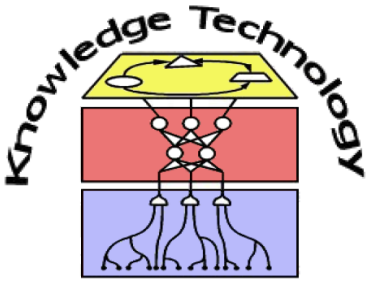
Research Methods

Human Participants

Dr. Sven Magg, Prof. Dr. Stefan Wermter



http://www.informatik.uni-hamburg.de/WTM/

Plan for today!

1. What is so special about humans?
2. Ethical considerations
3. Experiment preparation
4. Running experiments with human subjects
5. Data collection

content partly taken from:

- ERC presentation (Silvia Álvarez Santos)
- Adaptive Systems Research Group, University of Hertfordshire, UK

Why are humans different to robots?

From a roboticist's viewpoint, humans are…

- **…hard to control in terms of variability**
  - humans always have predispositions, prejudices, prior knowledge, intentions, biases…..
  - humans are affected by many cues (often not consciously perceived), changing their behaviour
- **…difficult to query about internal states or processes**
  - We are often unaware of own internal state
  - Conscious thoughts and the recollection and reports thereof often misleading
  - Memory recollection often erroneous (due to dynamic nature of memory)

Why are humans different to robots?

From a roboticist's viewpoint, humans…

- **…perceive themselves differently**
  - ""Know thyself"? If I knew myself I'd run away" – Goethe
  - The picture we get through self-reflection is often biased
  - Self-reflected personality traits often different to observed ones
  - Our decision-making process is constantly affected by our social environment (assumed expectations, etc)
- **…are constantly learning**
  - Two trials with same subject automatically lead to dependency between measurements
  - Order effects within one trial (one task affecting the next)

Why are humans different to robots?

From a roboticist's viewpoint, humans…

- **…have to be handled with special care**
  - They can't be easily repaired when broken
  - can't be easily extended with new sensors/actuators
  - always know lawyers in case you violate the previous points!

- **…in summary:**
  - Require especially rigorous experimental procedures
  - Always require the involvement of third parties (ethics board, independent testers/reviewers for procedures, lawyers)

Ethical Considerations

ERC / EU commission: Ethics in Horizon2020

- Ethics Issues:
  - Human embryonic stem cells or foetal tissue
  - Non-human primates
  - Research on animals
    - animal welfare
  - Intervention on human beings
    - Where the human body is physically involved in research
    - Involvement of children
  - Privacy and personal data
    - Coded data ≠ anonymous data
    - Informed Consent issues (healthy adult volunteers)

Ethical considerations

- Ethics Issues (cont.):
  - Research in international cooperation partner countries
    - Use of local resources
    - Benefit to local community
  - Dual use / Misuse
    - Potential military / terrorist application

- All are dealt with in several (legal) directives, e.g:
  - Data protection: Directive 95/46/EC
  - Charter of Fundamental Rights of the European Union
  - WMA Declaration of Helsinki
  - European Charter for Researchers

EU Charter for Researchers

- some quotes:
  - on Ethical Principles: *"Researchers should adhere to the recognised ethical practices and fundamental ethical principles appropriate to their discipline(s) as well as to ethical standards as documented in the different national, sectoral or institutional Codes of Ethics."*
  - on Accountability: *"Researchers need to be aware that they are accountable towards their employers, funders or other related public or private bodies as well as, on more ethical grounds, towards society as a whole."*
  - on data protection: "*They should also be familiar with the current national legal requirements regarding data protection and confidentiality protection requirements, and undertake the necessary steps to fulfil them at all times.*"

Declaration of Helsinki

- Ethical Principles for Medical Research Involving Human Subjects:
  - *"It is the duty of the physician to promote and safeguard the health, well-being and rights of patients"*
  - *"While the primary purpose of medical research is to generate new knowledge, this goal can never take precedence over the rights and interests of individual research subjects"*
  - *"It is the duty of physicians […] to* protect the life, health, dignity, integrity, right to self-determination, privacy, and confidentiality of personal information of research subjects."
  - *"The responsibility for the protection of research subjects must always rest with the physician […] never with the research subjects, even though they have given consent."*

