Privacy and Security Challenges in Doctoral Students' Research

Mary Anne Smart, Daniel Tan

Computer Science & Engineering, UC San Diego

Research Questions

Researchers who work with human subjects have a responsibility to try to protect participants from harm. One way that participants may be harmed is through violations of privacy. If we want to help researchers protect participants' privacy, it would be helpful to first understand the range of challenges that researchers are facing. We focus on PhD students, as researchers-in-training.

R1: What privacy and security-related challenges do PhD students face when conducting human subjects research?

R2: How do PhD students deal with privacy and security-related challenges they face in their research?

Methods

Online survey

- Eligibility: Graduate students at least 18 years of age with experience conducting research involving human subjects.
- n=18

Follow-up interview

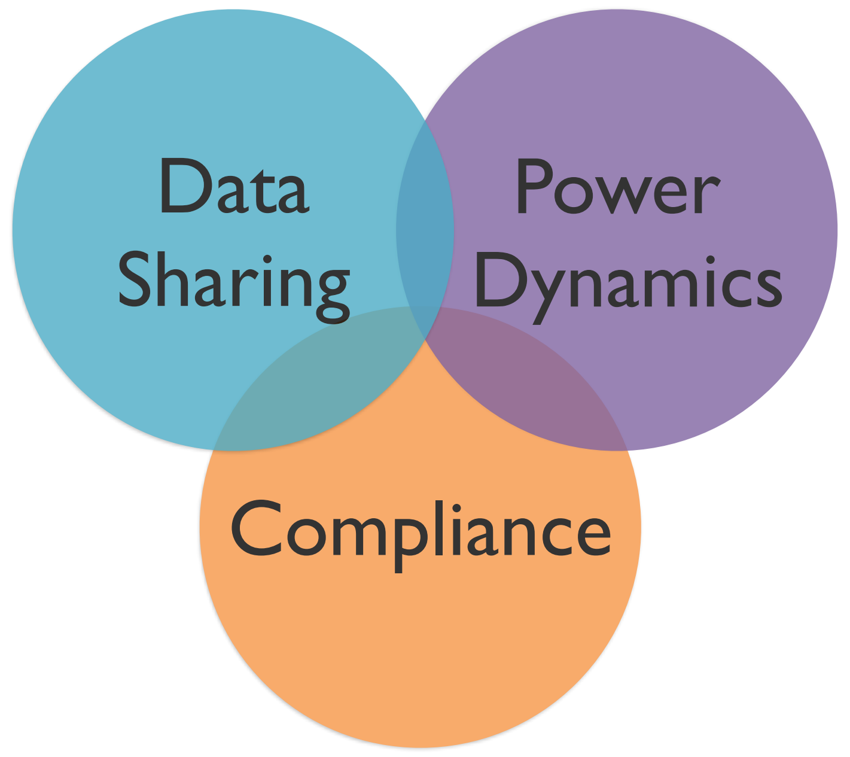
- Semi-structured
- Conducted via Zoom
- n=3

Represented Disciplines

- Biomedical sciences
- Cognitive science
- Computer engineering
- Computer science
- Economics
- Neurosciences
- Public health

Results: Challenges

Below we highlight three interrelated challenges mentioned by many of the students in our study.



When sharing data with other researchers, many students described power dynamics that complicated efforts to protect sensitive data. Uncertainty about how to comply with institutional policies also made sharing data difficult.

> *No matter how much I tried to avoid sharing files over WhatsApp, my team consists of much more senior researchers (boomers) in a context where privacy issues are of little concern generally (developing country).*

