The role of user studies in Computer Graphics and Virtual Reality

Carlos Andújar
Universitat Politècnica de Catalunya

Another kind of experimentation that it is also important in CS is the experimentation with **human subjects**.

As we shall see, conducting experiments involving humans such as user studies is quite challenging for a number of reasons.

In this talk, I'll focus on experimentation in my own research fields, which are CG and VR.

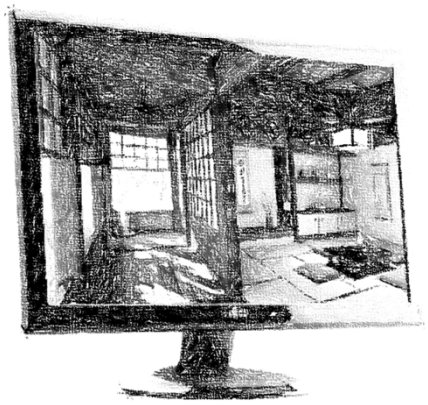
User-centric fields

- Computer Graphics
- Virtual Reality
- 3D User Interfaces

We can find good examples from these areas because they are user-centric.

User-centric fields

- Computer Graphics
- Virtual Reality
- 3D User Interfaces



The major goal of CG is to **render images of 3D models**, images that at the end will be seen by users. So an important topic in CG, beyond the technology and the algorithms, is how do users perceive these images.

User-centric fields

- Computer Graphics
- Virtual Reality
- 3D User Interfaces

VR goes a little bit further and combines multiple techniques (such as stereoscopic vision, and head tracking) to immerse users within the VE. So in VR, user's perception and interaction with the VE is a central topic.

User-centric fields

- Computer Graphics
- Virtual Reality
- 3D User Interfaces

And 3D UI is an emerging field which studies the interaction of humans with VEs, in particular when the interaction tries to benefit from the many DoFs of the human body. We are able to control many joints simultaneously and some apps might benefit from this.

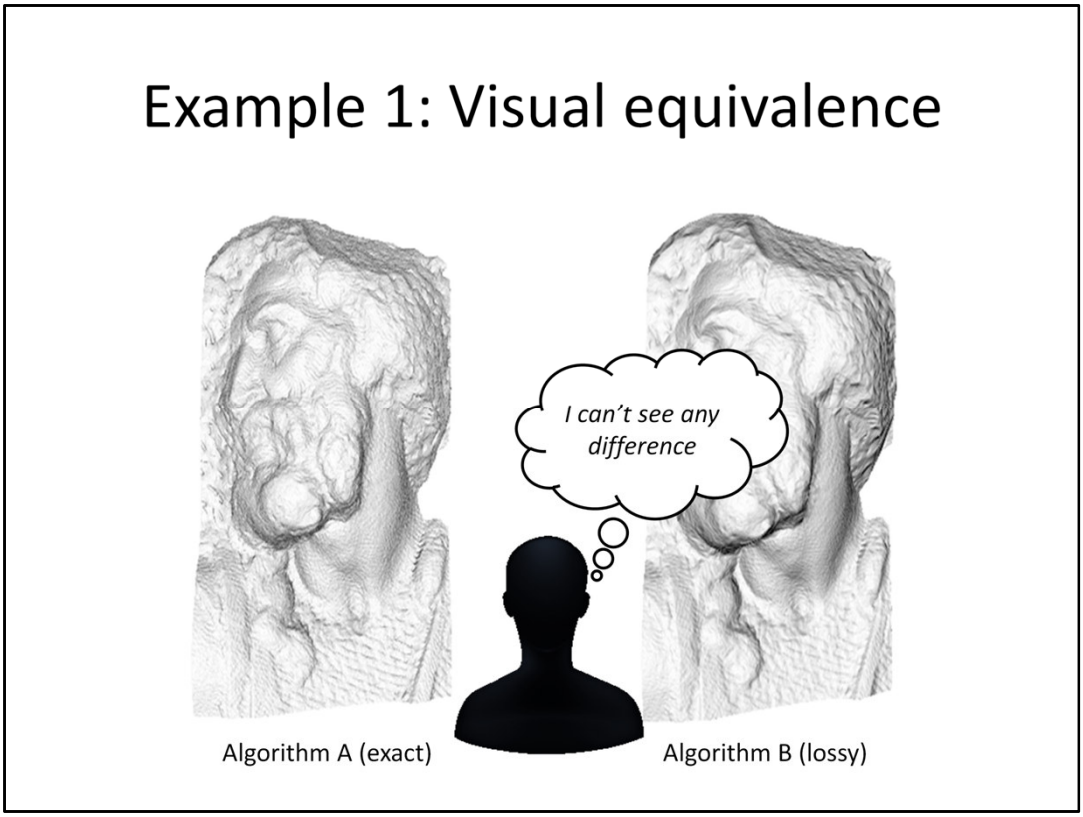
As you might guess, research in all these fields requires researchers to conduct **user studies** to evaluate their work.