EU Directive on Data Protection

- Directive 95/46/EC of the European Parliament:
  - difference between identifiable, pseudonymised (= coded), anonymised
    - **identifiable**: directly identifies the subject
    - **pseudonymised:** provided with an identifier that enables re-connection to identifiable data through connection of databases
    - **anonymised:** can not be connected to original subject record
  - **pseudonymised ≠ anonymised**
  - subjects have to be told
    - What is recorded, how it is processed, who is processing it
  - subjects have to give **explicit** consent
  - Individuals will have the right to refer all cases to their home national data protection authority

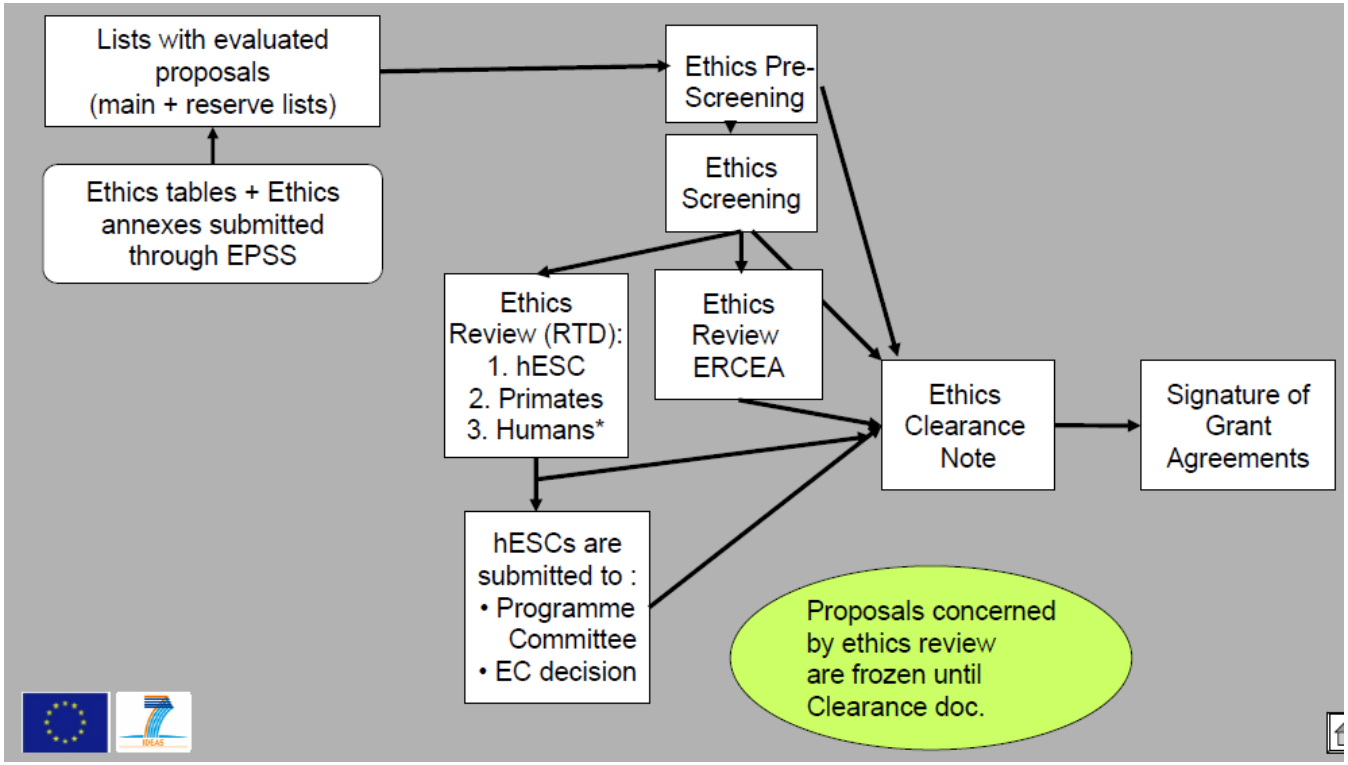
What do you have to do?

