The Cognitive Dissonance of Living in a World of “Big Data”
(And what the law is trying to do to regulate it)

Abe Hsuan
December 1, 2011
What is “cognitive dissonance”

*Wikipedia*: “Cognitive dissonance is a discomfort caused by holding conflicting ideas simultaneously.”

*Encyclopedia Britannica*: “the mental **conflict** that occurs when beliefs or assumptions are contradicted by new information.”
Why dissonance?

Virtual v. Physical World

Authoritarian v. Democratic

Public v. Private

Local v. Global

State v. Non-State

Published Media v. Social Media

Work vs. Personal Life
When people suffer from ‘cognitive dissonance’:

**Wikipedia:** “...people have a motivational drive to reduce dissonance...by **changing their attitudes, beliefs, and actions.** Dissonance is also reduced by **justifying, blaming, and denying.**”

**Britannica:** “... the person **rejects, explains away, or avoids the new information**, persuades himself that no conflict really exists, reconciles the differences, or resorts to any other **defensive means of preserving stability or order** in his conception of the world and of himself.”
The latest cause is “Big Data” – Large Datasets, Real Time, Diversity of Data Types

According to The Economist, ‘the amount of digital information increases 10X every five years (outpacing available storage).

(From “Data, data everywhere” The Economist - February 27, 2010)
We are seeing the benefits of how large datasets are revolutionizing whole fields and industry sectors.

Genomics
Health Care
Social Sciences
Crime Prevention
Epidemiology
etc. etc.
But we are also seeing (and can foresee) the possibilities of abuse

Identity Theft
Privacy Violations
Surveillance
Censorship
Human Rights Abuse
Etc.
To every man is given the key to the gates of heaven; the same key opens the gates of hell.

Buddhist proverb quoted by Richard P. Feynman
Facebook Wants to End Anonymity on Internet

8:30 PM - August 15, 2011 - By Kevin Parrish -
Source: Ubergizmo

In a recent panel, Facebook marketing director Randi Zuckerberg said that anonymity on the Internet must go away.

Randi Zuckerberg, the sister of Facebook CEO Mark Zuckerberg and the site's current marketing director, believes that anonymity on the Internet has to "go away." She made this revelation during a recent panel discussion on social media and cyberbullying hosted by Marie Claire magazine.

"The use of real names online could help curb bullying and harassment on the web," she told the audience. "I think anonymity on the Internet has to go away. People behave a lot better when they have their real names down. I think people..."
Does anonymity even exist in a world of Big Data?

Netflix Challenge

In October 2006, Netflix anonymized and released user records for 100 million records of over 500,000 users spanning 1999 to 2005.

• For each anonymous user, records contained the movies watched, any 1-5 star movie ratings, and the dates of the ratings.

• In total, records represented less than $\frac{1}{10}$ of Netflix data in 2005.
Turns out the dataset is very sparse

- 99% of subscribers can be uniquely identified by knowing 8 movie ratings and the dates of ratings (with a 14-day error).
- 68% based on two ratings and dates (with a 3-day error)

From: Arvind Narayanan and Vitaly Shmatikov, How to Break the Anonymity of the Netflix Prize Dataset (October 16, 2006)
Even knowing only the movies rated, without the dates when the ratings were made . . .

84% of subscribers can be uniquely identified with ratings for 6 out of 8 “long tail” movies (i.e., not top 100 or top 500 “blockbuster” movies)

Figure 8. Adversary knows exact ratings but does not know dates at all.
Netflix – IMDb mashup

It took Univ of Texas professors Arvind Narayanan and Vitaly Shmatikov just 2 weeks to match Netflix records with public information scraped from IMDb.com to identify 2 subscribers:
Mashups have been around for a long time

Donnelly, Polk, etc. using the most advanced computing power at the time took 5 years to mash 1970s U.S. Census data (aggregates of 1,500 households) into datasets that were demographically accurate for over 50% of all households.
But growth of Web 2.0 offers unprecedented amount of re-identification mash-up opportunities
And the virtual (and physical) world is only becoming more connected, not less
brilig.com – Open platform/exchange for businesses to analyze and share data openly
AudienceScience – Behavioral targeting using media consumption to ID segments
Blue Kai – Shopping activities matched with same user’s media consumption
Epsilon/Abacus – Matches people’s online identities with catalog/direct-marketing data
TargusInfo – Matches offline data about users with their online presence
ClearSight Interactive - Demographic data matched to 200 million-plus IP addresses
bluecava – Demographic data matched to 125+ million unique device IDs

Most companies do one thing well. We do at least two.

We target online advertising better. And we help fight fraud. With increasing frequency these days companies are saying to us, “Hey, that’s really great technology. I bet I can use it for [fill in the blank with something interesting].”

In the center of BlueCava’s brand is our patented way of identifying the unique fingerprint of any Internet-connected electronic device. Our universe is big. Like 10 Billion devices in the world big. That’s our sandbox.

