Module « Techniques d'interaction »

Monter une expérimentation permettant de tester l'*utilisabilité* des techniques d'interaction

Alain Giboin (Equipe Inria-I3S Wimmics)

Octobre 2014

Plan du cours

1. Principes d'une expérimentation
2. Application des principes dans une étude existante
3. Application des principes dans votre projet

1| PRINCIPES D'UNE EXPÉRIMENTATION

Principes d'une expérimentation

- *Source :* C. Appert. **Cours « Experimental Design »**
- **Expérimentation (en laboratoire) :** méthode utilisée pour l'**évaluation** (de l'utilisabilité) **des systèmes interactifs**

  En complément des méthodes :

  – Penser à voix haute (*Think aloud protocol*)

  – Etude de terrain

Principes d'une expérimentation

Steps

- **Design** (conception de l'expérimentation)
- **Run** (passation)
- **Analyze** (analyse des résultats)

(Source : C. Appert)

Principes d'une expérimentation

Experiment - definition

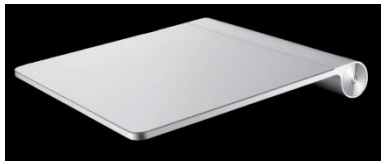
- "**A test under controlled conditions** that is made to demonstrate a known truth, examine the validity of a hypothesis, or determine the efficacy of something previously untried"

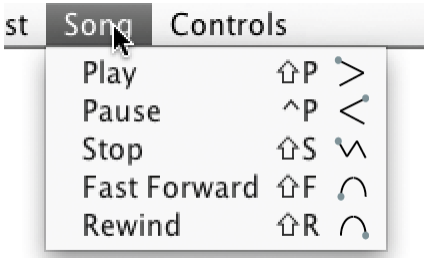
(Source : C. Appert)

Principes d'une expérimentation

Lab experiment

- **Propose and verify a research hypothesis** in a controlled lab setting
  - *Users point faster with a **mouse** than with a **trackpad***
  - ***Gesture commands** are easier to recall than **keyboard shortcuts***
  - *Users make more typing errors with **software keyboards** than with **physical keyboards***
  - …





(Source : C. Appert)

Principes d'une expérimentation

Hypothesis

- A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation
- A hypothesis should be:
  - ***testable***: the means for manipulating the variables and/or measuring the outcome variable must exist
  - ***falsifiable*** *(en français :* ***réfutable****)* : must be able to disprove the hypothesis with data
  - ***precise***: should be specific (operationalized)

(Source : C. Appert)

Principes d'une expérimentation

Testing a hypothesis

- The experimenter manipulates **independent variable(s)** and measures **dependent variable(s)**

(Source : C. Appert)

Principes d'une expérimentation

Manipulable cause

- An independent variable is a potential cause that can be *manipulable*
  - *the dose of a medicine is manipulable*
  - *genetic material is not manipulable*
- Experiments can explore the effects of things that can be manipulated
- That does not mean that a nonmanipulable thing cannot be a cause
- That just means that experimentation does have limits

(Source : C. Appert)

Principes d'une expérimentation

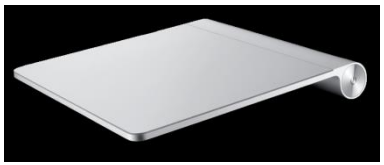
Falsifying a hypothesis *(Réfuter une hypothèse)*

- Statistics tests on observed data **can falsify a hypothesis**, it can **not validate a hypothesis**
- If measured data is coherent with a hypothesis, the experimenter can not conclude anything
- The experimenter research question is usually turned into the **null hypothesis** so as to reject it and support his research hypothesis

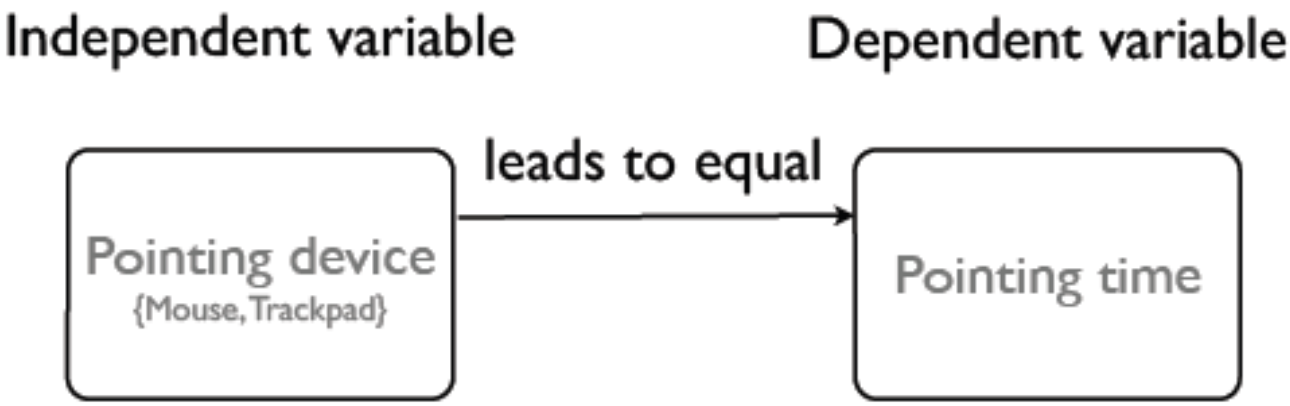
(Source : C. Appert)

Principes d'une expérimentation

Testing a hypothesis



- **research hypothesis:** Users point faster with a mouse than with a trackpad
- **null hypothesis:** Users point as fast with a mouse as they point with a trackpad

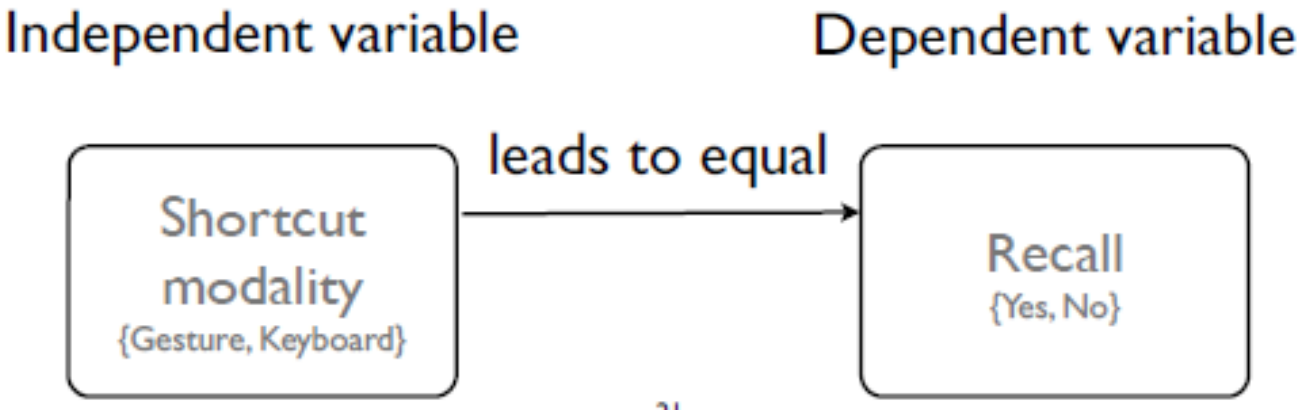


(Source : C. Appert)

