HUMAN-COMPUTER INTERACTION

EXPERIMENTAL DESIGN

Professor Bilge Mutlu

Computer Sciences, Psychology, & Industrial and Systems Engineering
University of Wisconsin–Madison

CS/Psych-770 Human-Computer Interaction
WHAT IS A HYPOTHESIS?
hypothesis |ˈhīpəθəsis|

noun (pl. -ses |-ˌsēz|)
a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation: professional astronomers attacked him for popularizing an unconfirmed hypothesis.

• Philosophy a proposition made as a basis for reasoning, without any assumption of its truth.

ORIGIN late 16th cent.: via late Latin from Greek hupothesis ‘foundation,’ from hupo ‘under’ + thesis ‘placing.’
HYPOTHESES

A statement of the predicted or expected relationship between at least two variables

A provisional answer to a research question

Has to define the variables involved

Has to define a relationship

Example:

**Question:** How does having information on the context of a caller affect whether the receiver picks up the call?

**Hypothesis:** Receivers will be more likely to pick up a call when they have information of their callers’ context than they will be when they do not.
HYPOTHESES

Variable 1 \rightarrow Relationship \rightarrow Variable 2

Information on Context \rightarrow Improves \rightarrow Call pick-up
GOOD HYPOTHESIS FORMATION

**Testable:** The means for manipulating the variables and/or measuring the outcome variable must potentially exist

**Falsifiable:** Must be able to disprove the hypothesis with data

**Parsimonious:** Should be stated in simplest adequate form

**Precise:** Should be specific (operationalized)

**Useful:**

- Relate to existing theories and/or “point” toward new theories
- It should lead to studies beyond the present one (often hard to determine in advance)
**VARIABLES**

Independent variable
- What is manipulated

Dependent variable
- What is measured

Variable 1
- Independent variable

Variable 2
- Dependent variable
VARIABLES

Control variables
  What is held constant
Random variables
  What is allowed to vary randomly
Confounding variable
  What correlates with the independent + dependent variable
WHAT IS THE DIFFERENCE BETWEEN CAUSALITY AND CORRELATION?
RELATIONSHIPS

Causal
   One variable depends on and is affected by the other

Correlational
   Two variables are affected by a third variable in the same direction
RELATIONSHIPS

Variable 1

Dependent variable

Variable 2

Causality

Causality

Variable 3

Dependent variable

Independent variable
CAUSAL

Heart rate

Condition

No drug
Sugar pill
Cocaine
CORRELATIONAL

Severity of cancer vs. Number of packs/day
EXPERIMENTAL DESIGN
EXPERIMENTAL DESIGN

Correlational research
Quasi-experimental research
Experimental research
CORRELATIONAL DESIGN

For studies examining the relationships between variables such as personality traits, work habits, gender, etcetera, the hypothesis is a specific statement about relationships

If when we observe an increase in X then we will also observe and increase (or decrease) in Y

Example questions:

Is there a relationship between smoking and lung cancer?

Is there a relationship between anxiety and test-taking performance?

Correlation does NOT imply causation
CORRELATIONAL DESIGN

Variable 1

Independant variable

Causality

Variable 2

Correlation

Variable 3

Dependent variable

Causality

Dependent variable
QUASI-EXPERIMENTAL DESIGN

Used when randomization is impossible and/or impractical

Separate participants based on some characteristic

   No random assignment

   E.g., Gender, occupation, verbal ability (VSAT)

Possible questions

   Do people with high verbal ability learn new languages faster?
QUASIS-EXPERIMENTAL DESIGN

- Group 1
- Group 2

Pre-experiment

Treatment

Post-experiment

Time
TRUE EXPERIMENTAL DESIGN

Studies in which variables are manipulated and outcomes measured, the hypothesis is a cause and effect statement

$Y$ will occur, when $X$ is manipulated

Examples

Students will remember more items from a word list if they learn the list in the quiet, rather than in the presence of intense music

Reading speed (words/minute) will change when font size is manipulated, such that reading speed will increase as font size is increased from 4 point to 20 point, but reading speed will decrease as font size is increased above 20 point
NUMBER OF VARIABLES

Single Variable
Only one independent variable
Cannot look at interactions

Multiple Variables
Two or more independent variables
If use **factorial design**, can look at interactions
Will require more participants (between) or time (within)
WITHIN VS. BETWEEN DESIGNS

Comparisons between conditions **within** participants

- Demands time
- Statistical power with smaller number of participants

Comparisons between conditions **across** participants

- Demands larger sample
- Avoids transfer effects
- Easier to avoid bias
FACTORIAL DESIGN

Suppose we are interested in the effect of both how gaming and computer use affect perception of robots.

