Méthodes d’évaluation empirique

Concevoir une expérience contrôlée

Experimental design

- Introduction and examples
- The elements of experiments
- Designing an experiment
- Conducting an experiment
- Gathering data

Experimental design

- **Introduction and examples**
- The elements of experiments
- Designing an experiment
- Conducting an experiment
- Gathering data

Controlled experiments

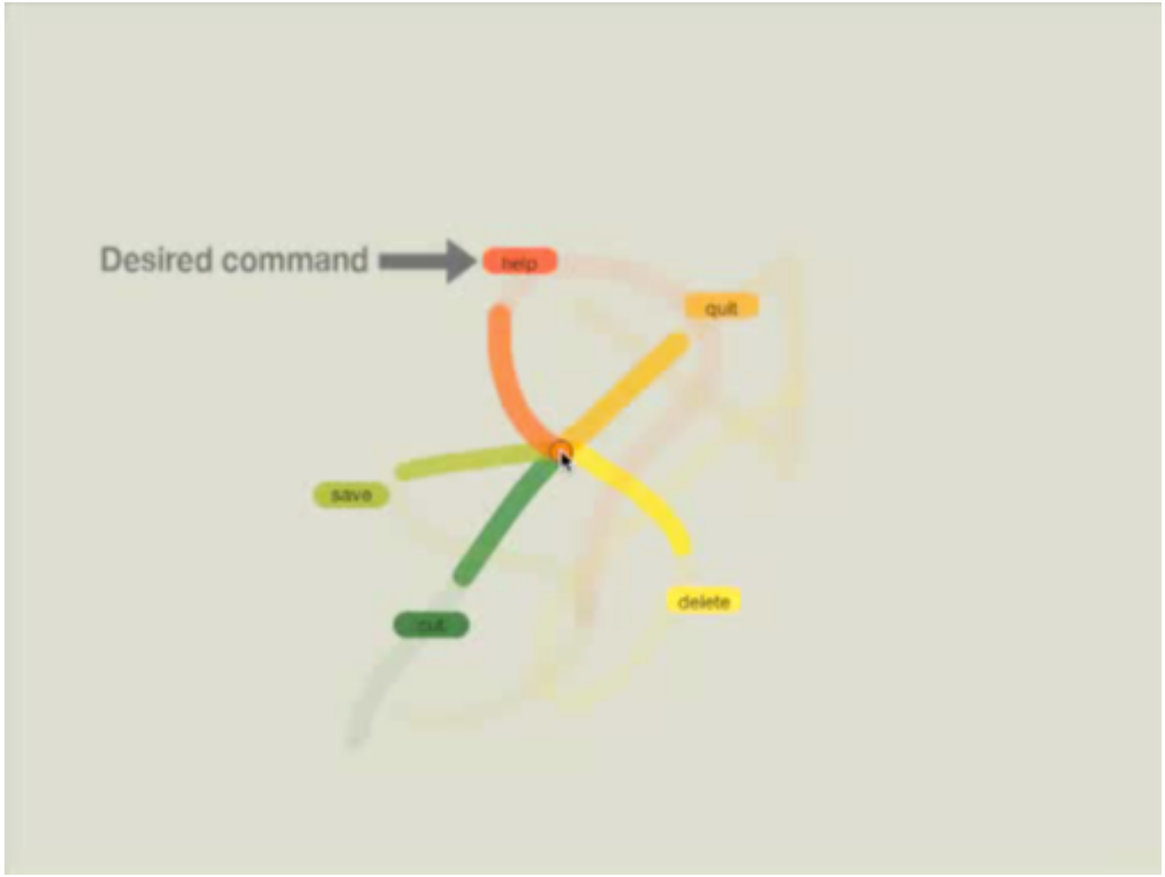
A scientific approach

- Answering specific questions with data
  - Performance
  - Learning
  - Satisfaction
- Providing basic knowledge generalizable across contexts.
- Demonstrate causality between different factors
  - correlation: show that a change in A occurs with a change in B
  - order: show that A takes place before B
  - no hidden cause: show that there is no C with C -> A and C -> B

Example 1: what is the best input device?

