EPFL

24.04.19
Virtual Reality

Experimental Design and experimentation in VR

Dr Bruno Herbelin

Laboratory of Cognitive Neuroscience
Center for Neuroprosthetics
Brain & Mind Institute
EPFL

Introduction

- People will trust what you say if you can prove it !
- Professional & scientific communities want evidences
  - Facts & observations
  - Measures
- Difficulties
  - How to prove that my method is better?
  - How to evaluate the change in performance / success?

- Pitfalls
  - Biased measurements
    - The new algorithm is faster (*on a new computer*) than the old algorithm *(on an old computer)*
  - Biased subjects
    - “I tested it myself !”

Experimental design principles

- Causal research
  - Experiment the causal relationship between one or more variables on one or more outcome variables
    - Variable; factor that can change and is measurable
  - **Causal relationship**
    - When one variable depends on the change of another

    e.g. Fever reduction depends on Paracetamol intake

`If` **`X`**`, then` **`Y`**

AND

`If` **`NOT X`**`, then` **`NOT Y`**

Experimental design principles

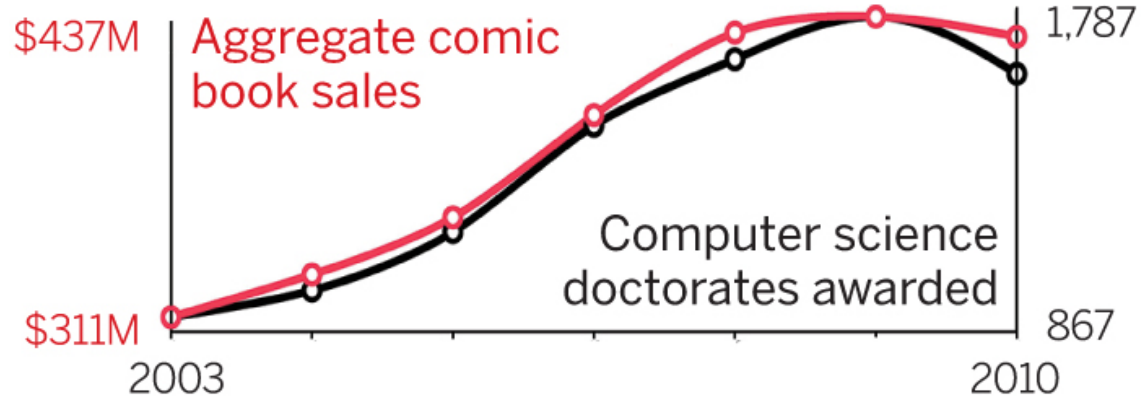
- Causal research
  - Experiment the causal relationship between one or more variables on one or more outcome variables
    - Variable; factor that can change and is measurable
  - **Causal relationship**
    - When one variable depends on the change of another

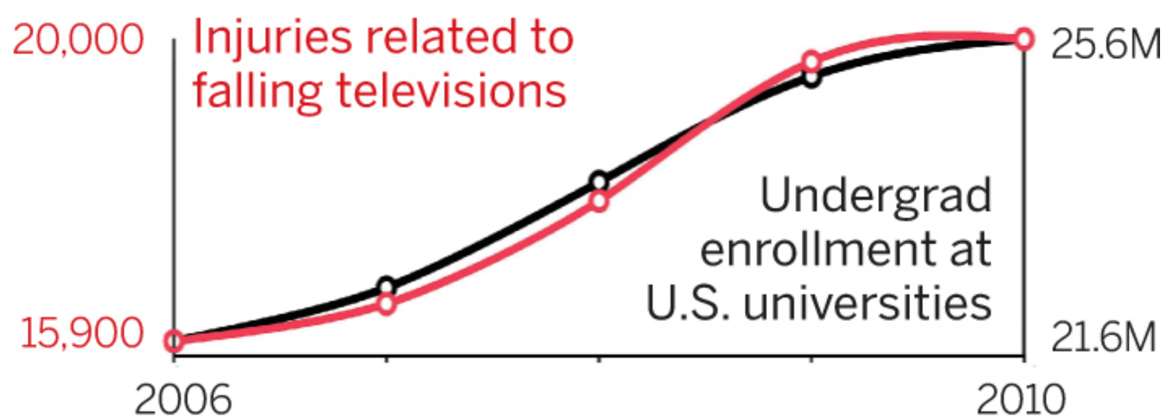
    e.g. Fever reduction depends on Paracetamol intake