- You have to
  - ensure, argue for the safety of subjects (risk analysis)
  - Inform the subject (before or maybe after) the trial about the use and processing of the data
  - Get written and explicit informed consent
  - Make sure that all recorded data is protected

- Questions you may have to consider:
  - How about data encoded within the systems?
  - Who has access to the data?
  - Does a robot reveal data to third parties in interaction?

Ethics Processes

- Check the procedures at your institute and follow them!
- Journals or funding bodies often have processes on top
- Universität Hamburg has no own ethics board or process!



Reasons for rejection

- **for clinical trials**
  - Safeguard data protection in terms of data transfer
  - Design of informed consent forms
- **Data protection and privacy**
  - Failure to describe informed consent procedures at community and individual levels
  - Codification, storage and anonymity of personal data
  - Secondary use of data
- **Issues related to children**
  - Failure to describe if children obtain a real and direct benefit

Consent Form

- What are the required components?

- **General information**
  - Purpose of the study
  - Who is conducting and funding the study?

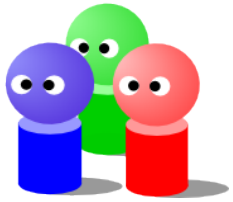
- **Information on the study**
  - What do you expect the participants to do?
  - How long will it take?
  - Risks and benefits
  - Compensation

Consent Form

- **Data**
  - What will be recorded
  - How will confidentiality be maintained?
  - How will the data be processed?
  - How do you offer anonymity?

- **Participation and rights**
  - Right to withdraw at any time without consequences
  - Participation is voluntary
  - Contact person
- **Signature with explicit statements on what one signs for**

Group Task!



3

10

How good is the consent form we have used?

How do we get data from humans?

- **External sensors**
  - video, audio, depth sensors, eye-tracker, EEG, fMRI,…
  - recording not biased by participant
  - Processing the data can be difficult
  - Data can again be manually processed/coded
    - Beware of new source for bias/variability
    - Procedure has to be fixed and tested
    - What are the variables and how do we capture them
    - Training of coders
    - Rigorous procedure and training can reduce variability

How do we get data from humans?

- **Questionnaires**
  - Can be prepared and tested beforehand
  - Easy to process, can be automated
  - Paper-based, computer-based, online
    - Paper-based is local and needs post-processing
    - Computer-based questionnaires offer additional options
      - Limited answering time
      - Limiting review of old answers
      - Randomization of order (questions, pictures, graphs, etc)
      - Automatic processing
    - Online questionnaires offer a large participant base
      - Strong sampling biases almost inevitable

LIREC Questionnaire

- Questionnaire from the study:

Syrdal, D. S., Dautenhahn, K., Koay, K. L., Walters, M. L., & Ho, W. C. (2013). Sharing Spaces, Sharing Lives–The Impact of Robot Mobility on User Perception of a Home Companion Robot. In Social Robotics (pp. 321-330). Springer International Publishing.

- Moving or stationary assistant
- Looking at spatial interactional dimension
- Is it important to share the physical space?
- 9 sessions (long-term study)
- Agents gave reminders, Skype, following user, etc

Group Task!

What are the variables that are measured by this questionnaire?

