Usabilidad en sistemas de RV
Experimentos

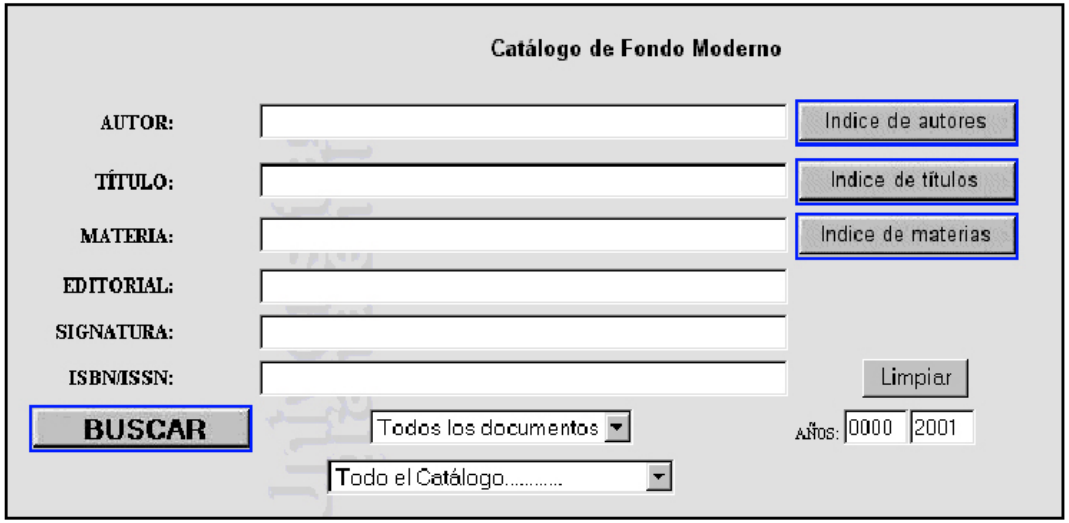
Pere Brunet
Carlos Andújar
Universidad Politécnica de Catalunya

Usabilidad

Velocidad y facilidad con la que las personas pueden realizar sus tareas mediante la interfaz

- Es preciso entender y trabajar con usuarios: actuales y potenciales
- Es preciso conocer y entender el contexto: objetivos y tareas de los usuarios en su aplicación
- Medición del grado de satisfacción, aumento de velocidad y aumento de productividad + calidad
- Quién determina si un producto será usable son los usuarios, NO los diseñadores.

Usabilidad: buenos y malos diseños



www.baddesigns.com

Usabilidad

Métodos para evaluar un prototipo:

- Inspección (por especialistas en usabilidad)
- Indagación: observación de los usuarios, grabación de su comportamiento, grupos de discusión dirigidos.
- Test (medición de prestaciones):
  - Usuarios reales. Selección y tamaño de la muestra. Contextual y uniforme
  - Tareas reales. Diseño del experimento. Hipótesis
  - Observación, medición y registro del comportamiento
  - Análisis estadístico de los datos

Usabilidad y Experimentos

Usabilidad y Experimentos

Human-subject experiments in virtual reality and 3DUI

Autor	Andujar Gran, Carlos A.
Tipus d'activitat	Presentació treball a congrés
Nom de l'edició	Virtual Realities (Dagstuhl Seminar 13241)
Any de l'edició	2013
Data de presentació	2013
Llibre d'actes	Dagstuhl Reports, Volume 3, Issue 6
Pàgina inicial	42
Pàgina final	42
DOI	10.4230/DagRep.3.6.38

Should we really believe the findings of published papers?

If so, are these findings really useful?