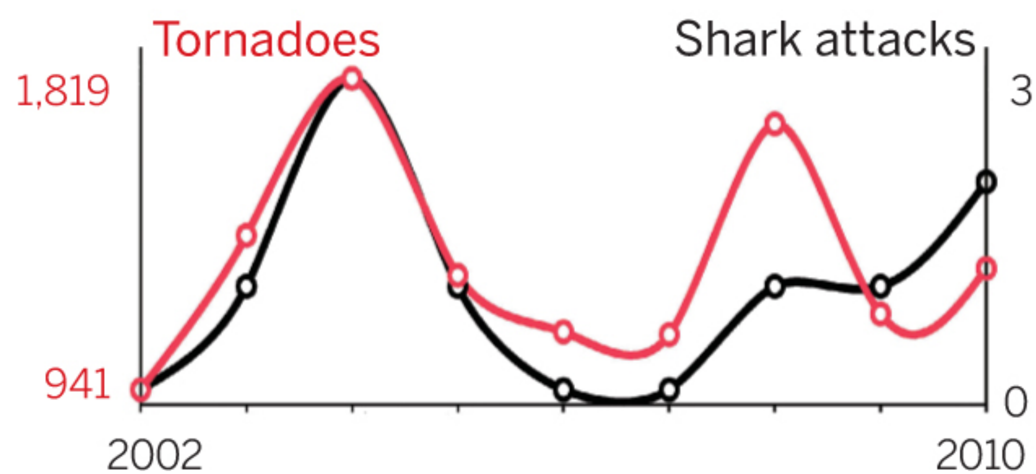
NB:

  - *Correlational relationship (non-experimental design)*
    - Measure of a correlation between two variables
      - Positive or negative (e.g. +/- 1.0 linear factor)
      - Non correlated (0)
    - Correlation **does not equal** causation
      - Unless your experiment uses control condition, you cannot infer the causation from a correlation

Spurious Correlations
Science 29 May 2015:
Vol. 348, Issue 6238, pp. 980







Experimental design principles

- Principle of **Randomization**
  - Protect experiment from factors not under consideration
  - Variations caused by extraneous factors are considered as noise (chance probability)

- Principle of **Replication**
  - Experiment should be done with more than 1 subject
  - Accuracy of mean effect increases with repetitions

- Principle of local **Control**
  - Measure (and suppress) variability of extraneous factors by deliberately testing them
  - Group or block design

Experimental design principles

- True experimental design obey the three principles
  - Random selection and assignment of subjects
    - **Selection** = how to find subjects
    - **Assignment** = how to use selected subjects
  - Repeat the experiment with many subjects
  - Control conditions or groups (e.g. Placebo group)

- If not, there are other approaches
  - Quasi-experimental design
    - No randomization, but with control
  - Non-experimental design
    - No randomization, no control
      - Survey, correlational studies

Experimental design principles

- **Independent** variables (IV)
  - Factors manipulated by the experimenter
- **Dependent** variables (DV)
  - Responses measured by the experimenter

- Experimental and control groups
  - Participants are randomly assigned in two groups
    - **Experimental group**; with manipulation
    - **Control group**; without any manipulation
  - This is because a truly causal relationship is bidirectional:

`If` **`X`**`, then` **`Y`**

AND

`If` **`NOT X`**`, then` **`NOT Y`**

Experimental design principles

- Experimental **hypothesis**
  - Statement predicting a cause and its effect(s)
  - Typically, how an independent variable will affect a dependent variable
    - e.g. “Increase of treatment will improve recovery”
- Null hypothesis
  - The hypothesis that the **experimental manipulation will have no impact on dependent variables**
    - e.g. “Increase of treatment does not improve recovery”
    - Does not mean there was no effect on the participant, but that the experiment does not measure it
- *An experiment is performed to validate an hypothesis*

i.e. prove that the null hypothesis is wrong

Experimental design principles

- Timing
  - Cross-sectional research
    - Experiment performed on a subject at one occasion
    - Focus on presently manipulated conditions
  - Longitudinal research
    - Data gathered repeatedly throughout the length of the study
    - Covers long periods, from few days to several years