Folks use BlueCava for ad targeting because targeting devices is an excellent way to identify and accumulate audiences. BlueCava also helps eliminate online fraud by knowing the identity and behavior of a device. Wouldn’t it be nice to know if that device trying to buy airline tickets from you is really Fred from San Jose or Boris from Bulgaria? BlueCava can help with it all.
Sense Networks – Indexes people’s real world habits via mobile and location-based sensors
Today ...
This is where all the data comes from
You need a hustler, a designer and a "prodineer."
But is there any harm?

U.S. Courts have been finding it difficult to identify the harm the laws are protecting against:

California court found plaintiffs in a class action against behavioral ad network (Specific Media) failed to articulate how use of “flash cookies” caused them economic injury.

Doubted plaintiffs could, even as a “class action,” meet the $5,000 threshold of injury required.
Professor Paul Ohm speaks of a “Database of Ruin”

In “Broken Promise of Privacy,”* an influential paper by leading privacy advocate and law professor Paul Ohm:

“For almost every person on earth, there is at least one fact about them stored in a computer database that an adversary could use to blackmail, discriminate against, harass, or steal the identity of him or her.”

For almost every one of us, then, we can assume a hypothetical “database of ruin”

Ohm’s answer is to regulate based on what some are calling a “Privacy Precautionary Principle”*

- Prosecute inadequate anonymization efforts
- Advocate total ‘opt-in’ consent model*
- Protect users against themselves (especially as to social media)

EU data laws revisited after a decade-plus

<table>
<thead>
<tr>
<th>E-Privacy Directive</th>
<th>Data Protection Directive</th>
</tr>
</thead>
</table>
Seven basic requirements to uphold

- **Notice**: notice when data is being collected
- **Purpose**: data used only for the purpose stated
- **Consent**: no data disclosure without consent
- **Security**: collected data to be kept secure
- **Disclosure**: disclosure about data collector
- **Access**: access and right to correct to data
- **Accountability**: data collectors are accountable
But the Web is no longer just about chat rooms, MUDs, etc.

- Mobile devices bridge the gap between virtual and physical world
- We expect (and count on) the Web to augment reality
Mashup of online and offline: Recent Carnegie Mellon Study released at Black Hat (Aug 2011)

First: Identified 10% of 5,000 users of U.S. dating site in one city (most using aliases) with face-recognition software 280,000 images found by search engine among Facebook profile pictures

Second: Identified 1/3 of 93 CMU students against 250,000 publicly available Facebook profiles:
  • Mining public sources (gov’t databases, etc.) then discovered at least one personal interest of each student and first five digits of Social Security Number
Immersive Labs – Digital signage that uses facial recognition

Adaptive Advertising

Artificial Intelligence Software for Digital Signs enables out-of-home networks, retailers and advertisers to deliver tailored messages to customers in real-time.

- Facial Detection and Machine Learning
- Targeted Ads with Reporting
- Increased CPM and ROI

Learn More | Demo Video
SceneTap – Facial recognition technology offered as an app, helping bar hoppers find where to go
Butterfly effect: dissonance caused when both law and citizens are fluttering at the same time
Sometimes a “Rule of Law” can have tragic results

**Couple with no gas found dead**

A coroner is to call for a change in the law after an elderly couple were found dead in their home weeks after their gas supply was cut off, due to the non-payment of a £140 bill.

George Bates, 89, and his 86-year-old wife Gertrude were found in a decomposed state in October in the south London house they had shared for 64 years.

British Gas said the Data Protection Act prohibited them from passing information on the situation to social services.

On Monday, coroner Dr Paul Knapman recorded a verdict of death by natural causes.
Unforgiving and unrelentless pace of proliferation

“Miss Kim” posts pics on South Korean social media site

South Korean bloggers crowdsource identity

South Korean mainstream press

Popular U.S. blogger blogs the story

BoingBoing picks up story

100s 10,000s 1,000,000s 10,000,000s

The Washington Post
Subway Fracas Escalates Into Test Of the Internet's Power to Shame
By Jonathan Krim
Thursday, July 7, 2005
Here’s what the “dog poop girl” wrote:

“I know I was wrong, but you guys are so harsh. I regret it, but I was so embarrassed so I just wanted to leave there. I was very irritable because many people looked at me and pushed me to clean the poop. Anyhow, I’m sorry. But, if you keep putting me down on the Internet I will sue all the people and at the worst I will commit suicide. So please don’t do that anymore.”
Here’s what the “dog poop girl” wrote:

“I know I was wrong, but you guys are so harsh. I’m regret it, but I was so embarrassed so I just wanted to leave there. I was very irritable because many people looked at me and pushed me to clean the poop. Anyhow, I’m sorry. But, if you keep putting me down on the Internet I will sue all the people and at the worst I will will commit suicide. So please don’t do that anymore.”
There is no “delete” button

By JEFFREY ROSEN
Published: July 21, 2010

Four years ago, Stacy Snyder, then a 25-year-old teacher in training at Conestoga Valley High School in Lancaster, Pa., posted a photo on her MySpace page that showed her at a party wearing a pirate hat and drinking from a plastic cup, with the caption “Drunken Pirate.” After discovering the page, her supervisor at the high school told her the photo was “unprofessional,” and the dean of Millersville University School of Education, where Snyder was enrolled, said she was
Courts have ruled that to have an expectation of privacy on social media sites is “wishful thinking”

- In personal injury and worker’s comp cases, courts have been typically allowing access to Facebook where photos and postings contradict plaintiffs’ claims of loss of quality of life, etc.