Device	Study	IP (bits/s)
Hand	Fitts (1954)	10,6
Mouse	Card, English, & Burr (1978)	10,4
Joystick	Card, English, & Burr (1978)	5
Trackball	Epps (1986)	2,9
Touchpad	Epps (1986)	1,6
Eyetracker	Ware & Mikaelian (1987)	13,7

Example 2: a better way do learn gestures?



Example 3: personalised news recommendation

Based on click behaviour (Google News)

- 10,000 users of the live traffic at Google News
- Users randomly assigned to a control group and a test group
- Experience running for 34 days, based on 12 months of user history
- Click-through rate (CTR) analysis over anonymized click logs
  - of the recommended news section,
  - of the Google News homepage, and
  - frequency of visiting Google News website.

Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces* (*IUI '10*). ACM, New York, NY, USA, 31–40. DOI: https://doi-org.docelec.insa-lyon.fr/10.1145/1719970.1719976

Experimental design

- Introduction and examples
- **The elements of experiments**
- Designing an experiment
- Conducting an experiment
- Gathering data

Method (1)

1. Define what you are looking for: write down an hypothesis
   - Circular menus reduce search time
2. Design the experiment pick variables and fixed parameters
   - Define the menu structure (1 level or 2 level? how many items?)
3. Conduct a pilot to test the experiment
   - Fine tune data collection adjust the number of repetitions

Method (2)

4. Recruit participants
   - Students spending more than 2h per day in front of their computer
5. Conduct experiment and gather data
6. Analyze and clean up data, to accept or reject the original hypothesis measure the effect (between conditions)
   - Average search time: 2.26 (Circular menu), 2.64 (Usual menu)
   - Difference is significant : $p<.05$

The elements of experiments

Factors (or independent variables)

- Variables we manipulate in each condition
- The quantity of menu items, the quantity of sub-menus

Levels (a.k.a. possible values for independent variables)

- Menu with 8 items or a menu with 12 items

Measures (or response): dependent variable(s)

- The measured outcome of the experiment
- Selection time of a menu item

Replication

- number of participants assigned to each level

Independent variables (factors)

The conditions of the experiment are set by independent variables

- The number of items in a list, text size, font, color

The number of different values used is the level

- The number of experimental conditions is the product of the levels
- E.g., font can be times or arial (2 levels), background can be blue, green, or white (3 levels). This results in 6 experimental conditions (times on blue, times, on green, ..., arial on white)

Dependent variables (measures)

The dependent variables are the values to be measured:

- Objective values: e.g. time to complete a task, number of errors, etc.
- Subjective values: ease of use, preferred option, etc.
- They should only be dependent on changes of the independent variables.

Exercise

Identify independent variables (factors) and dependent variables (measures) in each scenario + Give possible levels for independent variables.

- A study to learn whether people who have followed a digital security training use more secure password.
- A study to learn whether a joystick or a mouse is more efficient to select static targets or moving targets
- A study to learn whether teams using video Zoom are more productive than those using text-only chat.

Experimental design

- Introduction and examples
- The elements of experiments
- **Designing an experiment**
- Conducting an experiment
- Gathering data

Participants / subjects

Within-subjects design:

- Same participant exposed to all test conditions
- Fewer participants needed (noise minimisation)
- Learning effects

Between-subjects design:

- Independent groups of participants for each test condition (control and treatment)
- No learning effect
- Requires more users

Randomization and control

The order of presentation can have an impact measures:

- learning
- fatigue
- contrast (1st treatment leads to behavior/answers in 2nd treatment)

Solution

- Rest between treatment
- Counter-balancing, but can become complicated
- Latin square

Latin square (Carré Latin)

via http://hci.rwth-aachen.de/~chat/StatLecture/prerequisite.pdf

- Each condition appears at each ordinal position
- Each condition precedes and follows each condition one time
- Example: six treatments: A, B, C, D, E, F

1	A	B	F	C	E	D
2	B	C	A	D	F	E
3	C	D	B	E	A	F
4	D	E	C	F	B	A
5	E	F	D	A	C	B
6	F	A	E	B	D	C

Exercise

What kind of experimental design for:

- A study to learn whether people who have followed a digital security training use more secure password.
- A study to learn whether a joystick or a mouse is more efficient to select static targets or moving targets
- A study to learn whether teams using video Hangout/Skype are more productive than those using text-only chat.

