Intro to HCI Methods and Experimental Design

Michelle Mazurek

Some slides adapted from Lorrie Cranor, Blase Ur, Marshini Chetty, Elaine Shi, Christine Trask, and Yu-Xiang Wang, Howard Seltman
Today

- Reminder: reading reports should include both (all) readings, not just one
- Project logistics
- Presentation assignments, HW2 soon
- Finish up privacy, then start experimental design
Option 2b: Computer reads for you

- Platform for Privacy Preferences (P3P)
- W3C specification for XML privacy policies
  - Proposed 1996
  - Adopted 2002
- Optional P3P compact policy HTTP headers to accompany cookies
- Goal: Your agent enforces your preferences
Criticisms of P3P

• Too complicated, hard to understand
• Lacks incentives for adoption
  – Only major companies?
PrivacyFinder: P3P search engine

• Checks each search result for computer-readable P3P privacy policy, evaluates against user’s preferences

• Composes search result page with privacy meter annotations and links to “Privacy Report”

• Allows people to comparison shop for privacy

• [http://privacyfinder.org/](http://privacyfinder.org/)
Shoebuy.com, Inc. Privacy Practices

Privacy Policy Summary | Full Privacy Policy | Opt-Out | Contact Site | P3P Policy

Privacy Policy Check

Shoebuy.com, Inc.’s privacy policy does not match your preferences:

- Site may contact you to interest you in other services or products and does not allow you to remove yourself from marketing/mailing list

Shoebuy.com, Inc. may share your information with:

- Companies that help this site fulfill your requests (for example, shipping a product to you), but these companies must not use your information for any other purpose
- Delivery companies that help this site fulfill your requests and who may also use your information in other ways

Privacy Policy Summary

Policy Statement 1 - Access log information

Our Web server collects access logs containing this information.

Show data collection, use, and sharing details...

Policy Statement 2 - Cookies

Cookies are used to track visitors to our site, so we can better understand what portions of our site best serve you. We also use cookies to allow our server to maintain information about the contents of your shopping cart.

Show data collection, use, and sharing details...

Policy Statement 3 - Transaction info (required)

Information we collect in order to process your purchase.

Show data collection, use, and sharing details...

Policy Statement 4 - Transaction info (optional)
Impact on decisionmaking

• Online shopping study conducted at CMU lab
• Participants buy with their own credit cards
  – Bought batteries and a sex toy
• Pay them a fixed amount; keep the change
• Result: When information is accessible, many people will pay (a little) more for privacy

P3P in Internet Explorer

• Implemented in IE 6, 7, 8, 9, 10 …
• “Compact policy” (CP)
• If no CP, reject third-party cookies
• Reject unsatisfactory third-party cookies
No P3P syntax checking in IE

• Accepts bogus tokens, nonsense policies
• Valid:
  CAO DSP COR CURa ADMa DEVa OUR IND PHY ONL UNI COM NAV INT DEM PRE
• Also accepted:
  AMZN
  Facebook does not have a P3P policy. Learn why here: http://fb.me/p3p

Microsoft uses a “self-declaration” protocol (known as “P3P”) dating from 2002 .... It is well known – including by Microsoft – that it is impractical to comply with Microsoft’s request while providing modern web functionality.
Can policy agents ever work?

• Simplify the practices enough?
• Require users to specify their preferences?
  – Learn them via AI?
• Incentives for broad adoption?
Requirements for meaningful control

• Individuals must:
  – Understand **what** options they have
  – Understand **implications** of their options
  – Have the **means** to exercise options

• Costs must be reasonable
  – Money, time, convenience, benefits
Option 3: The power of math

• Can we provide strong guarantees that don’t rely on good behavior from the data collector?
• Sort of!
• Differential privacy, invented by Cynthia Dwork
Privacy and Justin Bieber

• Suppose you are handed a survey:
  – Do you like listening to Justin Bieber?
  – How many Justin Bieber albums do you own?
  – What is your gender?
  – What is your age?

