CHALMERS |  UNIVERSITY OF GOTHENBURG

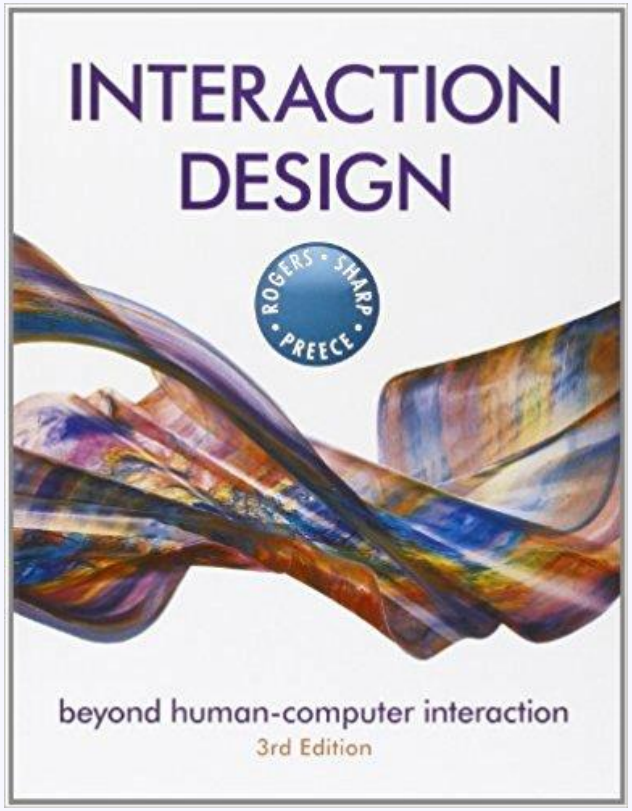
DIT045/DAT355
Requirements and User Experience

Lecture 10: Usability Experiments & RE, UX Videos

Jennifer Horkoff
jenho@chalmers.se

Interaction Design

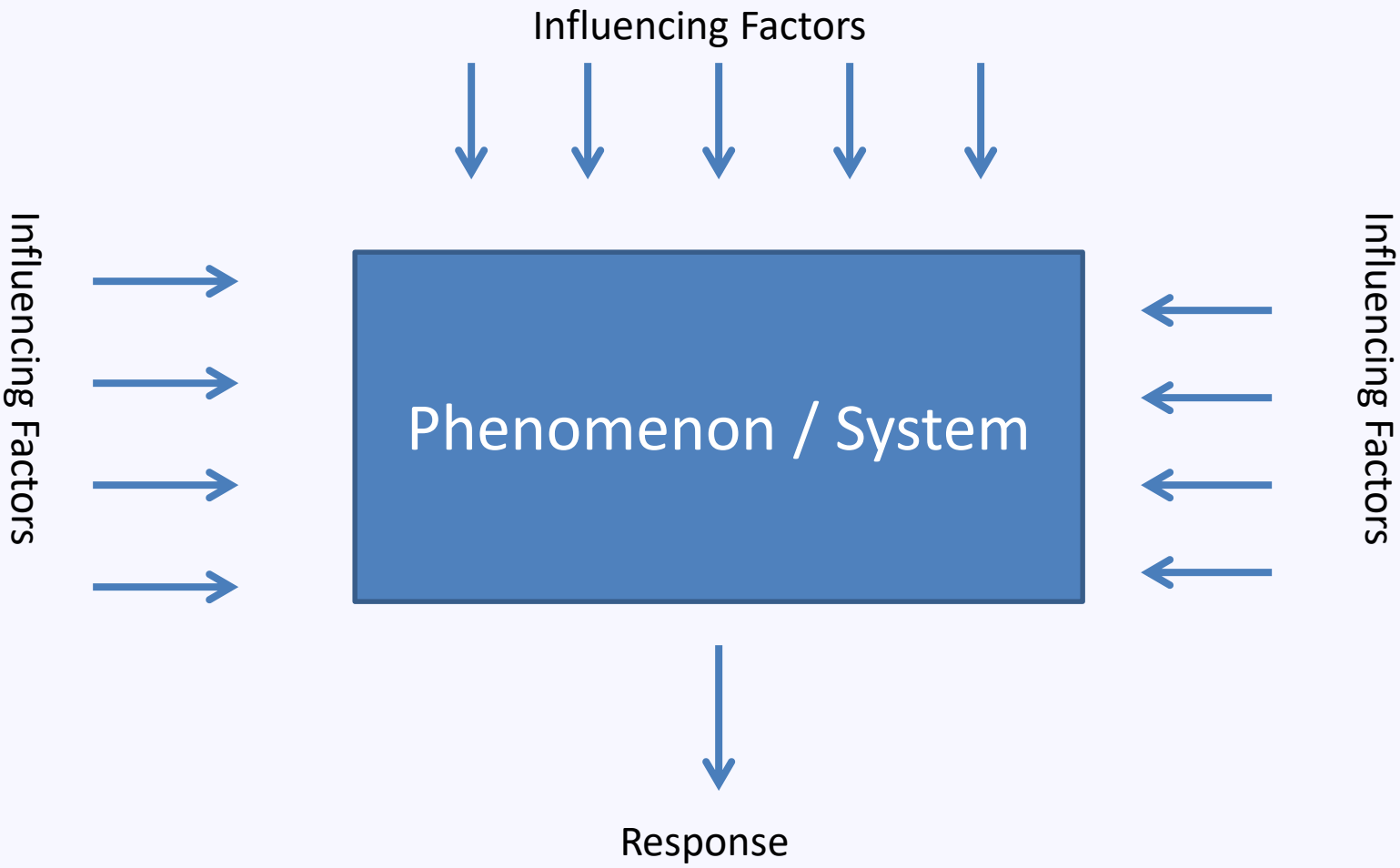
Evaluation: Usability Testing and Field Studies



Usability Experiments

- When you have to compare two candidate interfaces.
  - E.g. an old and a new one.
- There is a need for generalization for a larger class of systems (e.g. interested in statistical significance)
- Larger and representative sample sizes.
- More rigorous control of other influencing factors.

What are experiments?