Outline

- Examples from CG and VR
- Challenges in user studies for CG and VR
- Do we have enough background on experimentation?
- Conclusions

Examples from Computer Graphics and Virtual Reality

Let's start with some examples from my own research fields.



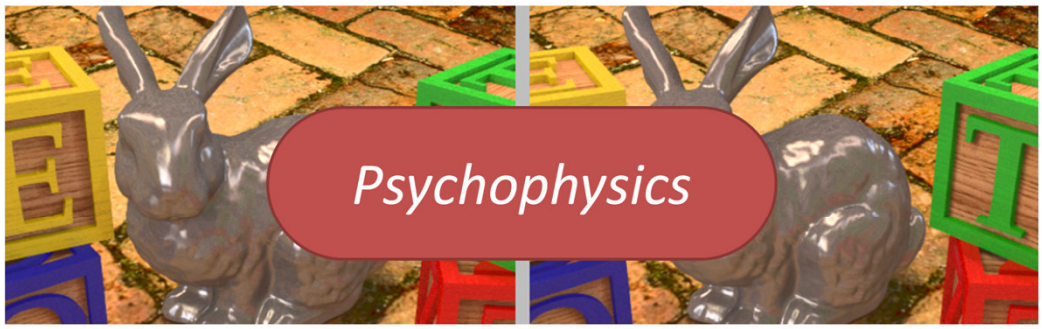
The first example is from CG and is **visual equivalence problem.**

These two images look pretty much the same, but the image on the right has been rendered with an algorithm which is **much faster** but **less accurate**.

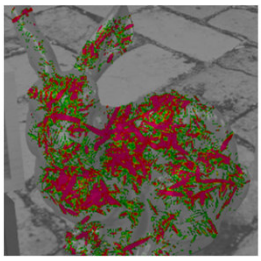
So the important question here is: under which conditions will users perceive these two images as the same?

And the answer to this question requires a user study.

Example 1: Visual equivalence



Visible Difference Predictor →



Visual Equivalence: Towards a New Standard for Image Fidelity
Ganesh Ramanarayanan, James Ferwerda, Bruce Walter and Kavita Bala
ACM Transactions on Graphics, Volume 26, Number 3 (Proceedings of SIGGRAPH 2007)

There are many metrics for comparing pairs of images, and the VDP is one of the best.

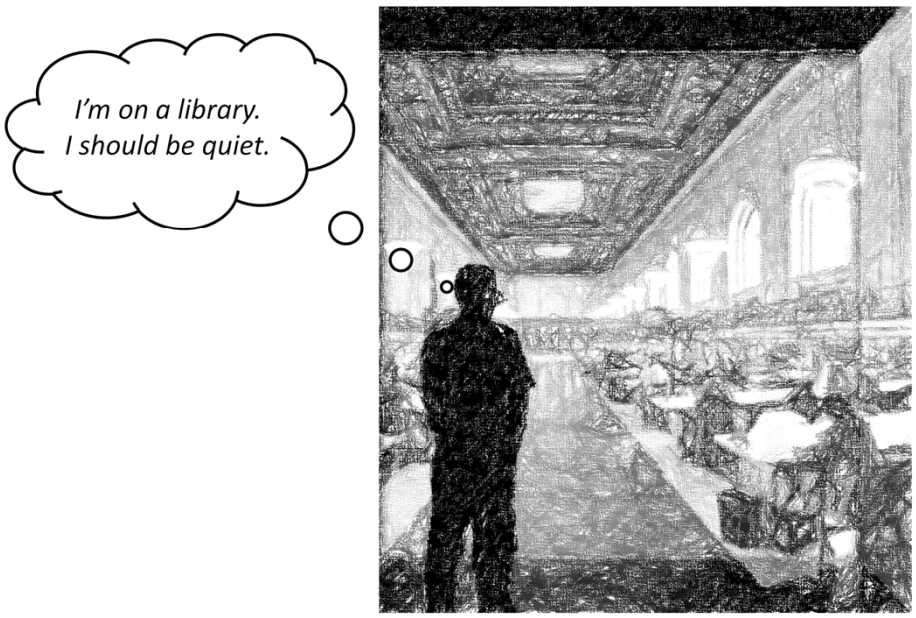
However, according to this predictor, these two images should be perceived as completely different.

