Ethics and consent; Understanding phishing attacks

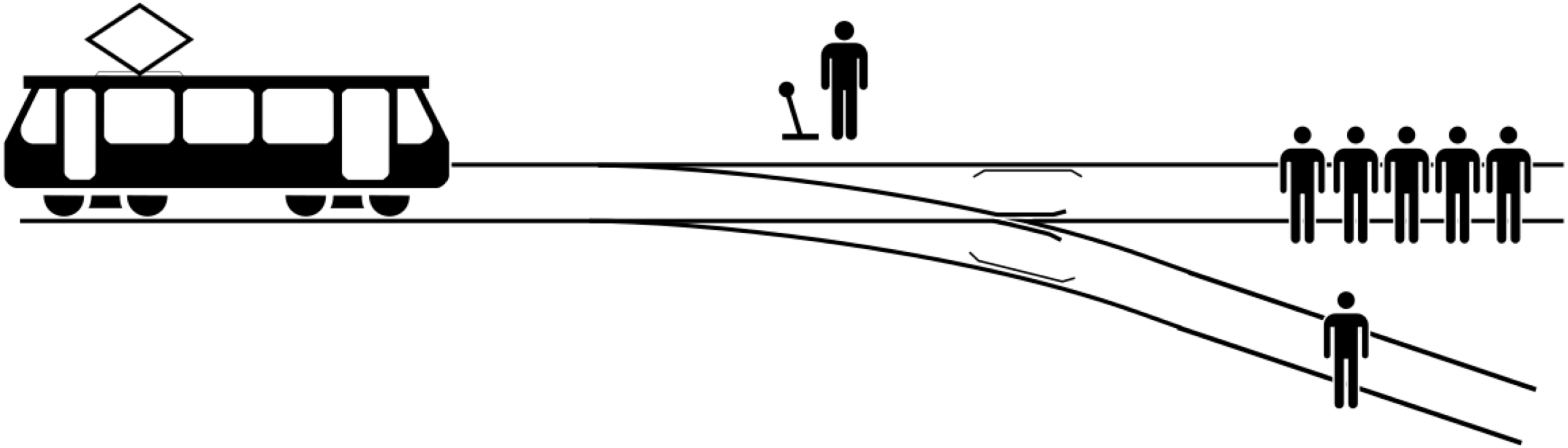
Mainack Mondal

CS 60081
Autumn 2022

Roadmap

- What is ethics in our case?
    - How to obtain consent
- Case study: ethical phishing experiments

Ethics



What we are interested in: Social ethics

- Why?

What we are interested in: Societal ethics

- Why?
    - In every study you do you are ultimately creating a computational system / observing a computational system and obtaining human feedback on how they interact with the system or how the system affected them

What we are interested in: Social ethics

- Why?
  - In every study you do you are ultimately creating a computational system / observing a computational system and obtaining human feedback on how they interact with the system or how the system affected them
  - Human subject research

why should we care?

Case 1: Tuskegee Syphilis Experiment

- Between 1932 and 1972
  - Experiment done by US public health service
- 600 African American males from Alabama
  - Told they are part of a “bad blood” study for 6 months
  - Promised free medical care and food
  - They did not know they had Syphilis
- In actuality
  - Went on for 40 years
  - Cure was found in 1940, they were NOT given
  - Finally:

Case 1: Tuskegee Syphilis Experiment

- Between 1932 and 1972
  - Experiment done by US public health service
- 600 African American males from Alabama
  - Told they are part of a “bad blood” study for 6 months
  - Promised free medical care and food
  - They did not know they had Syphilis
- In actuality
  - Went on for 40 years
  - Cure was found in 1940, they were NOT given
  - Finally: 28 dead, 100 died of related complications. 40 wives contacted the disease, 10 children born with congenital syphilis

Case 2: Stanford prison experiment

- Conducted in August 14 – 20, 1971 by Philip Zimbardo
- Sampled 24 healthy and mentally stable students
    - After only one day things went south
    - One prisoner acted “crazy”
    - Guards started abusing prisoners (serious physical abuse)
    - Prisoner No. 416, a newly admitted stand-by prisoner, expressed concern about the treatment of the other prisoners. The guards responded with more abuse
    - Continued even after the “prisoners” wanted to withdraw

A possibility of serious long lasting mental harm

Case 2: Stanford prison experiment

> *"The Stanford Prison Experiment led to the implementation of rules to preclude any harmful treatment of participants. Before they are implemented, human studies must now be reviewed and found by an institutional review board (US) or ethics committee (UK) to be in accordance with ethical guidelines set by the American Psychological Association. These guidelines involve the consideration of whether the potential benefit to science outweighs the possible risk for physical and psychological harm."*

https://en.wikipedia.org/wiki/Stanford_prison_experiment

Want more examples?

- Tea Room study:
    - Revealed sexual preferences of people in an public book
- Wilbrock Hepatitis study:
    - Infected children with Hepatitis virus
- Milgram shock study:
    - Asked people to kill with electric shock (and they believed)
- We are not even touching Nazi human experimentation

Experimentation involving human subjects require care and compassion

History of ethics

- 1972: End of Tuskegee study
- 1974: US congress created commission to study research ethics and create regulations
- 1978: Belmont Report is published detailing rules of “ethical” research
- 1981: These rules become US law
- 2010: To get US funded grants you need to to go through ethics training
- 2012: Menlo report published which updated Belmont report and include regulations around security research