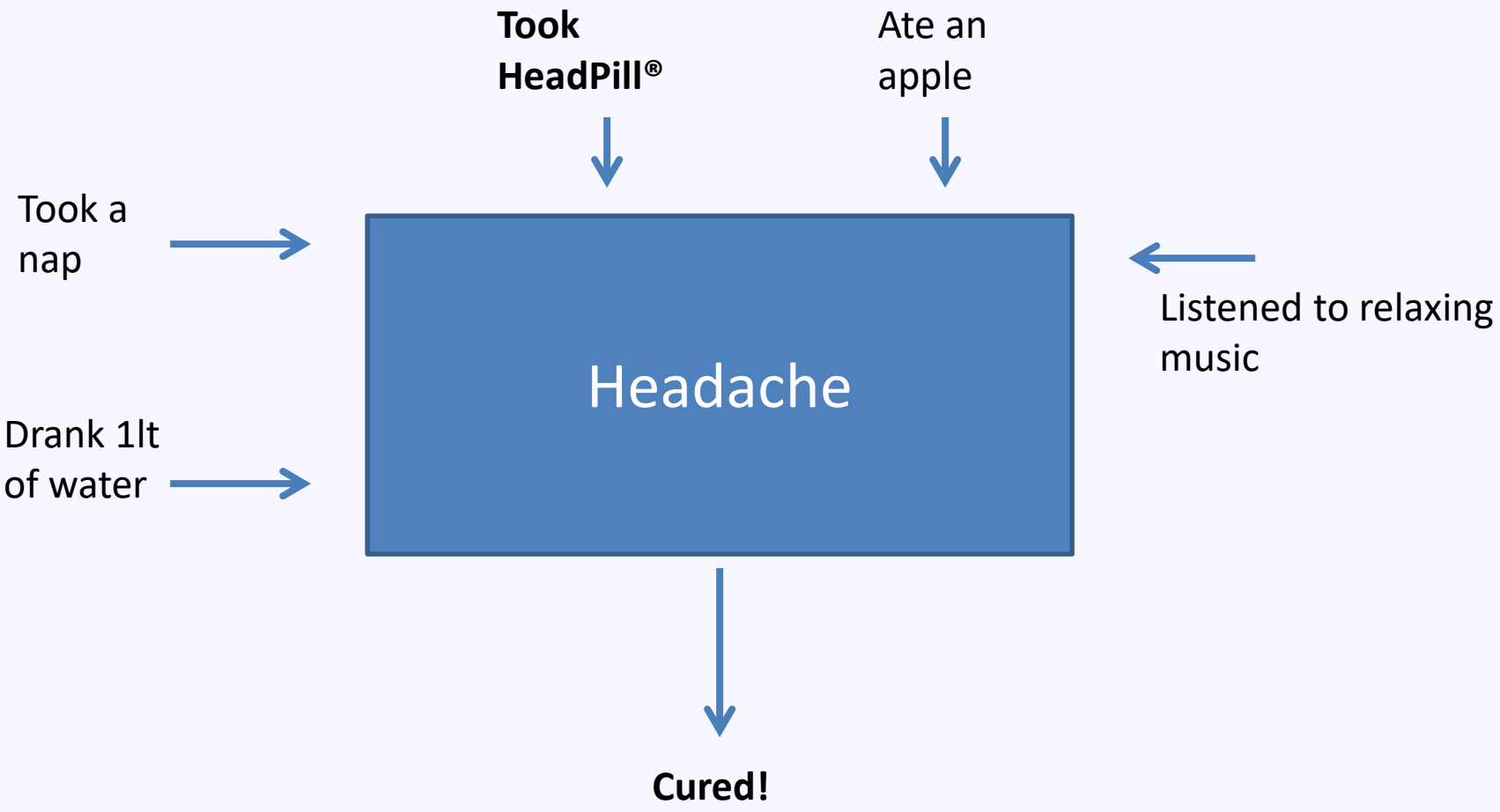
Experiments - Goal

- Establish whether the **response** can be actually **attributed** to one of the possible influencing factors.
  - I.e. say: “it was because of ABC factor that we observed XYZ response.”
- In usability experiments we want to say:
  - “it was because of the interface design that a user took less time to complete a task (and not because e.g. the user was smarter)”

Experiments - Example

- Assume that I produce the HeadPill® , which, **I claim**, cures headache.
  - Would you believe me and go buy the HeadPill?
  - What proof do you need in order to believe me?
- Assume that I give the pill to a participant.
- The participant comes back after a few hours and says he was *cured*!
- Are you convinced?

Experiments - Example



Experiments - Example

- The pill ***may*** have cured the headache…
- … **but** it may have also been **the apple** that cured the headache…
- … or the **1 litre of water** that the participant drank…
- … or the fact that he **took a nap**.

- So: we cannot attribute the response (that the headache was cured) to the pill.
- We need to **rule out all** other factors.
  - How??
  - By **controlling for** them.

Experiments - Technique

- Take 2 groups of participants A and B, all with headache.
- Participants in A eat an apple, drink 1lt of water, take a nap, listen to relaxed music **and take the pill.**
- Participants in B eat an apple, drink 1lt of water, take a nap, listen to relaxed music **but *do not* take the pill.**
- I.e. you **control for** apples, water, sleep, relaxed music.
- If A are cured and B are not cured:
  - You know it **must have been** the pill.
  - All else was equal for A and B! There is no other explanation. (*)

Experiments - Terminology

- **Independent Variable.**
  - The influencing factor you are tweaking in order to see if it has an effect to the **dependent variable**.
  - It is up to you to tweak it, that's why it is "independent".
  - **Example:** whether or not to take the pill.
- **Dependent Variable.**
  - The response/effect you are wondering if it should be attributed to the independent variable.
  - **Example:** whether headache goes away.
- **Confounding factors.**
  - Other influencing factors that, if left uncontrolled, they won't let you prove the connection.
  - **Example:** eating the apples, drinking the water, sleeping, etc.
- **Experimental Condition.**
  - The state of the independent variable at a given phase of the experiment.

Usability Experiments

- Same logic like any experiment.
- Used to **compare** two interfaces A (e.g. an existing) and B (a redesigned one).
- Have some subjects use A and some subjects use B.
- Keep all else equal:
  - The time of the day, the amount of sleep subjects had, the environment, the keyboard/mouse, the age/skill of the subjects.
- Observe: do subjects who use B perform better (e.g. faster) than those who used A?
  - If yes, and everything else is held equal (= is controlled for) then it is probably because of B being a better design that we observe the better performance.

Usability Experiments

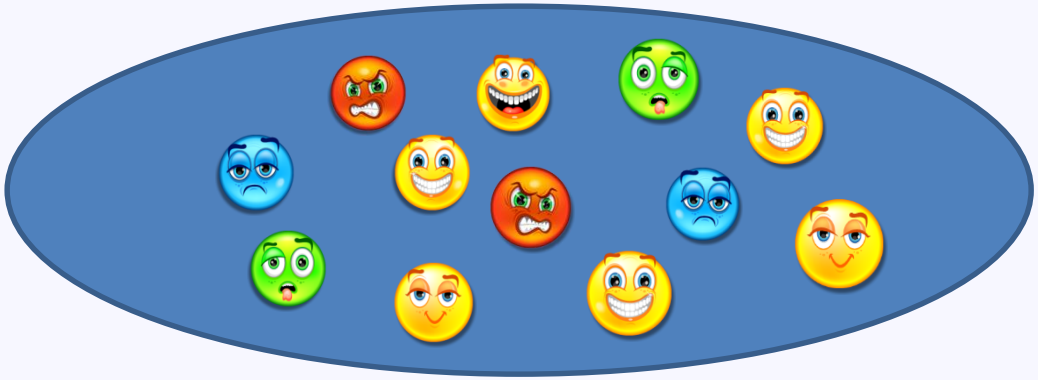
- **Dependent Variable:**
  - How long does it take to perform *a given* task (e.g. enrol a course?)
- **Independent Variable**
  - Use **System A** or **System B**
- **Confounding factors**
  - The time of the day, the amount of sleep participants had, the environment, etc.
- **Experimental Condition.**
  - Whether the given task is performed using System A or System B.

Samples and Individual Differences

- You want to test if a coin is fair.
- How many times do you toss it?
  - One time?
  - Two times?
  - More times?
- Clearly the more times the better. A bigger sample (of coin tosses) is more “representative” of the population of all possible coin tosses.
- The same logic applies to participants.
- Participants have different characteristics and capabilities:
  - Visual and motor skills, cognitive skills, attention levels, culture, background, education…
  - They have slept different amounts of time, eaten or not before they came to your experiment, may or may not be having a bad day, may or may not be stressed/distracted about something etc…
- Solution: choose many of them to level out these **individual differences.**

Assigning Participants to Experiments

- When comparing interface A with interface B we need to decide which participants will use A and which participants will use B.



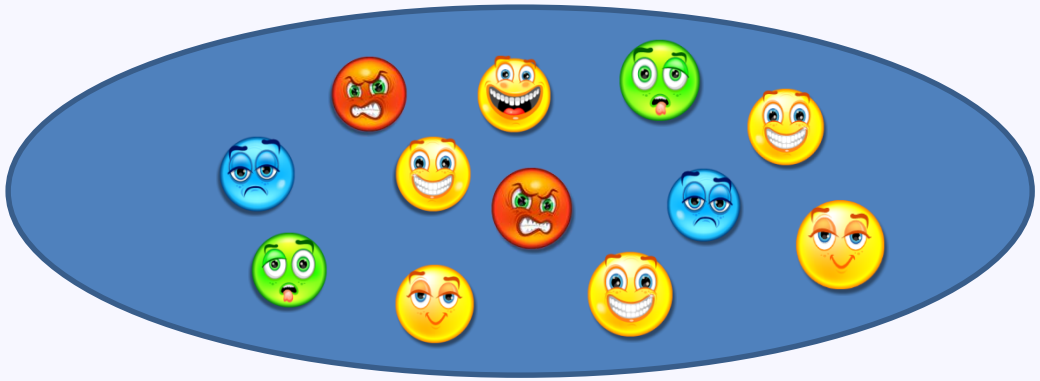
Interface A

Interface B

Experimental designs

- **Different participants** (“between subjects”) - single group of participants is allocated randomly to the experimental conditions.
- **Same participants** (“within subjects”) - all participants appear in both conditions. (**counterbalancing** to neutralize the learning effect)
- **Matched participants** - participants are matched in pairs, e.g., based on expertise, gender, etc.

