

# User Studies

Why and how to conduct

# Why User Studies are Necessary

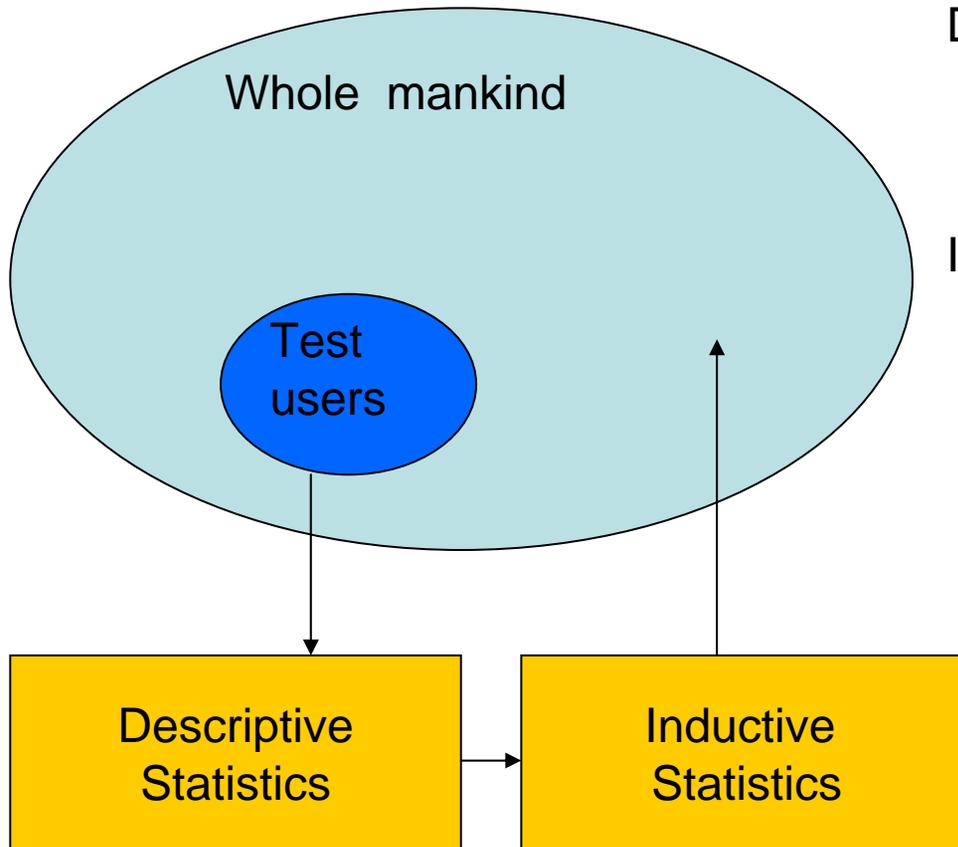
- To get a scientific statement (instead of personal opinion)
- To learn more (because of surprising results)
- To ensure quality in product development

## **Examples of scientific statements**

- Users are quicker using version A than using version B
- Users make 10% less errors when using version X than when using version Y
- 90% of the users can complete the transaction using version Y in less than 3 minutes
- On average users will be able to buy a ticket using version A in less than 30 seconds

# What Do We Get from a User Study?

- A result valid for all people



Test users:

must be representative

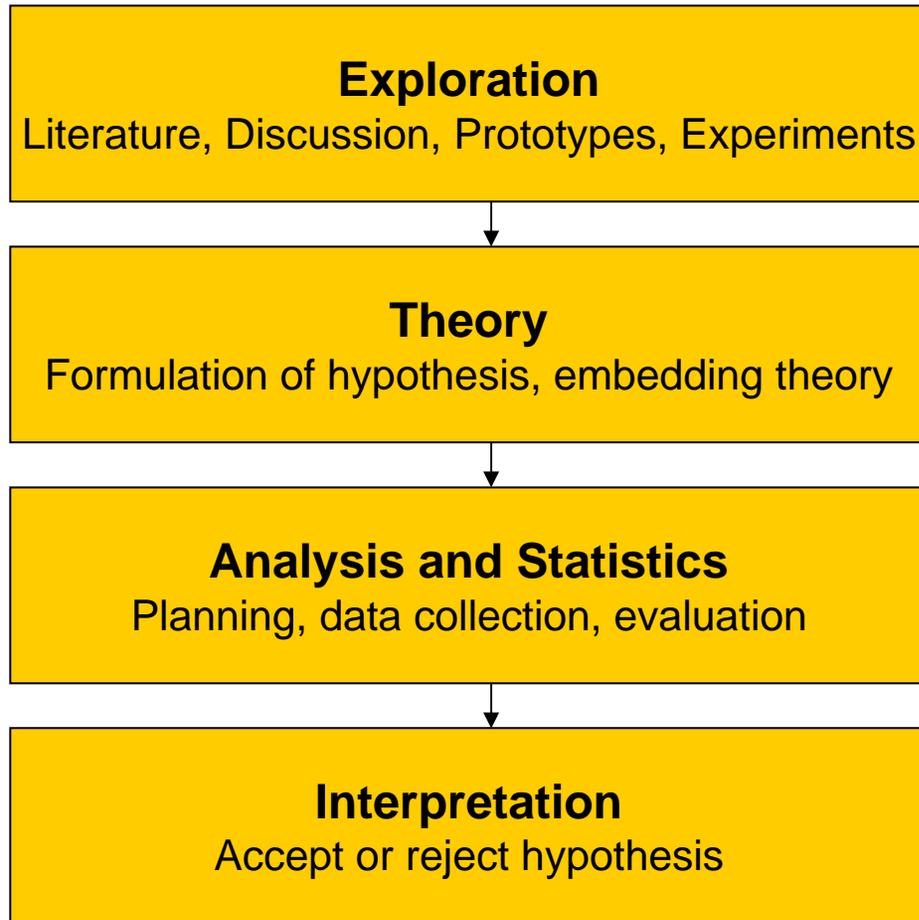
Descriptive statistics:

Tables, diagrams,  
means, etc.

Inductive statistics:

Ensure validity for the  
whole

# How to Do a User Study?



## Warnings

A user study takes a lot of effort – plan it carefully!

A user study without a hypothesis formulated BEFORE is worthless!

## Recommendations

Make sure all relevant data are recorded and friendly for evaluation.

Do a complete test with somebody – from data collection to evaluation

# What to Evaluate in HCI?

- The usability of a system!
- ... it depends on the stage of a project
  - Ideas and concepts
  - Designs
  - Prototypes
  - Implementations
  - Products in use
- Approaches
  - Formative evaluation – throughout the design, helps to shape a product
  - Summative evaluation – quality assurance of the finished product.

# Designing the Experiment

- System to test – specification of hard- and software, environment, ...
- Participants – representative for focus group, age, gender, culture, ...
- Variables – which variables have influence, which variables to measure, accuracy of measurements, disturbances
- Hypothesis – exact formulation
- Experimental Methods – Within groups, between groups, order of tests
- Statistical approach – how to evaluate the data

# Specification

- **The experiment should be set up to be reproducible!**

Describe and specify everything which is important for others to reproduce the experiment (leave out the unimportant things)

Typically: product name of commercial hard- and software in use, circuit schematics of self-build prototypes,

- **The environmental conditions must be equal for all users**

light conditions, atmosphere

- **The skills of the test users are part of the specification**

All participants have to be professional designers, the candidates should have no experience on using eye-trackers, ...

# Which Participants and how many

How many participants do we need?

- Depending on the project and the goals
- Depending on the set-up
  - measuring the login-in time of remote users vs. doing a full video observation for a 1 hour task
- Minimal size of about 10 participants

Participants should be representative for the user group

- Age, background, skills, experience, ...
- In most cases the other people on the team are NOT representative!

How to recruit participants

- Customer data base
- Market research services
- Volunteers (online, newspapers, etc) – this is risky because the people who will respond are often not representative

# Variables

## **Variables are manipulated and measured**

- Independent variables are manipulated
- Dependent variables are measured

## **The conditions of the experiment are set by independent variables**

- E.g. number of items in a list, text size, font, color,
- The number of different values used is called *level*
- The number of experimental conditions is the product of the levels

## **The dependent variables are the values that can be measured**

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

# Hypothesis

- Prediction of the result of an experiment
- Stating how a change in the independent variables will effect the measured dependent variables
- With the experiment it can be shown that the hypotheses is correct
- **Usual approach**
  - Stating a null-hypotheses (this predicts that there is not effect of the change in the independent variable on the measured variable)
  - Carrying out the experiment and using statistical measures to disprove the null-hypotheses
  - When a statistical test shows a significant difference, it is probable that the effect is not random