Principes d'une expérimentation

Testing a hypothesis

- **research hypothesis:** Gesture commands are easier to recall than keyboard shortcuts!
- **null hypothesis:** Gesture commands are as easy to recall as keyboard shortcuts are

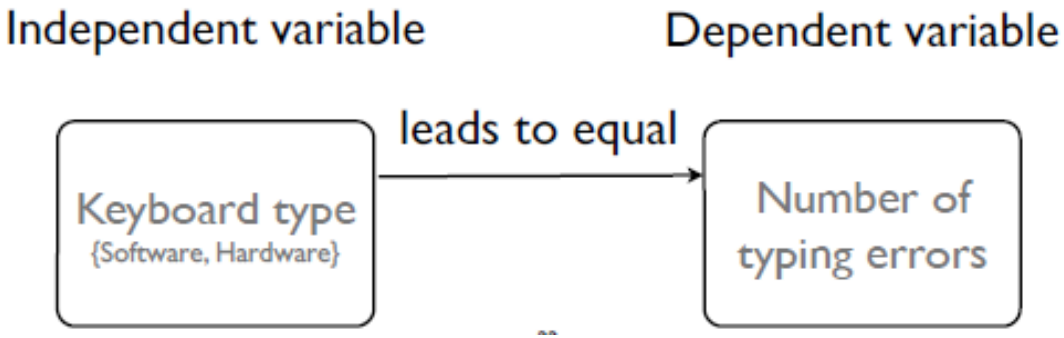


(Source : C. Appert)

Principes d'une expérimentation

Testing a hypothesis

- **research hypothesis:** Users make more typing errors with software keyboards than with physical keyboards
- **null hypothesis:** Users are as accurate with software keyboards as they are with physical keyboards



(Source : C. Appert)

Principes d'une expérimentation

Operationalization

- **Defining a precise way** (i.e. a task for users to do) **of observing dependent variables as a function of independent variables** in an experimental task
- The **task must be simplified** to its minimum so as to eliminate bias and effects from confounding variables (i.e. variables that are not of interest)

(Source : C. Appert)

Principes d'une expérimentation

Confounding variable *(Variable parasite)*

- Any variable other than the independent variable that can possibly explain the change in measures

  Learning

  – e.g. all participants are first tested with the physical keyboard and then with the software keyboard → software keyboard has the advantage that participants have learned the keyboard layout

  Prior experience

  – e.g. use conventional keyboard shortcuts (e.g. ctrl+V for paste) when comparing them to gesture shortcuts, which are a nonfamiliar type of shortcuts → keyboard shortcuts are favored because of participants' existing knowledge

(Source : C. Appert)

Principes d'une expérimentation

Operationalization

- Operationalization consists of turning a phenomenon into an experimental task that takes factors as input and outputs measures

  "Children grow more quickly if they eat vegetables"

  - Meaning of 'children'? (e.g. 4 < age < 8)
  - Meaning of 'grow more quickly'? (e.g. cm per year)
  - What are 'vegetables'? (e.g. quantity of vitamin C)

(Source : C. Appert)

Principes d'une expérimentation

Validity issues

- **Validity**
  - best available approximation to truth or falsity of propositions
- **Internal validity**
  - best approximate truth about inferences regarding causal relationships (i.e. did the treatment really cause the effect?)
- **External validity**
  - generalizability of the results
- **Operationalizing** requires to find the good tradeoff internal validity - external validity

(Source : C. Appert)

Principes d'une expérimentation

Causality and Correlation

- **Correlation**
  - **mathematical relationship** between two variables
- **Causality**
  - **physical relationship** between two variables. There is a chain of events when the first variable varies that causes the other variable to vary (involves time)

(Source : C. Appert)

Principes d'une expérimentation

Causality and Correlation

- **Correlation does not imply causality**
  - For example, we noted a high correlation between height and weight
  - This does not show that height ⇒ weight
- **"When does correlation imply causation?"**
  - When the experiment design is done with appropriate care to avoid confounding and other threats to the internal validity of the experiment

(Source : C. Appert)

Principes d'une expérimentation

Operationalization

Well-known standards

- If some well-known standards do exist, use them!
- **Pointing**
  - ISO. 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs)-Part 9: Requirements for non-keyboard input devices
- **Text entry**
  - MacKenzie et al.'s phrase set!
  - http://www.yorku.ca/mack/chi03b.html

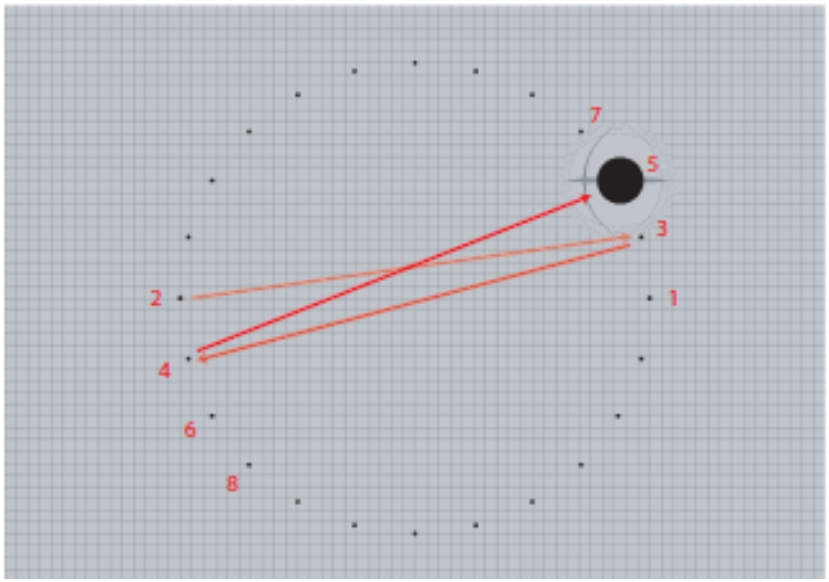
(Source : C. Appert)

Principes d'une expérimentation

Operationalization

Well-known standards - Pointing

- The circular layout and order of appearance of targets force participant to point in every direction



(Source : C. Appert)

Principes d'une expérimentation

Operationalization

Well-known standards – Text-entry

- Phrases that are moderate in length, easy to remember, and representative of the target language

```
video camera with a zoom lens!
have a good weekend!
what a monkey sees a monkey will do!
that is very unfortunate!
the back yard of our house!
I can see the rings on Saturn!
this is a very good idea!
...
```

