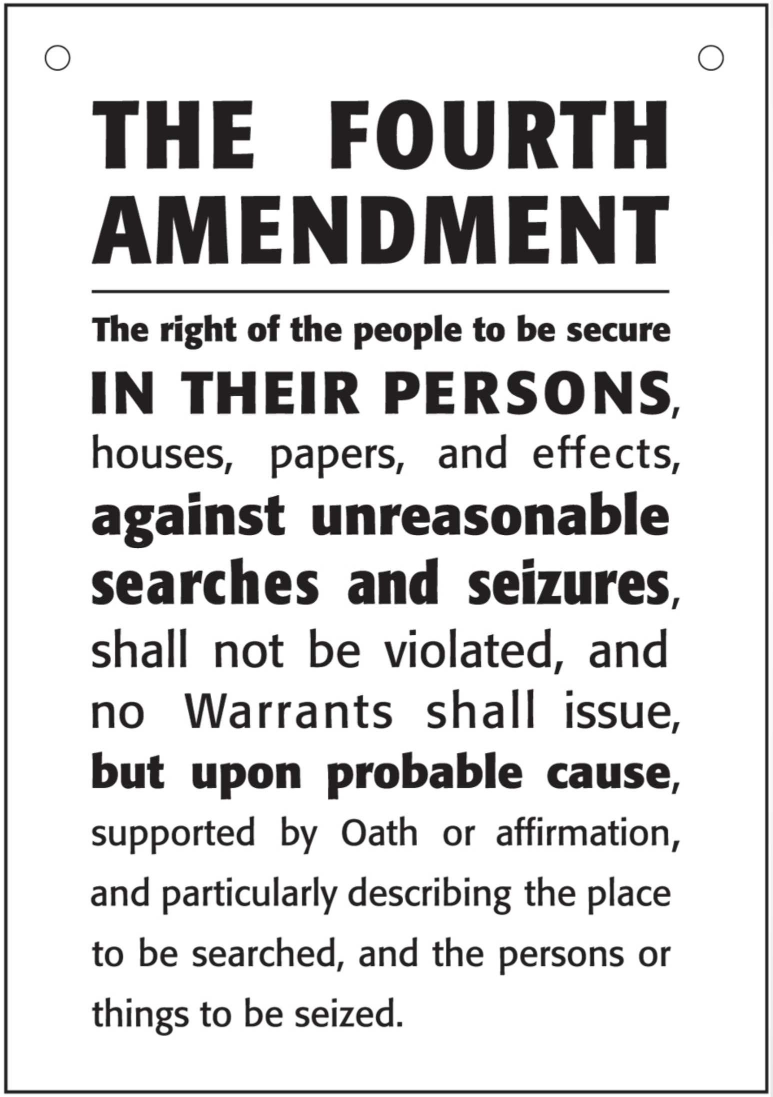
Historically speaking…

Privacy is a relatively new concept.

> *Google's Cerf Says "Privacy May Be An Anomaly." Historically, He's Right.* [TechCrunch]

Fourth Amendment to US Constitution is closest we get to founding fathers protecting individual privacy (1789).

The Right to Privacy (Harvard Law Review) –> first published legal document regarding privacy as a **human right** (1890).



THE FOURTH AMENDMENT

The right of the people to be secure **IN THEIR PERSONS**, houses, papers, and effects, **against unreasonable searches and seizures**, shall not be violated, and no Warrants shall issue, **but upon probable cause**, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

An opening example

Katz v. United States (1967)

Charles Katz used a **public pay phone** booth to transmit illegal gambling wagers from Los Angeles to Miami and Boston. Unbeknownst to Katz, the **FBI was recording his conversations** via an electronic eavesdropping device attached to the exterior of the phone booth. Katz was convicted based on these recordings. He challenged his conviction, arguing that the recordings were obtained in violation of his Fourth Amendment rights.

An opening example

Katz v. United States (1967)

Refined the 4th Amendment section defining “unreasonable search and seizure” to provide a legal definition to “search”; also extended 4th Amendment rights to include **“reasonable expectation of privacy**.”

But who defines what is **reasonable?**

Defining privacy

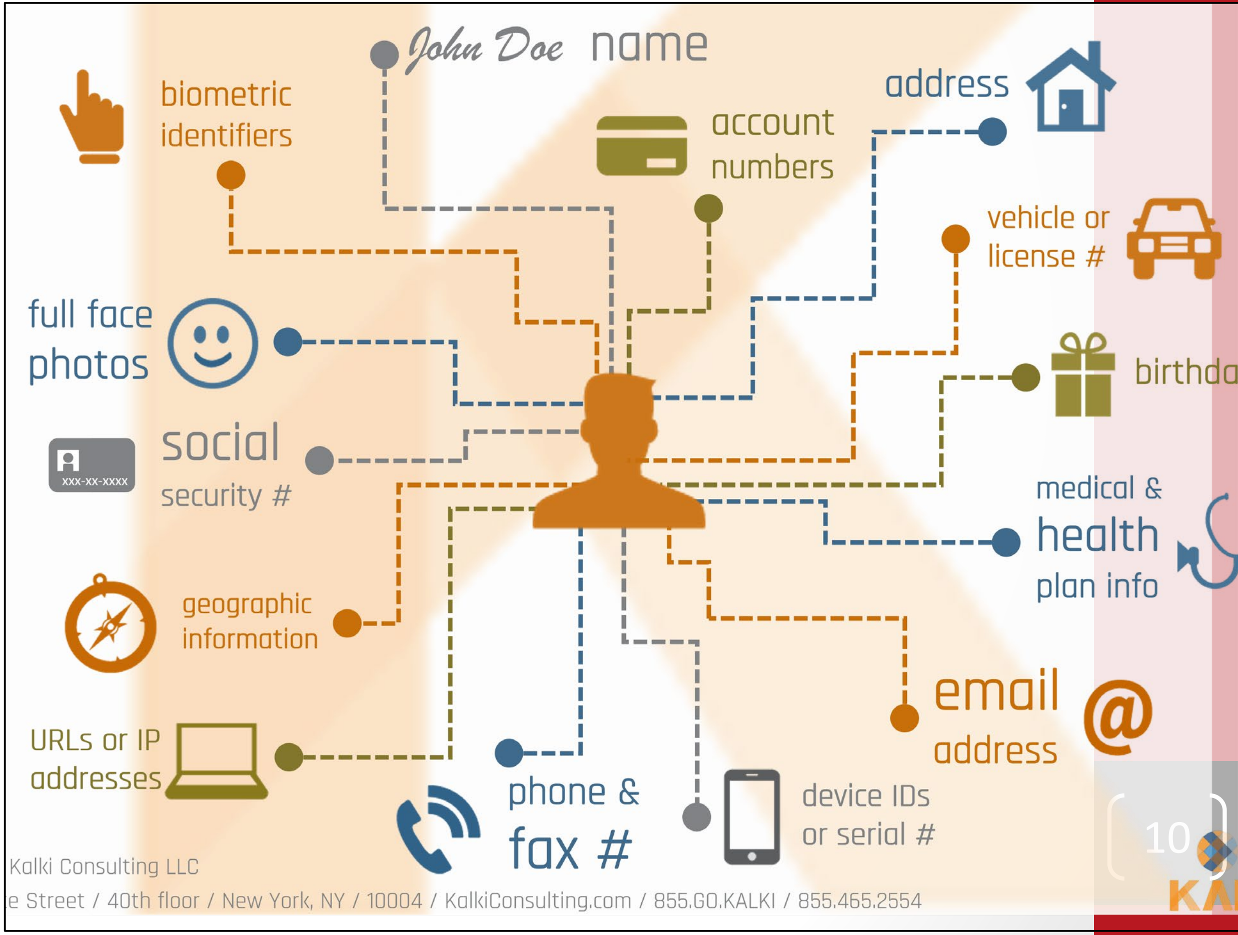
Four main frameworks for understanding privacy:

1. Privacy **as a right** (general): the “right to be left alone” (1890)
2. Privacy **as a state** (general): “limited access to a person” or “being apart from others”
3. Privacy **as control**: Altman’s “selective control of access to the self”; boundary regulation
4. Privacy **as a commodity:** we exchange some of our privacy for perceived benefits.

Common thread: **control** who has **access** to ”**me**”

What is "me"?