- Facebook responsible for 1 out of every 5 divorces in the U.S. (Loyola University study) and 4 out of 5 feature social networking evidence (American Academy of Matrimonial Lawyers)
Lawmakers To Facebook: Why Keep 'Likes,' Chats, Deleted Posts?

Austrian law student Max Schrems recently asked Facebook to turn over information on 1,200-plus page report that included not only obvious material like friend lists, “pokes,” people he defriended, material he “deleted,” and event invitations —

Schrems, who joined with some friends to launch Europe v Facebook in August, is using Facebook without having to worry about privacy.” He says in an FAQ about “user control.”

“After studying the privacy policy of Facebook for months I can still not tell you what we put on Facebook so we should have the right to do whatever we want.
Harvard Researchers Accused of Breaching Students' Privacy
Social-network project shows promise and peril of doing social science online

By Marc Parry

In 2006, Harvard sociologists struck a mother lode of social-science data, offering a new way to answer big questions about how race and cultural tastes affect relationships.

The source: some 1,700 Facebook profiles, downloaded from an entire class of students at an "anonymous" university, that could reveal how friendships and interests evolve over time.

It was the kind of collection that hundreds of scholars would find interesting. And in 2008, the Harvard team began to realize that

Jason Kaufman, of Harvard's Berkman Center for Internet & Society, says critics of his research on student Facebook profiles are acting like "academic paparazzi."

Enlarge Image
potential by publicly releasing part of its archive.
Quinnipiac University Professor Alex Halavais studied the tweets sent during the G20 riots.
Library of Congress, Facing Privacy Concerns, Clarifies Twitter Archive Plan

May 7, 2010, 12:00 pm
By Marc Parry

If you don’t want that tipsy 3 a.m. Twitter post preserved for posterity, then start deleting. Now.

Faced with privacy concerns, the Library of Congress is clarifying its plans to archive all public tweets posted since Twitter went live in March 2006. The database won’t contain deleted tweets or private account information, according to a list of frequently asked questions recently posted on the library’s blog.

And the Twitter database will only be made available to “qualified researchers,” Martha Anderson, director of the library’s National Digital Information Infrastructure and Preservation Program, tells The New York Times. The plan is to embargo messages for six months before making them available, but that period could be extended, she says.

“There’s concern about privacy issues in the near term, and we’re sensitive to these concerns,” she says. “We may have to filter certain things or wait longer to make them
Should there be a “Right to Be Forgotten”? 

EU justice commissioner Viviane Reding has suggested that the EU’s Data Protection Directive needs to be amended to provide EU citizens with a “right to be forgotten,” i.e., to have their information withdrawn (not just hidden) from websites.
91% of HR Departments using Social Media to evaluate job candidates


Study conducted by Reppler and Lab42*
College admissions

24% of college admissions officers look at social media and 20% conduct online searches*

*Kaplan Test Prep’s 2011 survey of college admissions officers
Zhao Jing or “Michael Anti”?

“I have used the name for more than a decade - my reputation is based on that. No one in the virtual world knows my real name.”

Quoted by Richard Water at TodayOnline.com “Identity and the Net: From pixels to persona” (Nov 23, 2011)
"Having two identities for yourself is an example of a lack of integrity."

Mark Zuckerberg quoted by David Kirkpatrick in “The Facebook Effect”
What role can or should social media sites play in democratic uprisings, human rights activism, dissent?

From February 2011 ...  

... to May 2011
The comings and goings of German Green party politician, Malte Spitz a la Deutsche Telkom

From Aug 31, 2009, to Feb. 28, 2010, Deutsche Telekom recorded and saved Spitz’s longitude and latitude coordinates more than 35,000 times.
Hasan Elahi: Professor and performance artist, once tracked by the FBI for 6 months in 2002
MIT Researcher Deb Roy records 120,000 hours of video footage of his infant son

Peter Yang, From Wired Issue 15:04 March 2007
Gordon Bell of Microsoft Research innovates around “lifelogging”
Jurisprudence around privacy developed because of this invention:

Invented in 1888, snapshot cameras altered our sense of when we should have a *reasonable expectation of privacy* . . . a “right to be left alone”

First “privacy” law enacted by NY in 1903, addressed uses of photographic likenesses
Frequency of the words “secrecy” versus “privacy” in books from 1890-2000
“Privacy” concerns spike when there is discomfort with what technology makes possible.
Are we merely talking about another kind of “photograph” – that of our “digital shadow”?

http://digitalslander.files.wordpress.com/2008/08/digital-shadow.jpg
Fact: It’s not possible to get “off the grid”
Are we suffering from “cognitive dissonance” or is this a step in adapting toward higher intelligence?

"The test of a first-rate intelligence is the ability to hold two opposing ideas in mind at the same time and still retain the ability to function."

F. Scott Fitzgerald:
Thank You

Abe Hsuan
AbeHsuanEsq@gmail.com