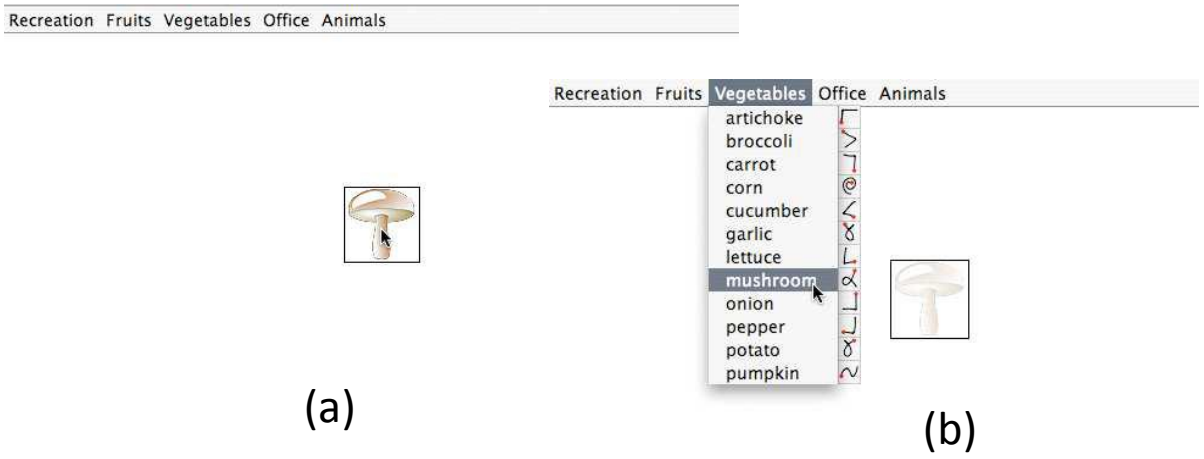
(Source : C. Appert)

Principes d'une expérimentation

Operationalization

No standard - Example#1

- Keyboard vs Gesture shortcuts
  - use arbitrary mappings (avoid the effect of prior experience)
  - use a number of shortcuts representative of expert usage (e.g. 14)



(a) (b)

ICON	Keys	Stroke	ICON	Keys	Stroke
	Shift+W			Ctrl+W	
	Shift+D			Ctrl+D	
	Shift+Q			Ctrl+Q	
	Shift+S			Ctrl+S	
	Shift+E			Ctrl+E	
	Shift+A			Ctrl+A	
	Shift+R			Ctrl+R	

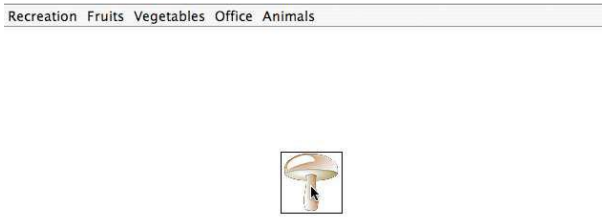
(Source : C. Appert)

Principes d'une expérimentation

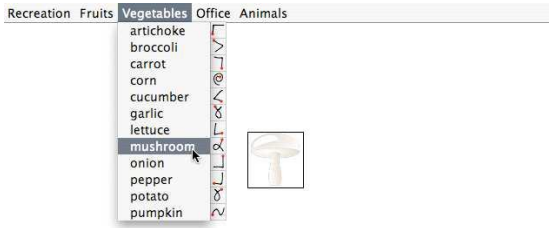
Operationalization

No standard - Example#1

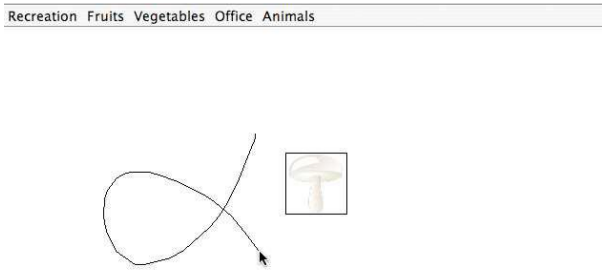
- Keyboard vs Gesture shortcuts
  - measure recall on two consecutive days



(a)



(b)



(c)

None	mushroom	Warm up (day 1)
Keyboard	mushroom ⇧S	Test (day 1)
Stroke	mushroom	Test (day 1)
Both	mushroom ⇧S	Re-test (day 2)

(Source : C. Appert)

Principes d'une expérimentation

Operationalization

No standard - Example#2

- **Discovering and learning Dwell-and-Spring**
  - The task consists in moving an icon to a target location
  - The participant is told before starting that she would be interrupted in the middle of her move by a pop-up message indicating how to finish the trial:
    - either put the icon back to its original location
    - or finish the current operation
  - The participant does not receive any indication about available techniques. She is simply encouraged to explore the interface

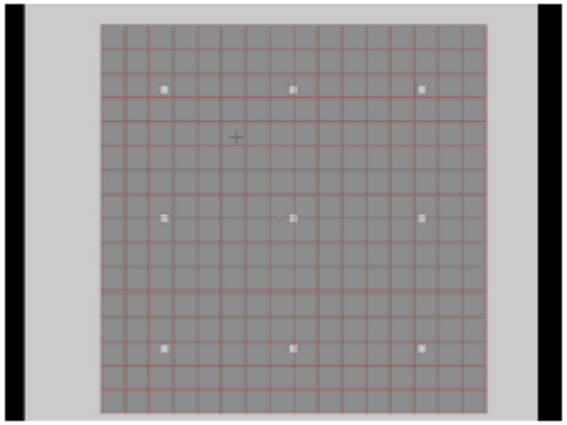
(Source : C. Appert)

Principes d'une expérimentation

Operationalization

No standard - Example#3

- **Multiscale searching in zoomable interfaces**
  - Ask participants to inspect *n* objects until they found the target (i.e. the square that has rounded corners)



(Source : C. Appert)

Principes d'une expérimentation

Operationalization

No standard - Example#3

- **Multiscale searching in zoomable interfaces**
  - Ask participants to inspect *n* objects until they found the target (i.e. the square that has rounded corners)
  - Visual acuity and chance can vary among participants and trials:
    - force a quantity of exploration a priori (at least *n* objects to inspect)
    - add an explicit action to unveil corners (press space bar)

(Source : C. Appert)

Principes d'une expérimentation

Dependent variable

Types of scales

- Four major scales of measurements
  - Nominal
  - Ordinal
  - Interval
  - Ratio (*de rapport*)

(Source : C. Appert)

Principes d'une expérimentation

Dependent variable

Nominal scale

- **Classification** into named or numbered **unordered** categories
  - country of birth, user groups, gender…
- **Allowable manipulations**
  - whether an item belongs in a category
  - counting items in a category

(Source : C. Appert)

Principes d'une expérimentation

Dependent variable

Ordinal scale

- **Classification** into named or numbered **ordered** categories (no information on magnitude of differences between categories)
  - preference, social status, gold/silver/bronze medals
- **Allowable manipulations**
  - as with nominal scale, plus
  - merge adjacent classes
  - transitive: if A > B > C, then A > C

(Source : C. Appert)

Principes d'une expérimentation

Dependent variable

Interval scale

- **Classification** into ordered categories with **equal differences between categories** (zero only by convention)
  - temperature (C or F), time of day
- **Allowable manipulations**
  - add, subtract
  - cannot multiply as this needs an absolute zero

(Source : C. Appert)

Principes d'une expérimentation

Dependent variable

Ratio scale (échelle de rapport)

- Interval scale with absolute, non-arbitrary zero
  - temperature (K[elvin]), length, weight, time periods
- **Allowable manipulations**
  - multiply, divide

Principes d'une expérimentation

Types of design

- Choose a **representative sample** of the population you want to study
- How to assign the different levels (i.e. values) of an independent variable?
  - Which participants will test pointing with a mouse and/or pointing with a trackpad?!
  - Which participants will test keyboard and/or gesture shortcuts?