Different Participants – Between Subjects

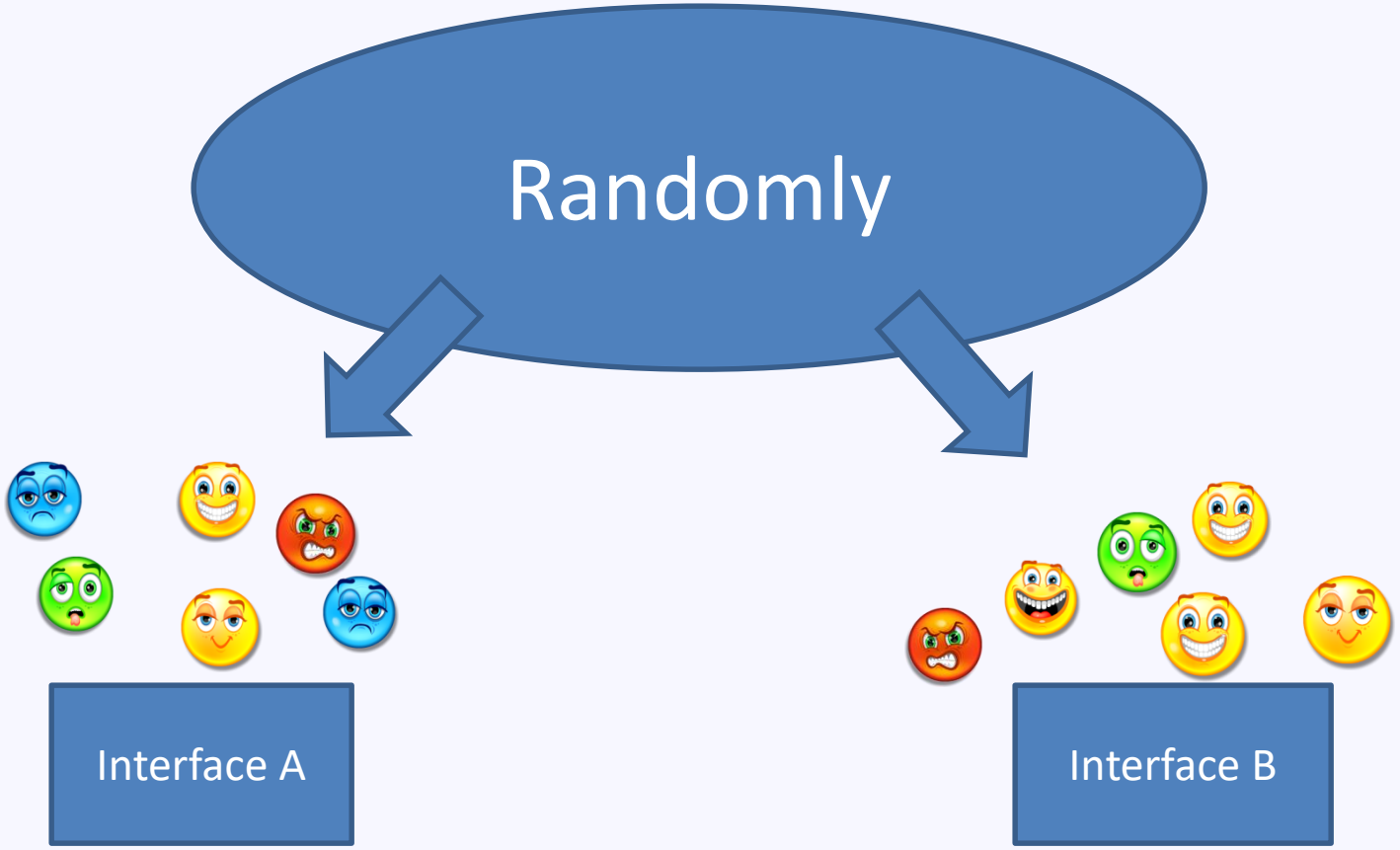


Interface A

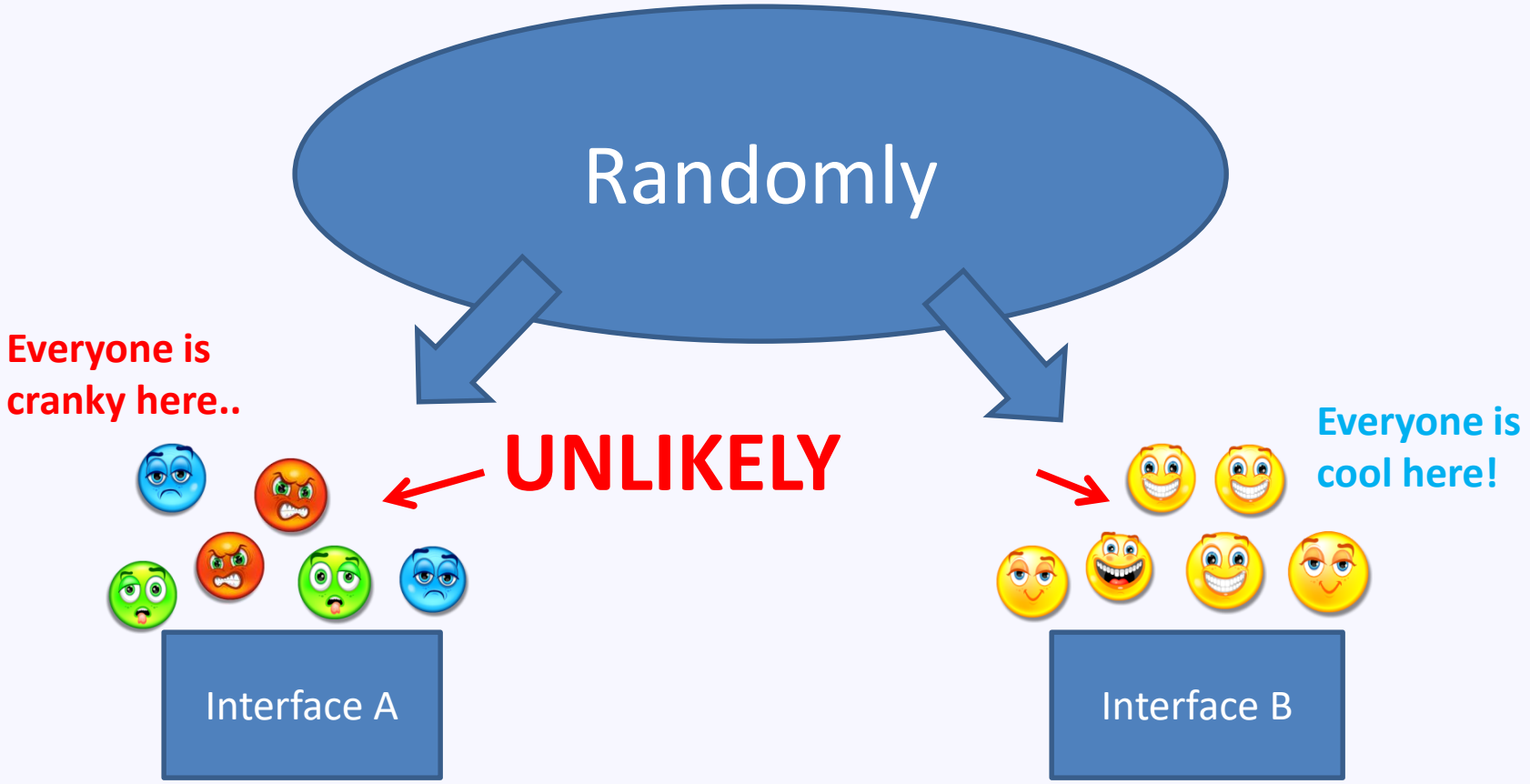
Interface B

Different Participants – Between Subjects

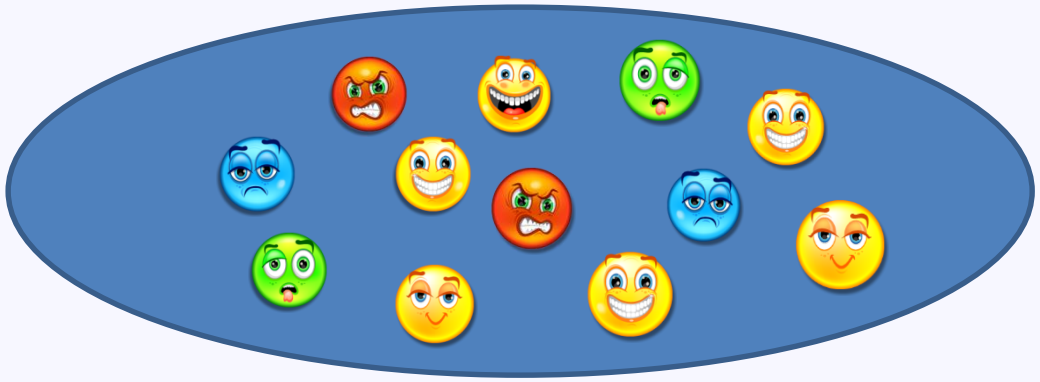
Through randomly distributing to conditions, when the sample is large enough, individual differences are unlikely to group together.