Personally Identifiable Information (PII): generally thought to include name, date of birth, SSN, address, etc. More broadly, PII includes "any **information that can be used to distinguish one person from another** and can be used for de-anonymizing anonymous data."



What Privacy **Is Not:** Related Concepts

Anonymity: ability to conceal one's identity fully.

> *Related:* ***pseudonymity,*** *or use of ID/pseudonym/handle instead of person's real name.*

Confidentiality: externalization of restricted but accurate information to a specific entity; controlled release of information (e.g., to a doctor).

Secrecy: intentional concealment of information (think "secret" and "top secret" government documents).

Security: focus on ***protecting*** information across three areas: *integrity, authentication, access* (more on Friday)

What Privacy **Is Not:** Related Concepts

right to be left alone

limited access to a person

selective control of access to the self

commodity

PRIVACY

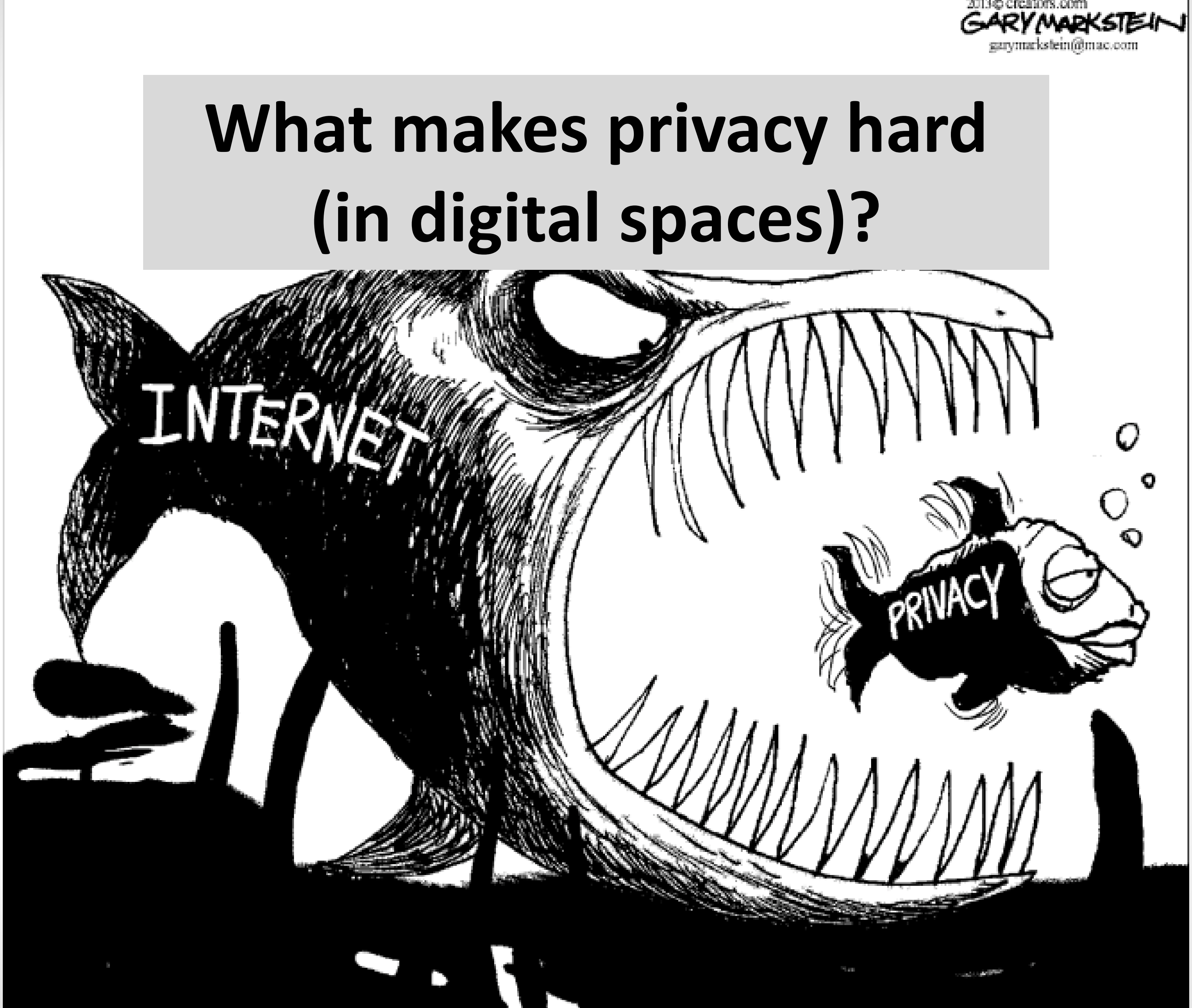
secrecy

confidentiality

anonymity

SECURITY

What makes privacy hard (in digital spaces)?



Public vs. (reasonably) private?

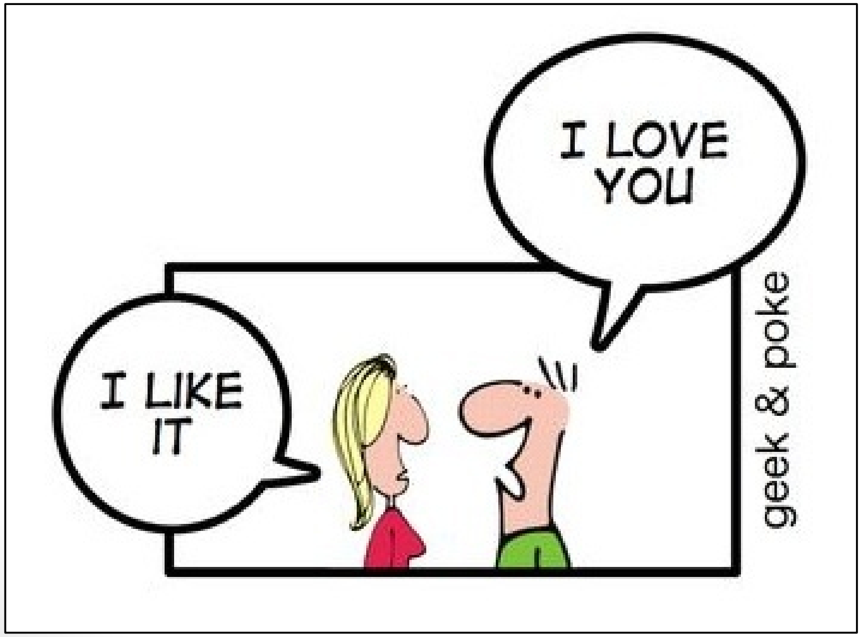
In most public spaces, we know we might be observed.

Public vs. (reasonably) private?

Digital spaces are increasingly **networked publics**

New technologies connect us more than ever…

...but they blur boundaries between public and private spaces.



This raises innumerable opportunities for communication misunderstandings, breakdowns, and norm violations.

Public vs. (reasonably) private?

Digital spaces are increasingly **networked publics**

Here, there are three main factors that make it harder to tell when you’re operating in a public space, a private space, or somewhere in between.

- **Invisible audiences**
- **Context collapse** due to the lack of spatial, social, and temporal boundaries
- **Blurring of public and private boundaries** due to lack of control over context