(Source : C. Appert)

Principes d'une expérimentation

Types of design

Between-subject design

- If the independent variable has *n* levels
- Randomly select *n* groups and assign a different level to each group
- Assumptions
  - Other non-controlled variables are randomly distributed between the *n* groups
  - The only systematic difference between the *n* groups is the independent variable

(Source : C. Appert)

Principes d'une expérimentation

Types of design

Within-subject design

- If the independent variable has *n* levels
- Successively expose each participant to the *n* different levels
- Automatically control of most of the other variables
- Allows you to use a smaller number of subjects

(Source : C. Appert)

Principes d'une expérimentation

Types of design

Factorial design

- Test several independent variables in the same experiment
  - Mouse vs. Trackpad: device (mouse, trackpad), target width (10, 20, 30), target distance (200, 400)
- Each variable can be distributed according to a between or within subject design
- A combination of a value for each independent variables defines an experimental condition
  - 2 device x 3 target width x 2 target distance = 12 conditions

(Source : C. Appert)

Principes d'une expérimentation

Controlling variation

- Replication and blocking are two mechanisms to reduce observed variation that is not due to difference between conditions
- Replication: A participant does several times the experimental task in the same condition
  - e.g. if the participant got distracted in a particular condition
- Blocking: Arranging the tasks into blocks of tasks that are similar to one another
  - e.g. eliminating the time due to successive changes between two pointing devices

(Source : C. Appert)

Principes d'une expérimentation

Controlling order effect

- May happen when an independent variable is presented according to a within-subject design
- Order effect problem:
  - if the mouse is always presented after the trackpad and observed time is shorter with the mouse, is pointing performance better because of the input device or because the user has become more familiar (thus efficient) with the task?
- Randomization: presentation the different conditions in a "random" order across the experiment

(Source : C. Appert)

Principes d'une expérimentation

Randomization

- Randomization is not haphazard
  - An experiment is randomized if the method for assigning levels of independent variables involves a deterministic probabilistic scheme
- Examples of bad randomization
  - assign mouse to the first half of participants who show up and trackpad to the other half (pb: the first half can be the most motivated participants)
  - assign mouse or trackpad depending on if start time in seconds is a odd or even (pb: can result in much more observations in one or the other condition)

(Source : C. Appert)

Principes d'une expérimentation

Counterbalancing

- Counterbalancing is a scheme to randomize a within-subject experiment design
- Consider an independent variable that has *n* levels and a sample of X participants
  - Compute the *n* possible orders and assign each order to X / *n* participants (requires a potentially high number of participants)
  - Compute n possible orders using a Latin Square and assign each to X / *n* participants

(Source : C. Appert)

Principes d'une expérimentation

Randomization

Latin square definition

- A Latin square is an $n \times n$ array filled with n different symbols, each occurring exactly once in each row and exactly once in each column
- Example: n=3 levels ({A, B, C})

A	B	C
C	A	B
B	C	A

(Source : C. Appert)

Principes d'une expérimentation

Quasi-experiments

- True Experiment
  - random assignment of units to conditions
- Quasi Experiment
  - units are not assigned to conditions randomly
- Quasi-experiments are very similar to true experiments but use naturally formed groups (sometimes called natural experiments)

(Source : C. Appert)

Principes d'une expérimentation

Quasi-experiments

- When randomization is not possible or unethical
  - e.g. Spanked children are less responsive to an educational class for reducing aggressiveness (ethics)!
  - e.g. Vaccine X is ineffective for people having the gene Y (not possible)
- External validity may be higher as quasi experiments can be seen as "more natural"
- Internal validity may be lower as other differences between groups may exist

(Source : C. Appert)

Principes d'une expérimentation

Quasi-experiments

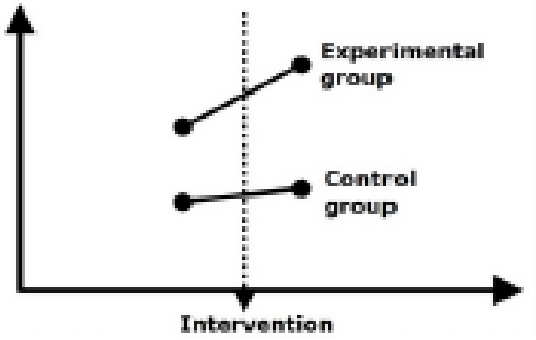
- Make explicit the "threats of internal validity"
  - history, selection of participants, participant loss, etc.
- Increase internal validity using design strategies
  - pretest-posttest
  - Interrupted time-series…

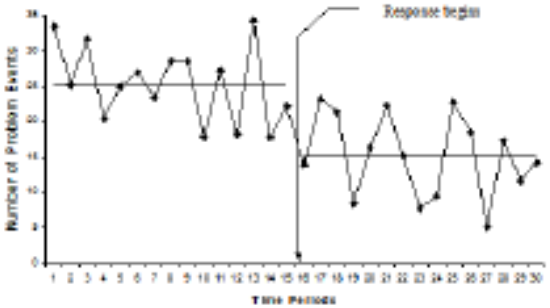
(Source : C. Appert)

Principes d'une expérimentation

Designs for Quasi Experiments

- Pretest-posttest
  - Measure before and after manipulating independent variables
- Interrupted time-series
  - Several measures before and after manipulating independent variables
  - Captures trends → higher internal validity





(Source : C. Appert)

Principes d'une expérimentation

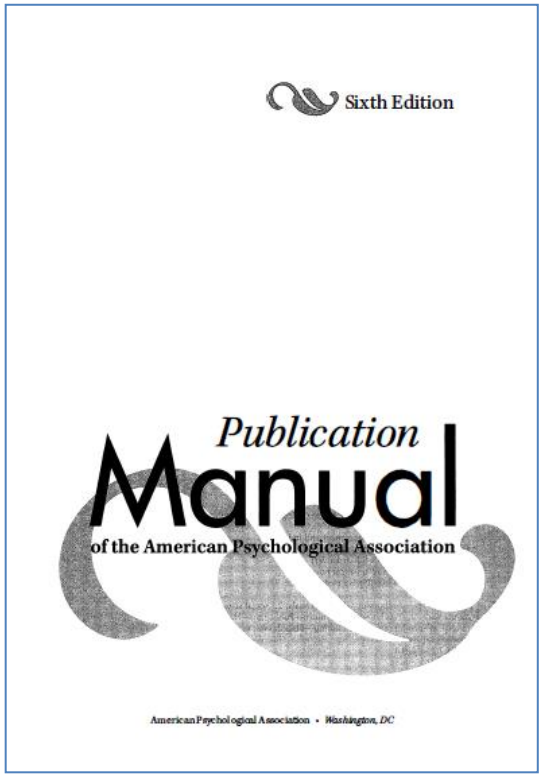
Replicability

- Any experiment should be replicable by others
- Always report:
  - a complete description of the experiment design
  - the hardware/software characteristics of the experimental environment (apparatus)
  - a description of the participants' characteristics that may impact the observed measures (gender, mean and variance in age, prior experience, social and occupational aspects...)