- Pre/post testing
  - **Pre-test**; measure dependent variable before manipulation
    - Not always applicable or relevant
  - **Post-test**; measure dependent variable after manipulation

Between subjects vs. Within subjects

- Between subject design
  - Subjects are divided into groups
    - Each group has only 1 condition of a variable
  - Results are compared **between groups**
  - Eliminates variations due to ordering of conditions

- Within subjects design
  - Subjects perform all conditions of a variable
  - Results are compared **between conditions** for all subjects
  - Eliminate variations due to user differences

Notations

- Observations or measures ; O
- Treatment or programs ; X
  - e.g. intervention, manipulation, etc.
- Assignment to groups
  - Random assignment; R
  - Non-equivalent groups; N
  - Cutoff assignment C

	pre-test	program	post-test
R	O	X	O
R	O		O

Time

Random Assignment in 2 groups

Two-groups experimental design

R	X	O
R		O

- Experimental and control groups
  - Subjects randomly assigned
- Post test only
  - Assume two groups are equivalent
  - Determine if two groups are different after the program
- Capable of assessing cause-effect relationships of one independent variable
  - Yes, it follows the three principles!

The simplest and a very often used research design method

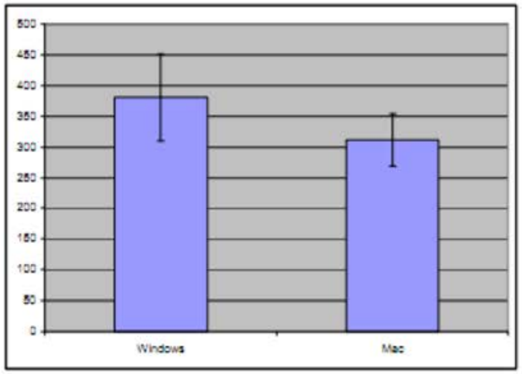
Analysis of results

- Compare the means of two groups
  - Compute the mean (or median) of DVs per group
    - They are equal; null hypothesis validated
    - One is higher than the other; the null hypothesis invalidated
  - Problem; how much variability in the data?
    - e.g. measures vary from -100 to 100, means differ by 2.0; does this really mean something?

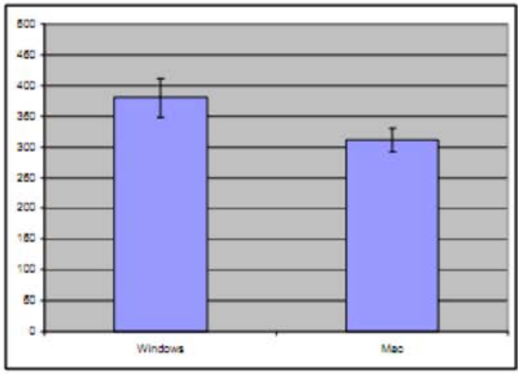
- Standard deviation
  - Measure the dispersion of data
  - = square root of Variance
- Standard error of the mean
  - Confidence interval for the computed mean
  - = standard deviation / number of measurements

Analysis of results

- Standard error of the mean (example)



4 users for each condition:
Error bars overlap, so can't conclude anything



16 users for each condition:
Error bars are disjoint, so Windows is significantly different from Mac

- Student's t-test
  - Tests if the mean of two groups are identical (null hypothesis)
  - **p value** = probability that the difference is chance
    - e.g. $p < 0.05 \rightarrow 95\%$ confidence that it is not chance!

Example in VR

- **Putting yourself in the skin of a black avatar reduces implicit racial bias**

  Tabitha C. Peck, Sofia Seinfeld, Salvatore M. Aglioti, Mel Slater. Consciousness and Cognition Volume 22, Issue 3, September 2013, Pages 779–787.