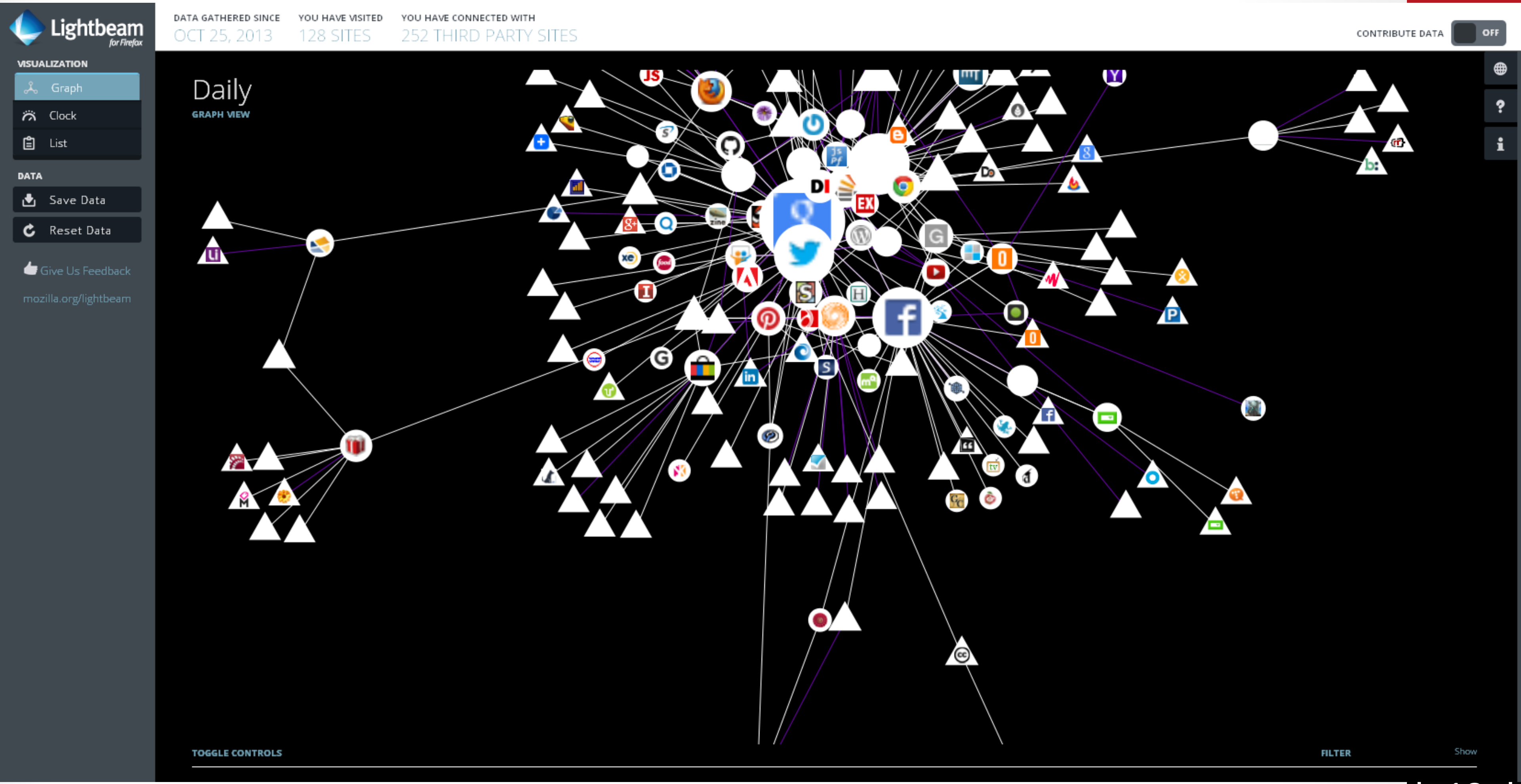
Invisible audiences

Imagined audience:
“a person’s mental conceptualization of the people with whom he or she is communicating” (Litt, 2012).

Influenced by:

- Environmental factors (e.g., social norms)
- Individual factors (e.g., motivation/purpose for using site)

Who is my "audience"?



DATA GATHERED SINCE OCT 25, 2013

YOU HAVE VISITED 128 SITES

YOU HAVE CONNECTED WITH 252 THIRD PARTY SITES

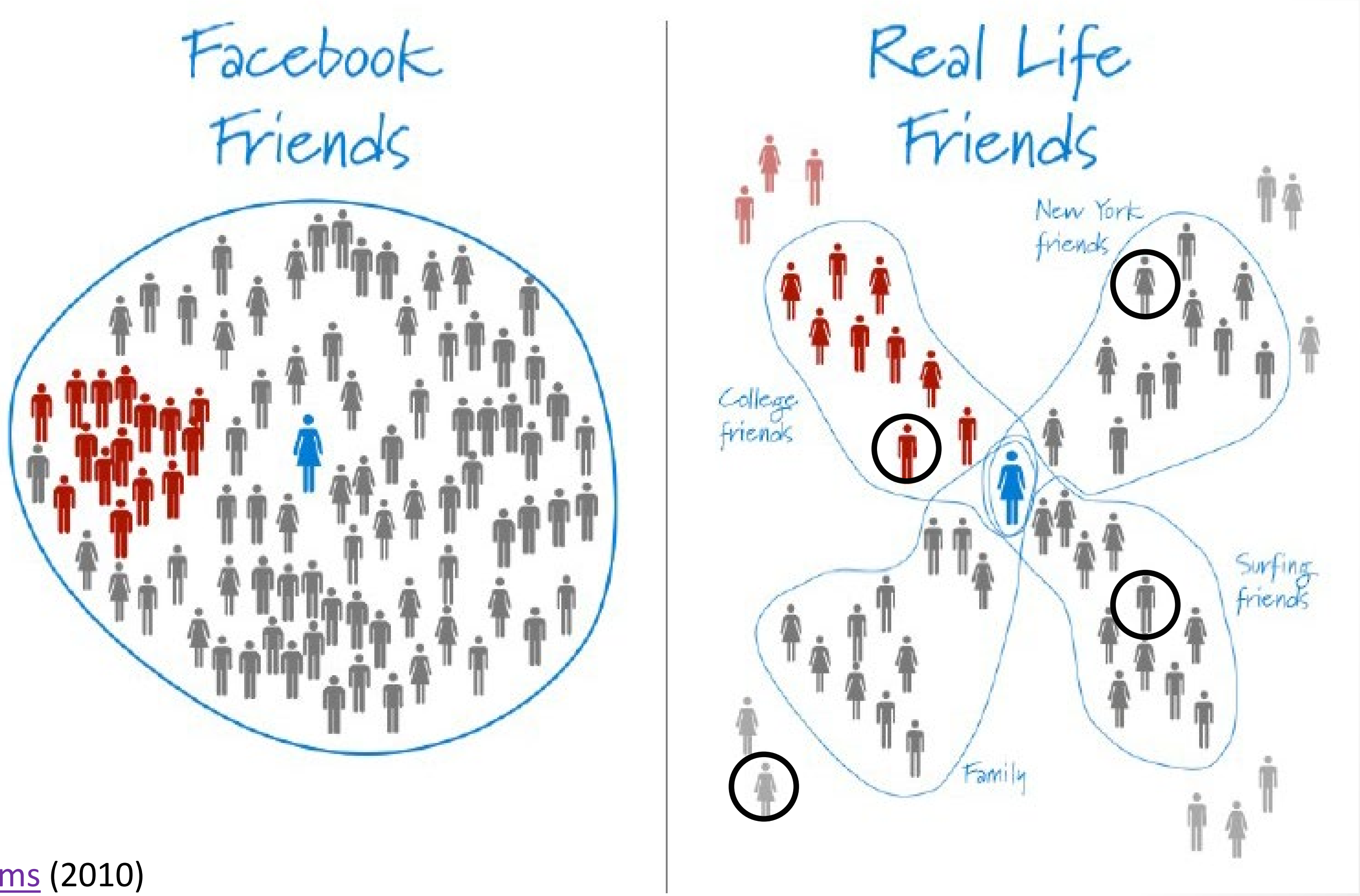
Selective self-presentation

We highlight certain aspects of our identities and minimize others; this varies based on audience (remember the editability affordance?).

The facebook version of you
The Realistic version of you
© comicalconcept.com

Why is this important?

