Ethics in Security and Privacy 1

INFR11158/11230 Usable Security and Privacy

Dr. Jingjie Li

12/03/2026

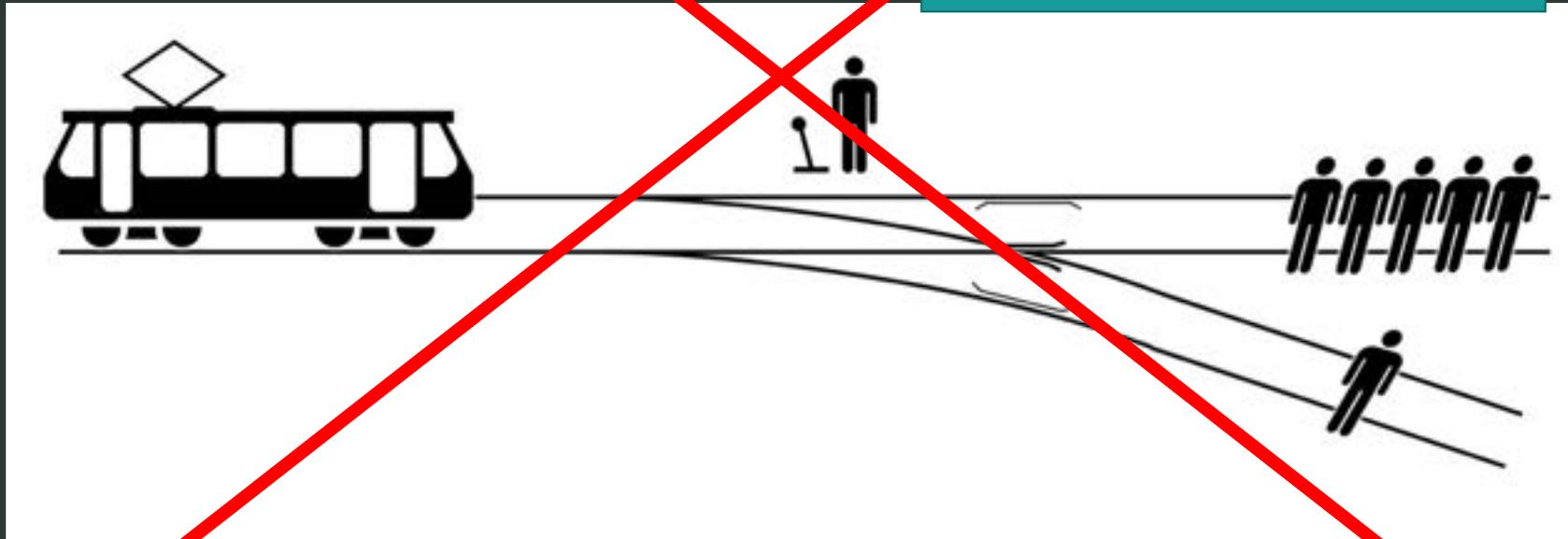THE UNIVERSITY of EDINBURGH

The trolley problem

Today we will *not* be discussing classical ethics.

Instead we are discussing ethics of study designs and how to behave in a way that society considers to be ethical.

A short history lesson on ethics

Tuskegee Syphilis Experiment

- Between 1932 and 1972 the US Public Health Service conducted a study on untreated syphilis

- 600 African American men in Alabama were given free medical care and food for participating

- They were told it was a 6 month study on "bad blood", but it actually lasted 40 years

- They were never told that they had syphilis, even though the researchers knew that they did

- A cure was found in the 1940's, but Tuskegee patients were prevented from gaining access to it

- 28 participants died of syphilis, 100 died of related complications, 40 wives contracted syphilis, and 19 children were born with congenital syphilis

History of ethics

- 1972 — Tuskegee study ended
- 1974 — US Congress created a commission to study research ethics and write regulations around ethics
- 1978 — Belmont Report officially released detailing a set of guidelines around what "ethical" research on humans looks like
- 1981 — Common Rule went into effect in the US regarding biomedical and behavioral research involving human subjects.
- 2010 — All researchers working on US funded grants must go through ethics training
- 2012 — Menlo Report officially published. Update to the Belmont Report focusing on ICT research