Example in VR

- **Putting yourself in the skin of a black avatar reduces implicit racial bias**
  - IV 1 : skin color of avatar in VR
    - DV: Racial Implicit Association Test

Group	Pre IAT	Skin change	Post IAT
R	O		O
R	O	X	O

  - IV 2 : synchrony of motion of avatar
    - DV: Questionnaire on body ownership

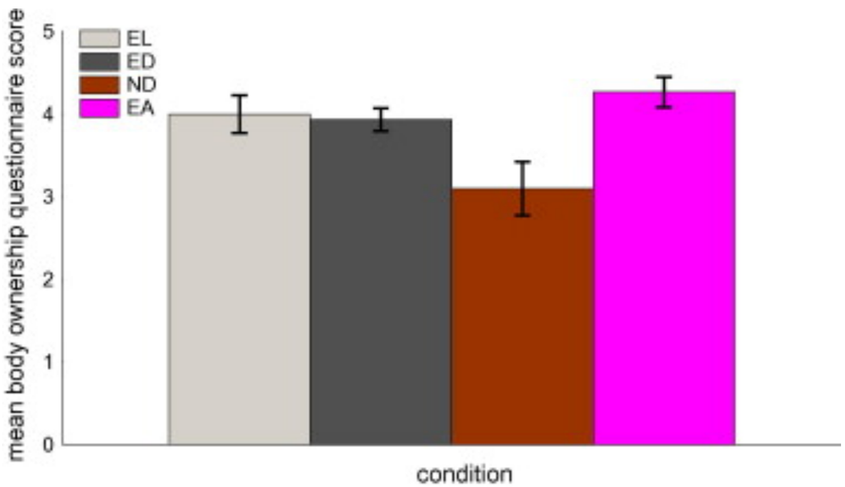
Group	Sync	Post Q
R	X	O
R		O

Experiment

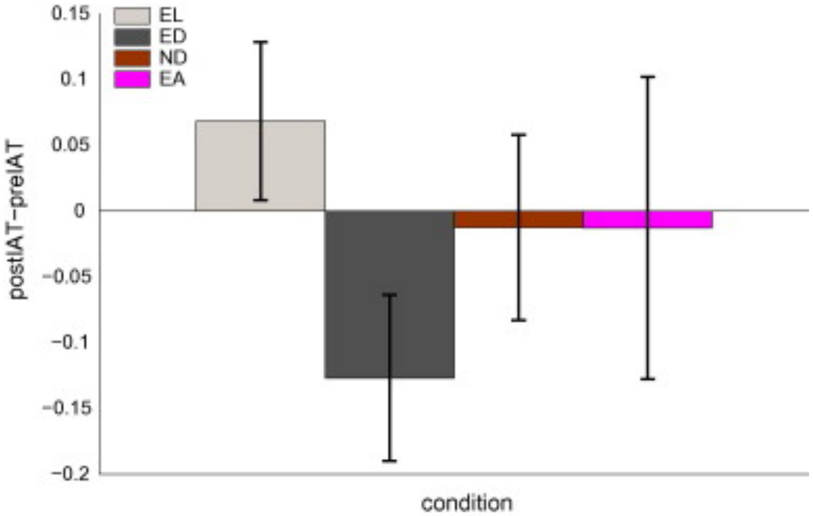
- 60 subjects light skin
- 4 Groups
  - Embodied-Light-Skinned (EL)
  - Embodied-Dark-Skinned (ED)
  - Non-Embodied Dark-Skinned (ND)
  - Embodied-Alien-Skinned (EA)

Group	Pre IAT	Program	Post IAT	Q
R	O	EL	O	O
R	O	ED	O	O
R	O	ND	O	O
R	O	EA	O	O

Results



Bart chart showing means and standard errors of average embodiment questionnaire score by condition.
The average embodiment score is obtained as (Q1 + 6 − Q4)/2.
(Q1:I felt as if the body I saw in the virtual world might be my body.
Q4 : I felt like the avatar was not me.)



Bar chart showing means and standard errors of Δ *IAT* by condition.

Null-Hypothesis “Synchrony of movement has no effect on body ownership” proven wrong.

Null-Hypothesis “Embodiment in a body of opposite color does not impact racial bias” proven wrong.