Context Collapse & Privacy Concerns



Paul Adams (2010)

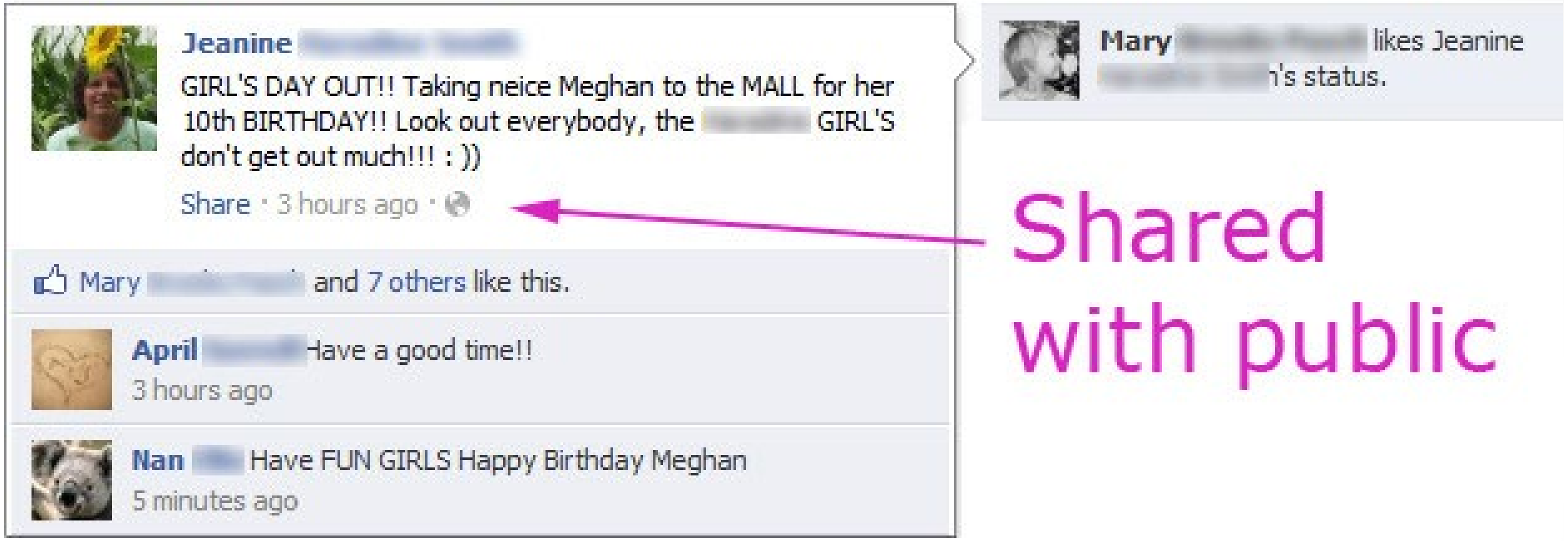
Networked Public Problems:
Context collapse can help or hurt your relationships

1) **Strength of weak ties** (Granovetter, 1973): users distribute content to entire network to increase likelihood that someone will see it and respond.

2) **Privacy settings:** users employ increasingly granular privacy settings to segment network into different audiences.

3) **Lowest common denominator** (Hogan, 2010): users only distribute content appropriate for **all** "friends."

How do I know who is my audience?



Jeanine [redacted]

GIRL'S DAY OUT!! Taking neice Meghan to the MALL for her 10th BIRTHDAY!! Look out everybody, the [redacted] GIRL'S don't get out much!!! :))

Share · 3 hours ago ·

Mary [redacted] and 7 others like this.

April [redacted] Have a good time!!

3 hours ago

Nan [redacted] Have FUN GIRLS Happy Birthday Meghan

5 minutes ago

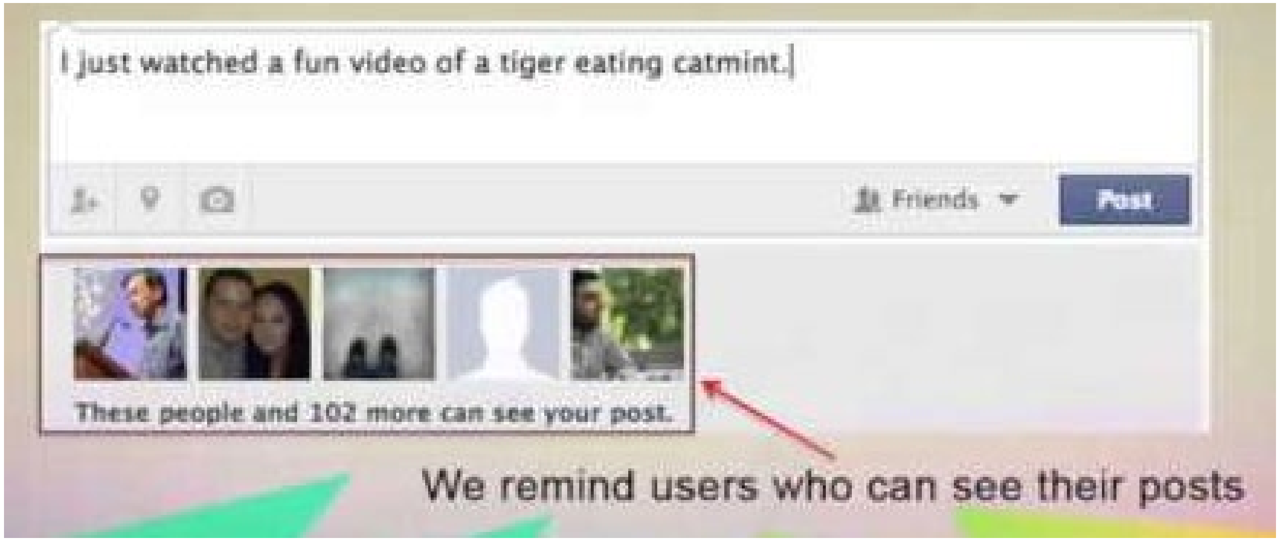
Mary [redacted] likes Jeanine [redacted]'s status.

Shared with public

What about:

- Likes
- Comments
- Friends of friends
- Tags
- Photos
- Etc.

How do I know who is my audience?

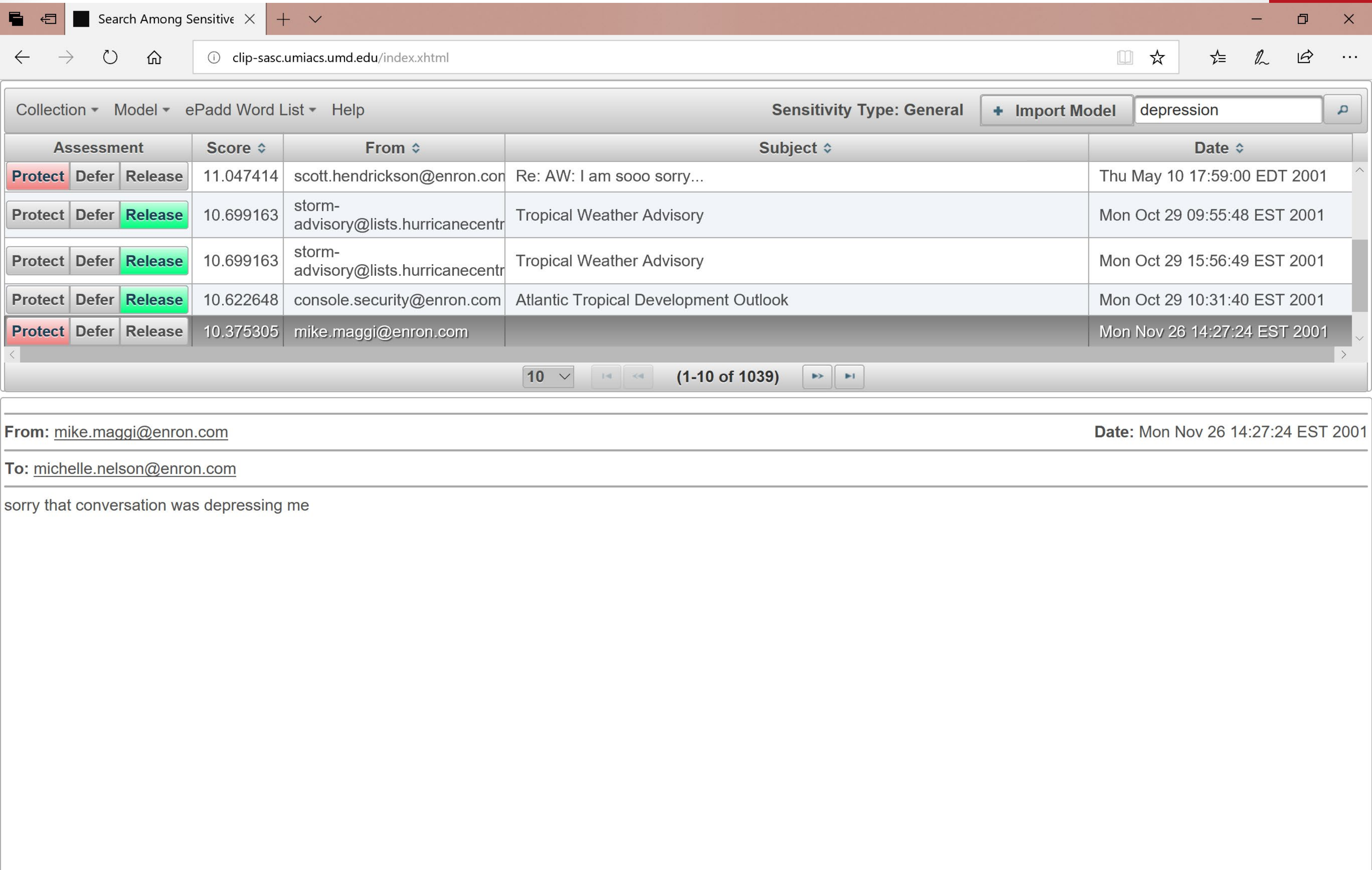


I just watched a fun video of a tiger eating catmint.