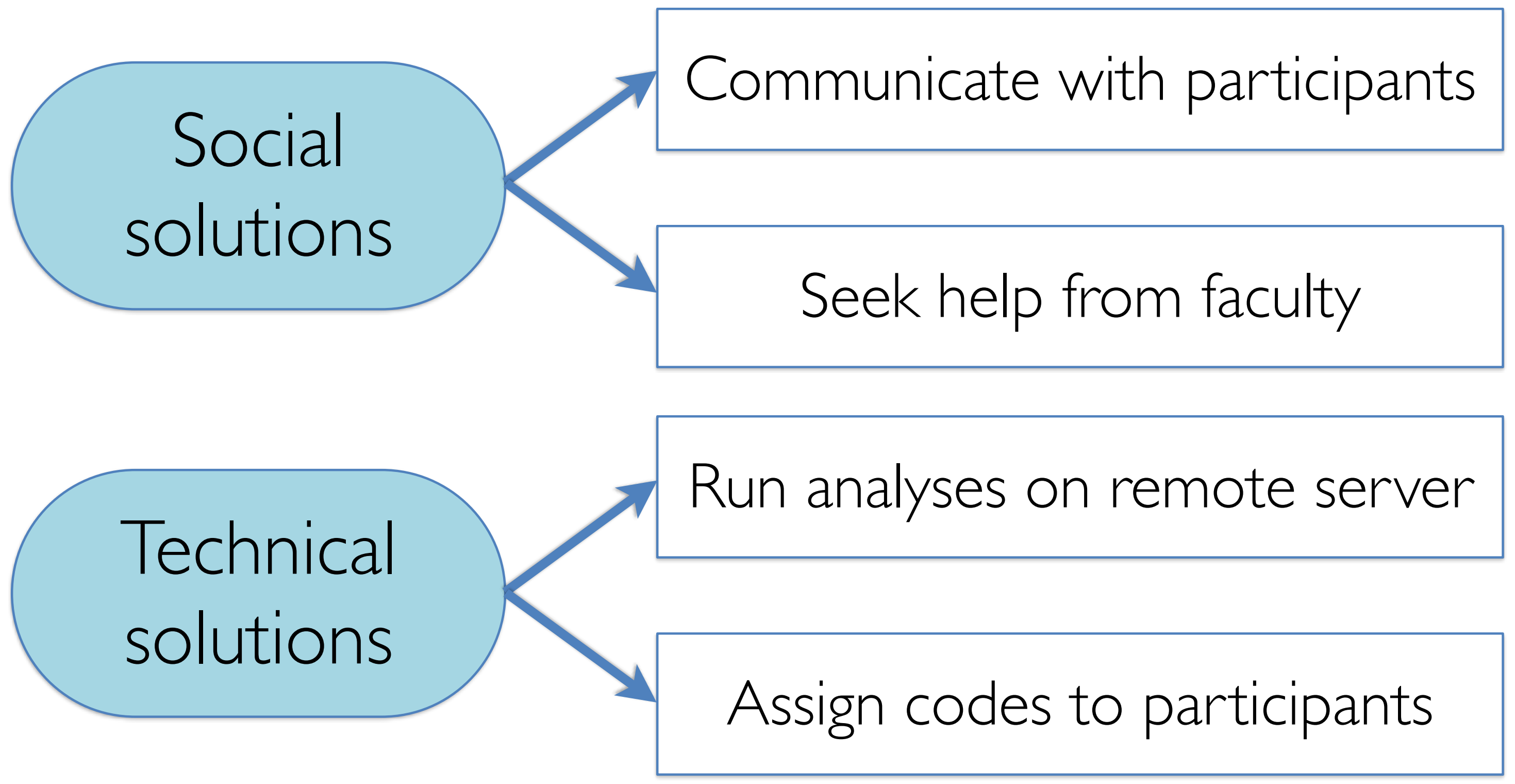
Other challenges included privacy concerns during participant recruitment, confusion about data security, proper anonymization of data, physical security issues, gaps in training, and technological challenges such as VPN connection issues.