More detail in the paper : http://www.sciencedirect.com/science/article/pii/S1053810013000597

Factorial design

- Extend 2-groups design for multiple variables
  - Each variable shall be tested with two groups, so two variables can be tested with four, three with six, etc..
  - Each independent variable is called a **factor**

R	X_{11}	O
R	X_{12}	O
R	X_{21}	O
R	X_{22}	O

- Consider two or more levels per factors
  - Variation of the variable
    - e.g. with / without manipulation + intermediate levels
- Mix the factors and levels
  - Subjects are randomly assigned to one level of one factor

Factorial design

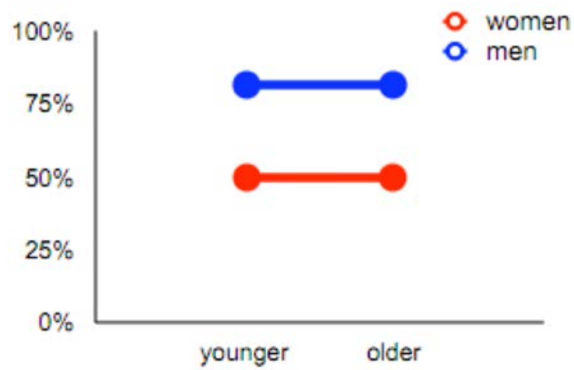
- Factorial notations
  - e.g. 2 x 3 (two by three) design
    - Two factors
      - First factor has two levels
      - Second factor has three levels
    - Total of 2x3=6 groups

- Mixed Factorial design
  - Between group measure on one independent variable
  - Repeated measure on the other

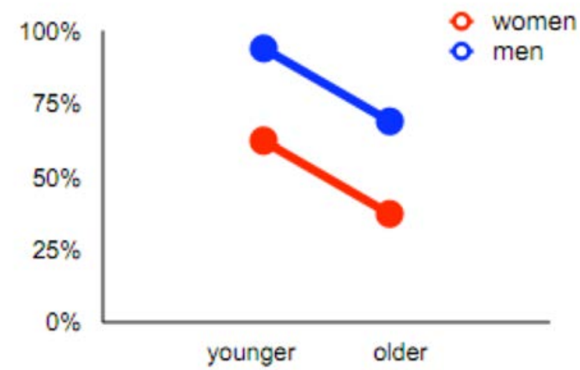
Analysis of results (examples)