The Belmont Report (1974)

- 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–and–policy/belmont–report/index.html

The Menlo Report

Ethical Principles Guiding Information and
Communication Technology Research

August 2012

Homeland
Security

Science and Technology

The Menlo Report (2012)

- Res
- Ben
- Jus
- Res

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 individuals as autonomous agents
- Give them the right to choose and the knowledge they need to make a good decision
- Persons with diminished autonomy are entitled to protection
- Applications
  - Participation should be voluntary
  - Participants should be fully informed of the costs and benefits of participation (consent)
  - Impacted non–participants should also be respected

Good Example

The survey for the security course.

Vulnerable population, so collected on paper to ensure right to choose.

Answer before you start

Please answer the questions below as best you can before starting the tutorial. If you don't know the answer then please select your best guess, or write "I don't know".

1. **We would like to use your answers in research publications and to improve this tutorial.**
   - ○ You may use my answers below in research publications
   - ○ Do not use my answers below in research publications

2. **Which of the following statements describe a POST request? Tick all that apply.**
   - ☐ retrieves information from the web server
   - ☐ sends information to the web server, most likely to be stored
   - ☐ data is enclosed in the body of the HTTP request
   - ☐ data is visible in the URL

3. **What is the status code of a successful HTTP request?**
   _____

4. **Which of the following statements best complete this description of a dictionary attack? An attacker performs a dictionary attack by systematically submitting:**
   - ○ all possible password combinations
   - ○ all the words in the English dictionary
   - ○ all the passwords in a pre-established list of passwords
   - ○ what is a dictionary attack?

5. **How would a developer secure their website against a brute force attack? Tick all that apply.**
   - ☐ caching the hashes of the users' passwords
   - ☐ account lock out if too many incorrect attempts
   - ☐ sanitising user input
   - ☐ rate limiting

Poor Example

Collected data that users had intentionally not published on purpose.

Most people would not expect Facebook to collect data before they post it.
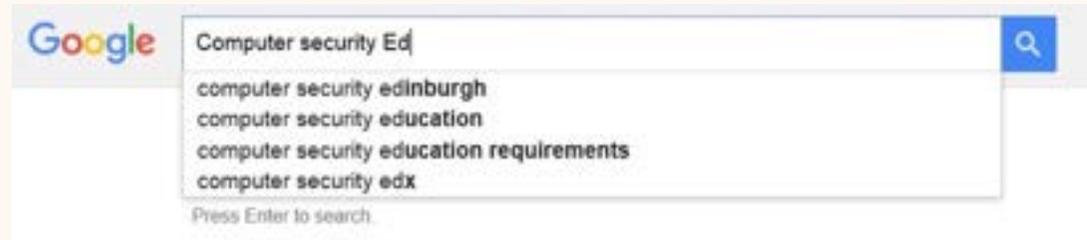
Self-Censorship on Facebook

Sauvik Das[1] and Adam Kramer[2]

[1]sauvik@cmu.edu   [2]akramer@fb.com
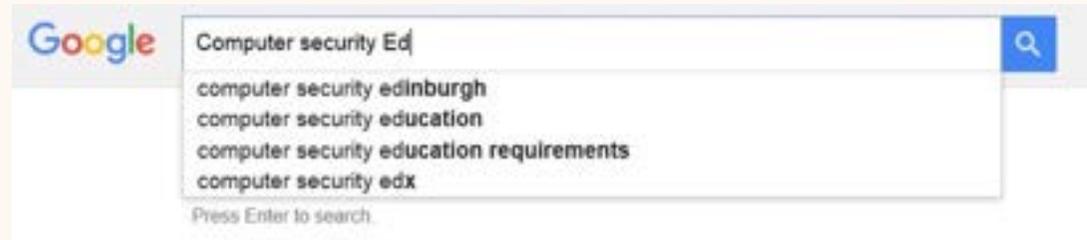Carnegie Mellon University   Facebook, Inc.

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 who exercise more control over their audience censor more content; and, users with more politically and age diverse friends censor less, in general.

