CS-200 - INTRODUCTION TO HUMAN-COMPUTER INTERACTION

Lecture 14
Evaluation

12/3/2021

HYPOTHESIS TESTING

HYPOTHESIS TESTING

- Make a prediction about the way users will perform with your design (a hypothesis)
- Test that prediction
- Helps design experiments and helps you answer research questions

HYPOTHESIS TESTING

The use of statistical procedures to answer research questions

Typical research question (RQ):

Are tasks quicker to complete using System A or System B?

Typical Hypothesis (H):

System A is faster to complete a task than System B

HYPOTHESIS TESTING

Null hypothesis:

There is no difference in the mean time to complete a task using System A vs. System B

Statistical tests seek to accept (sort of) or reject the null hypothesis i.e., a significant result means the null hypothesis is highly unlikely (see later in course)

ONE VS TWO TAILED

ONE TAILED

Makes a one sided prediction

System A is faster to complete a task than System B

TWO TAILED

Makes a prediction but leaves it open as to which side it will fall on

One of the systems will be faster than the other

VARIABLES

VARIABLES

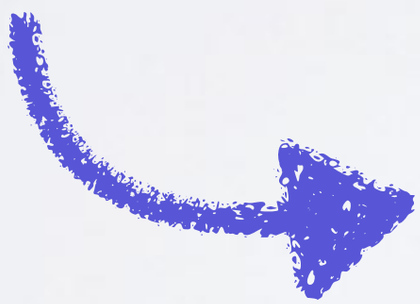
Independent variable (IV): what you are changing between iterations (e.g., interface, device, button layout etc.)

Dependent variable (DV): what you are measuring (e.g., task completion time, number of key presses, accuracy etc.)

VARIABLES

We want to evaluate the effect of **independent variable**(s) on the **dependent variable**

INDEPENDENT variable



Influences **CHANGE** in the

DEPENDENT variable

VARIABLES

In our example:

- System A is on a phone and System B is on a tablet. The app is identical on both systems.
- Independent variable: the screen size (i.e., phone or tablet)
- Dependent variable: time taken to complete a task

VARIABLES

Control variable: a circumstance (not being tested) that is kept constant throughout (e.g., background colour etc)

Random variable: a circumstance that is allowed to vary randomly (e.g., experience using a particular system, sunlight in the room, etc.)

TASKS AND MEASURES

TASKS AND MEASURES

Tasks: What participants will do during the study

- Access a representative set of functions in your design
- Make sure this is enough to allow you to measure

Measures: How you will measure performance

- Data capture methods
- What is success? (linked to hypotheses)

MEASURING PERFORMANCE

- Quantitative performance methods
    - Time to complete a task
    - Time to complete a task after a specified time away from the product
    - Number and type of errors per task
    - Number of errors per unit of time
    - Number of users making a particular error
    - Number of users completing a task successfully
    - Number of navigations to online help or manuals

EXPERIMENTAL DESIGN

EXPERIMENTAL DESIGN

- Which participants for which conditions in the experiment?
    - The experience of participating in one condition will affect the performance of those participants in other conditions
- Can avoid training effects by careful experimental design:
    - Between-groups design
    - Within-groups design
    - Pairwise design

BETWEEN-GROUPS DESIGN

- Also known as *between-subjects*, or *different-participant* design

- A single group of participants is allocated randomly to each of the experimental conditions

- So, different participants perform in different conditions

BETWEEN-GROUPS DESIGN

Condition 1

Condition 2

WITHIN-GROUPS DESIGN

- Also known as *within-subjects*, or *same-participant* design
- All participants perform in all conditions
- So, only half the number of participants are needed

WITHIN-GROUPS DESIGN

Condition 1

Condition 2

WITHIN-GROUPS DESIGN

- Need to think about bias (i.e., everyone using the designs in the same order could lead to learning effects.)
- So, split the participants into groups and rotate the order of designs for each group
- Counterbalance if possible. For example:
    - 2 designs == 2 possible order combinations
    - 3 designs == 6 possible order combinations

WITHIN-GROUPS DESIGN

Two designs (2x1 combinations)

Group 1

Group 2

WITHIN-GROUPS DESIGN

Three designs (3x2x1 combinations)

Group 1	A	B	C
Group 2	A	C	B
Group 3	B	A	C
Group 4	B	C	A
Group 5	C	A	B
Group 6	C	B	A

WITHIN-GROUPS DESIGN

Four designs (4x3x2x1 combinations)