These people and 102 more can see your post.

We remind users who can see their posts

http://clip-sasc.umiacs.umd.edu/



Sensitivity Type: General

depression

Assessment	Score	From	Subject	Date
Protect Defer Release	11.047414	scott.hendrickson@enron.con	Re: AW: I am sooo sorry...	Thu May 10 17:59:00 EDT 2001
Protect Defer Release	10.699163	storm-advisory@lists.hurricanecentr	Tropical Weather Advisory	Mon Oct 29 09:55:48 EST 2001
Protect Defer Release	10.699163	storm-advisory@lists.hurricanecentr	Tropical Weather Advisory	Mon Oct 29 15:56:49 EST 2001
Protect Defer Release	10.622648	console.security@enron.com	Atlantic Tropical Development Outlook	Mon Oct 29 10:31:40 EST 2001
Protect Defer Release	10.375305	mike.maggi@enron.com		Mon Nov 26 14:27:24 EST 2001

10 (1-10 of 1039)

From: mike.maggi@enron.com **Date:** Mon Nov 26 14:27:24 EST 2001

To: michelle.nelson@enron.com

sorry that conversation was depressing me

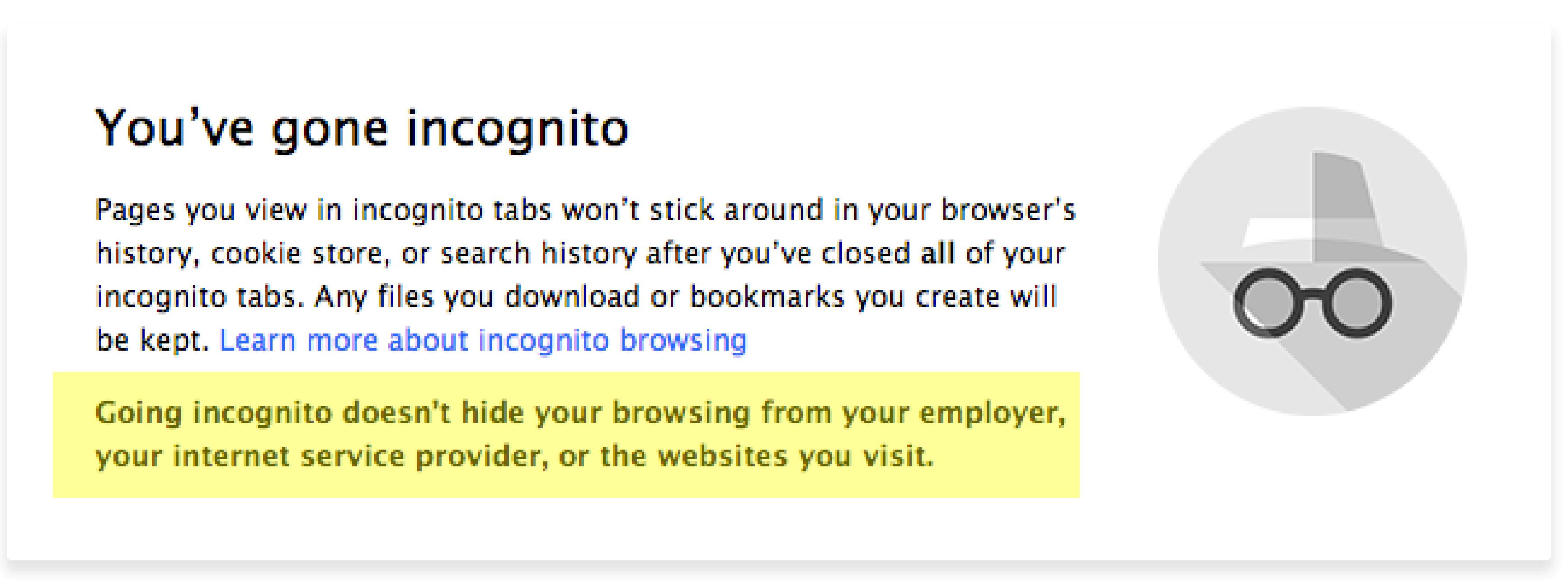
The public-private fallacy

Establishing privacy in public?

Technically possible, practically useless.

The public-private fallacy

Hiding online is even harder.
Think you’re anonymous online? You’re not.



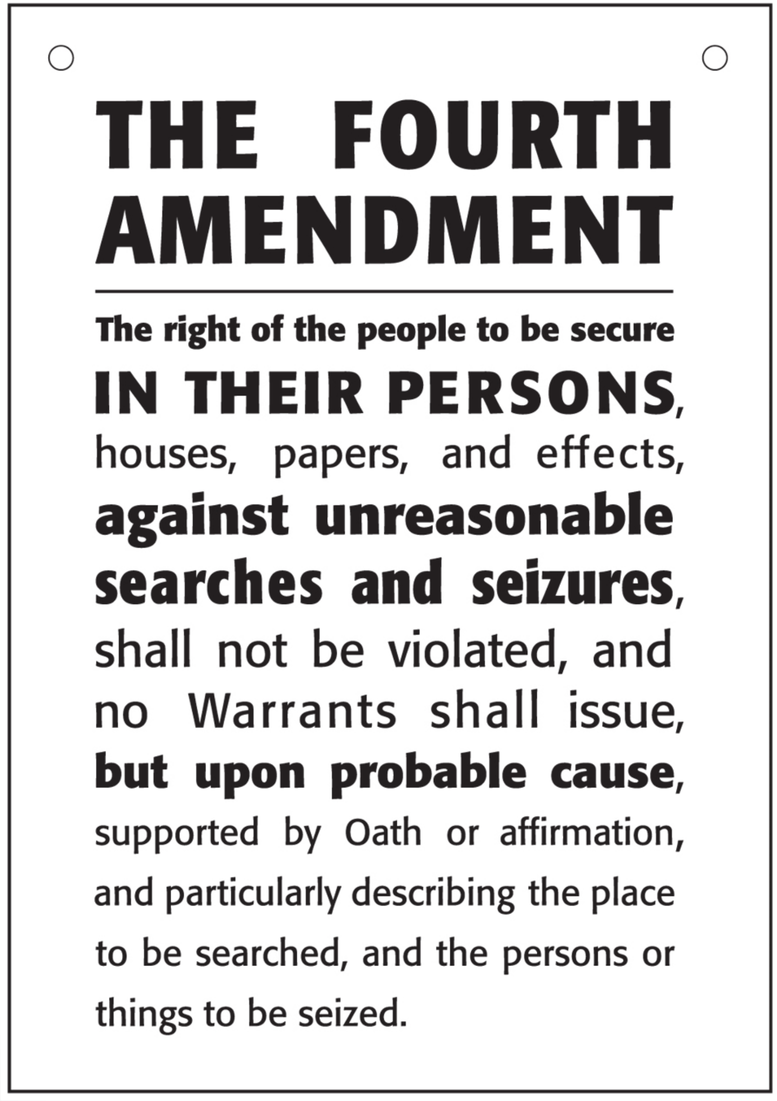
You’ve gone incognito

Pages you view in incognito tabs won’t stick around in your browser’s history, cookie store, or search history after you’ve closed all of your incognito tabs. Any files you download or bookmarks you create will be kept. Learn more about incognito browsing

Going incognito doesn't hide your browsing from your employer, your internet service provider, or the websites you visit.

So how can we manage our digital privacy (individually, as a society)?

Pass laws



THE FOURTH AMENDMENT

The right of the people to be secure **IN THEIR PERSONS**, houses, papers, and effects, **against unreasonable searches and seizures**, shall not be violated, and no Warrants shall issue, **but upon probable cause**, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

Katz v. United States (1967)

Extend 4th Amendment to include **"reasonable expectation of privacy.**"

Pass laws

Other really important privacy laws:

HIPAA (Health Insurance Portability and Accountability Act of 1996) provides data privacy and security provisions for safeguarding medical information. Requires electronic health records are properly secured.