Hypotheses

- Prediction of the result of an experiment
- Stating how a change in the independent variables will affect the measured dependent variables

Usual approach to hypotheses

- Stating a working hypothesis H_1
- Stating a null hypothesis H_0
  - intuition (naive) : if H_0 is false then H_1 must be true
- Carrying out the experiment and using statistical measures to disprove the null-hypothesis
- When a statistical test shows a significant difference it is probable that the effect is not random

Validity

Internal validity

- Manipulation of independent variable is cause of change in dependent variable
- Requires removing effects of confounding factors
- Requires choosing a large enough sample size, so the result couldn’t have happened by chance alone.

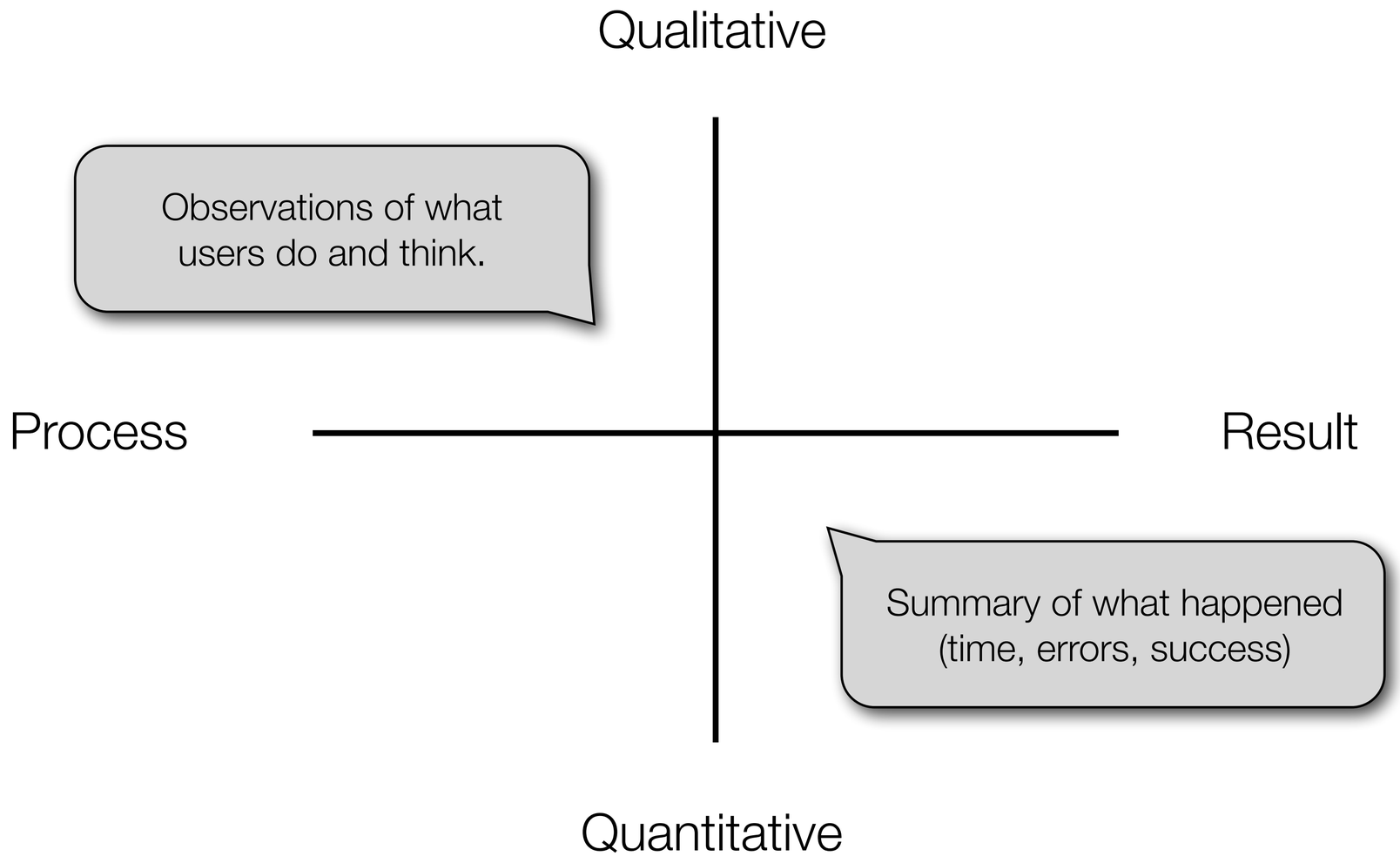
External validity

- Results generalize to real world situations
- Requires that the experiment be replicable
- No study “has” external validity by itself!