The project assignments are out

(please select a slot for discussuion

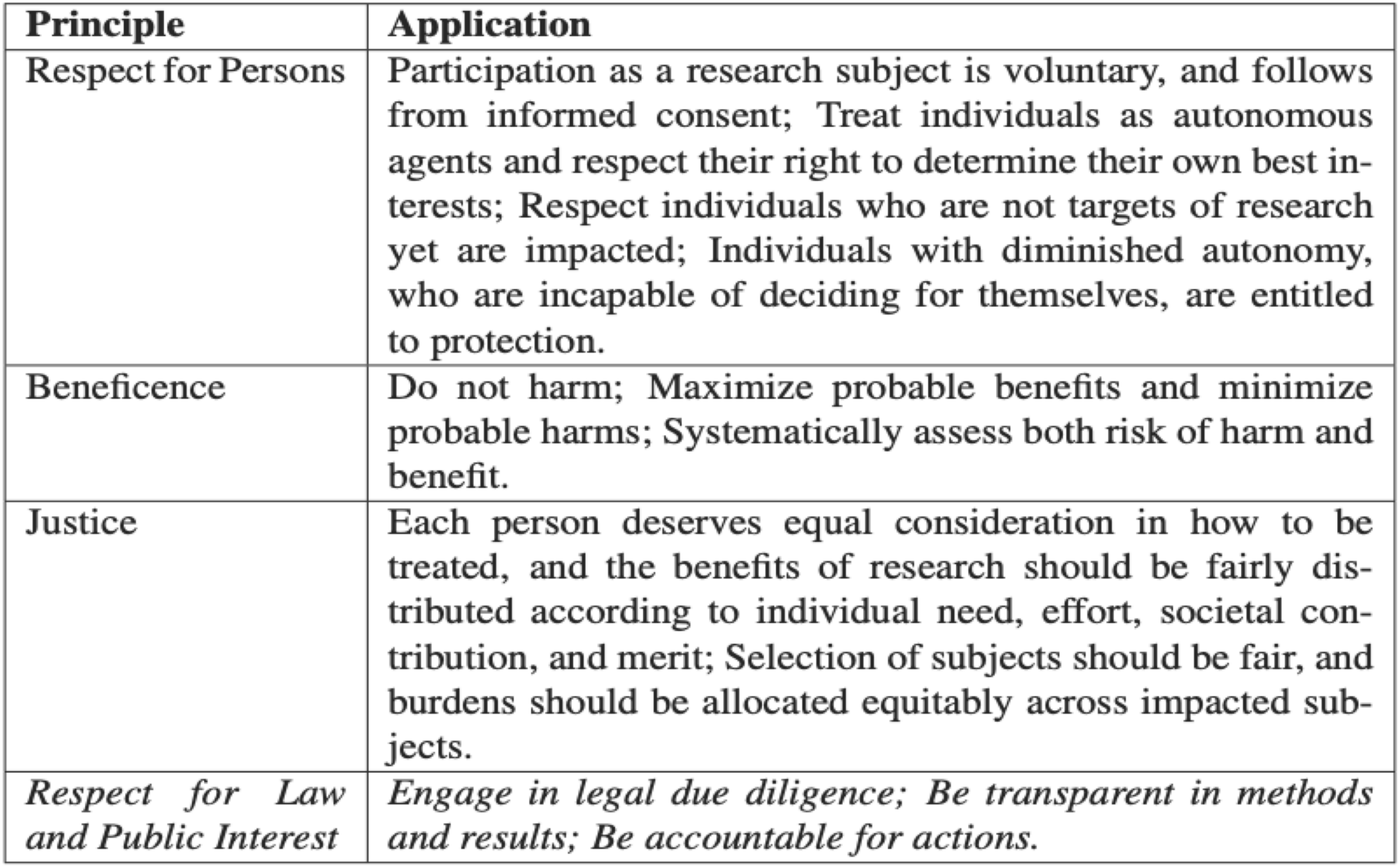
Later this week)

The Belmont report

- Respect for persons
  - Protecting the autonomy of all people and treating them with courtesy and respect and allowing for informed consent. Researchers must be truthful and conduct no deception.
- Beneficence
  - The philosophy of "Do no harm" while maximizing benefits for the research project and minimizing risks to the research subjects
- Justice
  - Ensuring reasonable, non-exploitative, and well considered procedures are administered fairly — the fair distribution of costs and benefits to potential research participants — and equally.

http://www.hhs.gov/ohrp/regulations 7 -and-policy/belmont-report/index.html

The Menlo report



Principle	**Application**
Respect for Persons	Participation as a research subject is voluntary, and follows from informed consent; Treat individuals as autonomous agents and respect their right to determine their own best interests; Respect individuals who are not targets of research yet are impacted; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection.
Beneficence	Do not harm; Maximize probable benefits and minimize probable harms; Systematically assess both risk of harm and benefit.
Justice	Each person deserves equal consideration in how to be treated, and the benefits of research should be fairly distributed according to individual need, effort, societal contribution, and merit; Selection of subjects should be fair, and burdens should be allocated equitably across impacted subjects.
Respect for Law and Public Interest	*Engage in legal due diligence; Be transparent in methods and results; Be accountable for actions.*

Respect for persons

- Treat people as autonomous individuals with free will
- Give them the right to choose and the knowledge so that they can make an *informed* decision
- Persons with diminished autonomy should be protected

- Concrete suggestion
  - Participation should be voluntary
  - Participants should be fully informed of the costs and benefits of participation

Good example of consent

- Let’s look at one of our lab studies:

Perceptions of Retrospective Edits, Changes, and Deletion on Social Media, *Günce Su Yılmaz, Fiona Gasaway, Blase Ur, Mainack Mondal.* ICWSM’21

Good example of consent

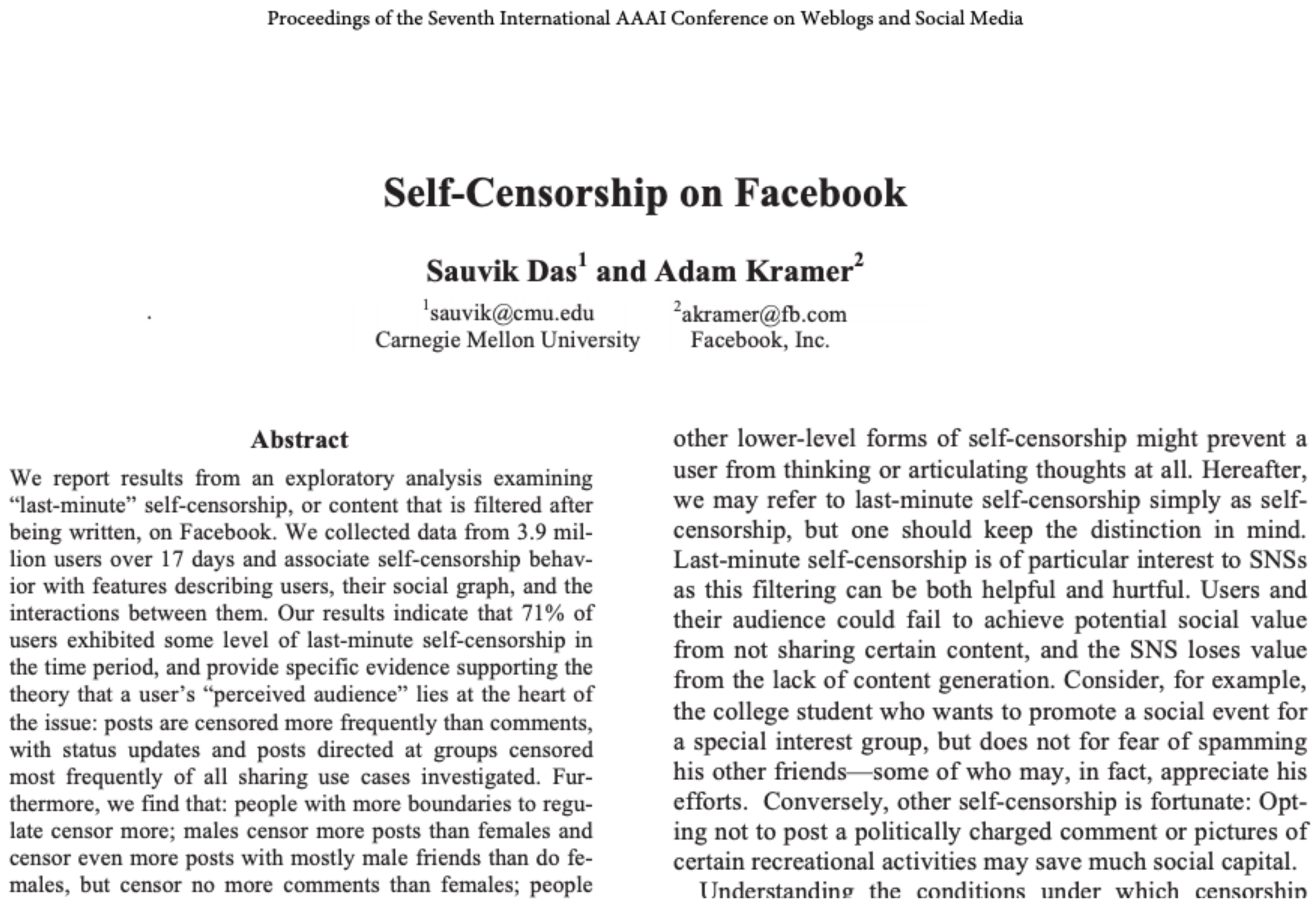
- Let’s look at one of our lab studies:

Perceptions of Retrospective Edits, Changes, and Deletion on Social Media, *Günce Su Yılmaz, Fiona Gasaway, Blase Ur, Mainack Mondal.* ICWSM’21

What about online studies or studies with Amazon mechanical Turk?

Ask their consent in an online form before the study

Bad example of consent: Case 1



Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media

Self-Censorship on Facebook

Sauvik Das[1] and Adam Kramer[2]

[1]sauvik@cmu.edu
Carnegie Mellon University

[2]akramer@fb.com
Facebook, Inc.

Abstract

We report results from an exploratory analysis examining "last-minute" self-censorship, or content that is filtered after being written, on Facebook. We collected data from 3.9 million users over 17 days and associate self-censorship behavior with features describing users, their social graph, and the interactions between them. Our results indicate that 71% of users exhibited some level of last-minute self-censorship in the time period, and provide specific evidence supporting the theory that a user's "perceived audience" lies at the heart of the issue: posts are censored more frequently than comments, with status updates and posts directed at groups censored most frequently of all sharing use cases investigated. Furthermore, we find that: people with more boundaries to regulate censor more; males censor more posts than females and censor even more posts with mostly male friends than do females, but censor no more comments than females; people

other lower-level forms of self-censorship might prevent a user from thinking or articulating thoughts at all. Hereafter, we may refer to last-minute self-censorship simply as self-censorship, but one should keep the distinction in mind. Last-minute self-censorship is of particular interest to SNSs as this filtering can be both helpful and hurtful. Users and their audience could fail to achieve potential social value from not sharing certain content, and the SNS loses value from the lack of content generation. Consider, for example, the college student who wants to promote a social event for a special interest group, but does not for fear of spamming his other friends—some of who may, in fact, appreciate his efforts. Conversely, other self-censorship is fortunate: Opting not to post a politically charged comment or pictures of certain recreational activities may save much social capital.

Understanding the conditions under which censorship

Bad example of consent: Case 1

- Did you know Facebook knows what you typed but not posted?

Bad example of consent: Case 1

- Did you know Facebook knows what you typed but not posted? What about Google?

Bad example of consent: Case 1

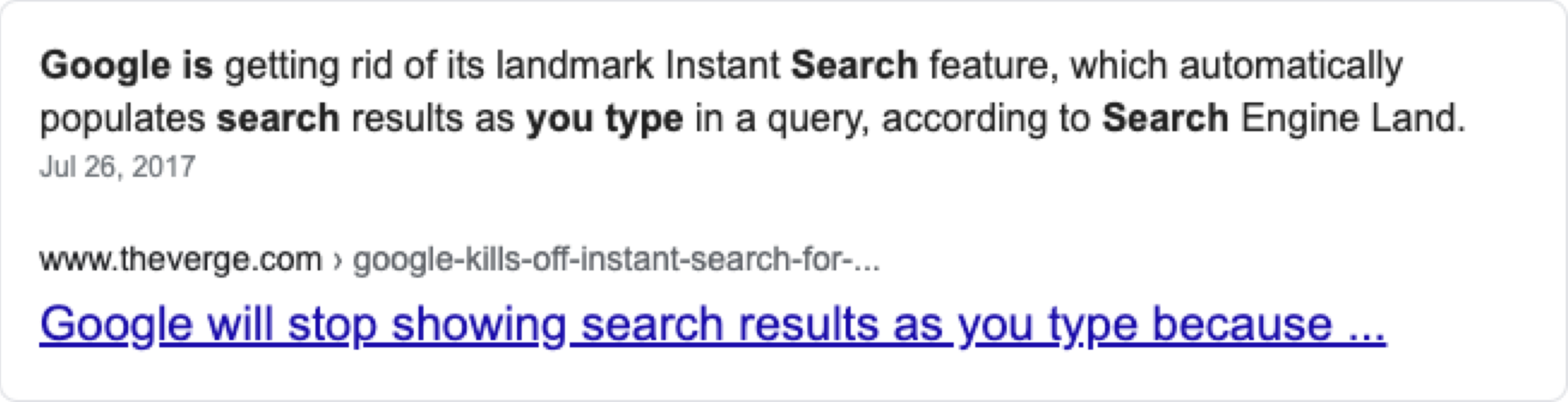
- Did you know Facebook knows what you typed but not posted? What about Google?

- Source: https://www.youtube.com/watch?v=1_Pt7UahrN0

Bad example of consent: Case 1

- Did you know Facebook knows what you typed but not posted? What about Google?

[redacted]



Google is getting rid of its landmark Instant **Search** feature, which automatically populates **search** results as **you type** in a query, according to **Search** Engine Land.

Jul 26, 2017

www.theverge.com › google-kills-off-instant-search-for-...

Google will stop showing search results as you type because ...