LIREC Questionnaire

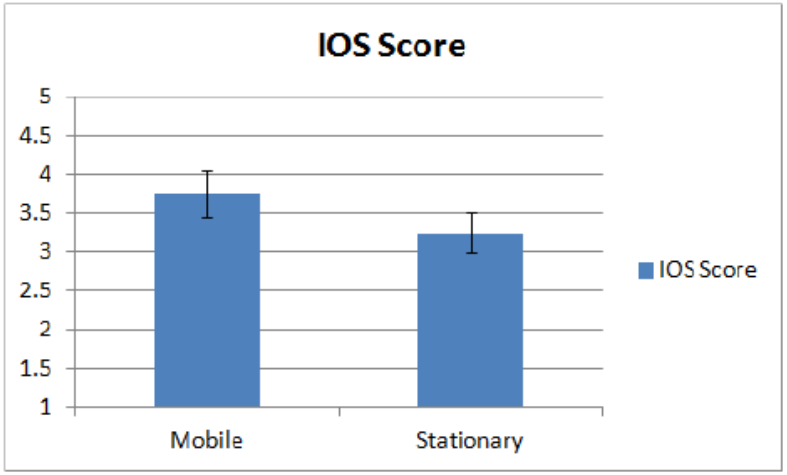
- Mixture of standardised and ad-hoc scales/measures
  - System Usability Scale (Brooke, 1996)
  - Scenario Acceptance Scale
  - Inclusion of Others in Self (IOS) scales (Aaron et. al 1992)
    - correlates with feelings of closeness in human-human relationships
  - GODSPEED questionnaire (Bartneck et al. 2008)
    - Measures user's perception of robots
  - Ad-hoc questions (Likert + follow-up)

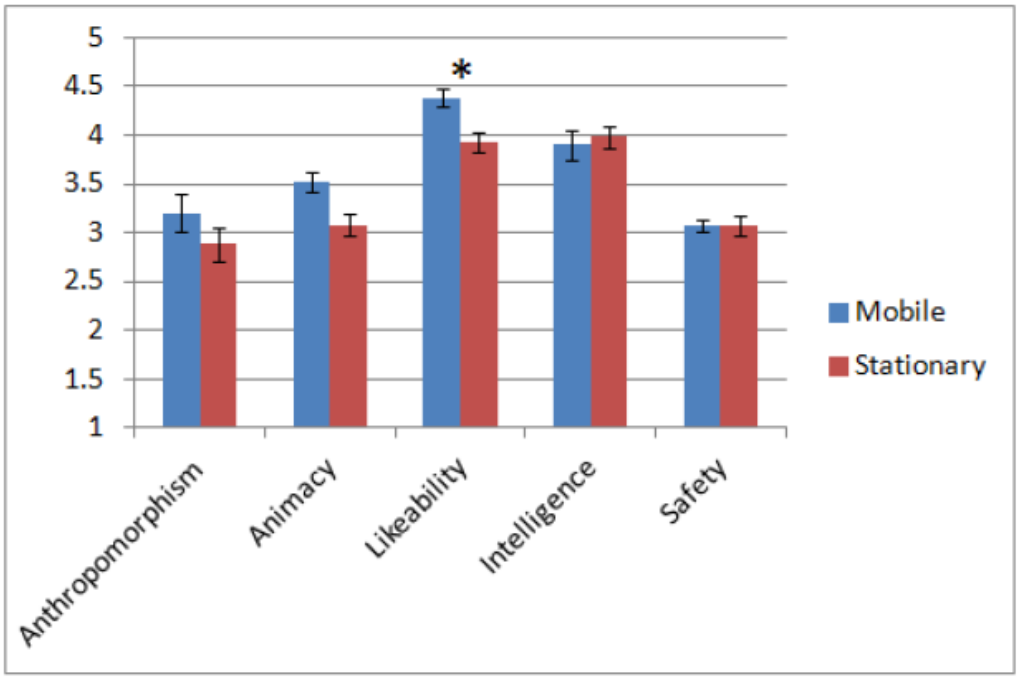
- Available scales already validated and used by others

$\Rightarrow$ Improved comparability of your own study!

Presentation of results

- 8 participants
- Tendency in the IOS score (p=.17) towards the mobile robot





- Godspeed results
- 5 dimensions
- No difference in rating, except for likeability and animacy

X-point scales

- Difference between items (1 line) and scale (sum of responses)

- **Semantic Differential Scales**
  - Use of bi-polar adjective pairs as items
  - Often used to measure attitudes
  - e.g. Lifelike-Artificial, Like-Dislike, Good-Bad, etc.…
  - Treated as ordinal or interval scale (careful!)

