Empirical Software Engineering Research

Ethics

Norman Peitek, Annabelle Bergum, Lina Lampel, Sven Apel

Learning Goals

- Gain awareness of ethical issues & considerations related to SE *research*
- Understand the role and goals of an ethical review board

- Limitation to typical SE research
  - Generally, free-willed human participants (= not animals, patients, drug development, …)
  - In other fields, the ethical considerations can be much more difficult

Motivating Example

- I invite you to an experiment that is horrible
  - Why is it bad?

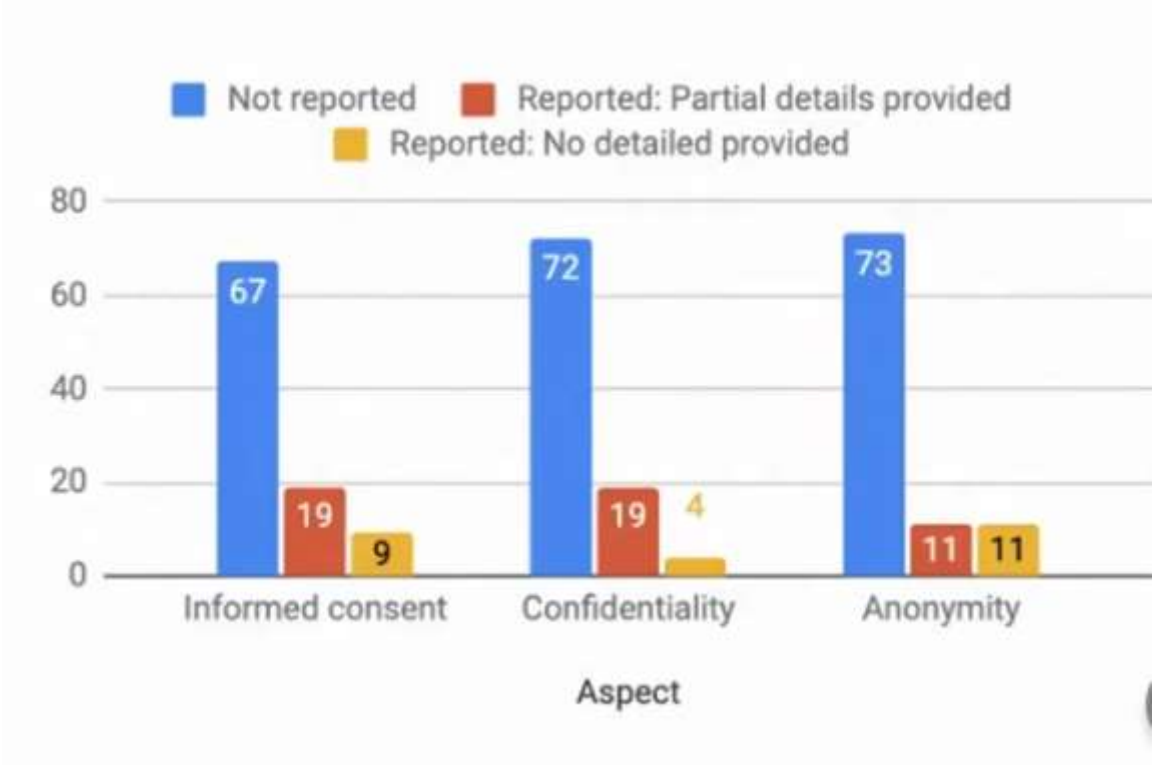
- Students and practitioners are most common subjects
  - They must understand consequences in terms of potential harm

Ground Rules for Research

- All (medical) research with humans must be conducted in accordance with the Declaration of Helsinki
  - Originally from 1964, several updates with the latest in 2013
  - https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Ground Rules for Research

- Adapted to software engineering is easy as we (typically) do not harm participants
- Researchers should:
  - Collect informed consent
  - Maximize confidentiality and privacy of personal data
  - Obtain ethical approval
- Reality in SE: in the past it was often not done (or not reported)



Informed Consent

- Informed Consent is **required** and should:
  - Disclose the purpose of the study, research approach
  - Who has access to the (raw) data and for what purpose
  - Risks (and benefits) of participation
  - *Voluntary participation*

  - Contact info
  - Individual consent
  - Withdrawal, anytime, without penalty (also partially, for example for interview statements)
  - Ask questions

Informed Consent – Example

Data consent form

For this EEG study by the *BrainsOnCode* research group, **personal demographic characteristics (age, gender, etc.), a self-assessment of programming skills, eye-tracking**, and **EEG data** are collected.

What data?