[redacted]

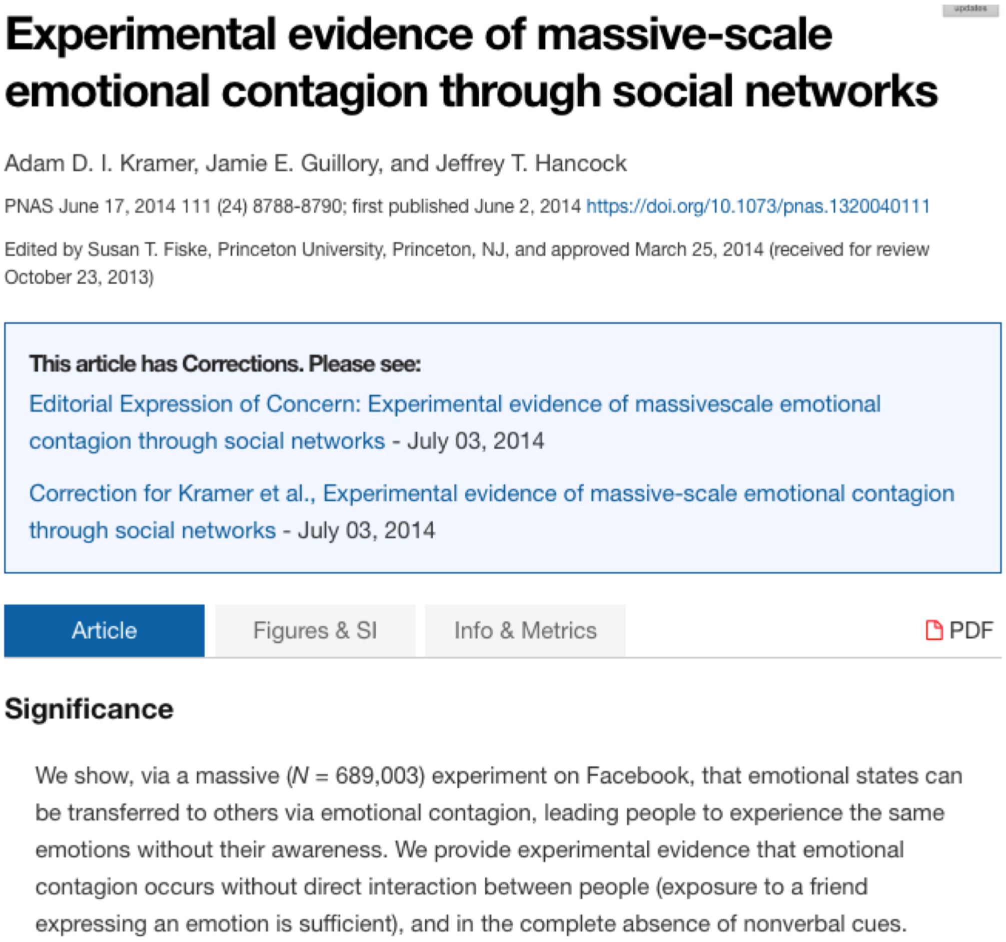
- Source: https://www.youtube.com/watch?v=1_Pt7UahrN0

Bad example of consent: Case 1

- In the “self censorship” study
    - The researchers at Facebook tracked “random sample of approximately 5 million English-speaking Facebook users who lived in the U.S. or U.K. over the course of 17 days (July 6-22, 2012)”
    - Never took user consent
        - Or did they?

Bad example of consent: Case 2

- Facebook strikes again



Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock

PNAS June 17, 2014 111 (24) 8788-8790; first published June 2, 2014 https://doi.org/10.1073/pnas.1320040111

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

This article has Corrections. Please see:

Editorial Expression of Concern: Experimental evidence of massivescale emotional contagion through social networks - July 03, 2014

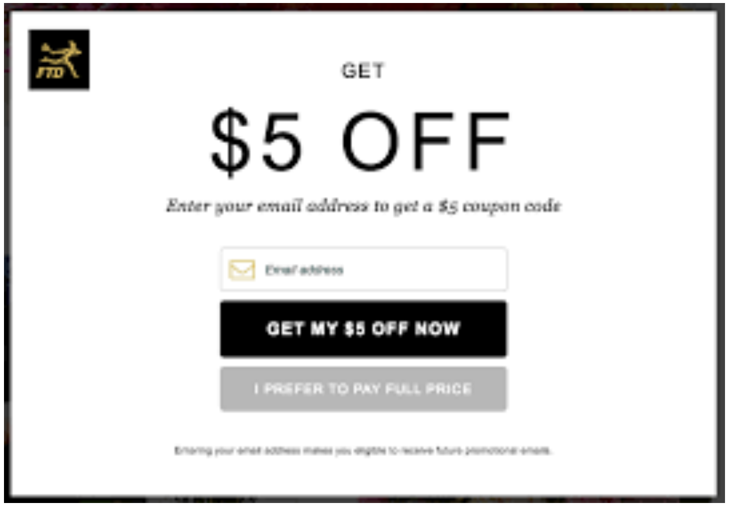
Correction for Kramer et al., Experimental evidence of massive-scale emotional contagion through social networks - July 03, 2014

Significance

We show, via a massive (N = 689,003) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. We provide experimental evidence that emotional contagion occurs without direct interaction between people (exposure to a friend expressing an emotion is sufficient), and in the complete absence of nonverbal cues.

Bad example of consent: Case 3

- Facebook is not alone

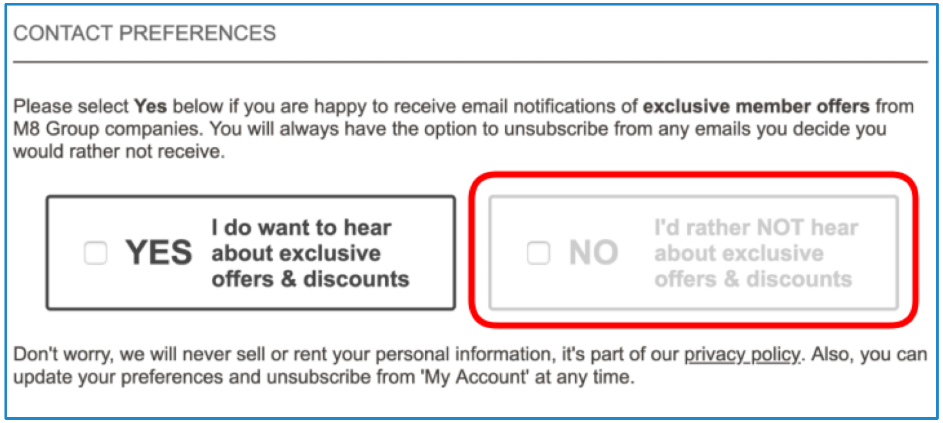


GET

$5 OFF

Enter your email address to get a $5 coupon code

Entering your email address makes you eligible to receive future promotional emails.



CONTACT PREFERENCES

Please select **Yes** below if you are happy to receive email notifications of **exclusive member offers** from M8 Group companies. You will always have the option to unsubscribe from any emails you decide you would rather not receive.

□ **YES** **I do want to hear about exclusive offers & discounts**

□ NO I'd rather NOT hear about exclusive offers & discounts

Don't worry, we will never sell or rent your personal information, it's part of our privacy policy. Also, you can update your preferences and unsubscribe from 'My Account' at any time.

