Methods to measure usability of secure/private systems (contd.)

Mainack Mondal

CS 60081
Autumn 2020

Avoid order bias

- Ordering of questions change responses
  - Online survey: people pick top choice
  - On phone: they pick last choice

- Randomize questions and answers

Demographic questions and stereotyping

- Don’t randomize the order in which demographic questions appear
  - Question affect the answer to later questions
  - Finding: Asking women and minorities about their demographics make them perform worse in maths

- Rule of thumb: Ask demographcs questions at the end

Length/compensation for the survey

- Longer survey -> worse response
  - 20 minutes is a good rule of thumb
  - Think: How hard are the questions
  - Do piloting

Grounding your survey (data/interface/situation)

- Ask “How do you comfortable with privacy settings of your Facebook posts”
  - Most common answer “Its ok”

- Ask “Do you want to change the privacy settings of post X”
  - Answer will be a better reflection

6. Survey and question creation

Questions to ask

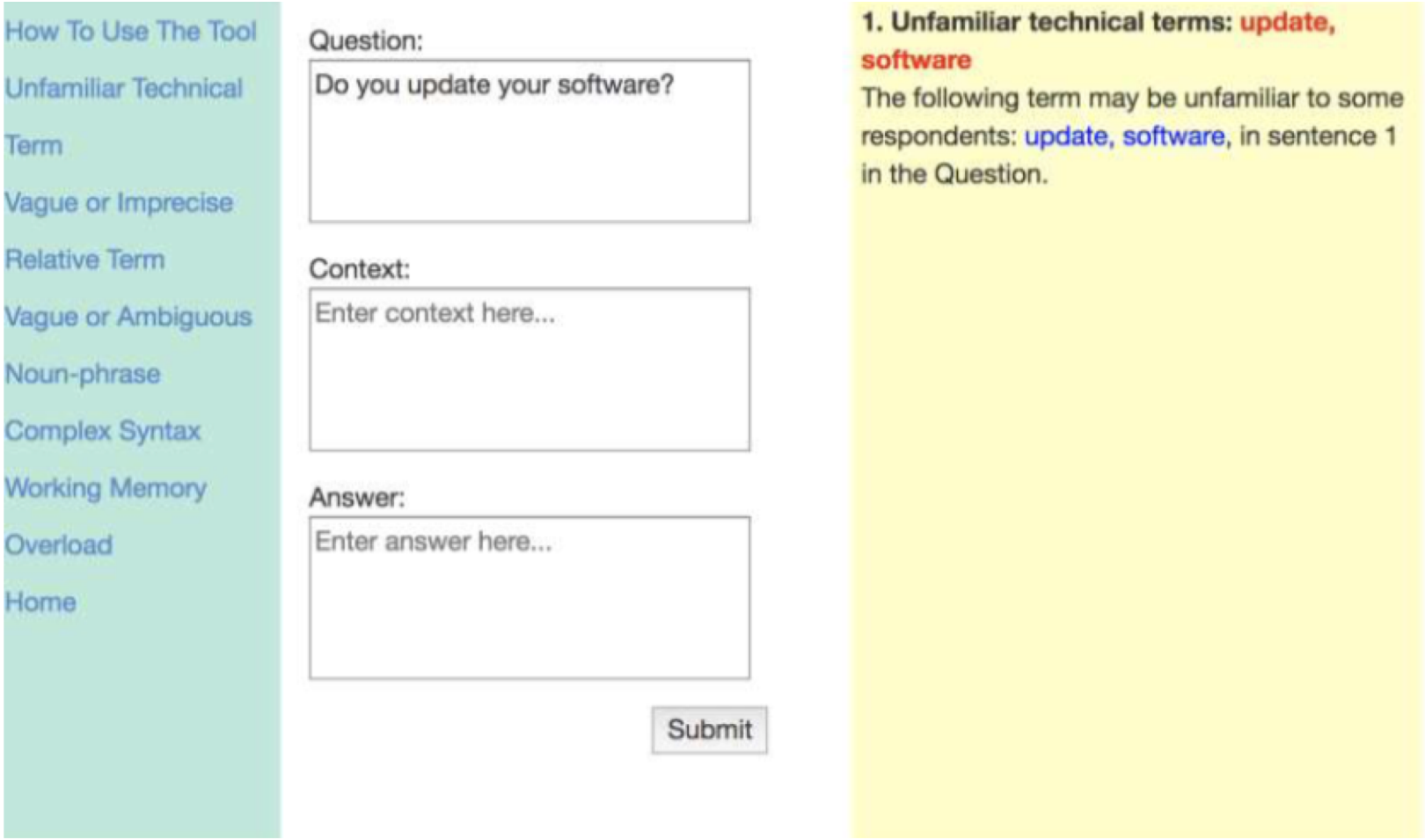
Biases to avoid

Pre-testing / piloting

Pre-test your questions

- Automated tools
  - QUAID: http://quaid.cohmetrix.com/

Pre-test your questions



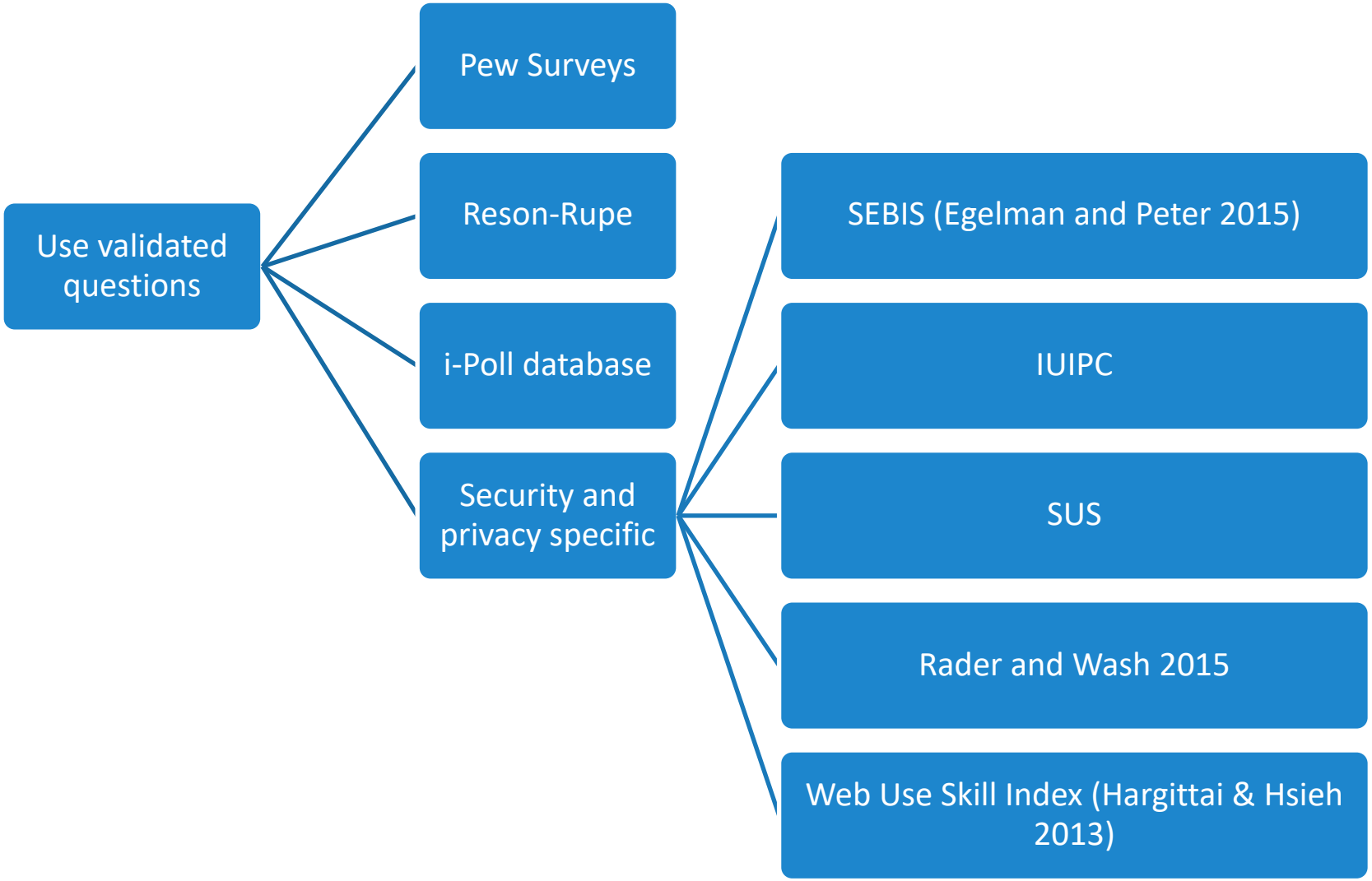
Pre-test your questions

- Automated tools
  - QUAID: http://quaid.cohmetrix.com/
- Cognitive interviews
  - Have respondents think aloud as they answer questions
  - Prompt them on terms that they may struggle with
- Expert review
- Piloting
  - Run a small sample of the survey
  - just because you get answers you like might not mean they are accurate

More on pilots

- Check wording
- Encourage pilot participants to tell you when there is ambiguity or uncertainty
- Verify that you're getting the measurements you thought and that your software works