FERPA (Family Educational Rights and Privacy Act protects the privacy of student education records. This requires written consent before student PII is released.

COPPA (Children's Online Privacy Protection Act of 1998) protects the privacy of children under 13 through additional requirements of websites that cater to young users. This is why most websites require users to be 13.

Pass laws

Strengths/weaknesses:

- Can provide quite strong deterrence (especially under criminal law)
- (good) laws are **specific** (to prevent abuse, enable judicious application), but privacy is **heavily contextual**
  - hard to exhaustively enumerate in advance where/when/with-what you have a reasonable expectation of privacy

Be obscure

What is obscurity?
The state of being hard to understand or interpret; unclear; murky

Woody Hartzog argues: when data are obscure, they are "safer."

People may engage in **data obfuscation.**

- For example, lying about your age when registering for a website.

Think about Where's Waldo as a good example of obscurity.

Be obscure

So what is obscurity?
The state of being hard to understand or interpret; unclear; murky

How do I make my data more secure?

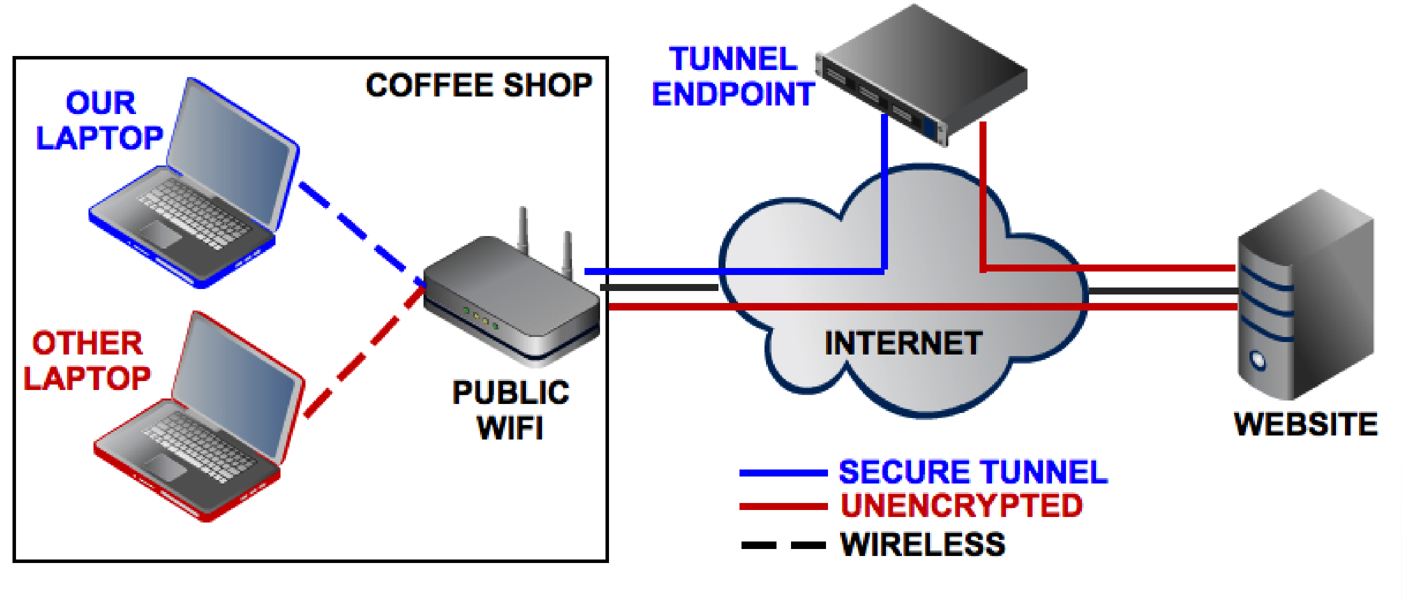
- Hide it from search engines
- Use privacy settings & pseudonyms
- Engage in social steganography/“vaguebooking”

“Many contemporary privacy disputes are probably better classified as concern over losing obscurity.” → Think Spokeo!

Be obscure with VPNs

VPN = virtual private network

- VPNs allow you to extend a private network (e.g., your home network) to public spaces and allows greater protection and anonymity of your data.
- VPNs can allow access to otherwise blocked content (e.g., censored websites in China)



Why use VPNs?

Increased interest in fallout of new legislation allowing ISPs to sell your browsing data to third parties (see this article).

Lots of news coverage encouraging people to use VPNs since they “hide” your traffic from your ISP.

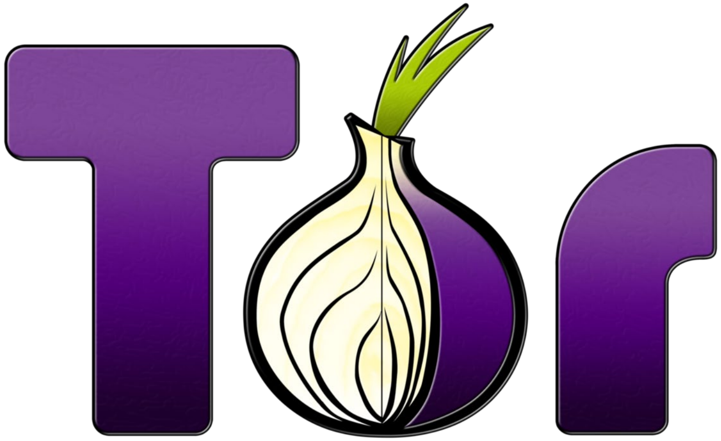
HOWEVER: VPNs are useful, but the aren’t going to “solve” the problem. They have limitations. For example:

- Netflix blocks most VPNs
- Your VPN could be selling your data.
- VPNs Won’t Save You from Congress’ Internet Privacy Giveaway [WIRED]

Read more:

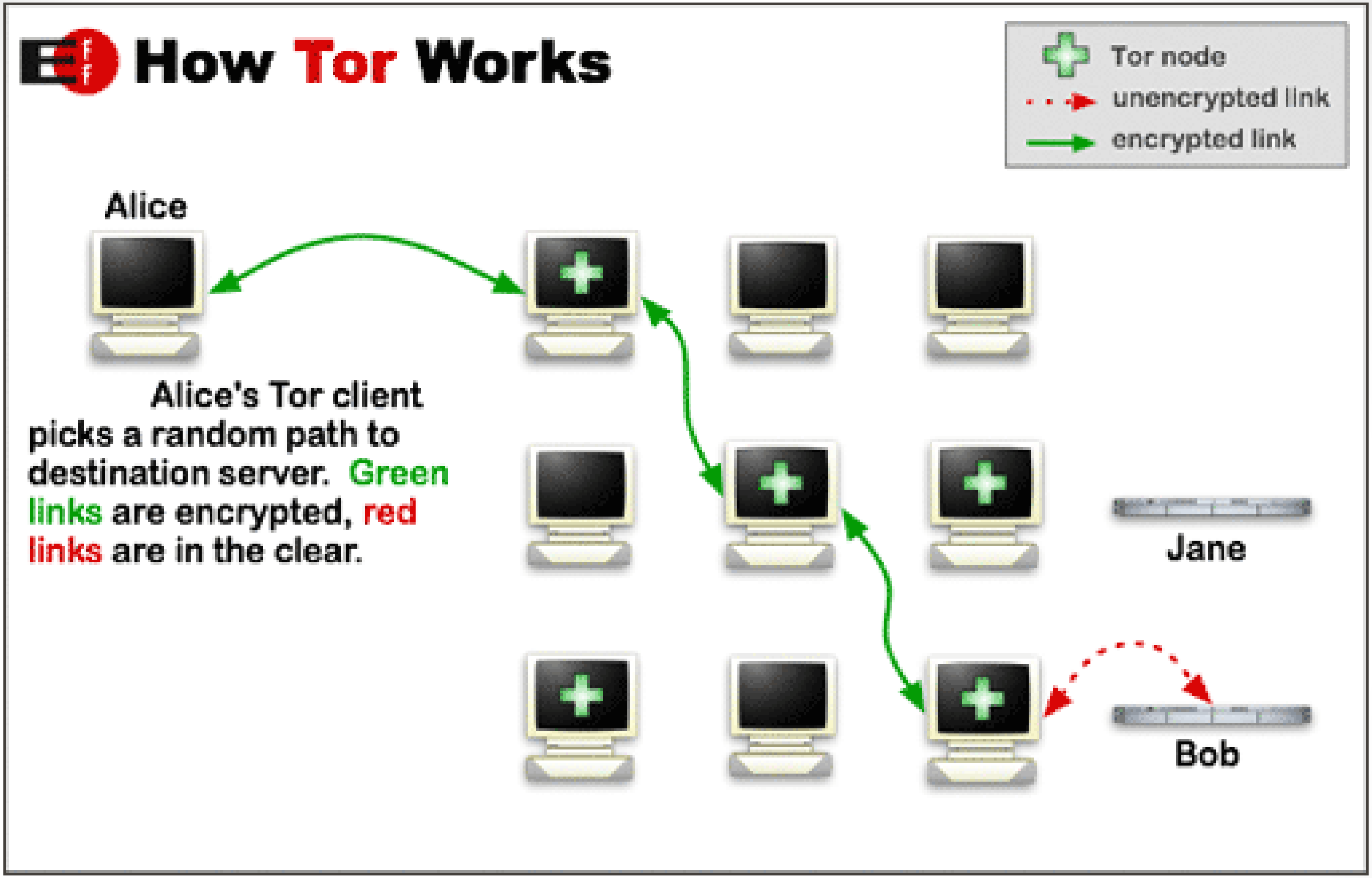
- The impossible task of creating a “Best VPNs” list today [Ars Technica]
- Online privacy? Forget it, even with VPN [USA TODAY]