Does Google know what you have typed before you click enter?

Google | Computer security Ed

computer security edinburgh
computer security education
computer security education requirements
computer security edx
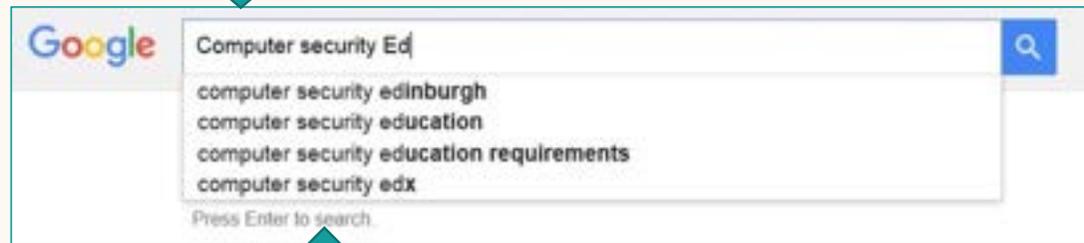
Press Enter to search.

A. Yes

B. No

C. Unsure

Does Google know what you have typed before you click enter?

A. Yes

B. No

C. Unsure

Most search engines send text to their servers as you write it

Google sends the letters you type to their server for processing.

Google | Computer security Ed

computer security edinburgh
computer security education
computer security education requirements
computer security edx

Press Enter to search.

The server then figures out what you are likely looking for and sends it back to make this list.

Poor Example

Collected data that users had intentionally not published on purpose.

Most people would not expect Facebook to collect data before they post it.

Self-Censorship on Facebook

Sauvik Das[1] and Adam Kramer[2]

[1]sauvik@cmu.edu   [2]akramer@fb.com
Carnegie Mellon University   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 who exercise more control over their audience censor more content; and, users with more politically and age diverse friends censor less, in general.

Beneficence

- Do not harm

- Maximize the possible benefits and minimize the possible harms

- Applications
  - Systematic analysis of the risks and benefits of the research to both the individual and to society at large

Good Example

Deception study where participants were asked to log into their actual bank accounts on a computer which had been "hacked" by the researchers but the security indicators were still accurate

Research question: will users enter their password if all the security indicators are missing?

- Notified participants that their actions would be recorded

- System did not record passcodes or other private data

- Care was taken with the technical design to make sure the participant's bank credentials remained safe

- Participant was debriefed after the study

- Participant was told how to protect themselves in the future

http://www.usablesecurity.org//emperor/emperor.pdf

Poor Example

Researchers knew before the study that being in the study might negatively impact the survival of a baby.

Research question: how much oxygen do premature babies need to prevent death or blindness?