- Have people talk through even protocols that will be conducted remotely

Use existing validated questions as elements



1. Define your research question
2. Identify your variables
3. Pick one/multiple study methods
4. Run your study
5. Evaluate the outcome

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

Validity

- Internally valid: To what degree are we confident that X causes Y
- Externally valid: To what degree can we generalize about our results
  - What biases does our sample introduce?
- Ecological validity: Does it mirror real-life conditions and context?
- Balancing all of these is hard!

There might be factors you are not considering which are the primary explanatory variable for your observations

Confounds / biases

Potential confounds (1)

- Measurement accuracy / resolution
- Differences caused by different experimental platforms and conditions
- Time of day for recruiting matters
- Failing to account for study dropout or nonparticipation (very subtle!)

Potential confounds (2)

- Learning effect
  - Randomize order of tasks
  - Consider learning effect as a covariate
- Different instructions for different participants
- Biases of recruitment / representativeness
- Self-report biases
  - Don’t ask people to rate expertise

Potential confounds (3)

- Different demographics in conditions
- Placebo effect
  - Why you need a control condition
- Hawthorne effect (changing behavior in response to being observed)
- Chilling effect
  - Fear of repurcassion
- Participants try to please experimenter
  - I like yours better!
  - Minimize knowledge of what's being tested

How to conduct the study?

- Deploy a study remotely (online), e.g., taking an online survey
- Ask participants to come to your lab
- Ask them to let you into their context