Experimentation allows researchers to …

… gain new knowledge

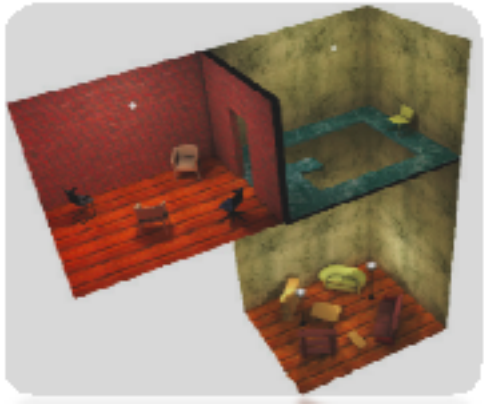
…proof or refute their hypotheses

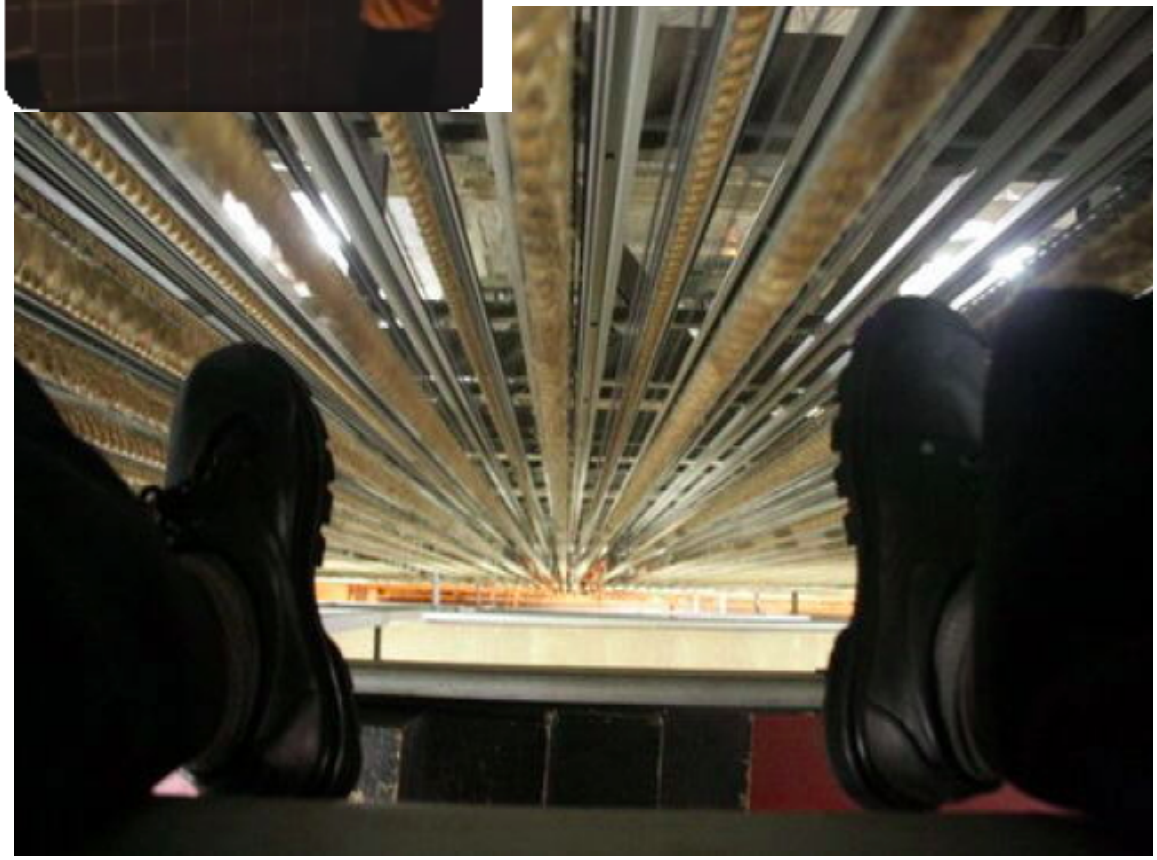
…validate their results

.. show that what they created has some benefit

…proof their work is worth publishing.