• After analysis, results will be released publicly
  – Do you feel safe submitting a survey?
  – Should you?
Brief notation

$\text{Pop} \subseteq \text{Pop}$

$D_I = \{d_i \mid i \in I\}$

$Q(\text{D}_I) = \text{R}$

$Q$ is the privatized query run on the data set, and $\text{R}$ is the result released to the public.
What do we want? (Privacy)

• My answer has no impact on the released results

• Any attacker looking at published $R$ can’t learn anything new about me personally (high probability)

• $Q(D_{(l-me)}) = Q(D_1)$

• $Pr[secret(me) | R] = Pr[secret(me)]$
Why can’t we have it?

• If individual answers had no impact, the results would be useless

• Trends in $R$ may be true of me too. (If I am 15, do I like Justin Bieber?)

• By induction, $Q(D_{(l)}) = Q(D_{\emptyset})$

• $\Pr(\text{secret(me)} \mid \text{secret(Pop)}) > \Pr(\text{secret(me)})$
Why can’t we have it?

If an attacker knows a function about me dependent on the general population:

• I’m 2x average age
• I’m the majority gender

Then the attacker knows things about me even if I don’t submit a survey!

• \( \text{age(me)} = 2 \times \text{mean}_\text{age} \)
• \( \text{gender(me)} = \text{mode}_\text{gender} \)
• \( \text{mean}_\text{age} = 16 \)
• \( \text{mode}_\text{gender} = F \)

• \( \text{age(me)} = 32 \text{ AND } \text{gender(me)} = F \)
What can we have instead?

• The chance that the released result will be $R$ is nearly the same, **regardless** of whether I submit a survey

• There is no (well, *almost* no) additional harm from submitting the survey
Differential privacy

\[ \frac{\Pr[Q(D_l) = R]}{\Pr[Q(D_{l \pm i}) = R]} \leq A, \text{ for all } l,i,R \]

• If A=1, there is 0 utility (individuals have no effect)
• If A >> 1, there is little privacy
• A should be chosen by collector to be close to 1
What this means

• Probability of result is nearly the same, regardless of whether I submit a survey

\[ \Pr[R] = X \quad \text{and} \quad \Pr[R] = Y \]

\[ X \approx Y \]

• How can anyone guess which world is true?

world where I submit a survey

world where I don’t submit a survey
Popular misconception

• The attacker can’t learn anything about me from the results (protection against all harms)

• NOPE: Background information still applies. Attackers can use aggregate results.
How to do it (high-level)

• Output perturbation: Return query answer plus some noise

• Input perturbation: Add noise to survey data before storing

• Perturbation of intermediate results

• Sample and aggregate
  – Ask Q over smaller samples; aggregate results
Challenges

• Utility / privacy tradeoffs
  – May require really large datasets
  – May result in ridiculous answers

• Privacy budget depletion
  – Each query reduces what else can be asked

• How can this fit in with personal privacy (as opposed to data protection?)

• What does a differential policy mean in practice?
Requirements for meaningful control

• Individuals must:
  – Understand **what** options they have
  – Understand **implications** of their options
  – Have the **means** to exercise options

• Costs must be reasonable
  – Money, time, convenience, benefits

*How can/should these approaches (laws, notice/choice, technical solutions) be balanced?*
HCI AND EXPERIMENTAL DESIGN
Human-Computer Interaction (HCI)

• You are not the user! You know too much.
• Think about the user throughout design
• Involve the user
500 OOPS: child died
What is usable?

• Intuitive / obvious
• Efficient
• Learnable
• Memorable
• Few errors
• Not annoying
• Status transparent
Difficulties

• Many systems and platforms
• Users are different from one another
• Required standards (or no standards)
• Documentation won’t necessarily be read
• Performance
• Legal / time pressures
• Social and external factors
Thinking about research studies

• What are you hoping to learn?
• What are your hypotheses?
  – Sometimes listed explicitly in a paper
• What are your metrics for success?
  – More secure, quicker to use, more fun, etc.
• What are you comparing to?
• What data might be helpful?
Broad types of studies

- Field study
- Laboratory study
- Online study
- (Measurement study)
Quantitative vs. Qualitative

• Quantitative: you have numbers (timing data, ratings of awesomeness)