- Observe people (take their consent, if not possible, consider necessity of design)

How to recruit participants

- Recruitment mechanism
  - Craigslist, participant pools, representative sample, Mturk, Prolific
- How to compensate?
  - Too little vs. too much
- How to get informed consent?
- How to handle their data ethically?

How to recruit participants

- Recruitment mechanism
  - Craigslist, participant pools, representative sample, Mturk, Prolific
- How to compensate?
  - Too little vs. too much
- ~~How to get informed consent?~~
- ~~How to handle their data ethically?~~

Overall considerations for running a survey

- How do we distribute it?
- How long should it be?
- One-time survey? Longitudinal survey?
- Will you use personalized data?
- What will participants learn?
  - What can we randomize to minimize this?
- Can we randomize the answer choices?

Cover all answer choices

- With whom do you regularly share posts on social media?
  - Family
  - Friends
- Allow multiple answers?
- Include “other” option?

- I connect to Facebook using https: ◯ Yes ◯ No
  - What about I don’t know?

Should we force an answer?

- What gender are you? (* required) ◯ Female ◯ Male

- What gender are you? ◯ Female ◯ Male ◯ I prefer not to answer

- With what gender do you identify? ◯ Female ◯ Male ◯ Non-binary ◯ I prefer to self describe___ ◯ I prefer not to answer

What demographics should we collect?

- Tech expertise, age, domain knowledge, gender, location, employment, etc.
- Don’t ask people to self-rate expertise
  - Ask questions with concrete answers
  - e.g., Have you earned a degree in, or held a job in, computer science, IT, or…
  - Include a knowledge test if you want to know about expertise

1. Define your research question
2. Identify your variables
3. Pick one/multiple study methods
4. Run your study
5. Evaluate the outcome