However, the HVS is quite tolerant to errors in the reflections of glossy objects.

This is why you can hardly see any difference.

If you want to answer this question, it is not enough to have a strong background on statistics, you also need to know about psychophysics, which is the area of psychology that studies how we perceive different stimuli.

Example 2: Presence evaluation



The second example is the evaluation of presence in VR systems, that is, to which extent users feel and behave as if physically present in the virtual world.

Users of IVR systems might forget about the real environment and the VE (the library in this case) can become the dominant reality.

Example 2: Presence evaluation

Burn patient using VR while getting wound care (HITLab)

Virtual Reality Pain Reduction. University of Washington Seattle and U.W. Harborview Burn Center.
http://www.hitl.washington.edu/projects/vrpain/

The evaluation of presence is very important in many VR applications, for example, for **pain relief**.

This burn patient has serious skin injuries.

While getting care, the patient is wearing this VR display that shows this kind of frozen VE, so that he gets distracted.

VR systems has been shown to be effective for pain relief.

Example 2: Presence evaluation

Behavioral and physiological response of the user

Physiological Measures of Presence in Virtual Environments
Michael Meehan, Brent Insko, Mary Whitton, Dr. Frederick P. Brooks, Jr.

These images here are from the virtual pit room experiment, which is very well know in our community.

Users had to wear a HMD showing this environment where one room had a large pit in it.

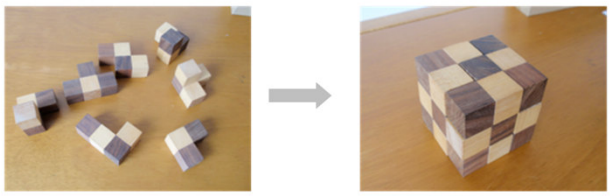
The task was to take one object to the chair.

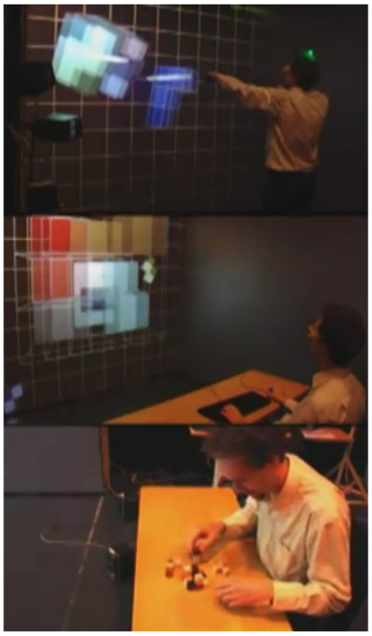
One of the most surprising results was that many participants didn't take a straight path, but they tried avoid the pit, despite they knew there was no real pit.
They also found a significant heart rate increase when in the pit room.

So a typical way of evaluating presence, beyond questionnaires, is by observing users' behavior and measuring their physiological response.

So that's a second example where user studies are required.

Example 3: Comparing 3D UIs





"Reality-based User Interface System (RUIS)".
RUIS homepage:
http://ruisystem.net
Department of Media Technology, Aalto University:
http://media.tkk.fi/en/

The last example is about the comparison of 3D UIs in terms of their usability.