The EEG measures the electrical activity of the brain with electrodes on the surface of the head. The procedure is completely harmless, painless, and medically safe to measure anyway, as the **electrodes are placed non-invasively on the surface of the head** with a cap and only some **water-based gel is applied** to improve conductivity.

Explain EEG method

Eye-tracking data is collected with a device attached to a screen that records the participant's face and tracks the position of the eyes on the screen. **The device does not store or provide a video recording of the participant, only the position of the eyes on the screen in the form of a coordinate.** For data collection, participants take part in the experiment with program comprehension and search tasks.

Explain eye tracking method

Data processing is anonymous as much as possible. The data will not be disclosed to third parties. All personal data are deleted after completion of the research activity, leaving only a completely anonymized data set that does not allow any conclusions to be drawn about the participating individuals.

Data privacy

Responsible office for the processing of personal data:

[redacted]

Informed Consent – Example

Voluntary

Can abort

Consent:

I am participating as a test subject in the research activities described above. My participation is voluntary. I may terminate my participation at any time without giving reasons. **With my signature, I agree to carefully complete the tasks of the experiments and to follow the instructions of the investigator to the best of my ability. I am aware that the collected data will be anonymized. Information about what kind of data was collected is possible at any time. However, due to anonymization, it is not possible to provide access to this specific data.** For the exercise of my rights to rectification, erasure and restriction of the processing of personal data concerning me, as well as to object to its processing, I may contact the aforementioned contacts of the controller. These rights are limited to the extent that they render impossible or seriously impair the achievement of the research purposes and the restriction is necessary for the fulfillment of the research purposes.

Data privacy

The undersigned has the right not to agree to this consent form - **however, since the research project described above relies on the collection and processing of the data mentioned at the beginning, failure to sign would preclude participation in this study.**

Voluntary consent

I consent to the publication of the personal data concerning me in the context of scientific publications. I may revoke this consent at any time. The legality of the publication made on the basis of the consent up to the revocation will not be affected.

Data publication

If I believe that the processing of personal data concerning me violates applicable data protection law, I have the right to lodge a complaint with a supervisory authority.

Informed Consent

Is it relevant how the informed consent was obtained?

For example, your supervisor tells you to participate in their study.

While more data may get collected this way, it highly questions the data quality

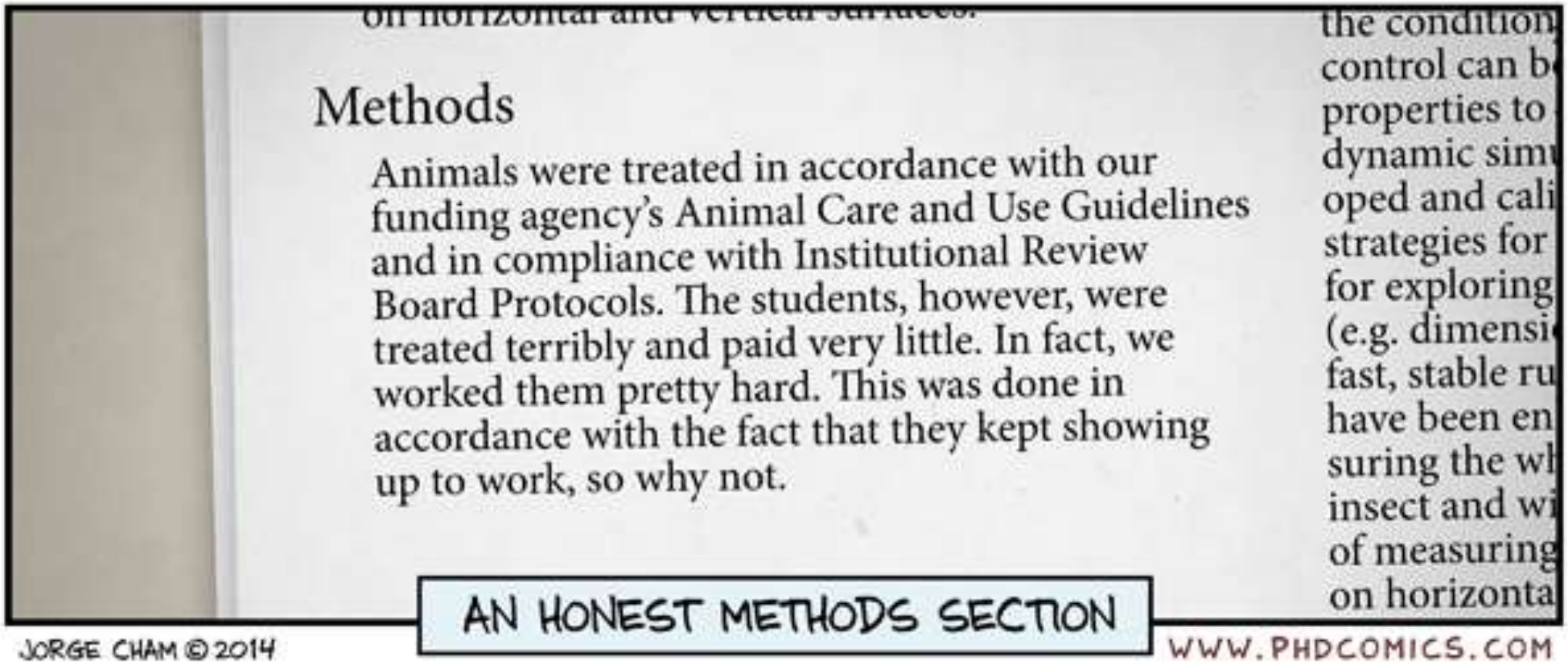
- → For example, an interviewee may not give honest answers if they feel unsafe