(Source : C. Appert)

Principes d'une expérimentation

Rapporter l'expérimentation



Principes d'une expérimentation

Rapporter l'expérimentation

- **Introduction**
  - Introduce the problem
  - Explore importance of the problem
  - Describe relevant scholarship
  - State hypotheses and their correspondence to research design

Principes d'une expérimentation

Rapporter l'expérimentation

- **Method**
  - Participant (subject) characteristics
  - Sampling procedures
  - Sample size, power, and precision
  - Measures and covariates
  - Research design
  - Experimenta l manipulations or interventions

2 | APPLICATION DES PRINCIPES
DANS UNE ÉTUDE EXISTANTE

L’étude existante

- **Pointing and Beyond: an Operationalization and Preliminary Evaluation of Multi-scale Searching**

  (E. Pietriga, C. Appert, & M. Beaudouin-Lafon, 2007)

L’étude existante

Multi-scale interfaces / Multi-scale searching

- *Also called:* **Zoomable User Interfaces** (ZUIs)
- *Role:* representing, navigating and manipulating large sets of data
- *Types of interfaces :*
  - From original pan & zoom
  - To various focus+context techniques

L’étude existante

Techniques tested:

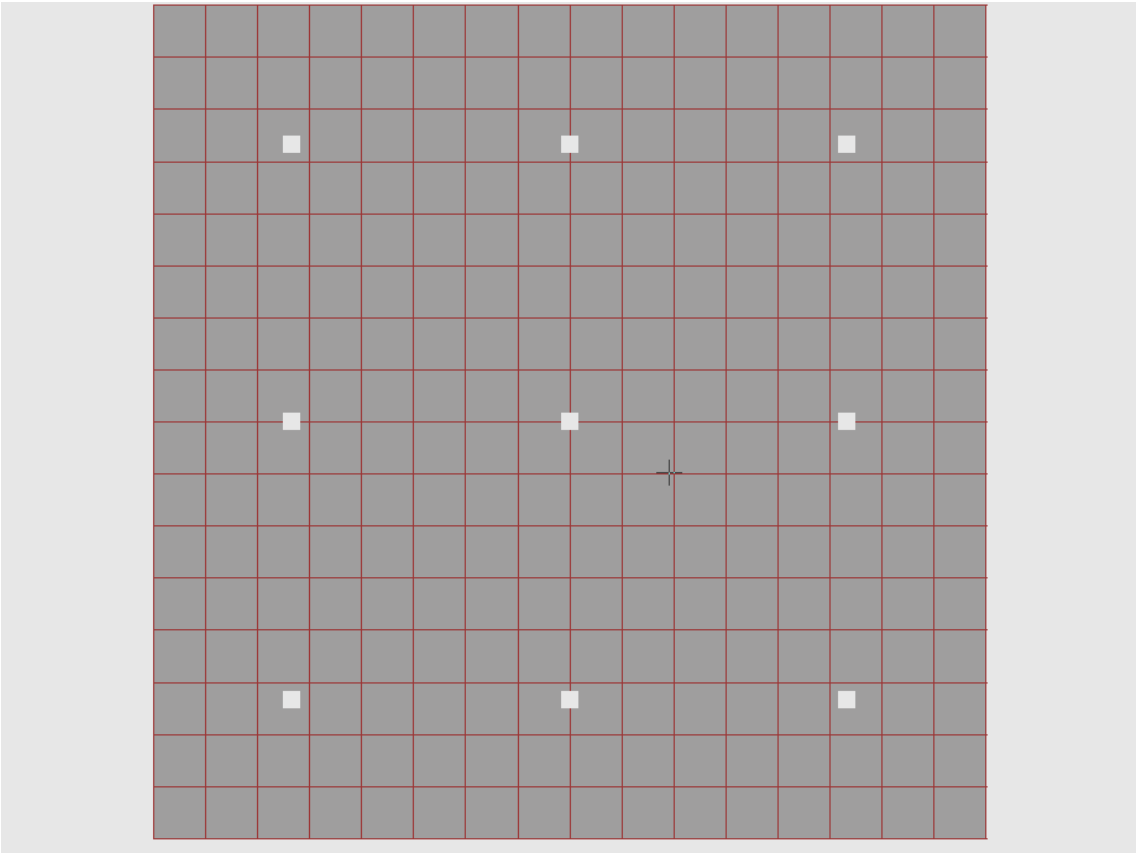
- Pan&Zoom (PZ)
- Pan-zoom+Overview + Detail (OD)
- Graphical Fisheye (Distortion) Lens (FL)
- DragMag Image Magnifier (DM)

L’étude existante

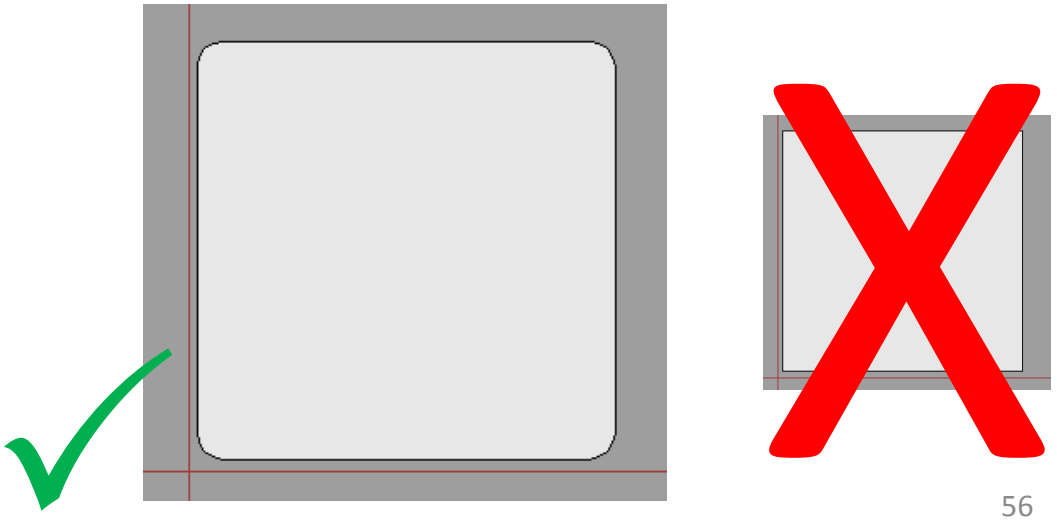
Task:

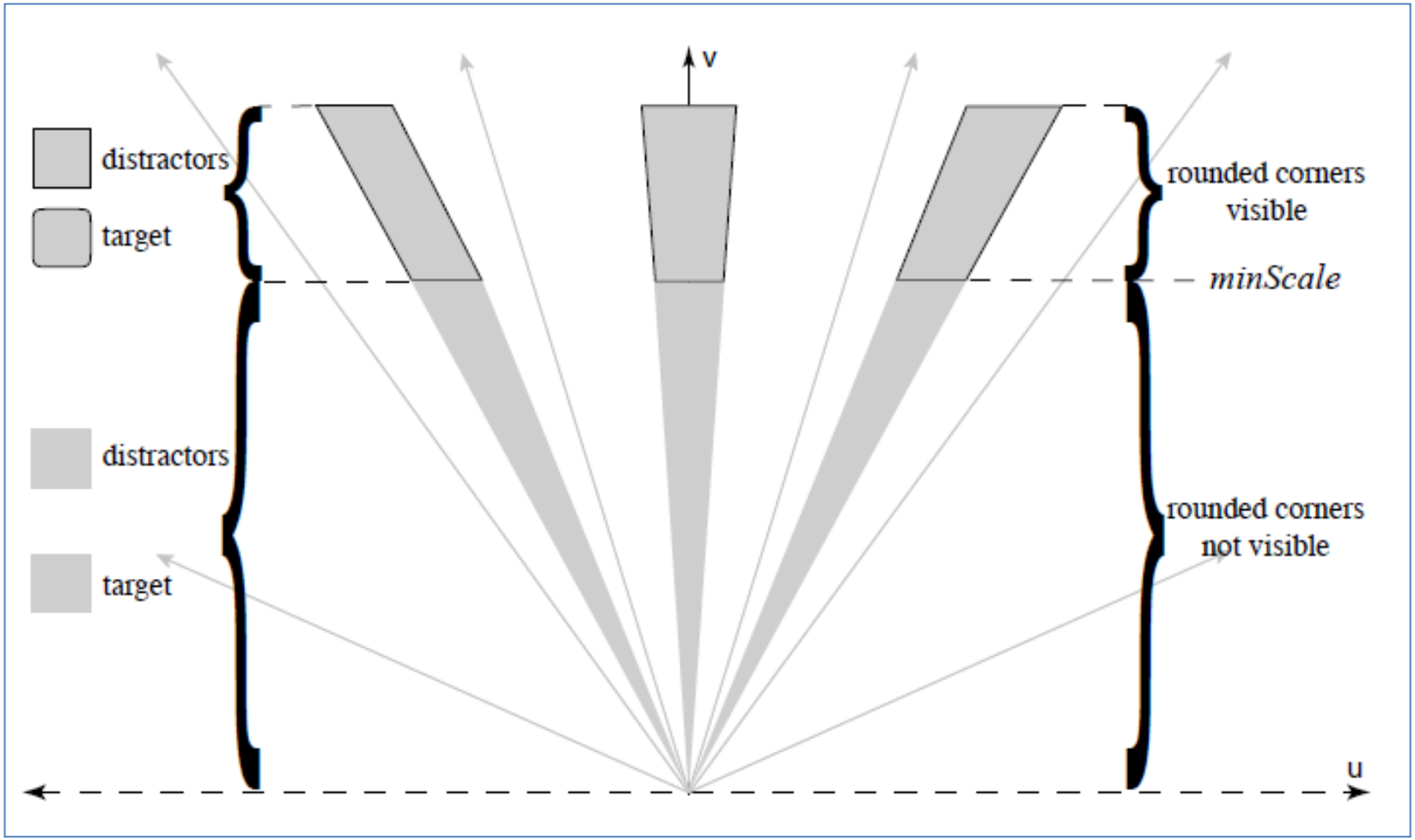
- **Finding a target** among a set of objects as quick as possible while minimizing the number of errors

- *Set of objects:* **9 white squares**

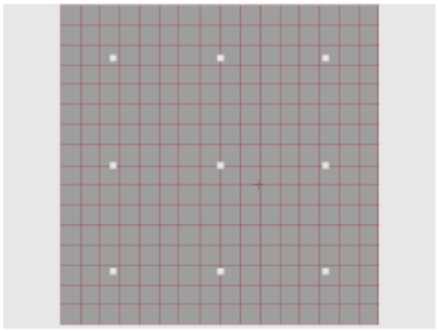


- *Target :* the square with **rounded corners**

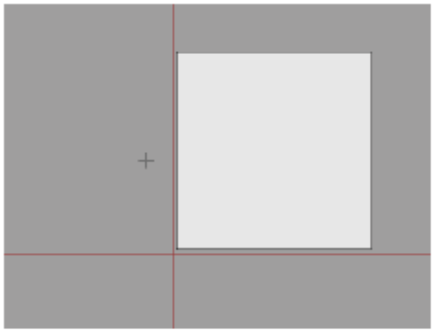



v
distractors
target
distractors
target
rounded corners visible
minScale
rounded corners not visible
u

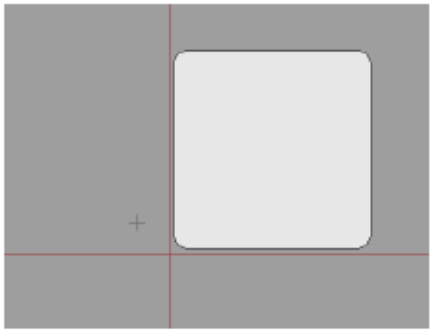
(a)



(b)



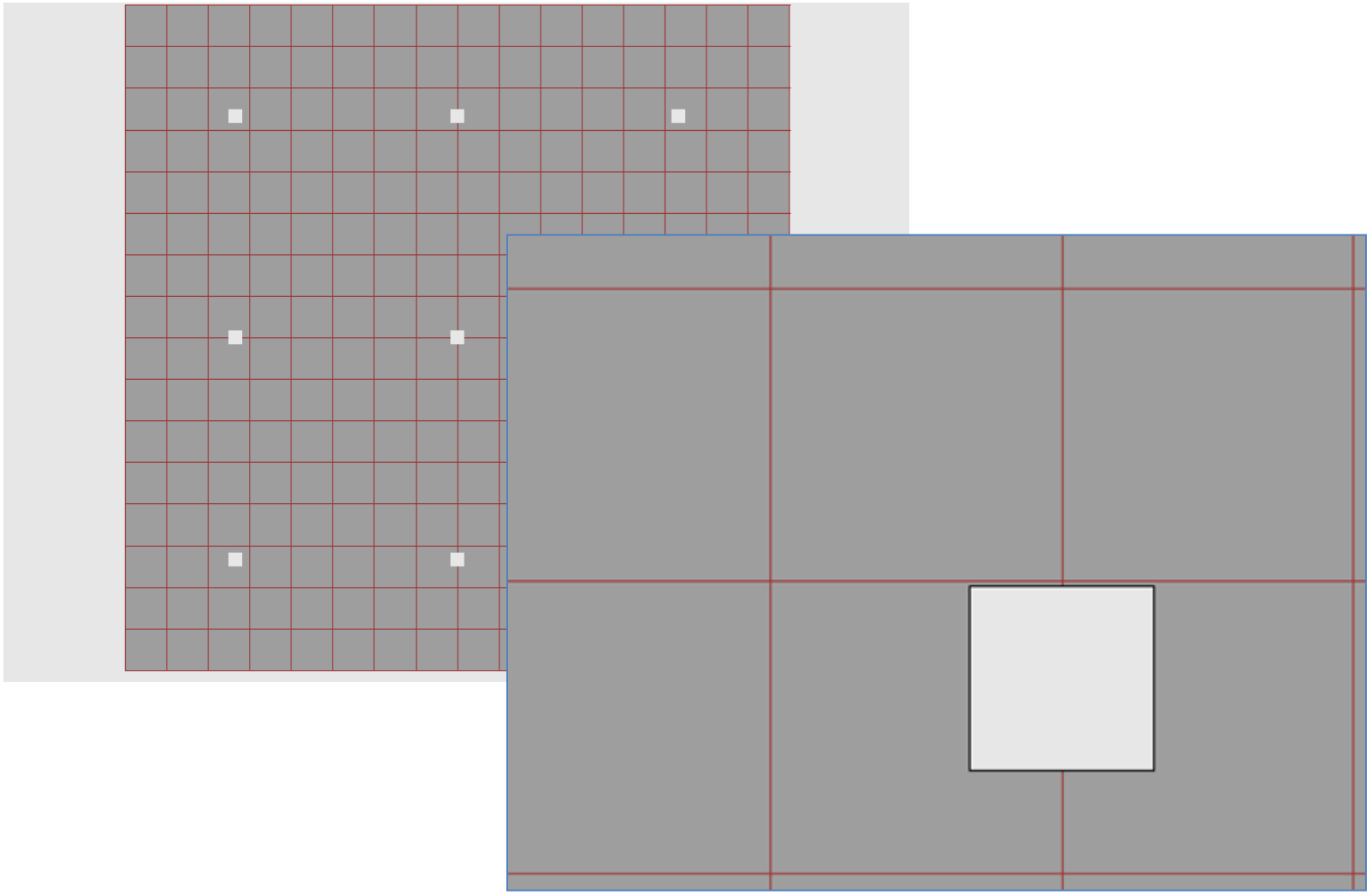
(c)