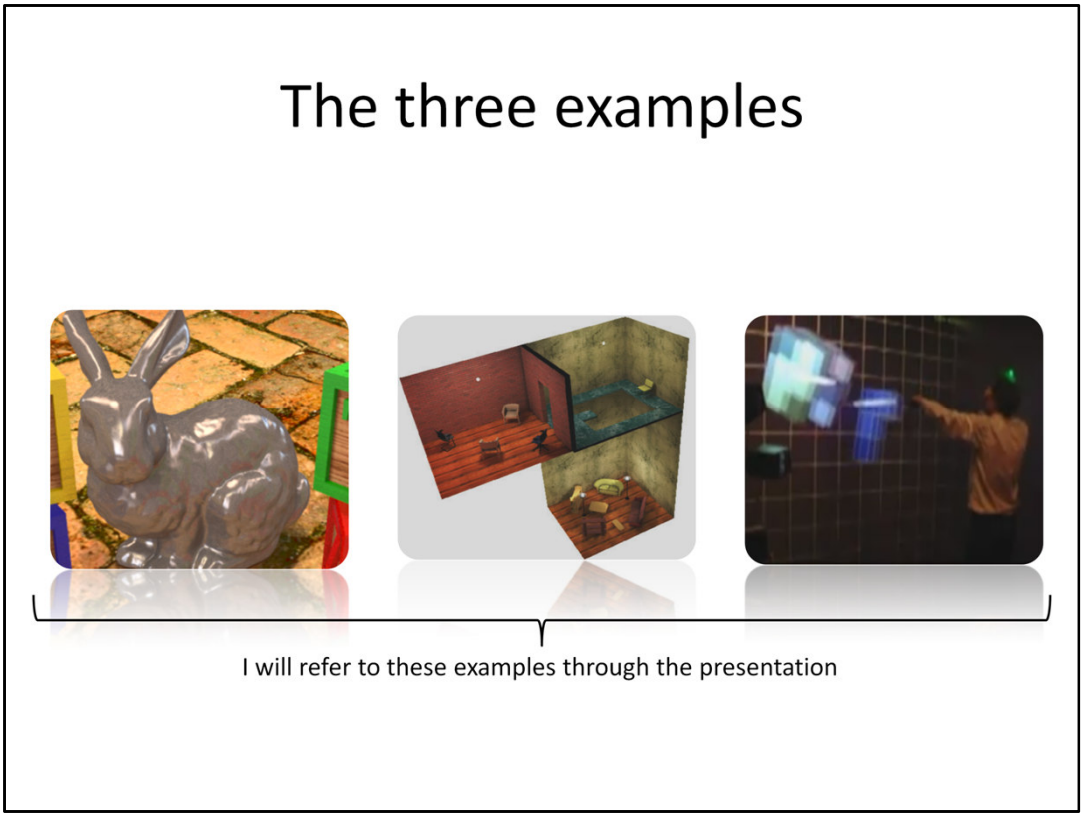
Imagine you want to solve this 3D puzzle.

You can solve the puzzle in multiple ways, with a real puzzle, using keyboard and mouse, or using a Wii controller, as in this video.

For this task, users must be able to select the pieces, manipulate them, and explore the model from different viewpoints.

So the important question here is: which UI is better in terms of usability.

The answer to this question requires again a user study.



We've seen three examples where the evaluation of the results **requires conducting user studies**.

I will refer to these examples through the rest of my presentation.

Challenges in user studies for CG and VR

The empirical method

- Formulate a hypothesis
- Make the hypothesis testable
- Design an experiment
- Get approval by ethics committee
- Recruit participants
- Conduct the experiment and collect data
- Pay participants
- Analyze the data
- Accept or refute the hypothesis
- Explain the results
- If worthy, communicate your findings

This slide shows the typical steps of the empirical method.

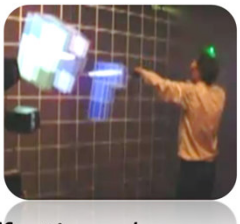
So I'll use these steps to guide this part of the presentation.

The empirical method

- Formulate a hypothesis
- Make the hypothesis testable
- Design an experiment
- Get approval by ethics committee
- Recruit participants
- Conduct the experiment and collect data
- Pay participants
- Analyze the data
- Accept or refute the hypothesis
- Explain the results
- If worthy, communicate your findings

First, you need to formulate a hypothesis, and make it testable.

Hypothesis formulation



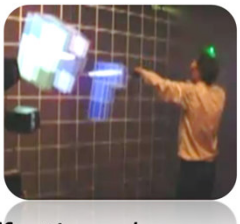
General hypothesis

Using the Wii controller will make people more effective when doing manipulation tasks.

Testable hypothesis

We measured the time it takes for users to solve a particular 3D puzzle, using either Wii or mouse. We hypothesize users will be faster using the Wii.

Hypothesis formulation



General hypothesis

Using the Wii controller will make people more effective when doing manipulation tasks.

Testable hypothesis

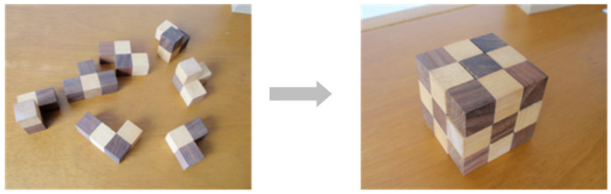
We measured the time it takes for users to solve a particular 3D puzzle, using either Wii or mouse. We hypothesize users will be faster using the Wii.