2×2 Factorial Design

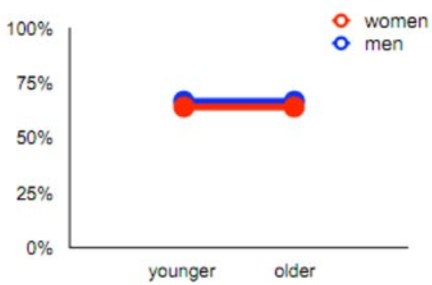
Main Effect of Sex



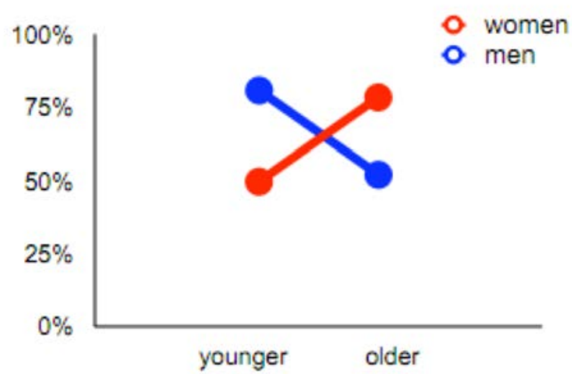
Main Effect of Age



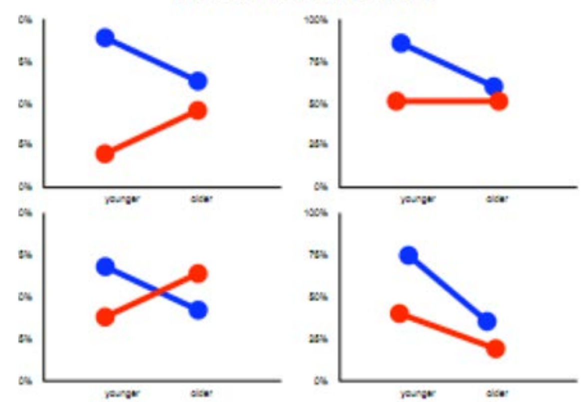
No Effects



Interaction between Sex and Age

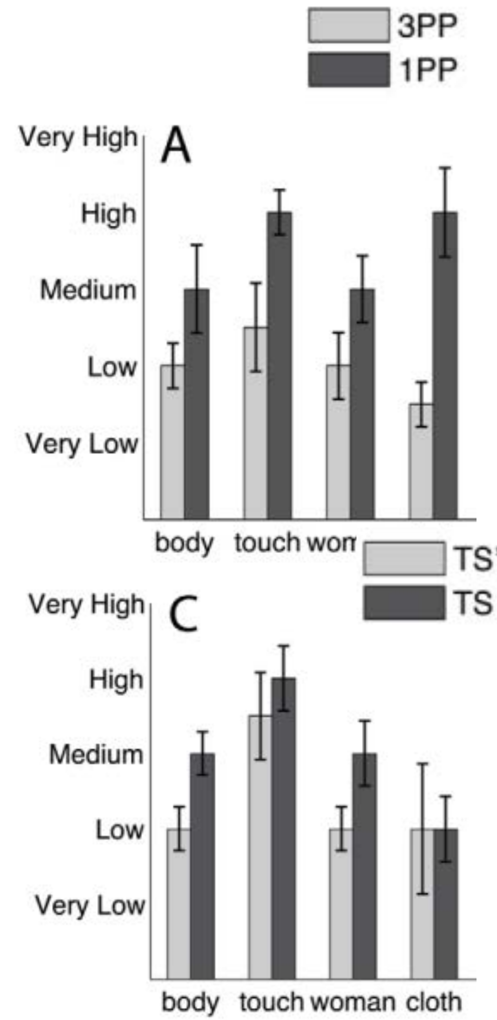


Interactions



Example in VR

- **First person experience of body transfer in virtual reality**
  - Slater, Spanlang, Sanchez-Vives & Blanke , PloS one, 2010.



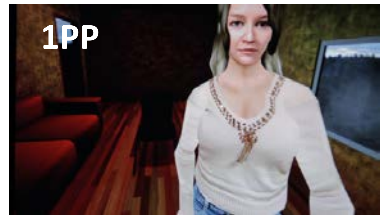
First-person experience of body transfert in VR

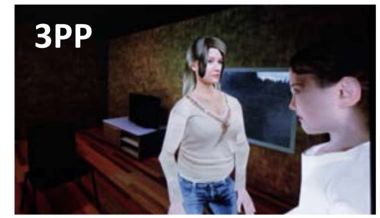
EXPERIMENTAL PROTOCOL for EEG RECORDING

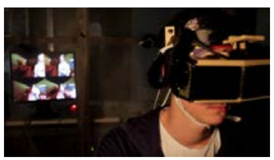
VERE-WP1 / LNCO & EventLab / Nov. 2010

Experiment : factorial design

X 2 sides
(left & right)