Different Participants – Between Subjects



Same Participants – Within Subjects



Interface A

Interface B

Same Participants – Within Subjects

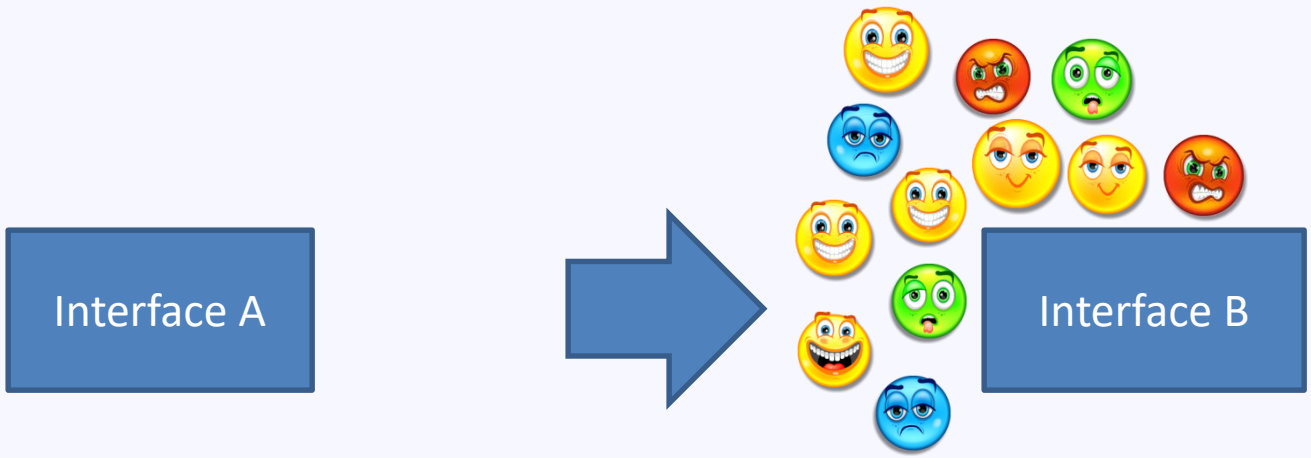
Stage 1: everybody uses interface A

Interface A

Interface B

Same Participants – Within Subjects

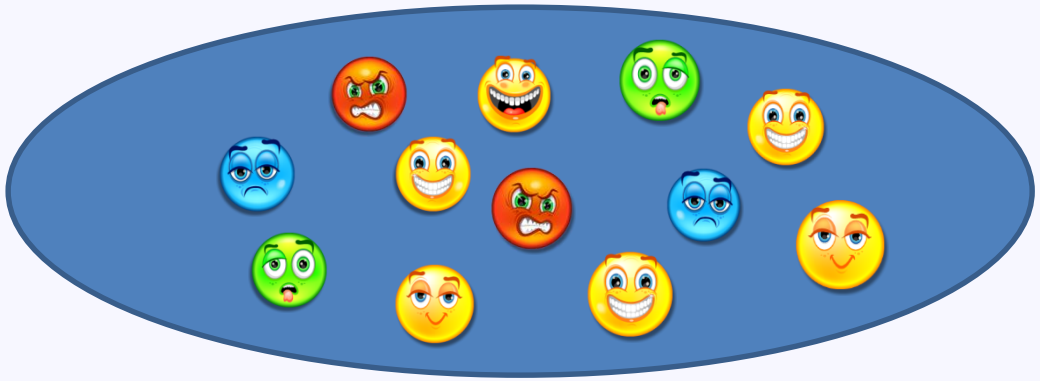
Stage 2: everybody uses interface B



Same Participants – Within Subjects: ISSUE

- **Learning Effect / Order effect**
  - When participants use interface A they familiarize themselves with it and the experimental process.
  - They are more likely to perform better using Interface B just because of this.
  - Thus ordering becomes a **confounding factor.**
  - Differences in the responses may not be due to the quality of the interface but the sequence in which it was tested.
- Solution:
  - Counterbalancing.
  - Split participants: half do A → B, and half do B → A.
  - Learning effects **still exist** but **cancel out**.
    - (Assuming they are symmetric)