Ideally: look at all 4 populations in one experiment.

High/Low computer use × Gamer/Non-Gamer

Why?

- We can learn more.
- More efficient than doing numerous single-factor experiments.
MAIN EFFECTS

Liking

Roomba: 7
Hoover: 1

Liking

Roomba: Decreasing
Hoover: Increasing
INTERACTION EFFECTS

Liking

Roomba

Women
Men

Hoover

Women
Men

Liking

Men

Women

Hoover

Roomba
RANDOM SAMPLING

Random Sampling

Choose participants randomly from the entire population

Allows generalization to population

Randomization allows the later use of probability theory and gives a solid foundation for statistical analysis

Avoid bias

The first six students who come in the lab might be highly motivated

Random Assignment

Random does not mean haphazardly

One needs to explicitly randomize

Random assignment at arrival, counterbalancing, matching
COUNTERBALANCING

Particularly important in within designs

Important because of transfer effects

  Taking part in earlier trials changes performance in the later trials

    Due to learning, fatigue, etc.

Makes within-subjects designs difficult to interpret
In within-subjects counter-balancing:

Possible **linear transfer effects**

Is the transfer from the 1st position to the 2nd position the same as the transfer from 2nd to 3rd position?

E.g., sometimes most learning happens in 1st trials

Always worry about asymmetrical transfer

Does A influence B more than B influences A?
Asymmetrical transfer

Effect of noise depends on order

People stick with the strategy they pick first

Or mix strategies

Use a between design
COUNTERBALANCING

Partial counterbalancing

Latin square

Every condition appears in every position equally:

Joe:   A  B  C
Mary:  C  A  B
John:  B  C  A
MATCHING

Try to reduce between-group differences

E.g., rank hearing as Good, Fair, Poor

Unmatched, could get;

Noisy:  Poor1, Poor2, Fair1

Quiet:  Good1, Good2, Fair2

Matched, get:

Noisy:  Poor1, Fair2, Good1

Quiet:  Poor2, Fair1, Good2
Suppose that some social measurements will be made in the morning and some in the afternoon.

If you anticipate a difference between morning and afternoon measurements:

- Ensure that within each period, there are equal numbers of subjects in each treatment group.
- Take account of the difference between periods in your analysis.

This is sometimes called “blocking.”
VERY BAD DESIGN

Week One

M | Tu | W | Th | F
---|----|---|----|---
C | C  | C | C  | C
C | C  | C | C  | C
C | C  | C | C  | C

Week Two

M | Tu | W | Th | F
---|----|---|----|---
T | T  | T | T  | T
T | T  | T | T  | T
T | T  | T | T  | T

T = treated, C = control, pink = female, blue = male
RANDOMIZED

<table>
<thead>
<tr>
<th>M</th>
<th>Tu</th>
<th>W</th>
<th>Th</th>
<th>F</th>
<th>M</th>
<th>Tu</th>
<th>W</th>
<th>Th</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

T = treated, C = control, pink = female, blue = male
STRATIFIED

Week One
M | Tu | W | Th | F
---|----|---|----|---
C | T  | T | C  | T
T | T  | C | C  | C
C | C  | T | T  | C
T | C  | C | T  | T

Week Two
M | Tu | W | Th | F
---|----|---|----|---
C | C  | T | C  | T
T | T  | T | T  | C
C | T  | C | T  | C
T | C  | C | T  | T

T = treated, C = control, pink = female, blue = male
FLOOR AND CEILING EFFECTS

Percent correct

2 times

4 times

# of study repetitions

Percent correct

2 times

4 times

# of study repetitions
BLINDING

Blinding

Measurements made by people can be influenced by unconscious biases

Ideally, dissections and measurements should be made without knowledge of the treatment applied

Single vs. double-blind designs

Internal controls

It can be useful to use the subjects themselves as their own controls (e.g., measuring the response after vs. before treatment)

Increased precision
REPRESENTATIVENESS

Are the subjects you are studying really representative of the population you want to study?

Ideally, your study material is a random sample from the population of interest.
SUMMARY

CHARACTERISTICS OF GOOD EXPERIMENTS

Unbiased
  Randomization
  Blinding
High precision
  Uniform material
  Replication
  Blocking
Simple

Protect against mistakes
  Wide range of applicability
  Deliberate variation
  Factorial designs
Able to estimate uncertainty
  Replication
  Randomization
QUESTIONS?
QUESTIONS?

Professor Bilge Mutlu
Computer Sciences, Psychology, & Industrial and Systems Engineering
University of Wisconsin–Madison

CS/Psych-770 Human-Computer Interaction