Group 1	A	B	C	D
Group 2	A	B	D	C
Group 3	A	C	B	D
Group 4	A	C	D	B
Group 5	A	D	B	C
Group 6	A	D	C	B
Group 7	B	A	C	D
Group 8	B	A	D	C
Group 9	B	C	A	D
Group 10	B	C	D	A
Group 11	B	D	A	C
Group 12	B	D	C	A
Group 13	C	A	B	D
Group 14	C	A	D	B
Group 15	C	B	A	D
Group 16	C	B	D	A
Group 17	C	D	A	B
Group 18	C	D	B	A
Group 19	D	A	B	C
Group 20	D	A	C	B
Group 21	D	B	A	C
Group 22	D	B	C	A
Group 23	D	C	A	B
Group 24	D	C	B	A

DATA GATHERING

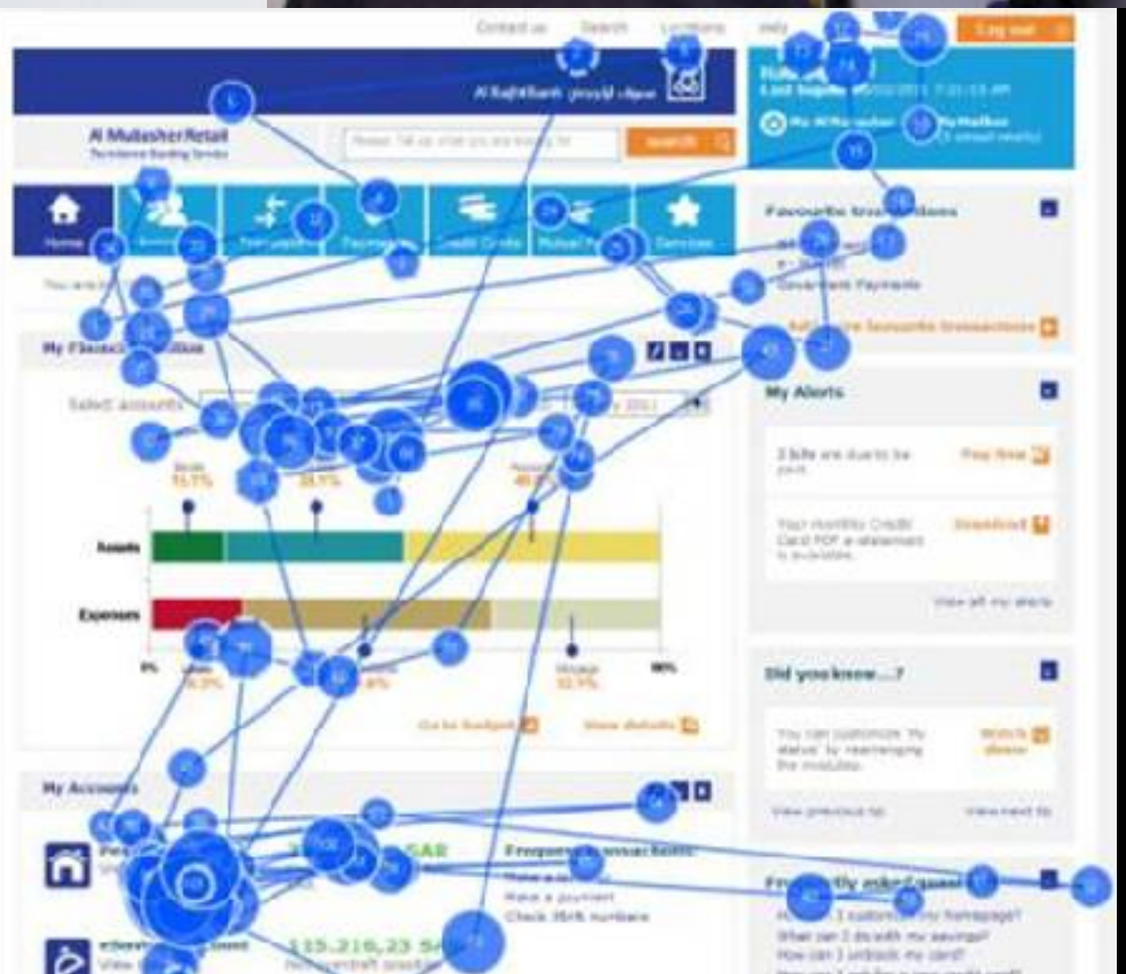
DATA GATHERING TECHNIQUES

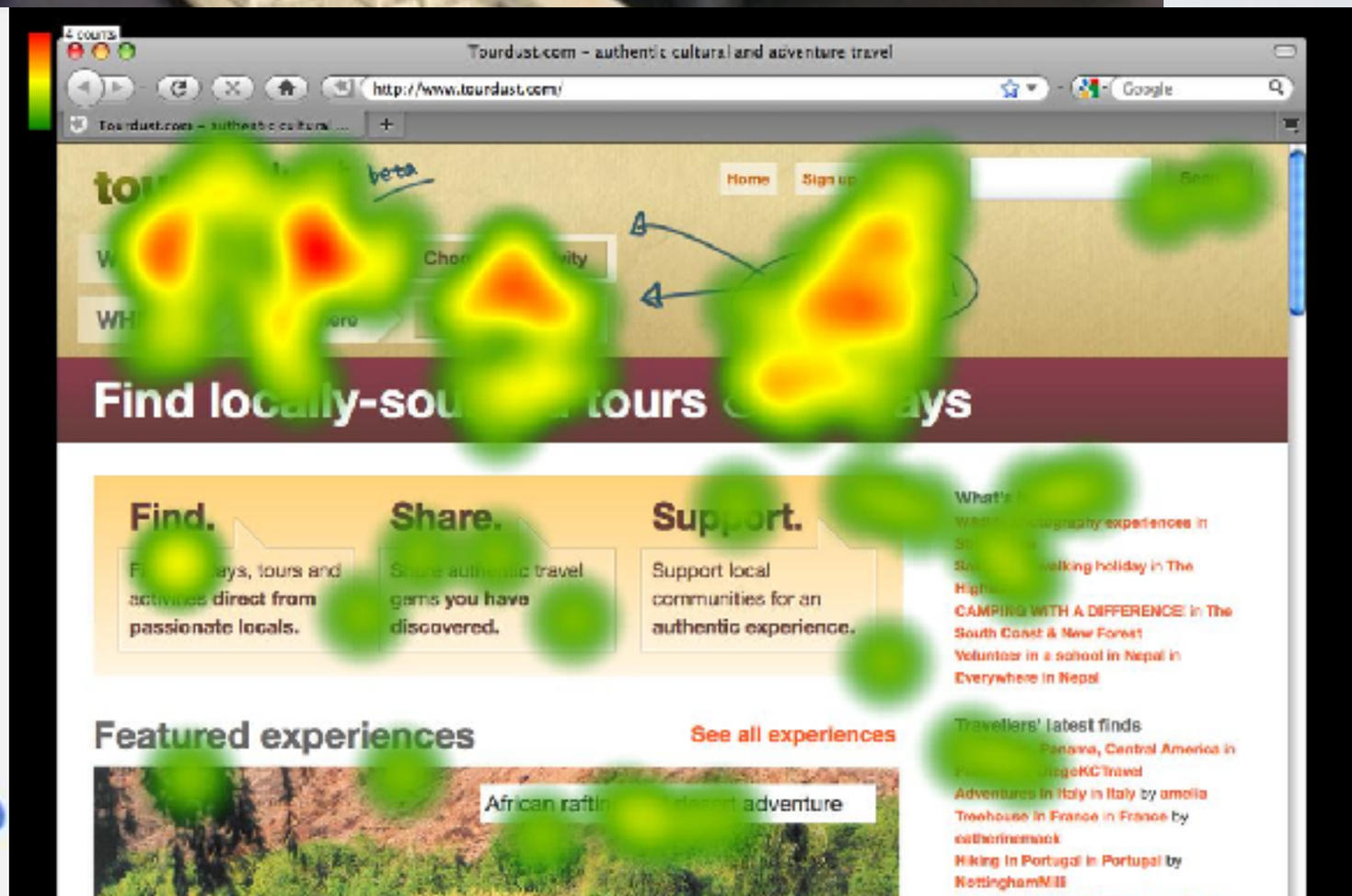
Covered this in previous lectures:

- Diary studies
- Interviews
- Questionnaires
- Observations
- Etc...

LOGGING

- Automatic logging of actions
- Easy to analyse (via scripts)
- But… watch out for privacy issues in longitudinal studies (later)




Tourdust.com – authentic cultural and adventure travel
http://www.tourdust.com/
Google
Home
Sign up
beta
Find.
Share.
Support.
Support local communities for an authentic experience.
Featured experiences
See all experiences
African raftin
adventure

PERSONAL LOGGING

ETHICS AND CONSENT

INFORMED CONSENT

- Participants need to be informed about what they are letting themselves in for (i.e., what the study entails)
- They also need to sign a consent form stating they are happy to proceed
- Two copies: one for you and one for them

ETHICS

- Ethical approval – required by the University when engaging with human participants
    - Bill of rights
    - Written or verbal instructions
    - Consent form
- Evaluated by the department's Ethics and Risk Assessment Committee:
    - Does the study protect participants' safety and personal data?
    - Are there any ethical or safety issues in its design?

INCENTIVES

It is common to provide participants with incentives to participate in a study

- Typically around £10 (gift voucher) per half hour
- Not a bribe – we still want honest answers!

OBSERVATION EFFECTS

- Hard to observe without skewing results
  - Hawthorne effect
- Withdraw to be inconspicuous
- Or build a relationship so presence is natural
  - Halo effect
- Be explicit in what you are studying
  - Ethics

NEXT TIME

Evaluation

- Data Analysis and Statistics