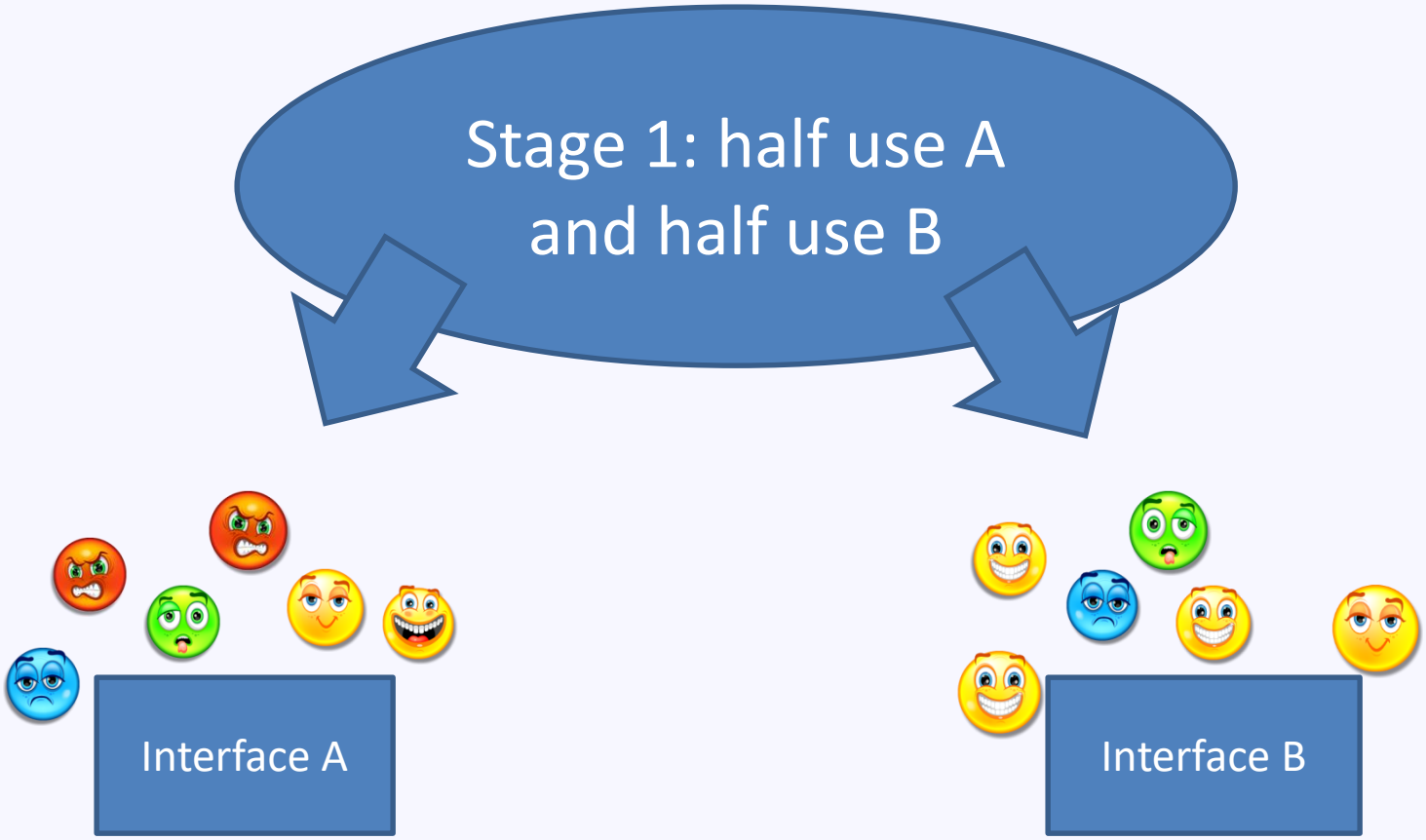
Within Subjects with Counter balancing



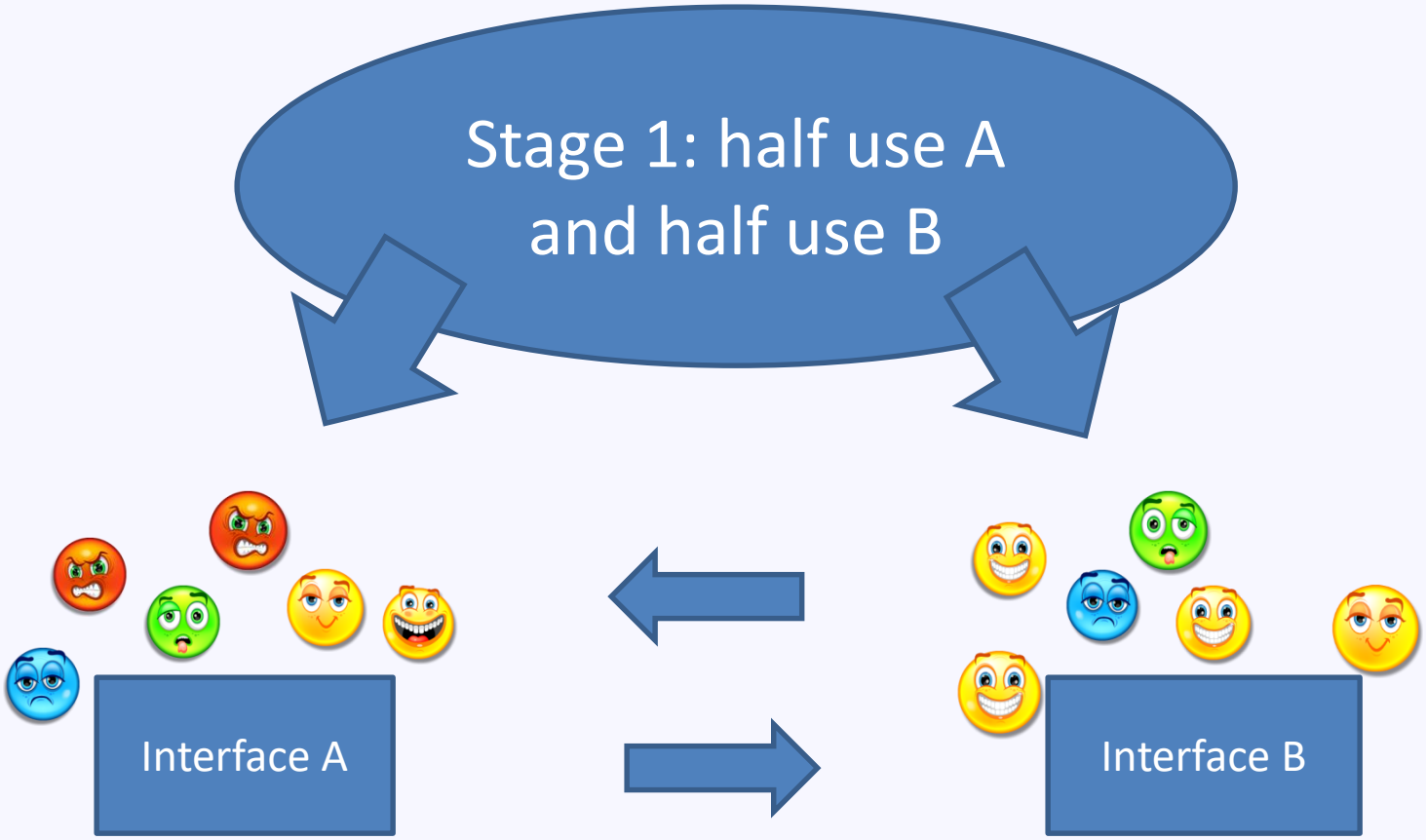
Interface A

Interface B

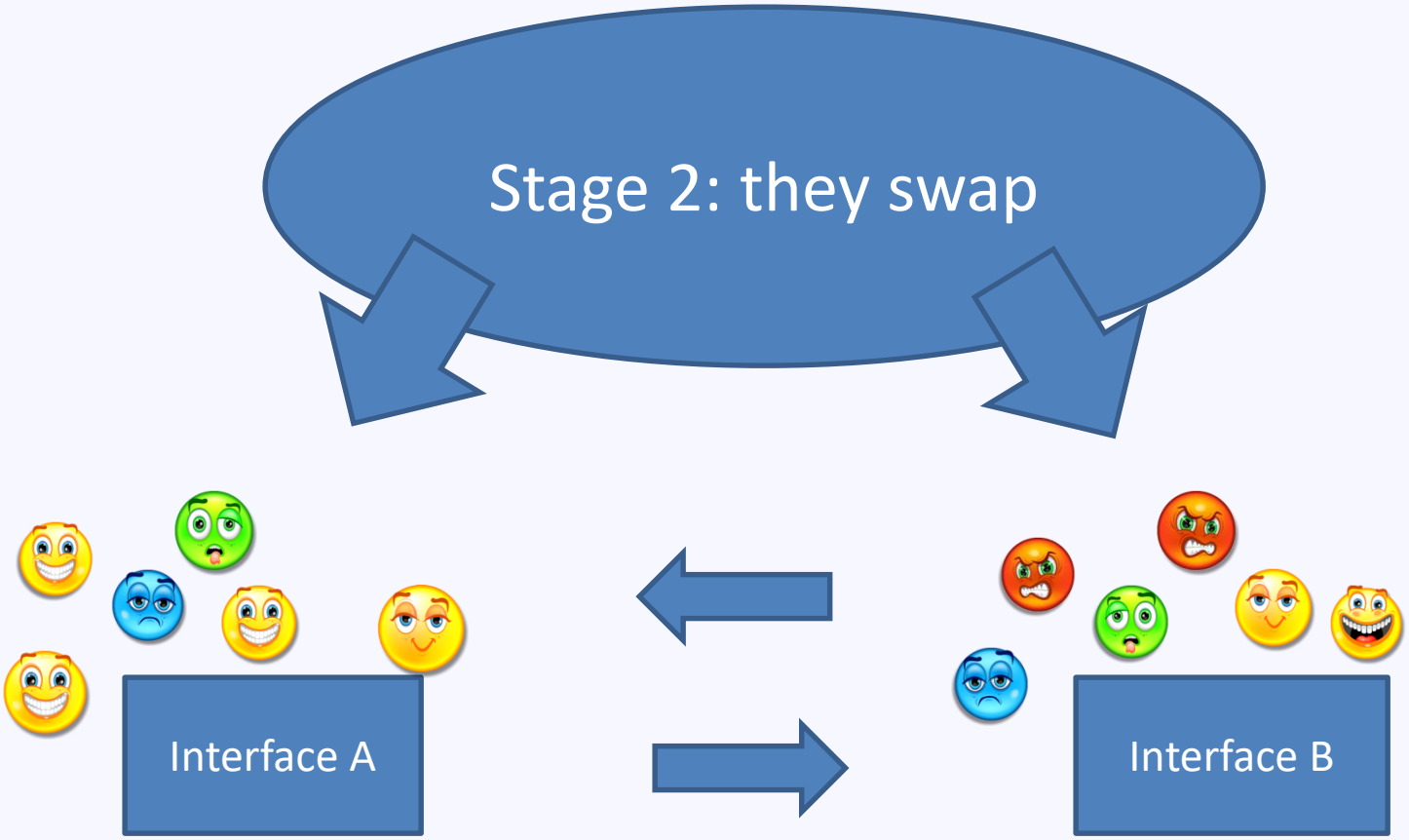
Within Subjects with Counter balancing



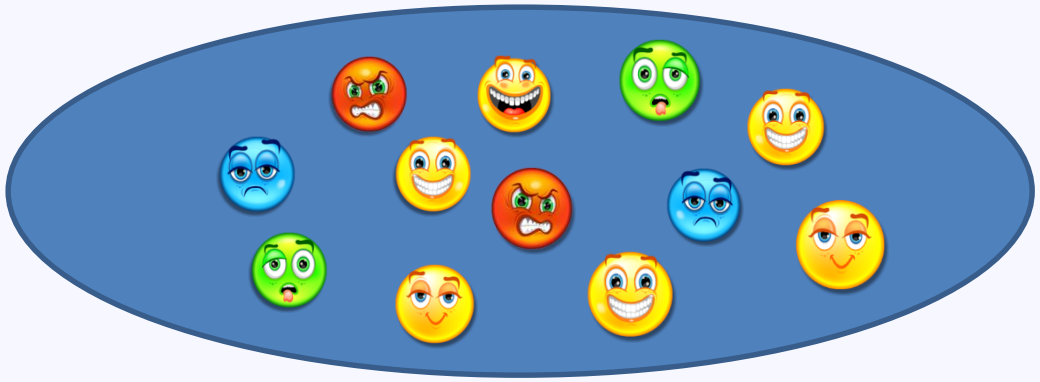