# Experimental Methods (1)

## **Within groups**

- Each user performs under all the different conditions
- Important to randomize the order of the conditions for each participant

### Problems and benefits

- Learning may influence results
- The effect of differences between individuals are lessened
- Fewer participants required

## **Between groups**

- One condition is selected for each participant
- Each user performs only under one condition
- Careful selection of groups is essential

### Problems and benefits

- Differences between individuals in different groups can play an important role (leads to large groups)
- No learning effects
- More user required
- Usually harder to show significance

# Experimental Methods (2)

- **State clearly how the experiment will be conducted**

Every test user must have same conditions. Don't learn how to conduct the experiment during the user study

- **State what to do in case of problems**

How to proceed, if the mobile phone of a user gets an incoming call during a test run? Stop the recording and repeat afterwards? Repeat the test run? Stop the test and don't use the data?

- **Randomize as much as possible**

If there are two test runs for two different text input methods, dice out the order. Perhaps people are a little bit faster in the second test run because they know the text already.

- **Specify the measurements**

Times can be recorded automatically by the testing software or stopped manually with a watch.

# Dealing with the Participants (1)

## **Consent form – get written consent from participants**

- Templates available
- May be checked with the legal department / review board

## **Let participants know what they are doing**

- What is the participant expected to do
- Procedure
- How long will it take, breaks
- What is the study for in general – but do NOT tell about the specific purpose or your hypotheses

## **Make sure they know**

- Quality of a UI / software is tested - they are NOT tested

## **Ethical and Privacy Issues**

# Dealing with the Participants (2)

## Participants Consent Form

**Study** \_\_\_\_\_ **Institution** \_\_\_\_\_

Name: \_\_\_\_\_ Date of Birth: \_\_\_\_\_

Email: \_\_\_\_\_

Phone: \_\_\_\_\_

I have been informed on the procedure and purpose of the study and my questions have been answered to my satisfaction.

I have volunteered to take part in this study and agree that during the study information is recorded (audio and video as well as my interaction with the system). This information may only be used for research and teaching purpose. I understand that my participation in this study is confidential. All personal information and individual results will not be released to third parties without my written consent.

I understand that I can withdraw from participation in the study at any time.

Date: \_\_\_\_\_ Signature: \_\_\_\_\_

# Dealing with the Participants (3)

For each user (typical example):

- Inform the user and sign the consent form
- Do a survey on demographics and questions of interest to the experiment
- Give the participant instructions on the task – do not reveal the hypotheses
- (optional) Make a training run - depends on the study
- Perform the actual run and measure variables
- (optional) do a survey on subjective measure
- Be available for questions of participants or for their (informal) feedback

# Typical Usability Testing Mistakes

- **Poor Task Design**

Ensure you use tasks that allow you to test core functionality and any areas you have identified as potentially problematic. Don't forget to record all relevant information.

- **Accidental Revelation**

Testers often provide too much information inadvertently. If users know what you want to evaluate, they try to fulfill your expectations.

- **Unprofessional Demeanor**

The appropriate attitude for a tester is one of professional detachment and neutrality. Using encouraging terms like 'Good' or 'Well done' may give the impression that the user, rather than the system, is being evaluated. Avoid the temptation to finish participants' sentences for them, or to verbalize what you think is in their minds. Instead, maintain your silence, listen, and be attentive.

(Parts are taken from [www.infodesign.com.au](http://www.infodesign.com.au))

# Example: T9 versus Multi-tap (1)

## Participants

- How many?
- Skills
- Computer user, Phone/T9 users?

## Independent variables

- Input method: 2 levels: Multi-tap and T9
- Text to input: 1 level: text with about 10 words

## Experimental conditions

- 2 conditions – T9 and Multi-tap
- User 1,3,5,7,9 perform T9 then Multi-tap
- User 2,4,6,8,10 perform Multi-tap then T9
- Different texts in first and second run?
- Particular phone model?
- Completion time measurement (e.g. stop watch or application?)
- Number of errors/corrections observed

# Example: T9 versus Multi-tap (2)

## Hypotheses

- H-1: Input by multi-tap is quicker than T9
- H-2: fewer errors are made using multi-tap input compared to T9

## Null-Hypotheses

- Assumes no effect
- H0-1: there is no difference in the input speed between multi-tap and T9
- H0-2: there is no difference in the number of errors made using multi-tap input compared to T9

## Experimental Method

- Within groups
- Randomized order of conditions