- Randomized assignment to high or low oxygen conditions

- Current best practice is to assign oxygen based on doctors opinion

- Existing research says that high oxygen levels can lead to blindness

- Primary outcome variable was if the babies developed sever eye disease or die

http://ahrp.org/an-experiment-designed-to-kill-

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

- Application
  - Selection of research participants

Good Example

Truly random sample of all students in the US that received a PhD degree.
If you don't "voluntarily" fill out this survey they will keep emailing you and sometimes send someone to your door to have you take it in person.
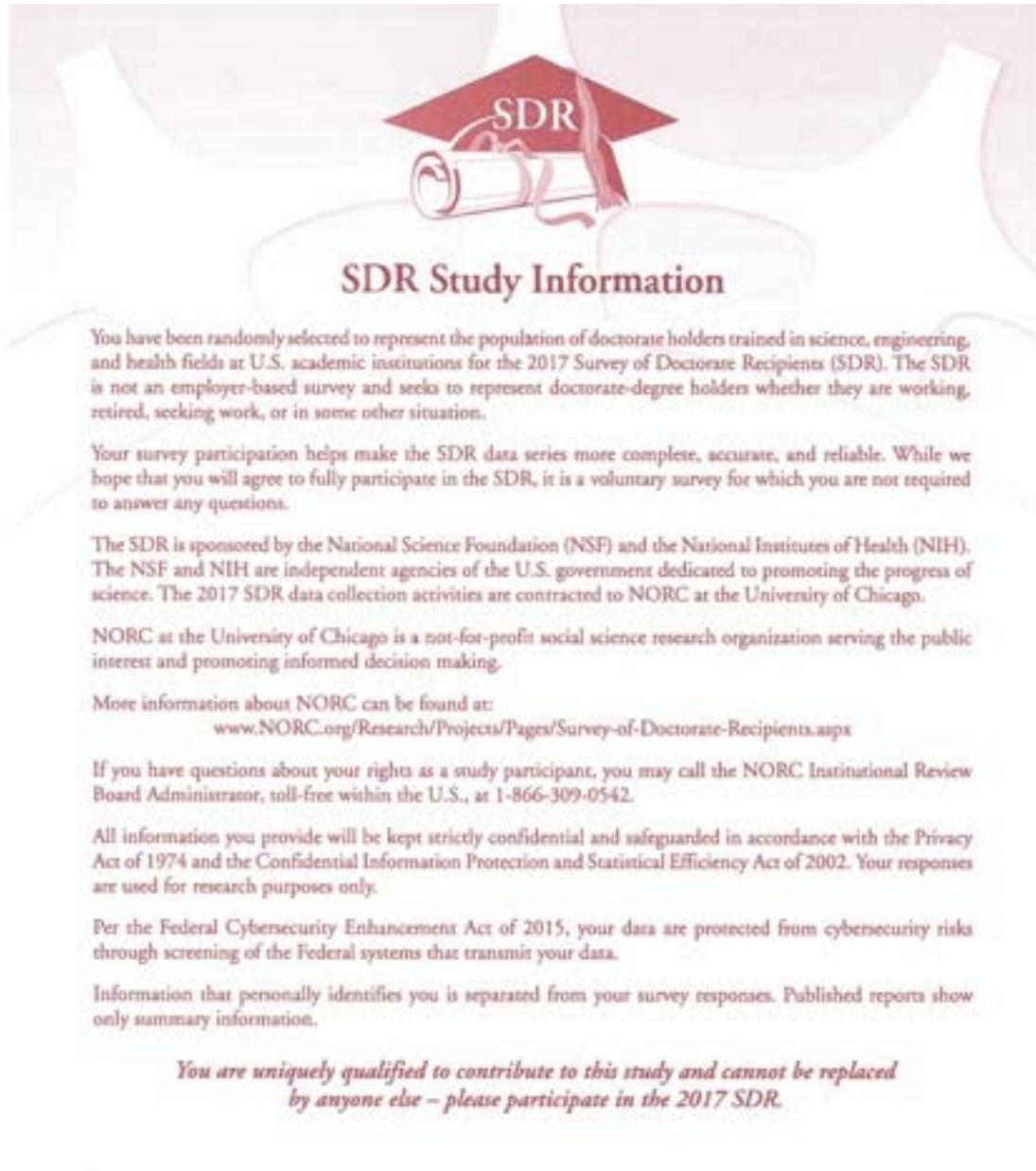
SDR Study Information

You have been randomly selected to represent the population of doctorate holders trained in science, engineering, and health fields at U.S. academic institutions for the 2017 Survey of Doctorate Recipients (SDR). The SDR is not an employer-based survey and seeks to represent doctorate-degree holders whether they are working, retired, seeking work, or in some other situation.

Your survey participation helps make the SDR data series more complete, accurate, and reliable. While we hope that you will agree to fully participate in the SDR, it is a voluntary survey for which you are not required to answer any questions.

The SDR is sponsored by the National Science Foundation (NSF) and the National Institutes of Health (NIH). The NSF and NIH are independent agencies of the U.S. government dedicated to promoting the progress of science. The 2017 SDR data collection activities are contracted to NORC at the University of Chicago.

NORC at the University of Chicago is a not-for-profit social science research organization serving the public interest and promoting informed decision making.

More information about NORC can be found at:
www.NORC.org/Research/Projects/Pages/Survey-of-Doctorate-Recipients.aspx

If you have questions about your rights as a study participant, you may call the NORC Institutional Review Board Administrator, toll-free within the U.S., at 1-866-309-0542.