Potential Ethical Problems for Participants

- Rewarding (large amounts of) money
- Rewarding extra credits for students

- Ethical dilemmas can also occur for the researcher
  - Example: a brain scan during an fMRI study reveals (by chance) that a participant has a small brain tumor that is easily treatable, but would be fatal in the long term
  - However, the participant indicated they did not want to know about potential health problems
  - → What should the researcher do?

Potential Ethical Problems



Data Anonymity

- Ideally, do not collect data that can be traced to an individual
  - For example, use random subject IDs (“ab12”, “dg83”, …) rather than their full name
    - Why can “Participant01”, “Participant02” be problematic?
  - Especially relevant when there is a dependency between researcher and participant
    - Example: students provide feedback with their full name on a new class before the exam, critics will get worse grades in the exam
- Only collect minimal necessary data
- If impossible, anonymize after data collection or some time period
  - Also consider external dependencies: transcription of interview data by a commercial service

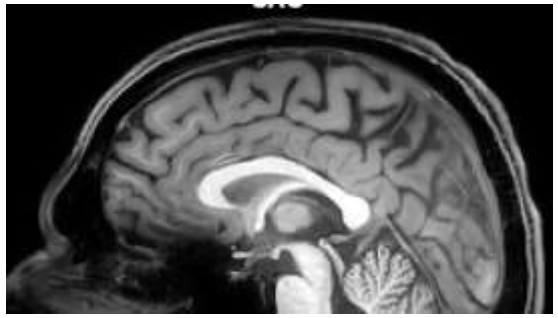
Data Anonymization vs. Deidentification

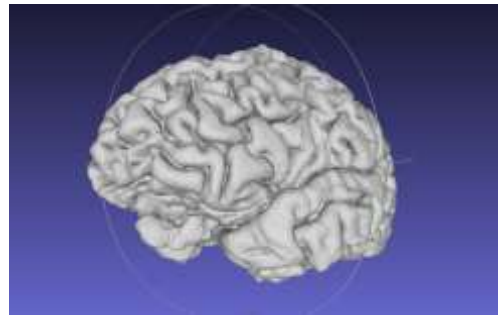
- Anonymization
  - Remove all individual information
  - Deduction of individual is impossible

Name	Birthday
Leonard Hofstadter	01.04.1980
Sheldon Cooper	27.02.1981

ID	Age
Participant ac08	43
Participant kg13	42

- Deidentification
  - Remove as much individual information as possible
  - Deduction of individual is reasonably difficult
  - Example: MRI data allows face reconstruction





Deviations from Ethical Principles

- Rather than anonymous publication, publicly name and thereby potentially damage the reputation of an organization…
  - **… if** there is clear and corroborated evidence for illegal, dangerous or unethical behavior

Data Confidentiality

- Participants have a right that their data is treated in confidence
- Protect raw data at all cost
  - For example, do **not** store experiment data on Google drive
  - Ideally, use an in-house solution of your institution (nextcloud, gitlab)
  - If you use physical medium (hard drive, USB stick), use encryption with password protection

Data Sharing

- Data confidentiality vs. open science
  - Publish aggregated results or raw data?
  - Tension of opposing goals
  - → Raw data must not allow tracing to individual person or organization

- Can we share non-anonymized data?
  - Yes, if it is was very clearly explained and consented to

Study Process

- (Step 0: send experiment proposal to ethical review board)
- Advertise study
- Inform participants, ideally before inviting them to the lab, if it's a non-trivial study
  - MRI consent form shared beforehand
- Informed consent at the lab
- Conduct study
  - What happens if participant wants to abort?