Usabilidad y Experimentos en RV





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

Hypothesis formulation

General hypothesis

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

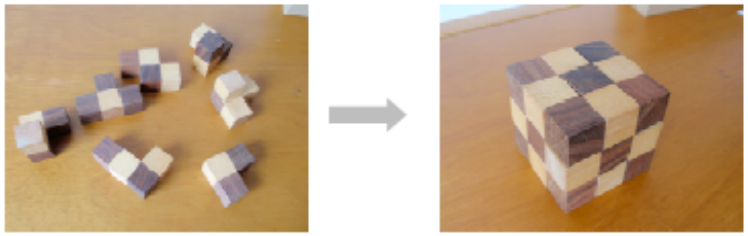
Testable hypothesis

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.

Choosing a representative task

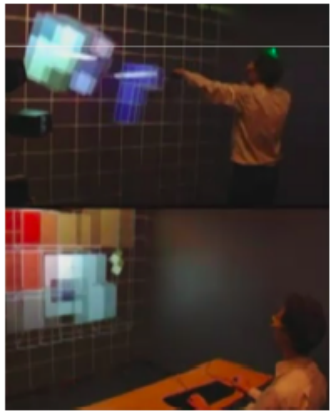
There exists such a representative task?

NO



Independent variables

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



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)

Confounding factors

- Factors that are not being studied, but will still affect experiment
- Examples:
  - stereo condition less bright than mono condition
  - More latency/less fps with stereo
- Important to predict and control confounding factors, or experimental validity will suffer

Experimenter issues

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

Hypothesis Testing Procedure

H_1: $\mu_w \neq \mu_m$ (*participants faster with Wii controller*)

H_0: $\mu_w = \mu_m$ (*logical opposite*)

- Construct sampling distribution assuming H_0 is true.
- Compute sampling statistic X
- Referring to sampling distribution, calculate $p = p(X \mid H_0)$
- If p low ($p \leq 0.05$, $p \leq 0.01$), we are unlikely to see X when H_0 is true. We reject H_0, and embrace H_1 (a result!)
- If p not low ($p > 0.05$), we are likely to see X when H_0 is true. We do not reject H_0 (no result; wasted time)

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?

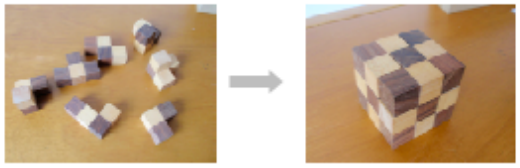
Personal answer: **NO**

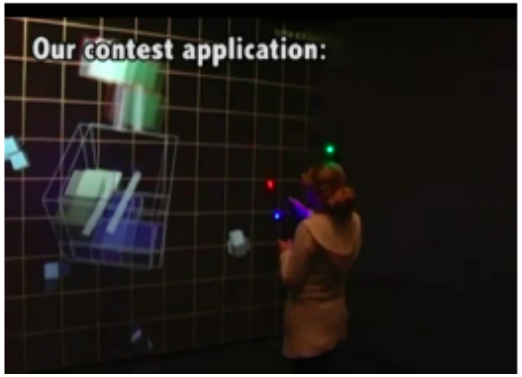
Arguments:

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

Common errors

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





Wrong hypotheses

We measure the time it takes for subjects to solve a 3D puzzle under Wii and Mouse conditions. We hypothesize that users will be slower with the Wii due to the lack of a physical support for free space interaction.

