Privacy, Security and Usability

Observation

Today's class

- Ways to observe real world activity
- Security behavior observatory
- Regrets and privacy nudging studies
- Selected observations from homework 2

Observe Real World Activity

- Many data collection challenges
- Usually not conducive to a controlled experiment
- For example: naturally occurring risks:
  - Events of interest may be infrequent

Ways to observe real world activity

- Ethnographic studies
- Short-term field observation
- Cameras, sensors
  - Longer-term observation
- Contextual inquiry – interviewing in context
- Instrumented software
- Logs (web server, help desk, etc.)
- Diaries

Observations

- Not real-time
  - Surveys and interviews about participants' experiences after the fact
- Hypothetical, but in context
  - Experience sampling
  - Paratyping
    - Experience prototyping

Advantages of real world observation

- More ecological validity than lab or online studies
  - Participants perform task in the context of their normal activities
  - Participants behave and respond to risk naturally
- Less chance of bias from experimenter
- More conducive to long term data collection
- Challenges: getting permission, not causing behavior changes when people feel they are being watched, instrumentation

Experience sampling

- Participants fill out questionnaires in response to periodic alerts, responses are based on what is happening now
- Often used to understand mood, time use, and social interactions
- Need to find way to alert participants and have them respond to short survey (< 2 minutes)
  - Beepers, email, SMS, diaries, etc.

Paratyping

- Measuring real-life experiences instead of testing the technology
- Paratypes
  - a simulation, or model, of interaction ("-type") with a technology which is evaluated alongside ("para-") real-world experience
  - "proxies" act as substitutes for researcher
  - As they go about their daily life they survey the people they interact with

USABILITY TESTING, LABORATORY STUDIES, AND ONLINE STUDIES

Topics

- Laboratory studies
- Online studies
- Post-test assessment of usability
- Users studies with simulated adversary/risk

Lab Studies - Advantages

- More controlled
- You can simulate software and products that don't exist yet
- You can trigger events that might normally be infrequent or hard to observe in the wild
- You can observe normally risky activities in a safe environment

Lab Studies - Advantages

- You can more easily instrument devices and the environment for data collection
- Some of this applies to online studies too!

What can you do in a lab study?

- Interviews, focus groups, surveys
- Observe participant reactions to various designs, prompts, stimuli
- Observe participants performing tasks (perhaps thinking aloud and/or with eye tracker)

What can you do in a lab study?

- Observe participant interaction
  - With devices, software, messages from “computer”
  - With researcher
  - With other participants
  - With actor posting as someone in the lab for a particular reason (participant, maintenance worker, etc.)

Wizard of Oz

- a research experiment in which subjects interact with a computer system
- subjects believe system is autonomous
  - it is actually being operated or partially operated by an unseen human being

What can you do in an online study?

- Surveys
- Observe participants' keystrokes, mouse movements, typed responses, etc. in response to various designs, prompts, stimuli
- Ask participants to do an online task and collect data on speed and accuracy

Pros and cons of online studies

- Collect data from a lot of participants quickly
- But may be a lot of work to setup automation and
- instrumentation
- You can't see if someone is
  - multi-tasking during study, gaming the system, etc.
- Real time limitations:
  - You usually can't prompt them, answer their questions, or ask follow-up questions

Post-test assessment of usability

- Questionnaires with rating scale
- Self-reported quantitative assessments
- Best used in conjunction with:
  - observed data
    - speed, accuracy, etc.
  - qualitative assessments
- Helps understand reasons behind ratings or how to improve
- Can be administered after each task, or at the end of test

The System Usability Scale (SUS)

- Series of 10 likert questions
  - usually 5-point or 7-point –
  - strongly disagree to strongly agree
- Produces a score from 0 to 100
- System Usability Scale

The System Usability Scale (SUS) – Potential Questions

- I think that I would like to use this system frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.

The System Usability Scale (SUS) – Potential Questions (cont.)

- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before
- I could get going with this system.