• Qualitative: you have non-numerical data (thoughts, opinions, types of errors)
Types of studies

• Find out what people want:
  – Contextual inquiry
  – Interviews
  – Focus groups
  – Surveys
  – Diary study (prompt people)

• Find out what/how people think:
  – Interviews
  – Surveys
Types of studies

• Controlled experiments to test causation:
  – e.g., A/B testing
    • Role-playing
    • Experiments in the field
    • Full-factorial design, or not
Types of studies

• Expert evaluation of usability:
  – Cognitive walkthrough
  – Heuristic evaluation

• Usability test:
  – Laboratory (“think aloud”)
  – Online study
  – Log analysis
Usability of Fruit

- https://www.youtube.com/watch?v=3Qg80qTfzgU
Data to collect during experiments

• Independent vs. dependent variables
• Performance (time, success rate, errors)
• Opinions and attitudes
• Audio recording, screen capture, video, mouse movements, keystrokes
• Formative (initial) vs. summative (validate)
Even more data to collect

• Demographics
  – Age, gender, technical background, income, education, occupation, location, disabilities, first language, privacy attitudes, etc.

• Open-ended questions

• Preferences and attitudes

Please respond to the following statements:
*This user interface was difficult to understand
1- Strongly disagree  2- Disagree  3- Neutral  4- Agree  5- Strongly agree

*This tool was fun to use
1- Strongly disagree  2- Disagree  3- Neutral  4- Agree  5- Strongly agree
Study designs

- Between subjects
  - Each participant tests 1 version of the system
  - You compare these groups
  - Groups should be similar (verify!)

- Within subjects
  - Every participant tests everything
  - Very important to randomize order!
  - Fewer participants
Logistics for a study

• How many participants?
  – Statistical power
  – Time, budget, participants’ time

• What kind of participants?
  – Skills, background, interests
  – Their motivations
  – Often not a “representative sample”

• What do you need to build, if anything?
  – Prototype fidelity
Prototype fidelity

High-fidelity, “Wizard of Oz,” low-fidelity
Think-aloud example (in pairs)

• Download and install software that lets you encrypt your email
  – Verify that it is installed

• Things you can ask:
  – What are you thinking now?
  – What do you expect to happen if you do X?
  – How did you decide to do that?
Paper prototype example (in groups)

• Draw a paper prototype of a tool to encrypt emails sent on Gmail
  – First step: Identify two tasks that you want to make sure are usable
Principles for research design

• “The goal of any research design is to arrive at clear answers to questions of interest while expending a minimum of resources.” – Ramsey and Shafer

• Goal: Identify sources of experimental variation and try to minimize/control them
1. Internal validity

• Avoid confounds!
  – Avoid criticism about causality

• How?
  – Randomize condition/treatment assignment
  – Change only one variable at a time per condition
  – Use blinding
  – Use a control group
2. Construct validity

• What do our metrics measure?
  – Is it what we intended?
  – Discuss: How would you measure tech-savviness?
3. External validity

• What population does your sample represent?
  – Race, gender, age, nationality, education, others

• What environment does your sample represent?
  – Carefully controlled study vs. real world

• This is especially critical for security research
  – Security as secondary task
  – What you think you should do vs. what you actually do
Promote power

• Covariates: Measure possible confounds, include in analysis

• Blocking: Group similar subjects, distribute across conditions

• Within-subjects promotes power
  – But has other drawbacks, esp. in security research

• Is what you are measuring strong enough?
  – Do you have enough participants?
  – Potential tradeoff: Generalizability for power
Determine use cases and goals

• What are the concrete tasks users should be able to accomplish?
  – Based on understanding of users!

• Set realistic metrics
Paper prototyping

• Don’t overthink, just make it
• Draw a frame on a piece of paper
  – Make all menus, etc.
• Redesign based on feedback
• “Think-aloud”
Paper prototype example

SCENARIO 1
"I want to listen to alternative music"
Paper prototype example
Usability prototyping for websites

Site Maps

1. home

2. login

3. forget pass

4. register new user

Storyboards

Schematics

Mock-ups