> *Participants don't always want to participate in a study that includes vocal signature data that cannot be de-identified (audio recorded data). There's not much that can be done about this.*

> *Keeping the process of assigning participant numbers instead of names completely to myself is all I can do, but that makes months of extra work for me as an individual.*

> *Some restricted data I have are not supposed to contain names. There are names. I ignore the names, but I don't do anything else. (The data provider is likely unable to solve the problem and may be legally required to renege on the agreement.)*

Results: Strategies

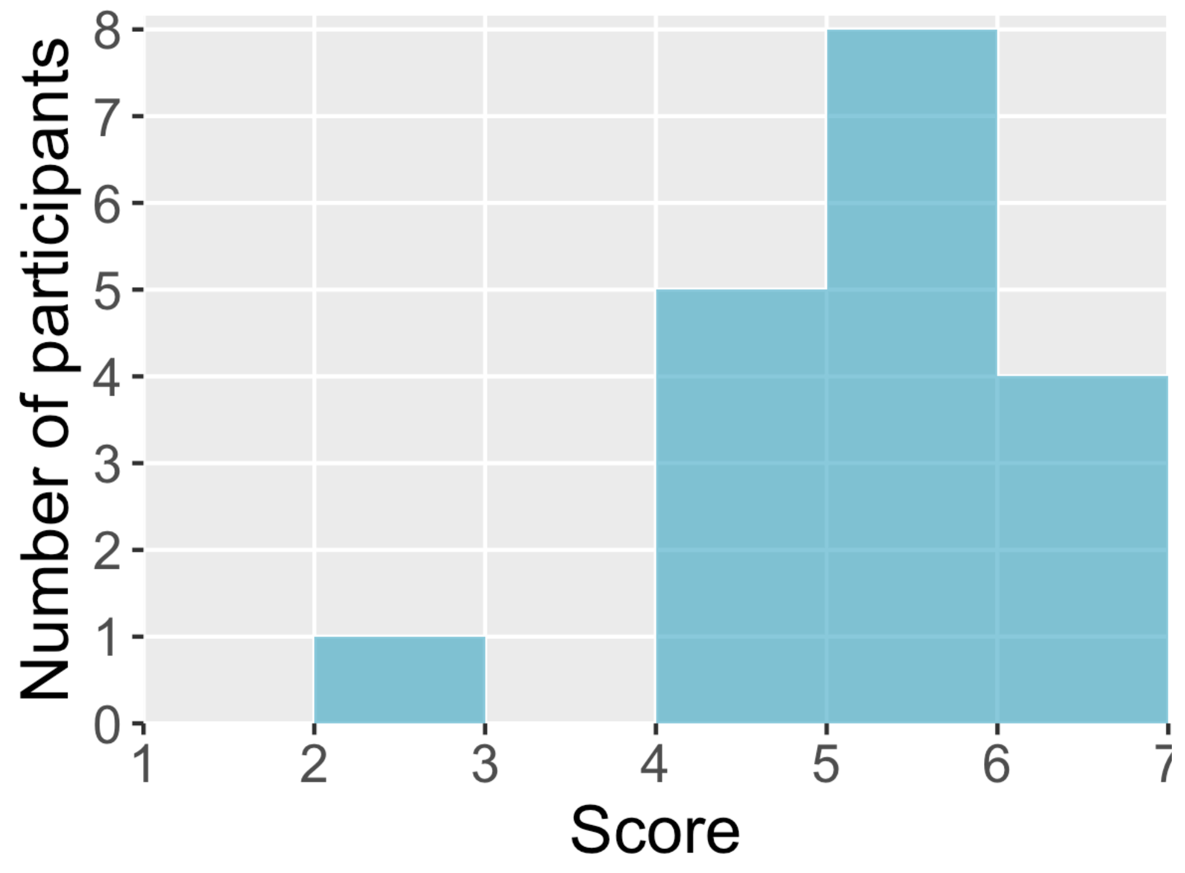


Social solutions
- Communicate with participants
- Seek help from faculty

Technical solutions
- Run analyses on remote server
- Assign codes to participants

Results: Self-efficacy

Most participants had relatively high self-efficacy scores. The scores range from one (low self-efficacy) to seven (high self-efficacy) and represent averages over five questions.



Future Work

- Looking beyond PhD students to also include faculty, librarians, IT departments
- Conducting observations in addition to surveys and interviews

Partnerships between campus libraries, IT departments, and usable security and privacy researchers could have a huge impact in helping PhD students address challenges.

msmart@ucsd.edu