Experimental design

- Introduction and examples
- The elements of experiments
- Designing an experiment
- **Conducting an experiment**
- Gathering data

What kind of data to record?



What to record: quantitative data

- Success rates
- Accuracy / Error rates:
  - How many mistakes did people make?
    And were they fatal or recoverable with the right information?
- Time on Task:
  - How long does it take people to complete basic tasks?
    (For example, find something to buy, create a new account, and order the item.)
- Pages visited, number of steps to reach goal...
- Recall:
  - How much does the person remember afterwards or after periods of non-use?
- Emotional Response:
  - Ratings on a satisfaction questionnaire, How does the person felt about the tasks completed?
    (Confident? Stressed? Would the user recommend this system to a friend?)

What to record: qualitative

- How people reacted to the system.
- How participants understood it.
- Which pathways participants took.
- Which problems participants had (critical incidents).
- What participants said as they worked.
- Participants’ answers to open-ended questions.

You need a plan!

A good plan for usability testing gives the participants:

- a goal/task (what to do or what question to find the answer for)
- data, if needed, that a real user would have when going to the site to do that task

You can give the scenario as just the statement of the goal/task or you can elaborate it a little with a very short story that adds motivation to get to the goal.

Participants

The participants must be like the people who will use your product.

Be ready to screen participants (do not grab the first person in the corridor)

Plan on a cost associated with finding the people

- you may still need to plan on incentives to get participants to participate ...

Test!

Make sure you have everything you need

- the prototype you are going to test
- the computer set up specified in your test plan
- consent forms
- questionnaires, if you are using any
- the participant's copy of the scenarios
- cameras, microphones, or other recording equipment
- folders to keep each person's paperwork in if using paper

Do a dry-run and a pilot test

Before starting

- You should know, and have written down
  - objective
  - description of system being testing
  - task environment & materials
  - participants
  - methodology
  - tasks
  - test measures
- Will help you design a good experiment plan
- Will help you figure out how to analyze your data

Laboratory space

Dedicated testing room

- Instrumented with data collection
  e.g. mic., cameras, loggers…

Keep variances among tests low

From C|Net “How Google tested Google Instant”
http://news.cnet.com/8301-30684_3-20019652-265.html

Measures

Situations in which numbers are useful

- time requirements for task completion
- successful task completion
- compare two designs on speed or # of errors

Measures

- time is easy to record
- errors or successful completion is harder
  define in advance what these mean

Do not combine efficiency measures with thinking-aloud.

- talking can affect speed & accuracy

Experimental design

- Introduction and examples
- The elements of experiments
- Designing an experiment
- Conducting an experiment
- **Gathering data**

Collecting data

- Records of what participants did
- Log data
  - Data of dependent variables (measures)
  - time, errors, etc.
- Table structure:

userid	group	trial_id	condition	executiontime	Errors

Experimental design

- Introduction and examples
- The elements of experiments
- Designing an experiment
- Conducting an experiment
- Gathering data

Mise en pratique

D’ici à la pause déjeuner

- Comparer l’efficacité de “smart keyboards”

- Se mettre d’accord sur les modalités d’entrée = types de clavier
- Description de l’expérience : variables, hypothèses, déroulé.

- Au retour de la pause / pendant la pause
  - Faire passer l’expérience à 2 personnes (chacun.e)

Mise en pratique

Après manger

- Analyse descriptive sur un échantillon de 24 personnes
  - Boxplot avec description des données, médiane, quartiles, distribution des données
  - présence (ou non) d’outliers
  - analyse des erreurs
- T-test et conclusion statistique