Be obscure with Tor



Tor is an
“anonymity network.”

From Wikipedia: *“Tor directs Internet traffic through a free, worldwide, volunteer network consisting of more than seven thousand relays to conceal a user's location and usage from anyone conducting network surveillance or traffic analysis.”*



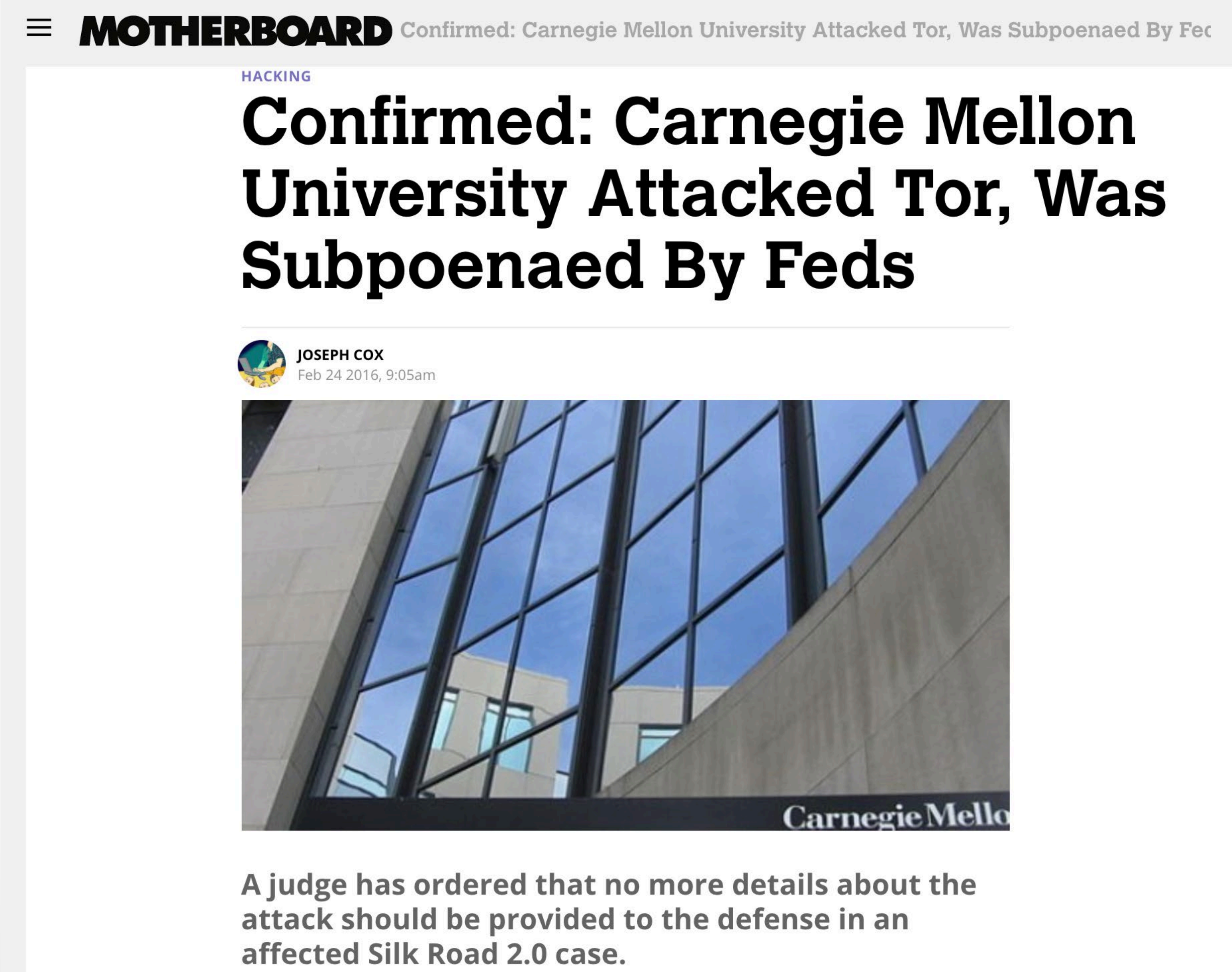
How Tor Works

Alice's Tor client picks a random path to destination server. Green links are encrypted, red links are in the clear.

Goal: Conceal identity of users from surveillance.

Process: Multilayer encryption (hence onion metaphor), then send content through a random network of relays.

Even Tor isn't a failsafe!



MOTHERBOARD Confirmed: Carnegie Mellon University Attacked Tor, Was Subpoenaed By Fec

HACKING

Confirmed: Carnegie Mellon University Attacked Tor, Was Subpoenaed By Feds

JOSEPH COX
Feb 24 2016, 9:05am

A judge has ordered that no more details about the attack should be provided to the defense in an affected Silk Road 2.0 case.

Be ephemeral

Apps highlighting ephemeral communication are becoming increasingly popular.

But are these a good substitute for other forms of communication?

See Bruce Schneier’s post on this topic.

