CS 498 KA
Experimental HCI & Interactive Technologies

Technical Introduction and Overview of Key Topics

Text Chapter 1: Introduction

"Unfortunately, most HCI researchers (and, in particular, students) do not have the luxury of time to wait and see if their novel idea takes off and is adopted at large." Though this happens – examples?

Yet as such:

"Proving that their HCI idea is a "good" one will need to be done in the context of a test that involves other people who try out the idea."

There are two different types of HCI test: <u>formal comparative experiments</u> and <u>exploratory usability evaluations</u> ."

Text Chapter 1: Introduction

"**Experiments** are objective tests that aim to demonstrate that the idea produces better results than an existing idea that performs the same function.

Experiments are more appropriate for ideas for which alternatives can readily be found and that are usually associated with well-defined tasks.

The outcome of an experiment is a conclusion indicating which idea results in better user performance."

NOTE: The above is a somewhat narrow view of the role that experiments play in scientific and engineering research (with its focus on "which is better?")"

In other areas of science (& even HCI), people often take a broader view. Why?

Text Chapter 1: Introduction

"**<u>Evaluations</u>** are exploratory tests that aim to show that the idea works in practice, in the context of typical uses.

Such evaluations (also called "usability studies") are more appropriate for larger, more complex pieces of interactive software for which it is difficult or impossible to find alternatives that fulfill identical functions and that typically support a wide range of user tasks.

The outcome of an evaluation is a list of suggestions for system improvement (as part of a formative test), or it can be confirmation that the system performs its function sufficiently well for it to be deployed."

Text Chapter 1: Introduction

<u>Conclusion</u>: The Organization and Focus of the (*E-HCI*) Text:

“[An] experiment is a one-off activity. It requires the cooperation of a large number of participants, all following the same experimental process.

Once it has started, it cannot be interrupted to correct any flaws in its design – not unless the whole experiment is to be redesigned and run again. It is important that an experiment be carefully designed because errors can cost a great deal of time and effort.

<u>The primary focus of this book is therefore on experiments</u> because their design is more complex and more risky than evaluations. In addition, good experimental design requires knowledge and skills typically not taught in a computer science or software engineering curriculum in the same way that usability evaluation is often covered as part of the iterative design cycle.”

Experimental Methods

- Three features of true experiments:

1. They test a hypothesis that makes a causal statement about the relations among variables

2. A comparison of the dependent measure is made at least for two levels of an independent variable

3. A high level of experimental control over relations between independent and dependent variables (and all other potential contaminating factors and noise sources).

Steps in Conducting an Experiment

- Define problem and hypotheses
  - Operationally define the **independent** (treatment) variables
  - Operationally define the **dependent** (response) variables
  - Define the expected relationship between these
- Specify the experimental plan
  - Check for the **exact definitions** of the above
  - Choose an experimental design
  - Determine the equipment, task(s), environment, subjects

Steps in Conducting an Experiment (cont'd)

- Conduct the study
  - Determine if a **pilot** study is necessary

- Analyze data
  - Exploratory data analysis (**look** at the data)
  - **Descriptive** statistical analysis
  - **Inferential** statistical analysis

- Draw conclusions
  - Ask **why?**
  - Formulate new questions for further research

Independent and Dependent Variables

- Operational definition:
  - Representation of ideas or concepts in terms of specific behaviors or concrete activities that anyone can witness or repeat

- Independent (treatment) variable(s):
  - Levels are established by the experimenter before the experiment and are thus independent of anything that happens in the experiment, i.e., it is **manipulated**

- Dependent (response) variable(s):
  - **If** a relationship exists, its value will depend on the level of the independent variable

Types of Experimental Designs

- Between-subjects design
  - Con: Large samples needed
  - Con: Higher noise (random error) across experimental conditions
- Within-subjects design
  - Pro: Increased sensitivity, smaller samples needed
  - Con: Carryover effects, practice and fatigue
- Factorial designs
  - Achieves economies by measuring values of dependent variables over combinations of independent variable values.