All information you provide will be kept strictly confidential and safeguarded in accordance with the Privacy Act of 1974 and the Confidential Information Protection and Statistical Efficiency Act of 2002. Your responses are used for research purposes only.

Per the Federal Cybersecurity Enhancement Act of 2015, your data are protected from cybersecurity risks through screening of the Federal systems that transmit your data.

Information that personally identifies you is separated from your survey responses. Published reports show only summary information.

You are uniquely qualified to contribute to this study and cannot be replaced by anyone else – please participate in the 2017 SDR.

Poor Example

Ignoring of social norms to do "good" work of catching criminals.

Government Set Up A Fake Facebook Page In This Woman's Name

A DEA agent commandeered a woman's identity, created a phony Facebook account in her name, and posted racy photos he found on her seized cell phone. The government said he had the right to do that. Update: Facebook has removed the page and the Justice Department says it is reviewing the incident.

Chris Hamby
BuzzFeed News Reporter

Posted on October 6, 2014, at 7:16 p.m. ET

🐦 Tweet   f Share   🔗 Copy

The government's response lays out an argument justifying Sinnigen's actions: "Defendants admit that Plaintiff did not give express permission for the use of photographs contained on her phone on an undercover Facebook page, but state the Plaintiff implicitly consented by granting access to the information stored in her cell phone and by consenting to the use of that information to aid in an ongoing criminal investigations [sic]."

posing on the hood of a BMW, legs spread, or, in another, wearing only skimpy attire. She was surprised; she hadn't even set up a Facebook page.

Poor Example

Artificial Intelligence systems are trained on available data, which can be biased.

Microsoft Kinect Can't Identify African-Americans?

By Usman
Nov 5, 2010

6 SHARES

f SHARE   🐦 TWEET   📤 SUBMIT

Houston... we have racism. Microsoft's recently launched Kinect device, while undergoing testing at GameSpot, had trouble identifying two dark skinned employees. Apparently, the employees in question had trouble getting the facial recognition features to work.

According to the website, the system recognized one person's face "inconsistently", and when it came to the second staff member, the device was "never able to properly identify the other despite repeated calibration attempts."

What's confounding is the fact that at the same time, the Kinect had no problems identifying a third dark-skinned employee, right after a single calibration. Lighter-skinned employees not of African-American lineage were all easily identified on the first try.

Fortunately, the problem seems to only be with facial recognition, and not with skeletal tracking as that worked fine for all three dark-skinned employees. Since skeltal tracking is the primary manner to play games with Kinect, it's somewhat reassuring that at least this feature works.

22

Respect for Law and Public Interest

- Compliance
  - Make sure you know what the laws are and don't break them
  - When breaking laws must be done, engage in due process

- Transparency and Accountability
  - Make the contents of research clear, including how data will be handled and used
  - Clearly communicate risks
  - Document the contents of your study and make that documentation public

Good Example

Asking relevant governing bodies for support in advance of conducting potentially prc wo

Crypto guru Matt Green asks courts for DMCA force field so he can safely write a textbook

Next move in EFF's plans to regain the right to tinker

"Researchers should be encouraged to educate the public and the next generation of computer scientists. Instead, they are threatened by an unconstitutional law that has come unmoored from its original purpose of addressing copyright infringement. We're going to court to protect everyone whose speech is squelched by this law, starting with Dr Green and his book."

as ATM machines, smart cars, and medical devices. But this could lead

Confusing Example

Password data breaches are not legal. But when they are made public should we use them for the good of the public?

RockYou settles FTC charges related to 2009 breach

Online gaming firm will pay $250,000, submit to independent audits for 20 years after exposing data on 30 million users

By Jaikumar Vijayan
Computerworld | MAR 27, 2012 5:00 PM PT

RockYou will submit to third-party security audits for the next 20 years as part of a settlement of charges filed by the U.S. Federal Trade Commission in connection with a Dec. 2009 data breach that exposed email addresses and passwords of more than 30 million people.