Ethical Review Board: Goals

- What is an “institutional/ethical review board” (IRB/ERB)?
  - Implementation and standards vary by institution and country (even name differs)
  - Typically a group of scientists (profs and non-profs), lawyer, external experts
- The goal is to review research regarding their accordance to ethical standards
  - They typically share their view/position in form of an “approve” or ”reject/revise” based on a risk-benefit analysis
  - They rarely monitor studies
- Not only review participant selection and study design, but more recently, they also review data protection, confidentiality

Ethical Review Board: Process

- Process:
  - Submit a proposal
    - Single study or an umbrella decision for series of studies
  - Review by (subset of) review board
    - Biased members are usually excluded from decision, and sometimes from discussion
  - Decision (if applicable, including recommendations)
    - Optional: renewed proposal
- Local ERB is responsible, unless multi-institution study
- https://erb.cs.uni-saarland.de/

Ethical Review Board: Proposal

- Aim of study
- Experiment design
  - Sample (human participants)
  - Describe potential risks to participants
  - Informed consent

- https://erb.cs.uni-saarland.de/wp-content/uploads/2016/11/Gesch%C3%A4ftsordnung-Ethikkommission-engl.pdf

Ethical Review Board: Responsibility & Law

- However, they are not responsible for the conducted research
- Is an application to the ERB required?
  - Generally, no*
  - Practically, sometimes yes (depending on the field)
  - When do you need to submit a proposal?
    - Highly dependent on the institution (some say, no need for a proposal when there are no humans involved)
    - Read/figure out the rules of your local institution. If in doubt, ask!

*Germany requires it for pharmaceutical products. Some countries, such as Switzerland, do require an ERB approval for all human research

Ethical Review Board: Problems

- A large time commitment for the members of the ERB
- Potential conflicts of interest
  - Do I reject an unethical study of my colleague/superior?
- Different standards and decisions across different ERBs

Ethical Review Board: Problematic Case

RWTH Aachen study on nitrogen dioxide in 2012

- The goal of the study was to understand how harmful nitrogen dioxide gas is for humans
  - At the time, it was already known to be harmful (EU had set thresholds)
  - A few years later, nitrogen dioxide gas was part of the diesel scandal
- Study funded by an interest group of the industry
- Participants were exposed to different levels of nitrogen dioxide gas
- When the study was criticized, the university responded with “ERB approved the study”
- → What are the problems?

Ethical Review Board

What do you do if you are at an institution that does not have a ethical review board?

- Check with your supervisor
- Keep a paper trail

Ethical Review Board

The ethical review board demands changes to your experiment. What do you do?

- They are usually correct and should be implemented
- If you have a good reason why the changes are not reasonable, ask for clarification
  - It is possible that the ethical review board misunderstood your proposal

Ethical Review Board

After getting approval from the ERB, you change the experiment design. Do you send it to the ERB again?

- Depends on the institution and the process
- Generally, I would try to be on the safe side: yes
  - Unless the change is completely irrelevant (for example, change statistical analysis from R to Python)

Ethics and Open-Source Data

- Are there ethical considerations related to research of open-source data?
  - Yes!

- Commit history of an open-source projects reveals individual humans
  - It is difficult/impractical to get informed consent
  - Despite the data being open source, they may personally do not want to be investigated
- Many other forms of data: email lists, build logs, …
- Researchers must always consider ethical considerations

Ethics and Non-Human Data

- Are there ethical considerations related to research not involving humans (e.g., code, organizations?)
  - Yes!

- Examples
  - Open bugs that reveal security issues
  - …

Reporting Standards

- Publication vendors now often require ERB approval
  - Not just “was approved”, but concrete ERB with application number
  - Empirical Software Engineering journal:
    - “…study-specific approval by the appropriate ethics committee for research involving humans and/or animals, informed consent if the research involved human participants…”
    - https://www.springer.com/journal/10664/submission-guidelines#Instruction%20for%20Authors_Ethical%20Responsibilities%20of%20Authors
- Reporting ethics are a completely different topic
  - Plagiarism, self-plagiarism, salami-sliced publishing, …
  - Do not falsify data (presentation)
  - …

Further Readings

- Ethics in SE: https://www.youtube.com/watch?v=usW6iISknzo
- Association for Computing Machinery (ACM)
  - Code of Ethics and Professional Conduct: https://www.acm.org/code-of-ethics
  - ACM Ethical standards for Empirical Research
    - https://acmsigsoft.github.io/EmpiricalStandards/Supplements/?supplement=EthicsEngineering
    - https://acmsigsoft.github.io/EmpiricalStandards/Supplements/?supplement=EthicsHumanParticipants