

06. Robust and Ethical Experiments (Part II); Survey Design; Privacy on Social Media

Blase Ur and Mainack Mondal

April 11th, 2018

CMSC 23210 / 33210



THE UNIVERSITY OF
CHICAGO



Security, Usability, & Privacy
Education & Research

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

Validity

- To what degree are we confident that X causes Y (**internally valid**)?
- To what degree can we generalize about our results (**externally valid**)?
 - What biases does our sample introduce?
- Is this study **ecologically valid**?
 - Does it mirror real-life conditions and context?
- Balancing all of these is hard!

What we conclude from studies

- It's very rare that we conclude something like “for all humans there is an $X\%$ effect of Y ” or “ $Z\%$ of people care about privacy”
 - Be clear what population you have sampled
- We often use proxies in measurement

What we conclude long-term

- **Repeatability:** findings consistent with same researchers and same infrastructure
- **Reproducibility:** findings consistent with different researchers and different (comparable) infrastructure
- Sadly, few studies are replicated
 - Bias against successful replication in peer review
 - (Also) bias against publishing negative results

Some potential confounds (1/3)

- Measurement accuracy / resolution
- Differences caused by different experimental platforms and conditions
- Order of recruiting matters
 - Round-robin (123123123, etc.), Latin squares
- Time of day for recruiting matters
- Failing to account for study dropout or non-participation (very subtle!)

Some potential confounds (2/3)

- 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

Some potential confounds (3/3)

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

Methodology sections

- Be clear and honest about what you did
 - Be honest about limitations
- Give enough detail for someone to replicate
 - Study materials as appendix if possible
 - Correctly report stats (e.g., APA guidelines)
- Release code if possible
- Release data if possible
 - Requires approval from IRB **and** participants

Pilot studies

- Conduct pilot studies!!!
- 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

An example study

- Research question: “Is UChicago the place where fun comes to die?”
- Recruiting participants: what can go wrong?
- Independent variable: assign a university
- Dependent variable: some proxy for fun
 - Hours not studying?
 - Hours not in the Reg?
 - Agreement with statement “We are having fun”

Participants, ethics, and deception

Participants (1)

- Recruit people to do something remotely (e.g., online)
- Recruit people to come to your lab
- Recruit people to let you into their “context”
- Observe people (if possible, get consent! If not possible, consider necessity of design)

Participants (2)

- What recruitment mechanisms?
 - Craigslist, flyers, participant pools, representative sample, standing on street
- How do you compensate them?
 - Ethics of paying \$0.00 vs. \$10.00 vs. \$100,000
- How do you get informed consent?
- What happens to their data?
- Prior knowledge / “what” are they?

Ethics

- How do we protect participants?
 - What risks do we introduce?
- Is there a less invasive method that would give equivalent insight?
- IRB is one arbiter of ethics; experimenters themselves are another crucial arbiter
- How do we make sure participation is voluntary throughout the experiment?

Deception

- Do we mind if participants know precisely what is being studied?
 - Sometimes, it's crucial that we observe their organic responses in context
- What “deception” or “distraction” task can we introduce?
- How do we **debrief** people at the end?

Institutional Review Board (IRB)

- Is it research? Are there human subjects?
- Full review vs. expedited vs. exempt
- Fill out and submit protocol
 - Include all study materials (e.g., surveys)
 - Include recruitment text and/or poster
 - Leave plenty of time

What to submit to an IRB

- Full consent form (use UChicago model)
- All scripts, survey questions, instructions
- Recruitment plan
- Recruitment materials
 - Don't emphasize compensation
- Information about how data will be handled
 - Password protection, encryption, etc.
 - Meetings to discuss

Social phishing (Jagatic et al., 2007)

- Use social networking sites to get information for targeted phishing
 - “In the study described here we simply harvested freely available acquaintance data by crawling social network Web sites.”
- “We launched an actual (but harmless) phishing attack targeting college students aged 18–24 years old.”

Social phishing (Jagatic et al., 2007)

- Control group: message from stranger
- Experimental group: message from a friend
- Used university's sign-on service to verify passwords phished

Ethics (Jagatic et al., 2007)

- How did they obtain consent?
- What ethical concerns are there?
 - What seemed to be done well?
 - What could have been done better?
- Who was potentially affected by the study?
- “The number of complaints made to the campus support center was also small (30 complaints, or 1.7% of the participants).”

Survey design

Overall survey considerations

- 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?

Are all answer options covered?

- With whom do you regularly share posts on social media?
 - Family
 - Friends
- Allow multiple answers?
- Include “other” option (write-in)?
- Do we care about previous use?

Are all answer options covered?

- I connect to Facebook over HTTPS
 - True
 - False
- What about “I don’t know”?

Are we biasing the answer?

- Strangers seeing your Facebook posts would cause you grave privacy concern.
 - Strongly agree
 - Agree
 - Neither agree nor disagree
 - Disagree
 - Strongly disagree

How will responses be distributed?

- For how long have you had Facebook?
 - Less than one day
 - Between one day and one week
 - More than one week

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

Likert-scale data?

- Respond to the following statement:
Companies collect too much private data.
 - 7: Strongly agree
 - 6: Agree
 - 5: Somewhat agree
 - 4: Neutral
 - 3: Somewhat disagree
 - 2: Disagree
 - 1: Strongly disagree

Likert-scale data?

- I feel that companies collect too much private data.
 - 7: Strongly agree
 - 6: Agree
 - 5: Somewhat agree
 - 4: Neutral
 - 3: Somewhat disagree
 - 2: Disagree
 - 1: Strongly disagree

What demographics do 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
- Consider why you are collecting this info

Creative survey designs

- Quantifying the invisible audience in social networking sites
 - M.S. Bernstein, E. Bakshy, M. Burke, B. Karrer. Quantifying the invisible audience in social networks. In *Proc. CHI 2013*.
- Compared actual data about what people saw (by working with Facebook) to survey questions about what they expected