  - 5-Point example item:

Like				X		Dislike

X-point scales

- **Likert-Scale**
  - Items are questions or statements that have to be evaluated
  - Usually between “Strongly agree” and “Strongly disagree”
  - Symmetric around neutral level
  - example 5-point item:

	Strongly agree	Agree	Neutral	Disagree	Strongly Disagree
Han shot first!	X				

  - Sometimes even scales with “forced choice” (no “Neutral”)
  - Treated as ordinal or interval scale

X-point scales

- **Evaluation**
  - Either as single items or group of items
  - Often evaluated on deviation from middle assuming an interval scale (e.g. in case of symmetric Likert-Scale)
  - Evaluated on central tendency (i.e. mean, median, etc.)

- **Known biases / problems:**
  - Central tendency bias: Avoidance of extreme values
  - Acquiescence bias: Tendency to agree with statement
    - Can be avoided by balancing positive/negative statements

Questionnaires

- If you have to **design a questionnaire**:

1. Define what you need to measure
2. Look for validated scales that measure what you need
3. If no scales exist:
   - Define items
   - Look for help from psychologists/social scientists/experts
   - Test the items and the evaluation
4. Run a **pilot study** to test the questionnaire and the evaluation, with a follow-up interview to get feedback

Interviews

- Interviews (before and/or after study)
  - Qualitative analysis
  - Very expensive (time for interview and tedious analysis)
  - Often difficult to quantify results
  - Always suffer from experimenter's biases!

- If you run interviews,
  - try to support them by quantitative measurements
  - Define clear aims for the interviewer (e.g. questions to answer)
  - Thoroughly train the interviewers!
  - Record the interview (for independent evaluation, review, etc)

The Experiment

- As with all experiments
  - define the experiment protocol and all procedures thoroughly
  - prepare consent form and questionnaire(s)
- Follow ethics approval procedure to get clearance
- Run pilot study and improve setting if necessary
- Recruit participants
- Run the session
  - Welcome and briefing (get informed consent)
  - Conduct the experiment
  - Debriefing and verifying records (i.e. check files and data)
  - Provide compensation (if promised) and reset for next

Pilot study

- There will be things you forgot!
  - Be test subject yourself to tests setup
  - Test setup with people with no prior knowledge on setup
  - Run interviews to get feedback on procedures

- Run test analysis on data and compare to interview feedback
  - Do the results match?
  - How large is the variability, can the data be used?
  - Retrain raters/coders if necessary, adjust procedures

Recruitment

- Before recruiting participants
  - think about biases that might affect your results and how to control for them
  - always assume there are effects through biases
    - Try not to avoid biases, but to accept them and minimise their effects
    - find measurements to identify them and check effects
    - do exploratory analysis to find confounding effects
  - carefully design groups
    - think about effective randomisation
    - or include biases in the design (e.g. especially selecting for a bias to minimise variability, careful grouping)

Recruitment

- Where and how to recruit?
  - After you have identified who you want to recruit, think about how to recruit those participants
  - Maybe different, tailored announcements/channels?

- After recruitment
  - Make sure you gather data to check if you have succeeded
  - Include analysis to confirm your success

Summary

- Humans are difficult to deal with
- Human participants will include unnecessary variability
  - Take extra care when designing your experiment
  - Always have detailed procedures/scripts
  - Test and re-test your setup until you are sure the analysis can test your hypotheses

- Accept biases and include them in design and/or analysis
  - Often variability can be greatly reduced due to intelligent design, e.g. phrasing of instructions or questions
- Same rule as in software engineering:
  - Early errors in design $\Rightarrow$ expensive/impossible to correct later

What have we learned?

1. Human participants require very carful experiment design and rigorous procedures/scripts
2. With human participants, not running a pilot study is an act of negligence
3. Think about ethical issues and explicitly define and write down all measures that will be taken to ensure good conduct
4. Whenever possible, resort to already available scales and the help of psychologists or social scientists