Within Subjects with Counter balancing



Within Subjects with Counter balancing



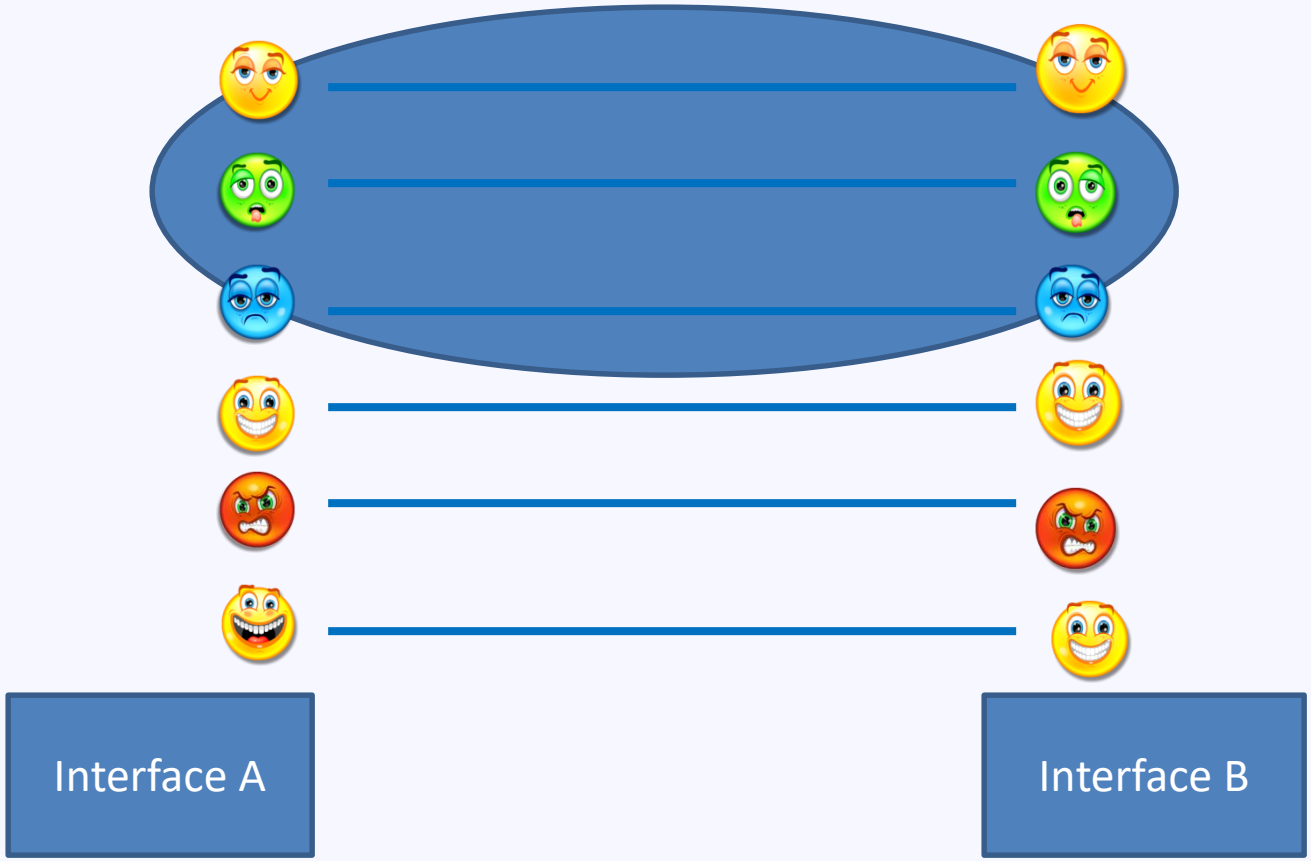
Matched Participants



Interface A

Interface B

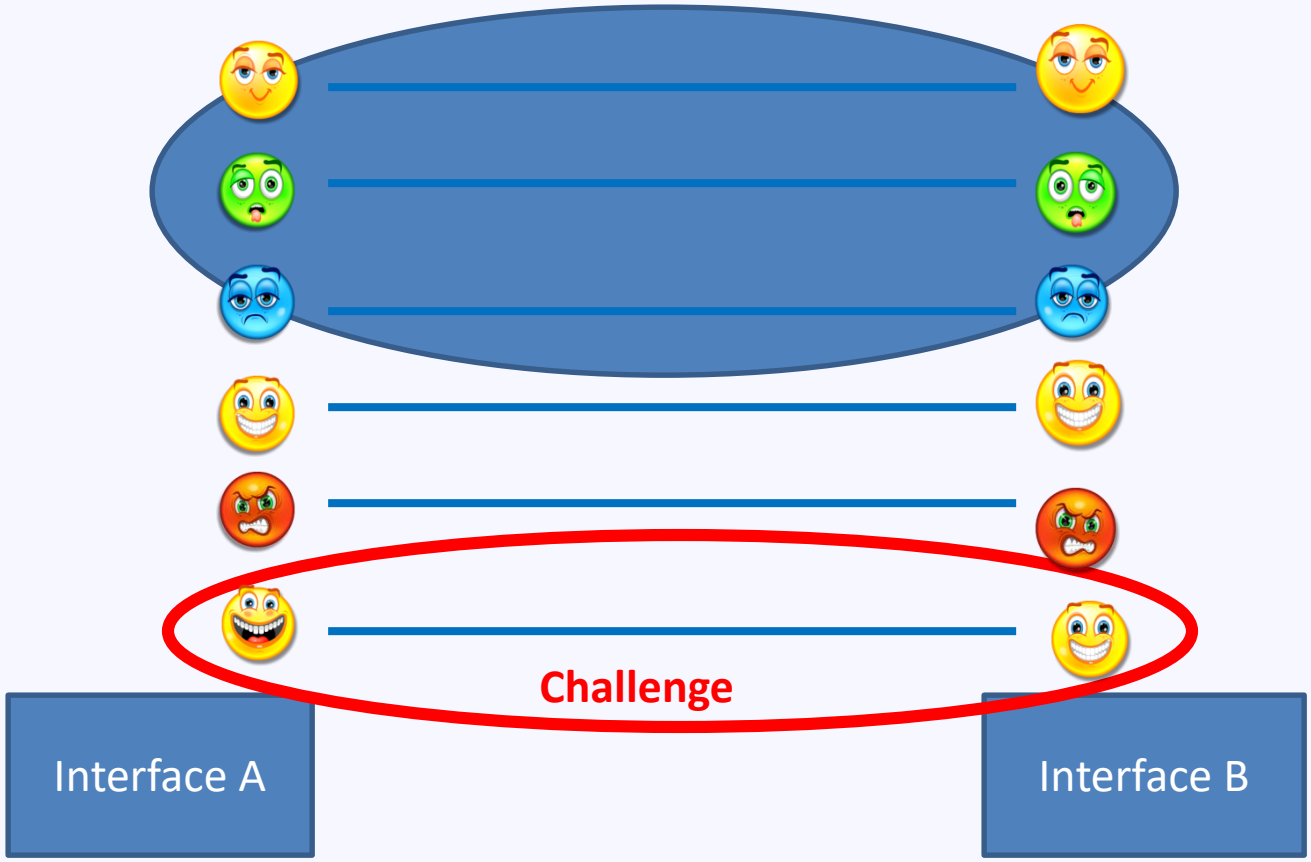
Matched Participants



Like between subjects, but not leave it to randomness.