As part of the settlement announced Tuesday, the online social gaming company will also pay a $250,000 civil penalty to settle charges that it violated the Children's Online Privacy Protection Act (COPPA) by knowingly collecting email information from about 180,000 underage children without first getting parental consent.

The proposed settlement also requires RockYou to maintain a formal data security program and prohibits it from making 'deceptive claims' about its privacy and security practices.

[Further reading: What is blockchain? The most disruptive tech in decades]

In a statement, RockYou CEO Lisa Marino called the settlement a "fair" one.

"We appreciate the work the FTC has done in this process as they have been fair, reasonable and timely throughout," she said. "The events that led to this

— CURRENT JOB LISTINGS —

https://www.science.org/content/article/unethical-ai-research-reddit-under-fire

Consent

Consent in General Data Protection Regulation

The basic requirements for the effectiveness of a valid legal consent are defined in Article 7 and specified further in recital 32 of the GDPR. **Consent** must be freely given, specific, **informed** and unambiguous. In order to obtain freely given consent, it must be given on a voluntary basis. The element "free" implies a real choice by the data subject....

Right to be informed

There is a need for transparency regarding the gathering and use of data in order to allow EU citizens to exercise their right to the protection of personal data. Therefore, the General Data Protection Regulation (GDPR) gives individuals a **right to be informed** about the collection and use of their personal data, which leads to a variety of information obligations by the controller.

Informed consent in real life and in research

An advertisement should:

- Be short and easy to read or decide to ignore

- Explain the main content of what participants will be asked to do

- Explain what the costs, benefits, and risks of participating are

- State who to contact about the research in case of concern

- State if the research has been through ethical review

A consent form should:

- Who you are

- What the study involves, what they will be asked t do

- What kind of data will be collected and how it will be used

- What rights the participant has

- Compensation, if any

We are students in the Human–Computer Interaction course. For our first coursework we are studying how students at the University of Edinburgh use calendaring systems such as paper calendars, Google Calendar, and Office 365 Calendar.

In this survey we are investigating how people use their online calendars so that we can better understand their calendar–related needs and choices. We will ask you for some information about yourself, about the way in which you use computers and the internet, about the tools you use to manage your timetable and other events.

Completing the survey will take about 10 minutes. You can interrupt the survey at any time and return to finish it later. All the data that you provide will be stored on SurveyMonkey and user–level access will be restricted to our group. Questions marked with a red star are mandatory – you will need to answer them in order to complete the survey. Data you provide will be deleted two months after the last day of this school term.

This project has undergone ethical screening in accordance with the University of Edinburgh School of Informatics ethics process (RT1432).

Do you agree to take part in this study, and do you agree that I can use your data for my HCI student project?

Ethics: if we are free from ethical issues when no human subject is involved?

Ethics in Social media research: Social media has been a great resource for people to do a wide range research; But people are becoming more and more careful nowadays. What are the ethical considerations in using social media **data** for research?

Reddit Is a Public Platform

Much of the information on the Services is public and accessible to everyone, even without an account.

When you submit content (for example, a post, comment, or chat message) to a public part of the Services, any visitors to and users of our Services will be able to see that content, the username associated with that content, and the date and time you originally submitted that content. That content and information may also be available in search results on internet search engines like Google or in responses provided by an AI chatbot like OpenAI's ChatGPT. You should take the public nature of the Services into consideration before posting. By using the Services, you are directing us to share this information publicly and freely.

https://www.reddit.com/policies/privacy-policy

Are we, researchers, good enough under Reddit's privacy policy?

Some ethical practices

- Follow the terms of use

- Obtain informed consent when possible

- Check our ethics guidelines for more!
https://resource.ppls.ed.ac.uk/lelethics/index.php/frequently
–asked–questions/research–with–social–media–data/

Ethics: from human to systems

BROWN UNIVERSITY P2P

Andy Pavlo

https://hardware.slashdot.org/story/09/04/13/0120226/grad-student-project-uses-wikis-to-stash-data-miffs-admins

"Two graduate students at the Ivy League's Brown University built a P2P system to use abandoned wiki sites to store data. The students were stealing bandwidth from open MediaWiki sites to send data between users as an alternative to BitTorrent. There was immediate backlash as site operators quickly complained to the University. The project appears to be shutdown, but many of the pages still remain on the web. The project homepage was also taken down and the students posted an apology this afternoon."

https://hardware.slashdot.org/story/09/04/13/0120226/grad–student–project–uses–wikis–to–stash–data–miffs–admins

(More) Case studies in Ethics in Security

EXPERIMENTAL EVIDENCE OF MASSIVE-SCALE EMOTIONAL CONTAGION THROUGH SOCIAL NETWORKS

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

Aka Facebook emotion contagion study

"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."