The Belmont report

- Respect for persons (Informed consent)
  - Protecting the autonomy of all people and treating them with courtesy and respect and allowing for informed consent. Researchers must be truthful and conduct no deception.
- Beneficence
  - The philosophy of "Do no harm" while maximizing benefits for the research project and minimizing risks to the research subjects
- Justice
  - Ensuring reasonable, non-exploitative, and well considered procedures are administered fairly — the fair distribution of costs and benefits to potential research participants — and equally.

http://www.hhs.gov/ohrp/regulations 7 -and-policy/belmont-report/index.html

Beneficence

- Do not harm
- Maximize benefits and minimize harms

- Concrete suggestion
  - Create the consent form very carefully
  - It should describe risks and benefits to the participants

Good example of beneficence

- Study: *The Emperor's New Security Indicators An evaluation of website authentication and the effect of role playing on usability studies, IEEE S&P, 2007*

- A deception study
    - Did not tell participants what the goal of the study
    - Participants recruited using on-campus flyers
    - Flyers said the participant could "earn $25 and make online baking better"
    - No mention of security or privacy in any advertising materials or consent form
    - Needed debriefing at the end of the study

The emperor’s study

- RQ: Will users enter their real bank account password even if some/all the security indicators were missing?
    - “Our consent form notified participants that we would be observing their actions. (To obscure the purpose of the study, we did not detail that we were specifically observing password behavior)”
    - “Our observation system did not record user IDs, passcodes, or other private information”
    - “We did not introduce risks to participants beyond those inherent to accessing their bank from a university-managed computer. We took additional technical precautions to protect sensitive information revealed by participants during study tasks”
    - “At the end of the study, we provided participants with a debriefing that explained the purpose of the study, the attack clues that we had presented, the precautions we had taken, and how participants could protect themselves from real site-forgery attacks in the future”

Bad example of beneficence

- RQ: How much oxygen do premature babies need to prevent death or blindness?
    - https://ahrp.org/an-experiment-designed-to-kill-babies/

The Belmont report

- Respect for persons (Informed consent)
  - Protecting the autonomy of all people and treating them with courtesy and respect and allowing for informed consent. Researchers must be truthful and conduct no deception.
- Beneficence
  - The philosophy of "Do no harm" while maximizing benefits for the research project and minimizing risks to the research subjects
- Justice
  - Ensuring reasonable, non-exploitative, and well considered procedures are administered fairly — the fair distribution of costs and benefits to potential research participants — and equally.

http://www.hhs.gov/ohrp/regulations 7 -and-policy/belmont-report/index.html

Justice

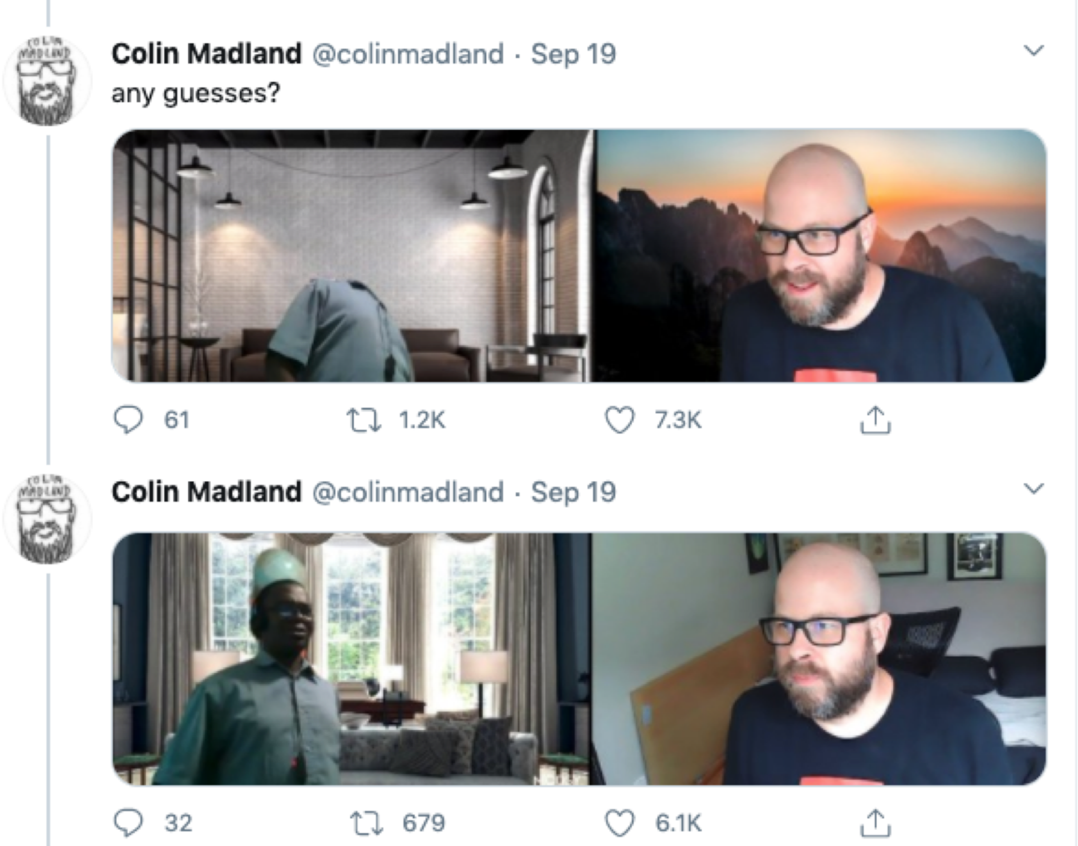
- Who should bear the burdens of research and who should receive the benefits?
  - To each person an equal share
  - To each person according to individual need
  - To each person according to individual effort
  - To each person according to societal contribution
  - To each person according to merit
- Concrete suggestion
  - Selection of research participants
  - Compensation of research participants in consent form

Good example: Refugee study

- “Computer Security and Privacy for Refugees in the United States”, Simko et al., IEE S&P, 2018
- Interviewed case managers, teachers, refugees to US
    - A vulnerable population
    - Did a focus group not to intimidate the refugees
    - Understood the need and barriers of better security practices for refugees

Bad example: Racial bias in AI

- Artificial Intelligence systems (like facial detection) are trained on available data, which can be biased.
  - That data is labeled by often crowd workers