Matched Participants

Like between subjects, but not leave it to randomness. That might be challenging.



Usability Experiment - Example

- You have two course enrolment interfaces: the Old one and the New one.
- You claim that the New one allows users to enrol courses much faster than the Old one.
- Design an experimental study that could support or refute your claim.
- Specify:
  - Independent Variables
    - Old/new interface
  - Dependent Variables
    - Speed to enrol courses/time
  - Nuisance Variables / Confounding Factors
    - Internet speed
  - Experimental Conditions (what participants actually do)
    - Between subject: group 1 uses new, group 2 uses old
    - Within: group 1 uses new then old, group to uses old then new
    - Matched: pre-experiment survey, try to balance participants
- How do you split participants for:
  - Between Subjects?
  - Within Subjects?
  - Matched Participants?

Different, Same, Matched Participant Design

Design	Advantages	Disadvantages
Different (between)	No order effects	Many subjects & individual differences a problem
Same (within)	Few individuals, no individual differences	Counter-balancing needed because of ordering effects
Matched	Same as different participants but individual differences reduced	Cannot be sure of perfect matching on all differences

Usability Testing vs. Research

Usability testing

- Improve products
- Few participants
- Results inform design
- Usually not completely replicable
- Conditions controlled as much as possible
- Procedure planned
- Results reported to developers

Experiments for research

- Discover knowledge
- Many participants
- Results validated statistically
- Must be replicable
- Strongly controlled conditions
- Experimental design
- Scientific reported to scientific community

Field Studies

- Field studies are done in natural settings.
- The aim is to understand what users do naturally and how technology impacts them.
- Field studies can be used in product design to:
  - identify opportunities for new technology;
  - determine design requirements;
  - decide how best to introduce new technology;
  - evaluate technology in use.

Field Studies

- Good for understanding **appropriation**:
  - Understanding how users, integrate and adopt technology to their needs, desires and culture.
- Data collection:
  - Notes, pictures, recordings
  - Video
  - Logging (and often prompting)
- Analysis:
  - Qualitative analysis of various types. (e.g. activity theory)

Key Points

- Testing is a central part of usability testing.
- Usability testing is done in controlled conditions.
- Usability testing is an adapted form of experimentation.
- Experiments aim to test hypotheses by manipulating certain variables while keeping others constant.
- The experimenter controls the independent variable(s) but not the dependent variable(s).
- There are three types of experimental design: different-participants, same- participants, & matched participants.
- Field studies are done in natural environments.
- Typically observation and interviews are used to collect field studies data

RE, UX & Videos

RE and Videos

- “Three of the most important goals of requirements engineering (RE) are to create a clear scope, **shared understanding**, and high specification quality”
- Using videos are a great way to achieve shared understanding

- Problem: software professionals are not directors
- Solution:
  - Outsource video (not in this course)
  - Make it yourself…

Karras, Schneider, 2018

Vision Videos for RE

Terms

- **VisionVideo**
  - Video about a vision; here: about a software-based, future product
    - *Video = format or representation*
    - *Vision = content*

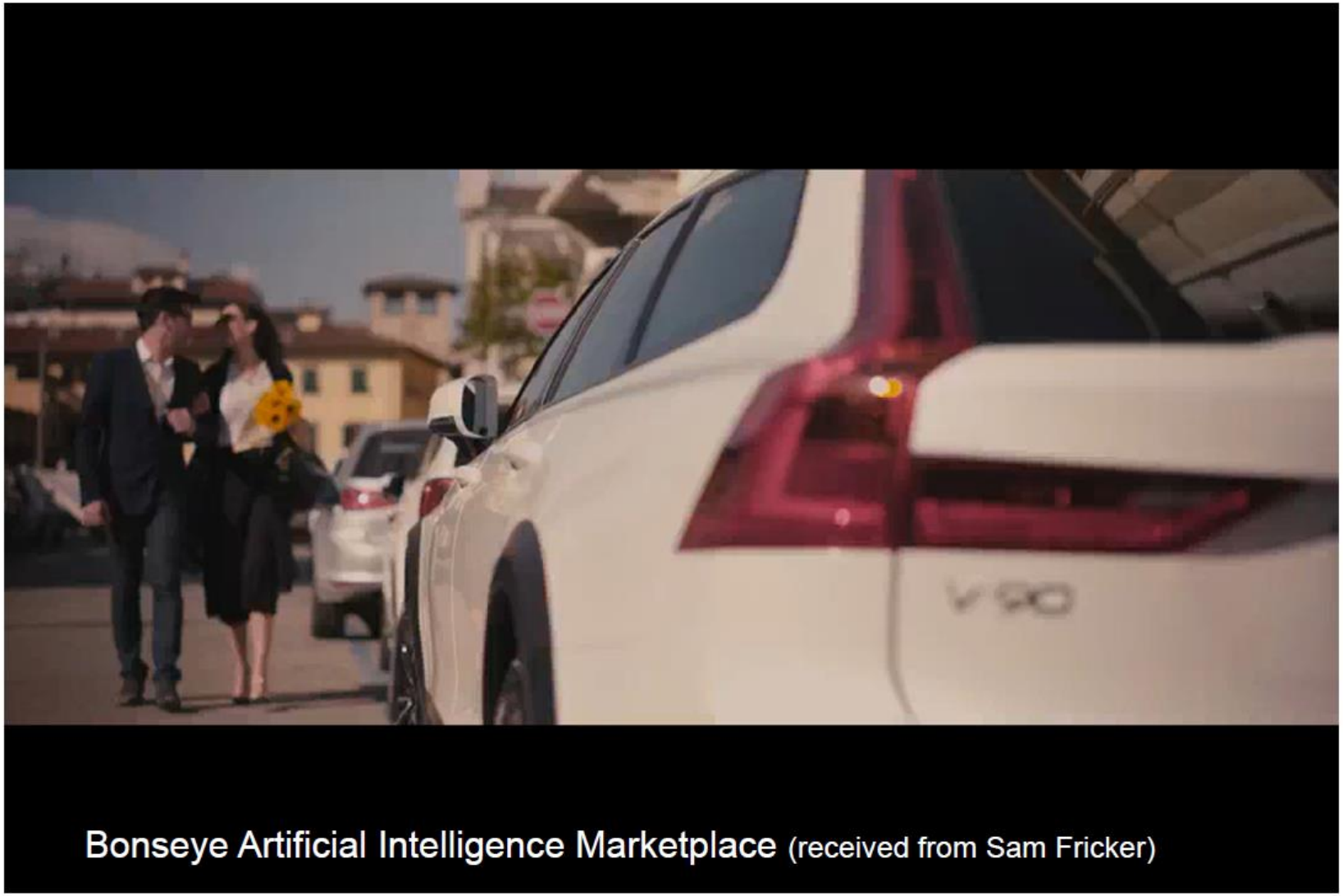
- **Reasons to create Vision Videos (expectations)**
  - Fast, easy, and low-effort to watch – *as opposed to text or models*
  - Short and concrete – *not abstract and vague*
  - Comprehensible *without training, introduction or learning a notation*

… but

  - it takes time to design and to create
  - it requires creativity to do right
  - many requirements/software people are hesitant to use

- **An example Vision Video can tell more than text …**

An Example Vision Video



https://vimeo.com/234685357

Another (Boring) Vision Video

https://www.youtube.com/watch?v=u5V_VzRrSBI&feature=youtu.be

Making Videos

- “According to Brill et al. [19] and Broll et al. [23], software professionals can create affordable videos with sufficient quality for purposes in RE”
- “Nevertheless, videos produced by software professionals have a lower quality since the work is done by amateurs.”
- But (according to Karras and Scheider):
  - 1) Lower Quality Videos are Not a Problem
  - 2) There is No Need for Better Equipment
  - 3) We Need to Know How to Visually Communicate

Making Videos

- “A comprehensive specification and evaluation of the quality of a video requires the definition of the necessary and desired quality characteristics associated with the producers’ and viewers’ goals and objectives for a video”
- Who is the audience?
- Why are you making the video?
  - Not: because she told me to