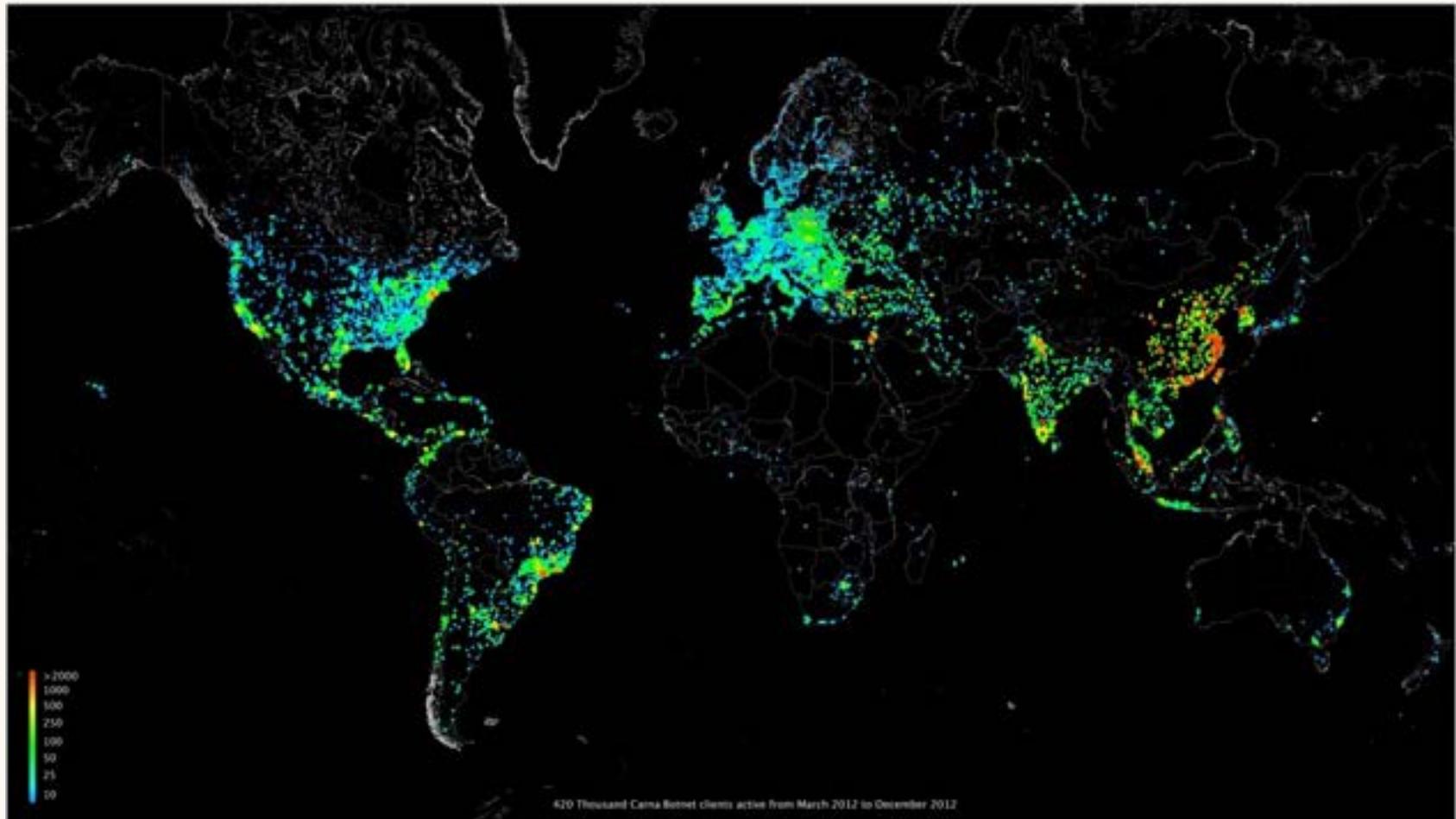http://www.pnas.org/content/111/24/8788.full