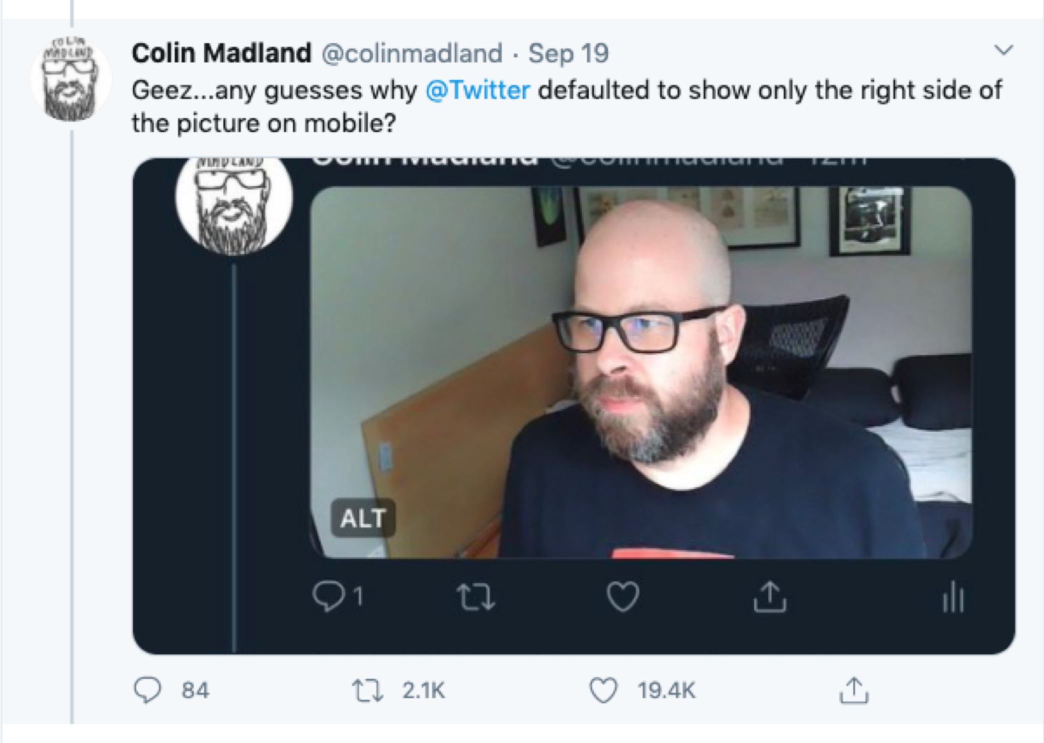
Colin Madland @colinmadland · Sep 19

any guesses?

Colin Madland @colinmadland · Sep 19

Bad example: Racial bias in AI

- Artificial Intelligence systems (like facial detection) are trained on available data, which can be biased.
    - That data is labeled by often crowd workers



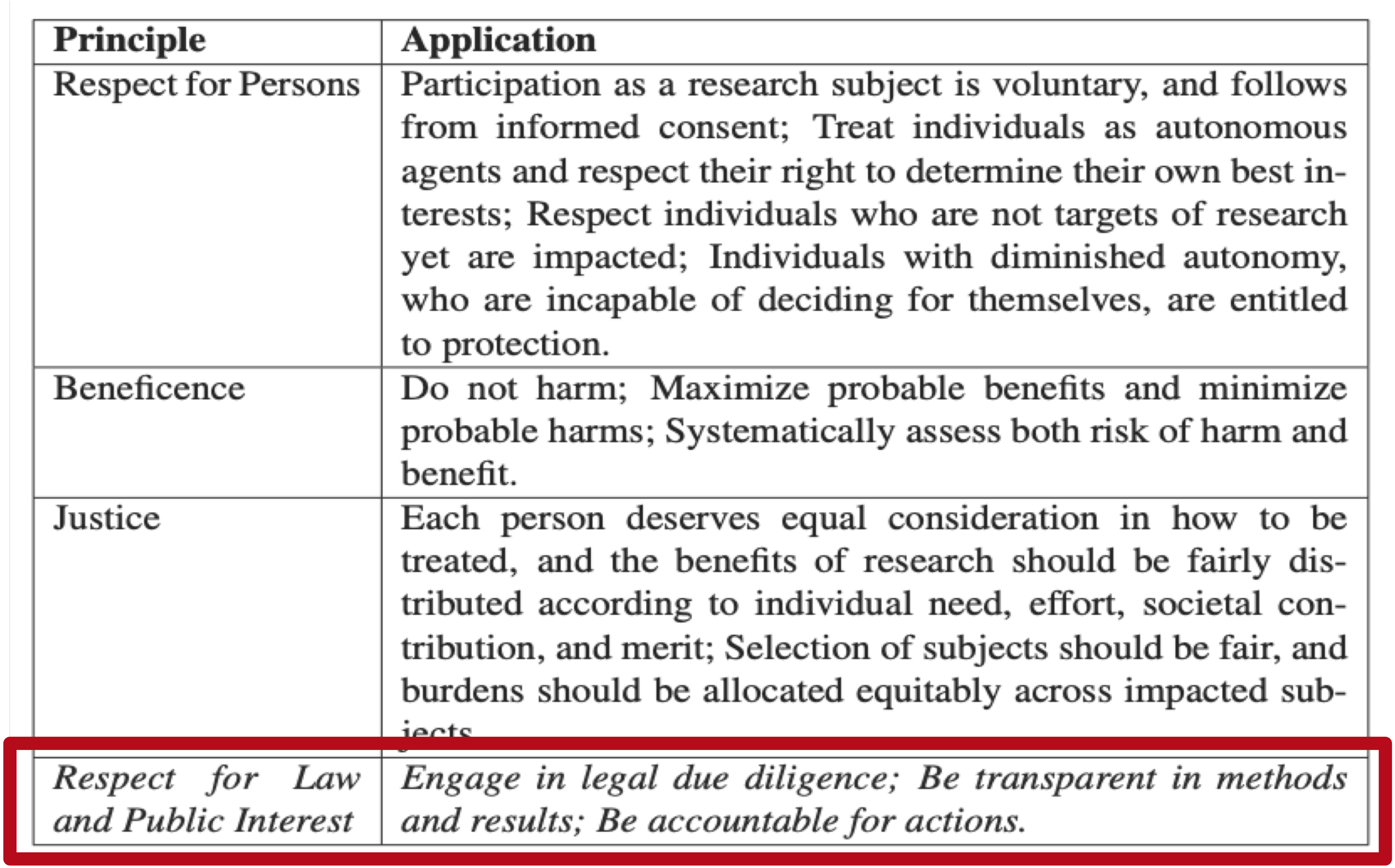
Colin Madland @colinmadland · Sep 19

Geez...any guesses why @Twitter defaulted to show only the right side of the picture on mobile?

Bad example 2

- Tuskegee Syphilis Experiment

The Menlo report



Principle	**Application**
Respect for Persons	Participation as a research subject is voluntary, and follows from informed consent; Treat individuals as autonomous agents and respect their right to determine their own best interests; Respect individuals who are not targets of research yet are impacted; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection.
Beneficence	Do not harm; Maximize probable benefits and minimize probable harms; Systematically assess both risk of harm and benefit.
Justice	Each person deserves equal consideration in how to be treated, and the benefits of research should be fairly distributed according to individual need, effort, societal contribution, and merit; Selection of subjects should be fair, and burdens should be allocated equitably across impacted subjects.
Respect for Law and Public Interest	*Engage in legal due diligence; Be transparent in methods and results; Be accountable for actions.*

Respect for law and public interest

- Compliance
  - Make sure you know what the laws are and don’t break them
  - When breaking laws is necessary go to university/organization and seek counsel

- Transparency and accountability
  - Make the objective and procedure of research clear
  - Include how data will be handled
  - Clearly mention risks to participants
  - Document the procedure, results of your study and make it public