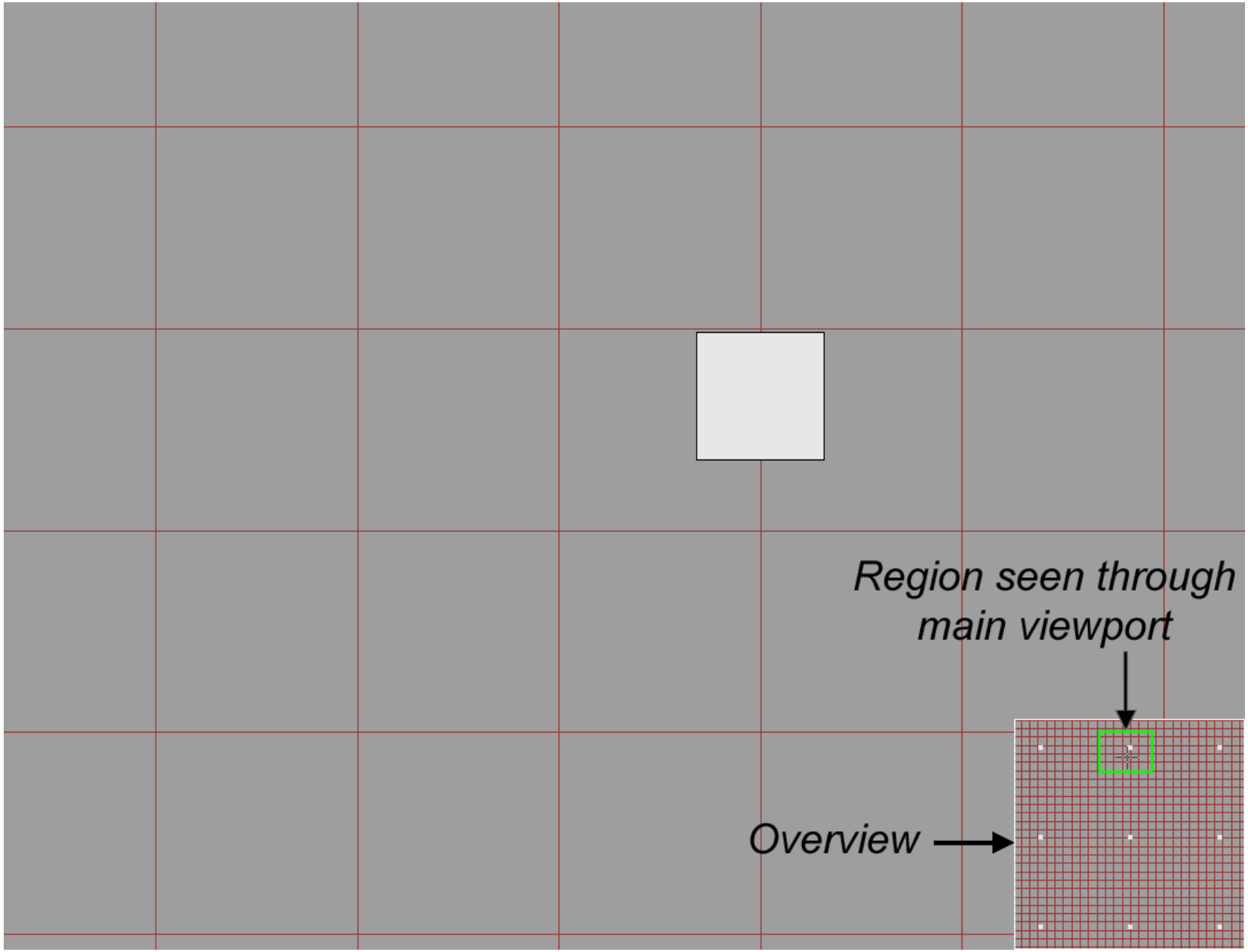
(d)

Storyboard of the task:
(a) start of trial
(b) navigation to the set of objects
(c) inspection of an object (before unveiling)
(d) after unveiling the target

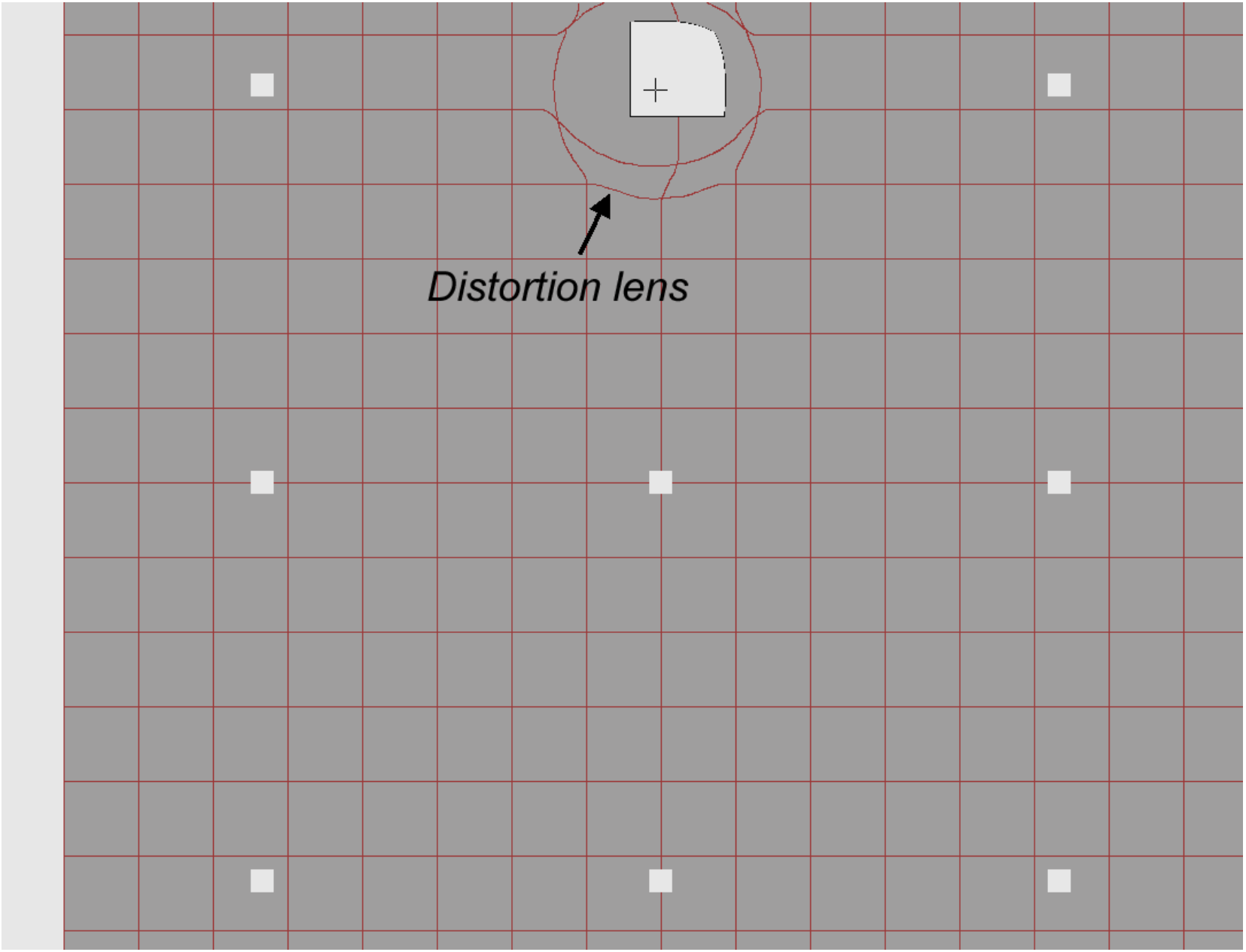
Pan & Zoom (PZ)