The problem is that you had to choose a particular task, which you take as representative.

Choosing a representative task

There exists such a representative task?

NO



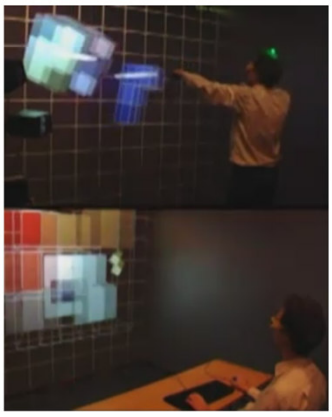
So the question is: there always exists such a representative task?

Unfortunately this is not the case; because manipulation tasks include almost everything humans can do, from cooking a dish to repairing a car.

The task space is so large and heterogeneous that you cannot find a small set of representative tasks.

Independent variables

- **Interaction technique:** {Wii, mouse}
- Stereo viewing: {stereo, mono}
- Graphics quality: {wire, shaded, textured}



Continuing with the UI example, the most important variable is the interaction technique, with at least two conditions, Wii and mouse.

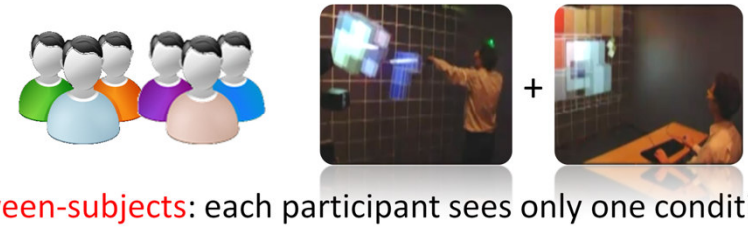
But there are also other variables that could certainly affect the experiment, such as Stereo viewing and CG quality, just to name a few.

And the more variables you control, the more difficult the experiment will be.

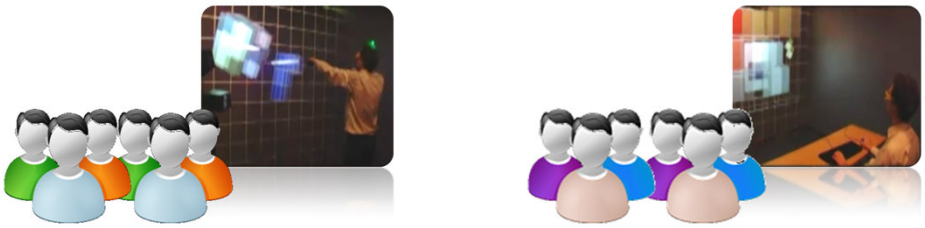
Experiment design

Independent variables can vary in two ways:

- Within-subjects: each participant sees all conditions



- Between-subjects: each participant sees only one condition



Experiment design

Independent variables can vary in two ways:

- Within-subjects: each participant sees all conditions
  - More time consuming for the participants
  - Need to counterbalance for **learning effects** and **fatigue effects**

- Between-subjects: each participant sees only one condition
  - Need to recruit more participants
  - Less power (less chances to proof your hypothesis)

The decision is always **controversial**.

Experiment design

Independent variables can vary in two ways:

- Within-subjects: each participant sees all conditions
  - More time consuming for the participants
  - Need to counterbalance for **learning effects** and **fatigue effects**

- Between-subjects: each participant sees only one condition
  - Need to recruit more participants
  - Less power (less chances to proof your hypothesis)

The empirical method

- Formulate a hypothesis
- Make the hypothesis testable
- Design an experiment
- Get approval by ethics committee
- Recruit participants
- Conduct the experiment and collect data
- Pay participants
- Analyze the data
- Accept or refute the hypothesis
- Explain the results
- If worthy, communicate your findings

You need to get the approval by the ethics committee, recruit participants, and get their informed consent.

Some ethical issues in user studies.

Another issue is frustration handling. Some users might not be able to complete the task, specially with novel VR equipment; it is important to tell users that you are testing the technology, not the participants' abilities or education.

The empirical method

- Formulate a hypothesis
- Make the hypothesis testable
- Design an experiment
- Get approval by ethics committee
- Recruit participants
- Run the experiment and collect data
- Pay participants
- Analyze the data
- Accept or refute the hypothesis
- Explain the results
- If worthy, communicate your findings

The next step is to run the experiment and collect data.