Example: Password breaches

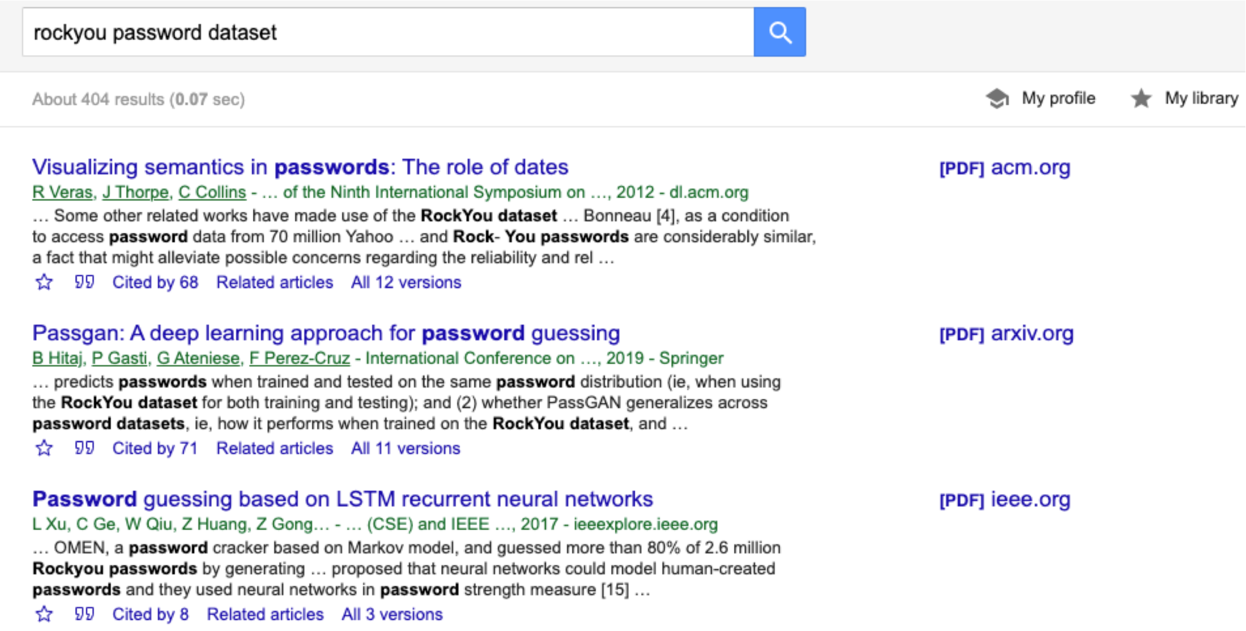
- People break into systems
  - Then make the passwords public

- Which principles might be violated?

Example: Password breaches

- People break into systems
  - Then make the passwords public

- Which principles might be violated?



rockyou password dataset

About 404 results (**0.07** sec)

Visualizing semantics in **passwords**: The role of dates
R Veras, J Thorpe, C Collins - … of the Ninth International Symposium on …, 2012 - dl.acm.org
… Some other related works have made use of the **RockYou dataset** … Bonneau [4], as a condition to access **password** data from 70 million Yahoo … and **Rock- You passwords** are considerably similar, a fact that might alleviate possible concerns regarding the reliability and rel …
[PDF] acm.org

Passgan: A deep learning approach for **password** guessing
B Hitaj, P Gasti, G Ateniese, F Perez-Cruz - International Conference on …, 2019 - Springer
… predicts **passwords** when trained and tested on the same **password** distribution (ie, when using the **RockYou dataset** for both training and testing); and (2) whether PassGAN generalizes across **password datasets**, ie, how it performs when trained on the **RockYou dataset**, and …
[PDF] arxiv.org

Password guessing based on LSTM recurrent neural networks
L Xu, C Ge, W Qiu, Z Huang, Z Gong… - … (CSE) and IEEE …, 2017 - ieeexplore.ieee.org
… OMEN, a **password** cracker based on Markov model, and guessed more than 80% of 2.6 million **Rockyou passwords** by generating … proposed that neural networks could model human-created **passwords** and they used neural networks in **password** strength measure [15] …
[PDF] ieee.org

So what should you do concretely?

- Step 1: Design the recruitment text and consent form
  - Consent form template: https://sbsirb.uchicago.edu/templates/
  - Recruitment: https://www.irb.northwestern.edu/recruitment-materials-and-guidelines/
- Step 2: Fill up an IRB form (include all materials)
  - Keep in mind, you want consented data, you don't want to do harm, you want to abide by law
  - For most of you this if the form to fill https://irb.northwestern.edu/docs/social-behavioral-protocol---protocol---583.docx
- Step 2: The IRB (IEC in our institute) comes back with questions
  - You answer them and/or change your study design

Roadmap

- What is ethics in our case?
  - How to obtain consent

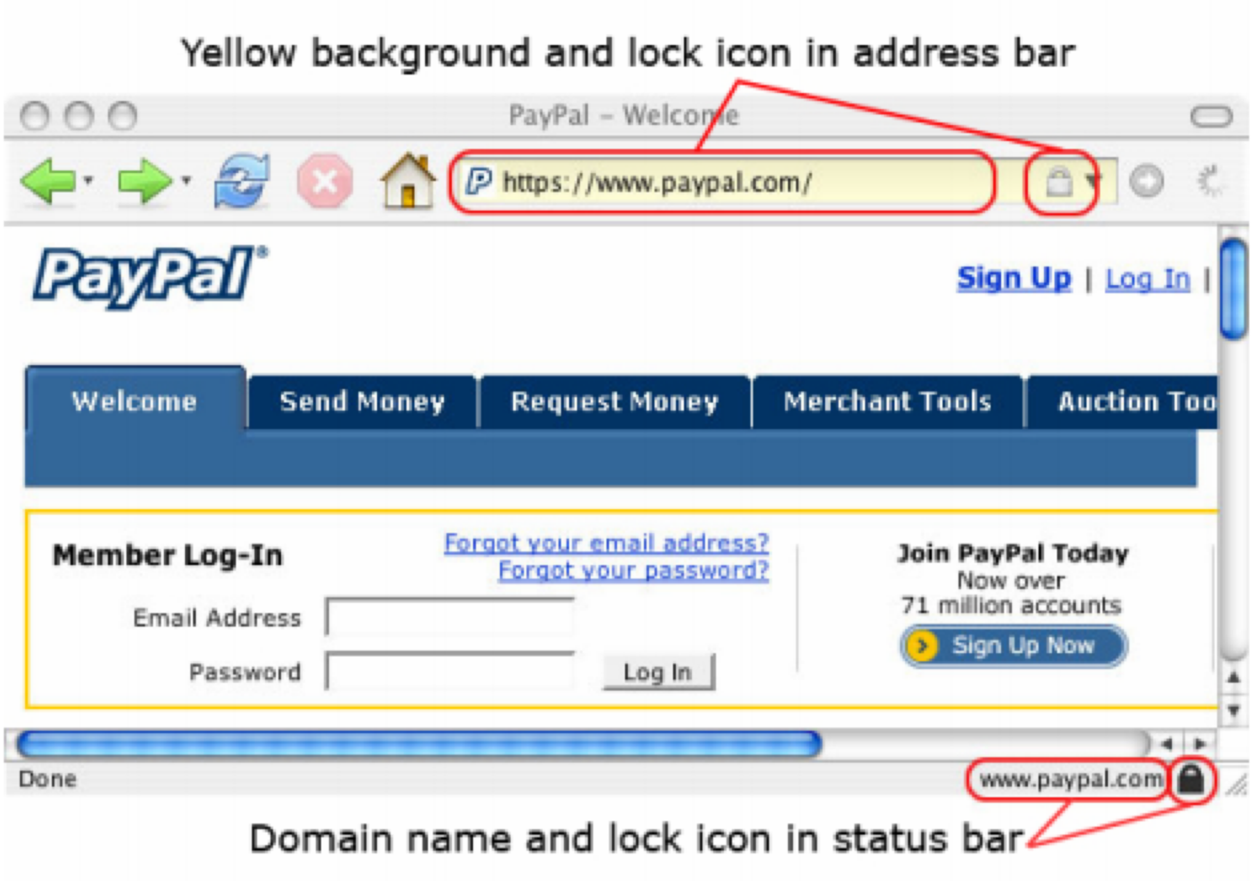
- Case study: ethical phishing experiments
  (slides from Markus Jakobsson)