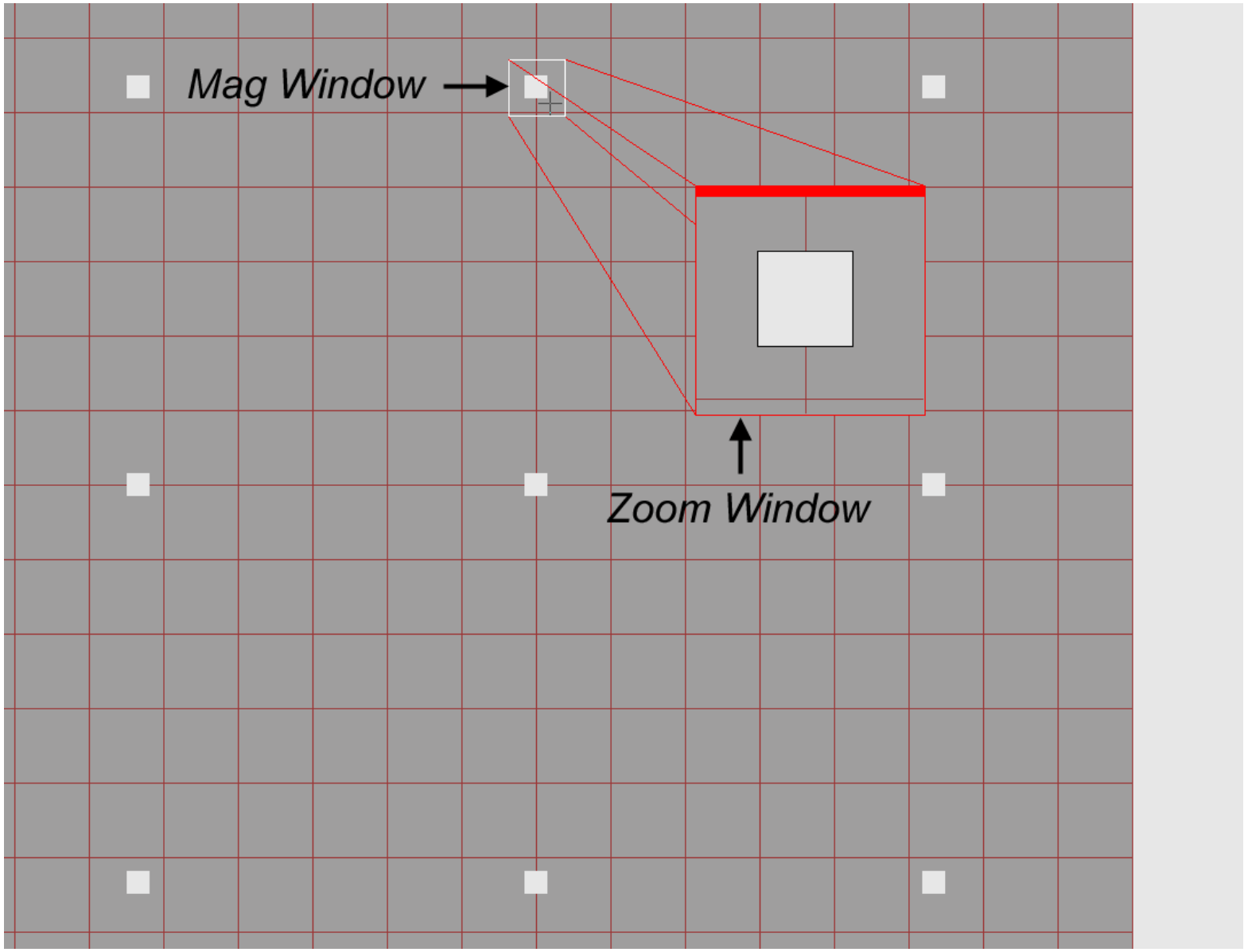
Pan-zoom+Overview + Detail (OD)



Graphical Fisheye (Distortion) Lens (FL)



DragMag Image Magnifier (DM)



L’étude existante

Objectif : Continuer vous-mêmes l’analyse de l’expérimentation

Identifier dans l’article :

- Quelles hypothèses ?
  - Quelles V.I. et quelles V.D. ?
- Quels participants ?
- Quelle procédure ?
- …

3| APPLICATION DES PRINCIPES DANS VOTRE PROJET

Votre projet

- Quelles techniques vont être testées ?
- Quelles hypothèses ?
  - Quelles V.I. et quelles V.D. ?
- Quels participants ?
- Quelle tâche ?
- Quelle procédure ?
- …

Références

- C. Appert. Cours « **Evaluation of Interactive Systems », Cours n° 3 « Experimental Design »**
  - Slides: https://www.lri.fr/~appert/m1hcid-eval/3-experimental-design.pdf
- **Publication Manual of the American Psychological Association** (cf. chapter « Manuscript Structure and Content », pp. 21-) :
  - www.uai.cl/images/carlos_escarate/citas_bibliograficas/APA6th.pdf
- E. Pietriga, C. Appert and M. Beaudouin-Lafon. (2007). **Pointing and Beyond: an Operationalization and Preliminary Evaluation of Multi-scale Searching**. In Proc. ACM Conference on Human Factors in Computing Systems, CHI 2007, San Jose, CA, USA. ACM Press, New York, NY, April 2007, pages 1215-1224
  - Article: http://hal.inria.fr/inria-00158867/PDF/main.pdf
  - vidéo : http://pages.saclay.inria.fr/emmanuel.pietriga/2007/04/chi2007_pointing_and_beyond.mov