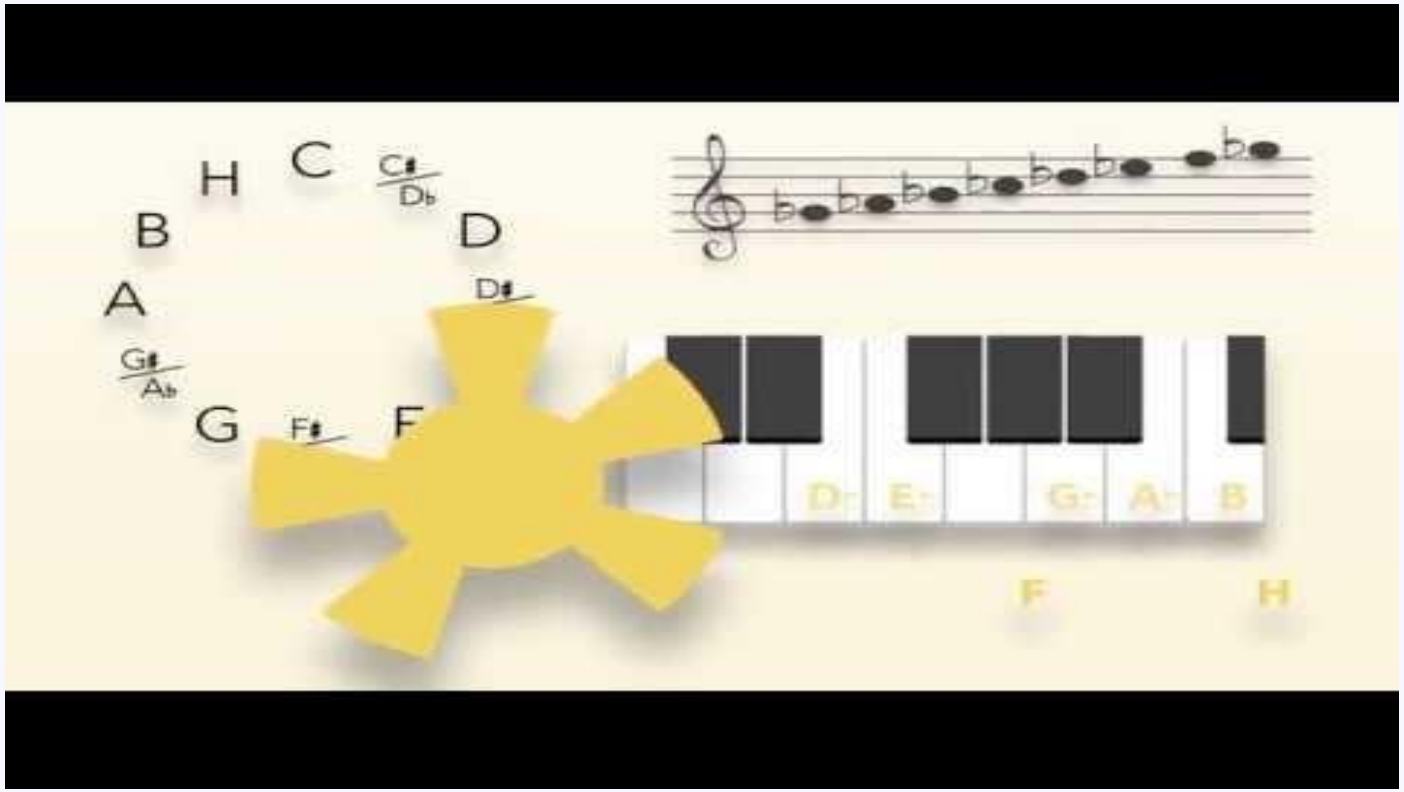
Video Production Guide

- Useful, but far more info than you need
- In canvas
- http://web.mit.edu/techtv/videoprodguide/videoprodguide.pdf
- In your case you can take screenshots
- Can also do some live filming if you like
- Or screenshots of images/drawings
- Consider:
  - Storyboard
  - Timing
  - Narration, subtitles?
  - Labelling
  - Etc.

UX Videos

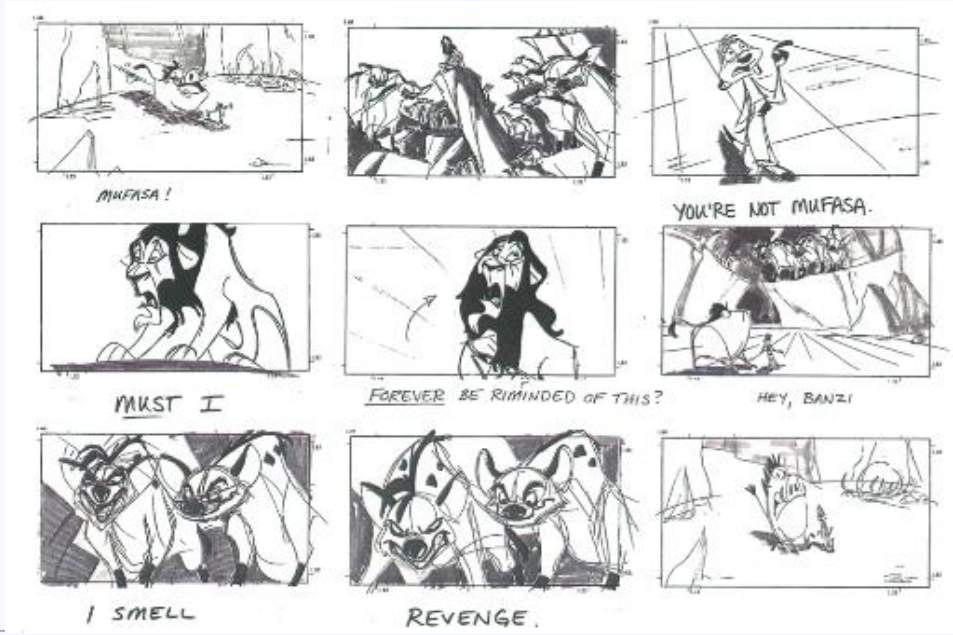
- https://www.youtube.com/user/UxAwards

  Look for “winning projects” here, lots of great examples



Storyboarding

- Used in both RE and UX
- Comes from film industry
- Source: https://uxplanet.org/storyboarding-in-ux-design-b9d2e18e5fab



Storyboarding

- Scenarios should have: characters, scenes, plot
- “All too often you designers jump straight into explaining the details of their design without first explaining the back story. Don’t be one of them—your story must be created with a structure in mind, there should be an obvious beginning, middle, and end. The narrative that unfolds in your storyboard should focus on a goal for the character. Plot should start with a specific trigger and end with either the benefit of the solution, or a problem that the character is left with.”
- Quite similar to customer journey maps

Template

Title _____ Scene _____ Page ____

Storyboarding Software

- Storyboarder
- Make storyboard
- Canva
- Boards (free trial)
- Etc.

Our Past Paper

- Comparing storyboards (for example) for use for elicitation with Children (with SEM Bachelor students!)
- Horkoff, J., Ersare, J., Kahler, J., Jörundsson, T. D., & Hammouda, I. (2018, August). Efficiency and Effectiveness of Requirements Elicitation Techniques for Children. In *2018 IEEE 26th International Requirements Engineering Conference (RE)* (pp. 194-204). IEEE.

Video Creation/Editing Software

- You can use whatever you want
- Make sure the output is easily viewable on all OSes (e.g., mp4)
- I was using Camtasia Studio (which is not free)
- Now Zoom
- Also Powerpoint
- Screen Recording:
  - OBS Studio?
  - Flashback express?
  - Apowersoft Free Online Screen Recorder?
- Video Editing (can be same tool)
  - Blender
  - Lightworks
  - …

Summary

- Basic message:
  - **plan your video before you start making it**

- Videos are common in UX
- Becoming more common in RE
- You may run into them again

Sources

- Karras, Oliver, and Kurt Schneider. "Software Professionals are Not Directors: What Constitutes a Good Video?." *arXiv preprint arXiv:1808.04986* (2018).
- Brill, Olesia, Kurt Schneider, and Eric Knauss. "Videos vs. use cases: Can videos capture more requirements under time pressure?." *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer, Berlin, Heidelberg, 2010.
- What kind of creativity do software engineers need for vision videos?, Prof. Dr. Kurt Schneider CreaRE 2018