What is phishing?

- Phish: Fraudulent email that looks real
  - Usually try to extract credentials (e.g., password), financial information (e.g., bank account), or other private information
- Spear Phish: Targeted phishing email

Why does phishing work?

- “Why phising works”, Dhamija et. al., CHI 2006
  - Prime users to look for security indicators



Yellow background and lock icon in address bar

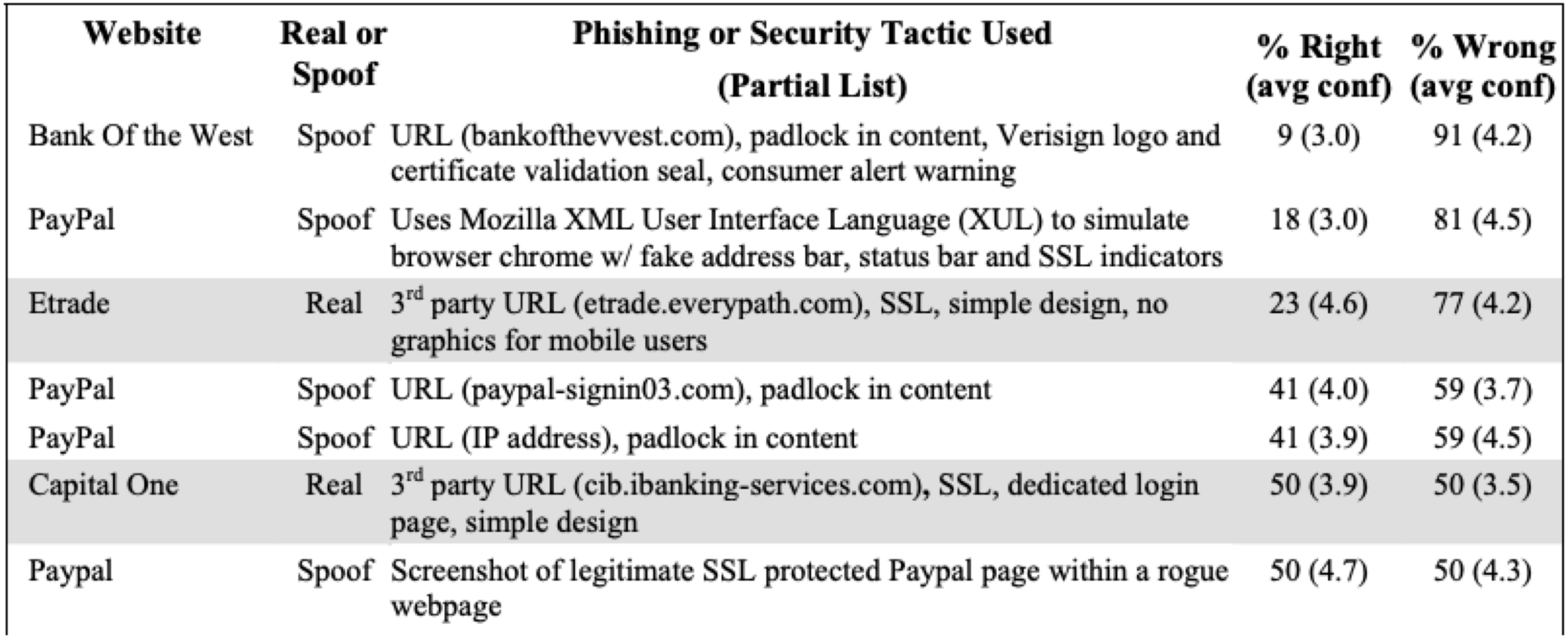
Domain name and lock icon in status bar

Why does phishing work?

- “Imagine that you receive an email message that asks you to click on one of the following links. Imagine that you decide to click on the link to see if it is a legitimate website or a "spoof" (a fraudulent copy of that website).”

- They informed participants any website may be legitimate or not, independent of what they previously saw.

Why does phising work?



Website	Real or Spoof	Phishing or Security Tactic Used (Partial List)	% Right (avg conf)	% Wrong (avg conf)
Bank Of the West	Spoof	URL (bankofthevvest.com), padlock in content, Verisign logo and certificate validation seal, consumer alert warning	9 (3.0)	91 (4.2)
PayPal	Spoof	Uses Mozilla XML User Interface Language (XUL) to simulate browser chrome w/ fake address bar, status bar and SSL indicators	18 (3.0)	81 (4.5)
Etrade	Real	3rd party URL (etrade.everypath.com), SSL, simple design, no graphics for mobile users	23 (4.6)	77 (4.2)
PayPal	Spoof	URL (paypal-signin03.com), padlock in content	41 (4.0)	59 (3.7)
PayPal	Spoof	URL (IP address), padlock in content	41 (3.9)	59 (4.5)
Capital One	Real	3rd party URL (cib.ibanking-services.com), SSL, dedicated login page, simple design	50 (3.9)	50 (3.5)
Paypal	Spoof	Screenshot of legitimate SSL protected Paypal page within a rogue webpage	50 (4.7)	50 (4.3)

Why does phishing work?

- Good phishing websites fooled 90% of participants
- Existing anti-phishing browsing cues (address bar, status bar, or security indicators) are ineffective for 23% of participants
- On average, participants made mistakes 40% of the time
- Popup warnings about fraudulent certificates were ineffective
- None of education, age, sex, previous experience, hours of computer use had a statistically significant correlation with vulnerability to phishing

- Required reading: social phising, Jagatic et. al.