Dependent variables

- Time
- Errors
- Preference score

- Path followed by users
- Heart rate
- Galvanic skin response (sweat)

!

Data collection for VR experiments is often challenging.

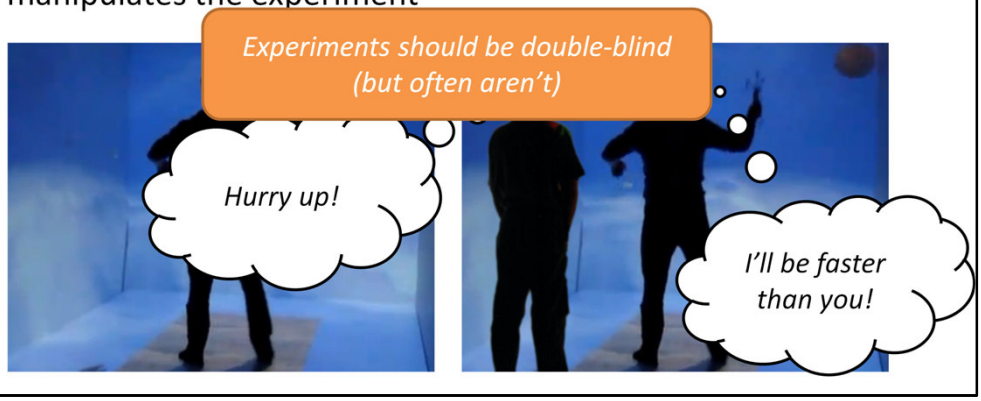
Some experiments require you to measure easy variables such as time to complete the task and error counts,

But other experiments require hard-to-collect data.

This is another difficulty.

Experimenter issues

- Placebo effect
- Hawthorne effect: increased attention from superiors or colleagues increases worker performance
- Observer-expectancy effect: the researcher unconsciously manipulates the experiment



And you have to be careful about **experimenter issues**.

The placebo effect is very well know,

The Hawthorne effect occurs when …

The performance of this participant might change is somebody else, e.g. the previous participant, is observing.

And if the observer is the researcher, who has a conflict of interest, then you can get the observer-expectancy effect: the researcher might manipulate the experiment unconsciously, using eg body language.

The empirical method

- Formulate a hypothesis
- Make the hypothesis testable
- Design an experiment
- Get approval by ethics committee
- Recruit participants
- Conduct the experiment and collect data
- Pay participants
- Analyze the data
- Accept or refute the hypothesis
- Explain the results
- If worthy, communicate your findings

The empirical method

- Formulate a hypothesis
- Make the hypothesis testable
- Design an experiment
- Get approval by ethics committee
- Recruit participants
- Conduct the experiment and collect data
- Pay participants
- Analyze the data
- Accept or refute the hypothesis
- Explain the results
- If worthy, communicate your findings

Do we have enough background on experimentation?

Do we (as CS engineers) have enough background **to conduct the user studies required to evaluate our own research**?

Do we have enough background?

Personal answer: **NO**

Arguments:

- Too many submitted papers with serious evaluation errors
- Too many **published papers** with serious evaluation errors!

And my personal opinion is that this is not the case.

Common errors

- No user evaluation
- User study but no analysis
- Wrong analysis
- Wrong interpretation of the results

Conclusions

Conclusions

- Shift of focus
  Implementing technology → using technology
- Validation through user studies is critical
- Major challenges
  - Lack of background (e.g. experimentation, psychophysics)
  - User studies are really time consuming
  - Difficult to fulfill all requirements (double-blind experiments, informed consent, representative users)
  - Difficult to find representative datasets/models/tasks
  - Will **executable papers** help?

Bibliography

J. E. Swan, Stephen Ellis, Bernard Adelstein: *Conducting Human-Subject Experiments with Virtual and Augmented Reality*. IEEE Virtual Reality 2007

DC Howell: *Statistical Methods for Psychology*, 5th ed. Pacific Grove, CA, 2002.

GA Gescheider (1997): *Psychophysics: The Fundamentals*. Lawrence Erlbaum As.

ETSI EG 201472 (2000): *ETSI Guide: Human Factors*; European Telecommunications Standards Institute

American Psychological Association (1992a): *Ethical principles of psychologist and code of conduct*. American Psychologist, 47, 1957-1611.