The study

- All Facebook users who spoke English qualified

- Two groups: positive and negative emotions

- Positive/negative posts where then suppressed from the news feed

- 689,003 participants randomly selected by user id

- Saw an impact
  - When positive posts withheld the participant's posts got more negative
  - When negative posts withheld the participants posts got more positive
  - Withdrawal effect: people who saw less emotion posts less likely to express themselves for several days

Think–pair–share

- Does the Facebook Emotion Contagion study fit the requirements of the Belmont / Menlo Report?

Mapping the Internet: Someone made the most detailed map of the internet ever by hacking into just under half a million computers

http://motherboard.vice.com/blog/this-is-most-detailed-picture-internet-ever

420 Thousand Carna Botnet clients active from March 2012 to December 2012

Is it ethical to use this data to do good things?

THE EMPEROR'S NEW SECURITY INDICATORS:
AN EVALUATION OF WEBSITE AUTHENTICATION AND THE EFFECT OF ROLE PLAYING ON USABILITY STUDIES

Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer

http://www.usablesecurity.org//emperor/emperor.pdf

Will bank customers enter their passwords even if their browsers' [security UI element] is missing?

Study design

- 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 (deception study)
- Participants came to the lab and used a lab computer
- Computer was pre-setup to attack the connection between the bank and the user

To handle ethics the researchers:

- Notified participants that their actions would be recorded

- System did not record passcodes or other private data

- Care was taken with the technical design to make sure the participant's bank credentials remained safe

- Participant was debriefed after the study

- Participant was told how to protect themselves in the future

Ethics analysis and design in security research

USENIX Security Ethics Framework

- **Disclosures.** Vulnerabilities, if known to adversaries, can expose people to negative outcomes, such as harms or rights violations. Publicly disclosing vulnerabilities before they have been privately disclosed to the responsible parties, and hence before they have been mitigated, can therefore expose people to negative outcomes.

- **Experiments with live systems without informed consent.** Researchers testing live services (e.g., for vulnerabilities) such as web services or APIs that give access to otherwise non-public algorithms or models must also consider ethics. Such experiments should only be performed after carefully analyzing the potential negative outcomes to the service provider, which may include cost (of CPU cycles or of human effort) or corrupting system state, and to end users who are using the same service provider for non-research purposes.

- **Terms of service.** If experiments violate terms of service, the justification for violating them should be discussed in the paper.

- **Deception.** In most cases, participants should be fully informed of the purposes and risks (among other things) of participating in experiments. If deception is to be used, the necessity of doing so should be carefully considered; participants should be debriefed afterward to explain the necessity of the deception, even when the deception was mild.

- **Wellbeing for team members.** In some cases, research activities have the potential to negatively impact team members. For example, research on hate speech could expose team members to disturbing content and negatively impact their psychological wellbeing.

- **Innovations with both positive and negative potential outcomes.** Technologies that can positively impact one stakeholder group may negatively impact those same or other stakeholder groups. For example, advancements in anonymity systems could positively impact people that need anonymity under repressive regimes or excessive surveillance.

- **Retroactively identifying negative outcomes.** While research teams should strive to proactively identify and address all ethics-related concerns *before* commencing their research and proactively address any new concerns that arise about the project's next steps *during* the research, in some cases research teams may discover *post facto* that their *past* research activities had unexpected and previously unknown (to the researchers) negative outcomes.

https://www.usenix.org/conference/usenixsecurity26/
call-for-papers#ethics

Take–home

- Kohno, T., Acar, Y. and Loh, W., 2023. Ethical frameworks and computer security trolley problems: Foundations for conversations. In *32nd USENIX Security Symposium (USENIX Security 23)* (pp. 5145-5162).