Flickr: jessycat_techie

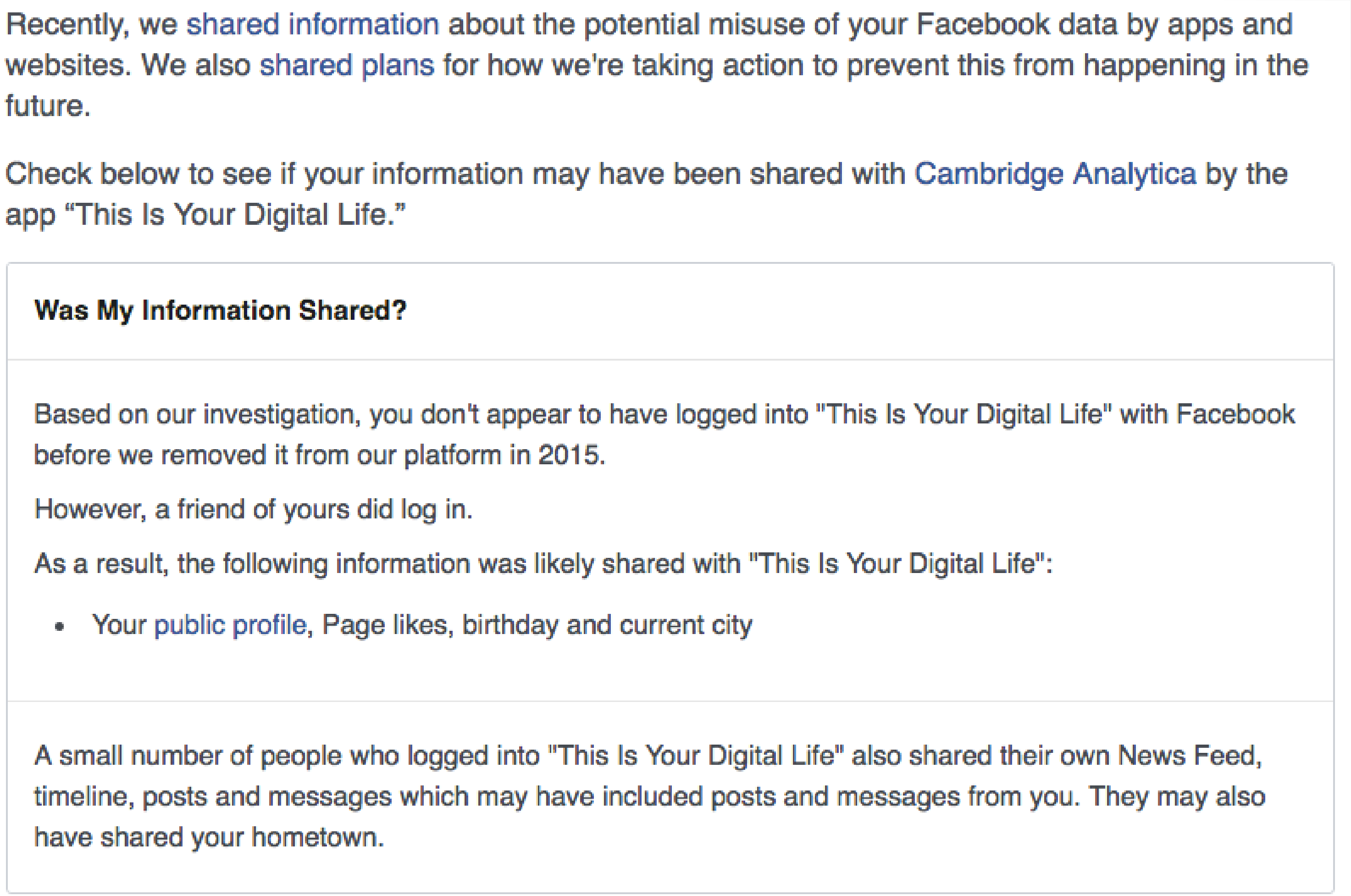
Be judicious… but how?

The public-private fallacy

1. People will sacrifice privacy & security for usability.
2. Companies structure their services to encourage sharing information with a wide audience.

Don't believe me? Watch this.

MORDAC, THE PREVENTER OF INFORMATION SERVICES.
SECURITY IS MORE IMPORTANT THAN USABILITY.
scottadams@aol.com
www.dilbert.com
IN A PERFECT WORLD, NO ONE WOULD BE ABLE TO USE ANYTHING.
©2007 Scott Adams, Inc./Dist. by UFS, Inc.
11-16-07
To complete the log-in procedure, stare directly at the sun.



Recently, we shared information about the potential misuse of your Facebook data by apps and websites. We also shared plans for how we're taking action to prevent this from happening in the future.

Check below to see if your information may have been shared with Cambridge Analytica by the app "This Is Your Digital Life."

Was My Information Shared?

Based on our investigation, you don't appear to have logged into "This Is Your Digital Life" with Facebook before we removed it from our platform in 2015.

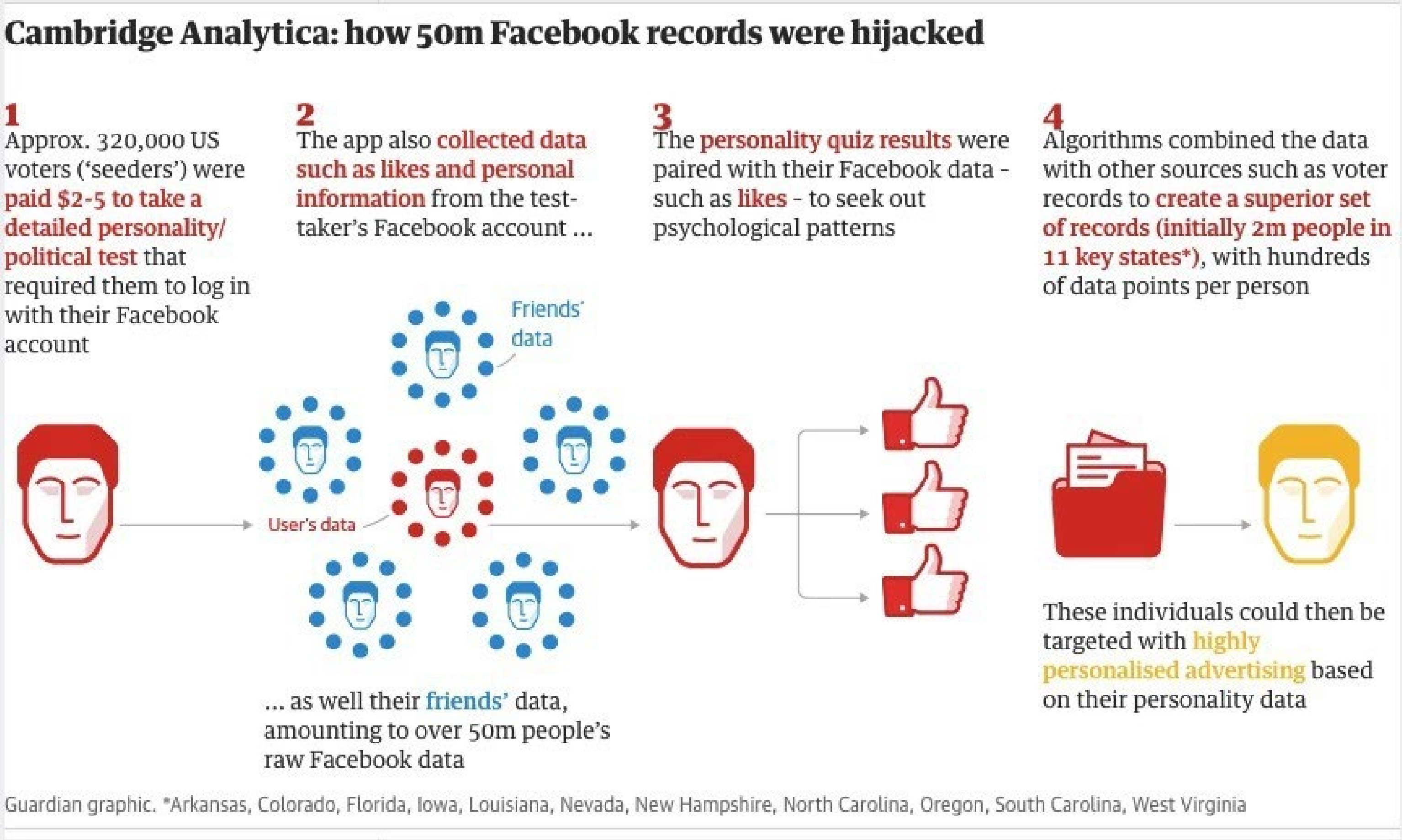
However, a friend of yours did log in.

As a result, the following information was likely shared with "This Is Your Digital Life":

- Your public profile, Page likes, birthday and current city

A small number of people who logged into "This Is Your Digital Life" also shared their own News Feed, timeline, posts and messages which may have included posts and messages from you. They may also have shared your hometown.

Be judicious… but how?



Cambridge Analytica: how 50m Facebook records were hijacked

1
Approx. 320,000 US voters ('seeders') were **paid $2-5 to take a detailed personality/ political test** that required them to log in with their Facebook account

2
The app also **collected data such as likes and personal information** from the test-taker's Facebook account …

… as well their **friends'** data, amounting to over 50m people's raw Facebook data

3
The **personality quiz results** were paired with their Facebook data - such as **likes** - to seek out psychological patterns

4
Algorithms combined the data with other sources such as voter records to **create a superior set of records (initially 2m people in 11 key states*)**, with hundreds of data points per person

These individuals could then be targeted with **highly personalised advertising** based on their personality data

Guardian graphic. *Arkansas, Colorado, Florida, Iowa, Louisiana, Nevada, New Hampshire, North Carolina, Oregon, South Carolina, West Virginia

So how do we protect our data?

We’ll talk about that on Friday:

- What counts as “me”?
- Sociotechnical solutions for <u>securing</u> “me”