Short usability scales

- Net Promoter Scale (1 question, 10-point scale)
  - How likely are you to recommend this website/product/service to a friend or relative?
- Single Ease Question (1 question, 7-point scale)
  - Overall, this task was? (very difficult to very easy)

USER STUDIES WITH SIMULATED ADVERSARY/RISK

Observe hypothetical security tasks

- Issue: users may be more alert to security issues than is natural
- Simulated risk:
  - May use deception + debriefing

Observe non-security tasks

- users still doing tasks they have been told to do as part of a study
  - May be different than they would otherwise
    - Sequence, choice of variables

CROWDSOURCED STUDIES

Topics

- Mechanical Turk
- Let's launch a live study!
- Why crowdsource?
- Limitations of crowdsourcing
- A few other considerations
- Let's check on our study results…

Mechanical Turk

- Crowdsourcing service run by Amazon.com
- Requesters setup account and deposit money
- Requesters post Human Intelligence Tasks (HITs)
- Workers ('Turkers') accept HITs, complete them, and request payment
- Requesters review work, approve payment
- Amazon takes money from requester's account and deposits it in worker accounts

Logistics of collecting data on Mturk

- Mturk provides basic survey capabilities but is very limited
- You can redirect participants to a survey website
  - E.g. Google Forms, SurveyGizmo, Qualtrics, SurveyMonkey, etc.
- You can redirect participants to your own custom study website
- When participants finish study on another website, they will return to Mturk to request payment

Completion codes

- Use your survey tool to generate a random completion code
- Use scripts to check that codes are not reused and match the code in the survey
- For quick manual check insert a fixed digit or two in your random code
- for example: [4-digit-random]41[2-digit-random]
  - 19464252
  - 82034246
  - 50344269
  - 22654261

Paying participants

- When the participant has finished, you can review their work and pay them from your account
- Pay workers promptly or they will complain (to you, to other workers, to IRB)
- Pay workers fairly
  - Figure out how long task should take and pay at least minimum wage, probably more ($10/hour often suggested as fair)
  - Remember you also have to pay fees to mturk

Other things to know about Mturk

- Screen workers with Mturk settings or your own screening survey
- You can send workers notifications through Mturk
- Prevent repeated worker for same task if you offer it in batches over time
  - Tell workers if they participate more than once they won't get paid
  - Write scripts to check worker IDs and reject workers who try again
- Turkers communicate with each other

Why crowdsource?

- Inexpensive
- Access to fairly diverse participant pool
- Fast participant recruitment
- Easy to pay participants
- Can contact participants for follow-up without collecting their PII

Alternative: online study with recruitment ads or participant pool

- Advantages
  - Depending on where/how you recruit, may have more control over type of participants
- Disadvantages
  - Usually takes longer to recruit participants
  - May be more complicated to pay participants (sometimes uses raffle to simplify payment)
- May require collection of PII to follow-up and/or send payments or prizes

Alternative: paper survey in public place

- Advantages
  - Depending on where/how you recruit, may have more control over type of participants
- Disadvantages
  - Time consuming and labor intensive
  - Participant pool limited to geographic location
  - Requires data entry (unless done on a tablet)

Alternative: paper survey by postal mail

- Advantages
  - Can randomly select participants
- Disadvantages
  - Tends to have low response rate
  - Expensive to mail surveys
  - Requires data entry

Alternative: lab study

- Advantages
  - Allows for direct observation of participants
  - Facilitates physical interaction with artifacts
  - Allows for direct interaction between participants and experimenter
- Disadvantages
  - Requires a lot more time and physical space
  - Participant pool limited to geographic region of lab

Limitations

- Can't directly observe participants
  - Pilot in person to debug study
  - Sometimes you can collect keystrokes, mouse movements, etc.
- Some users not motivated to pay attention, or have become professional survey takers
  - Use attention checks to filter out inattentive participants
  - Pay participants fairly
  - Recruit extra participants to get what you need

Limitations

- MTurk population younger, more tech-savvy, but more diverse than typical lab study

Attention checks

- Advice used to be to include attention check questions in every survey
- However studies find that attention checks can reduce quality
  - Introduces bias – Removed participants are not demographically distributed
  - Hawthorne effect – participants feel they are being watched
  - Participants may perform worse on questions after attention check if they recognize it as such

Alternatives to attention checks

- Check open response questions for meaningful answers
- Include duplicate question near beginning and end of long survey
- Include questions that verify understanding rather than paying attention
- Check survey and/or individual question completion time

Questions?