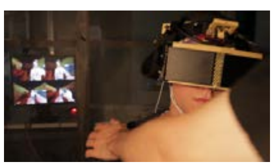




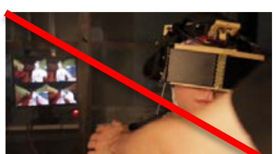


No stroking

2x control conditions



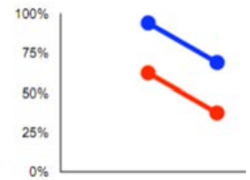
Synchronous



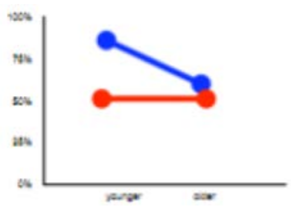
A-Synchronous

Group	Sync	Post Q
R	X 1,1	O
R	X 1,2	O
R	X 2,1	O
R	X 2,2	O

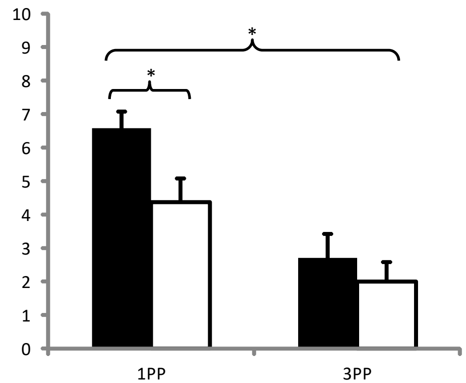
Result



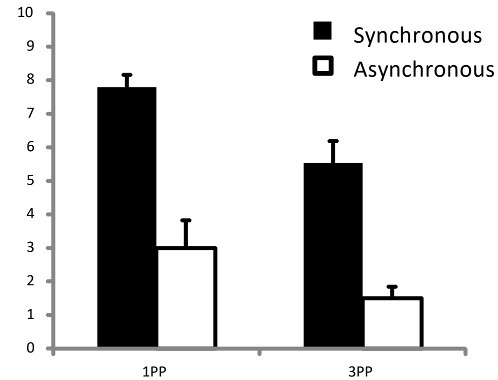
Main effect of synchrony



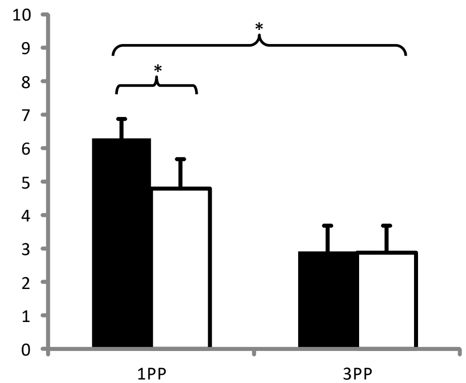
Interaction of the two factors (synchrony and perspective) for the subjective self identification (Q1 and Q3)



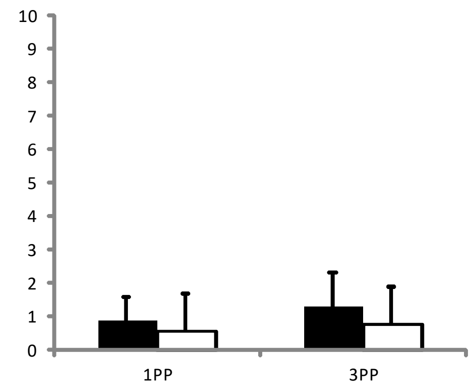
Self identification (Q1); How much did you feel that the seated girl's body was your body?



Illusory touch (Q2); How strong was the feeling that the touch you felt was caused by the woman that you saw?



Self identification (Q3); How strong was the feeling that you were wearing different clothing?



Control (Q4); How strong was the feeling that the color of the room changed during the experiment?

Experimental design for VR

- In VR like in HCI, human factor determines the validity of the system
  - Need to perform valid experimentation on subjects
  - Non-experimental design is not strong enough to conclude on the causality

- Pitfalls
  - User experience and acceptation of technology strongly influence results
    - Can't compare a teenage gamer with elderly
    - People like / dislike high-tech gadgets
  - High habituation and learning during testing
    - Past the surprise of immersion, people change behavior

Experimental design for VR

- Testable hypothesis
  - Modify one feature = create an independent variable
  - Dependent variable = measurement of user

- Techniques for controlling IVs
  - Enable / disable a feature (→groups)
  - Change hardware (→groups)
    - e.g. HMD vs. CAVE
  - Compare with / without training
  - Vary the influence or magnitude of a feature
    - Change values of parameters (→factorial levels)
    - e.g. low, medium, high
  - Vary the difficulty or complexity of the task
    - e.g. show that a factor is better for complex tasks

Experimental design for VR

- Tools for measuring user (DV)
  - Simulation
    - Time (to reach, to accomplish a task, etc.)
    - Distance (from A to B, length of a path, etc.)
  - User
    - Physiological
      - Eye tracking, heart rate, etc.
    - Cognitive
      - Questionnaire scores, scales, etc.

- Pitfalls
  - Do not change instructions! (influence performance)
  - Do not ask directly about the IV

Conclusion

- Experimental design is powerful
  - Proves the causal relationship between variables
  - High scientific value (statistics)
- Experimental design has strong constraints
  - Needs several subjects
  - Every little difference in conditions changes everything
- Many more combinations & possibilities
  - e.g. ANOVA for factorial within subjects
  - e.g. Covariance design with ANCOVA analysis
- In short
  - You test people, not system
    - results indicate what subjects do / think, not if the system is better…