User study but no analysis

Participant	Mouse	Wii	Real
1	0.62	2.67	
2	2.39	1.64	
3	1.59	3.85	
4	0.58	2.75	3.00
5	2.69	1.70	2.20
6	1.68	3.91	1.01
7	0.70	2.65	3.10
8	2.58	1.90	2.00
9	1.58	3.86	1.01
10	0.50	2.86	3.01
11	2.58	1.61	2.00
12	1.69	3.91	0.91
13	0.49	2.60	3.01
14	2.69	1.69	2.00
15	1.70	3.82	1.00
Mean	**1.61**	**2.76**	**2.04**

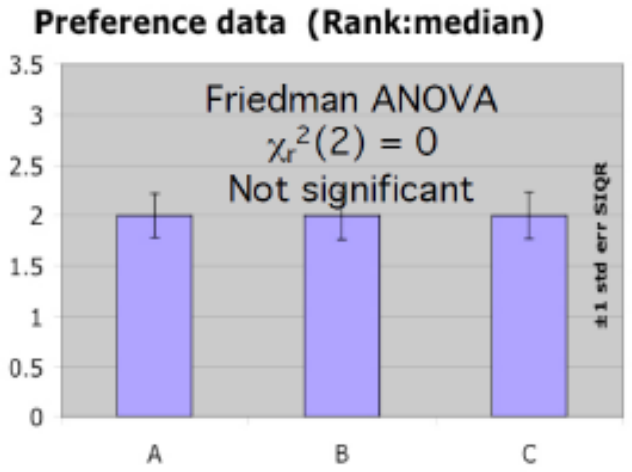
The Wii controller is the winner!

Analysis, but wrong

More correct analysis

Participant	Mouse	Wii	Real
1	3	1	2
2	1	2	3
3	2	3	1
4	1	2	3
5	3	1	2
6	2	3	1
7	1	2	3
8	3	1	2
9	1	2	3
10	2	3	1
11	3	1	2
12	2	3	1
13	3	1	2
14	1	2	3
15	2	3	1
Median	**2**	**2**	**2**

1 means "least preferred"
3 means "most preferred"

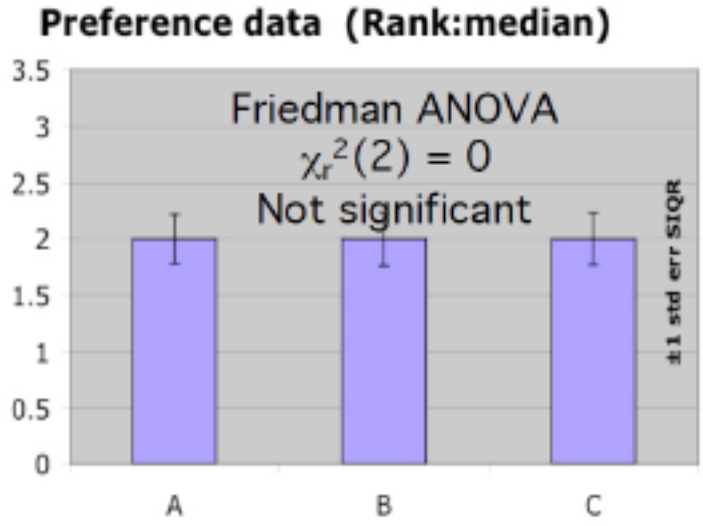


Wrong interpretations

"The three techniques are equally preferred by users"

wrong!

$p(X|H_0) > 0.05$ means **no result**; does not imply H_0 is true!



Not only in 3DUI

Table 1. The frequency of statistical errors in manuscripts submitted to Biochemia Medica during 2006-2009. Errors are sorted according to their frequency.

Error	Error rate N (proportion)
Power analysis not provided	55/55 (1.0)
Incorrect use of statistical test for comparing three or more groups for differences	21/28 (0.75)
Incorrect presentation of P value	36/54 (0.66)
Incorrect choice of the statistical test	34/55 (0.62)
Incorrect interpretation of correlation analysis	11/20 (0.55)
Incorrect use or presentation of descriptive analysis	19/55 (0.35)
Incorrect interpretation of P value	12/54 (0.22)

"There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims"
John Ioannidis (epidemiologist)

Summary

Validation through user studies is critical

Major difficulties

- 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)

Facilitate fair comparisons

Share your code, share your data

Share your code

Executable papers

Make sure your students get enough background on experimentation

Enough experimental design & statistical analysis courses?

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