Types of Experimental Design

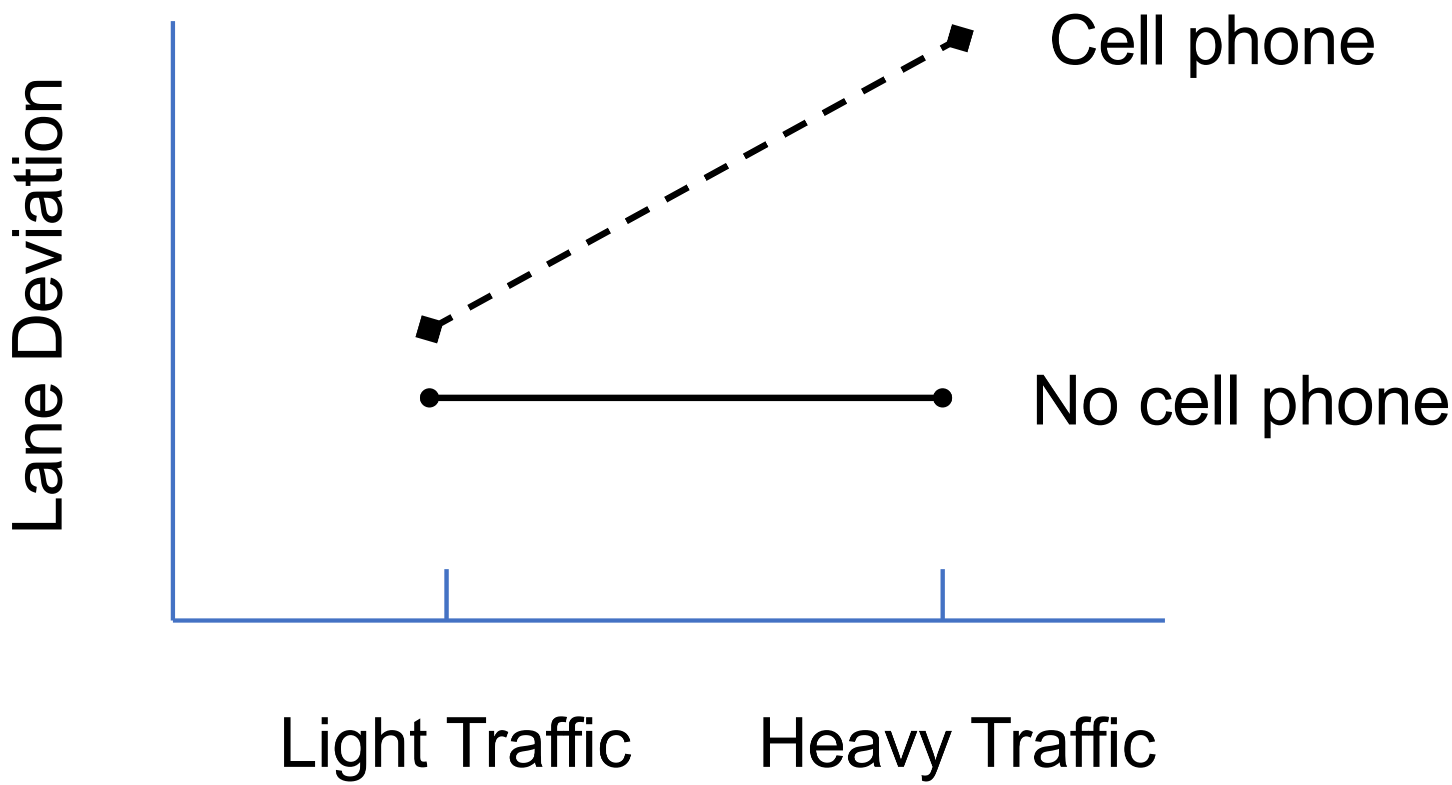
- Factorial designs
  - Combinations of **independent variable levels**
  - Examination of the effect of **each** independent variable (main effects)
  - Examination of **interactions** between independent variables (interaction effects)
  - The interaction effects **must be interpreted first**, because interactions effects may explain what appear to be main effects when viewed individually
  - In an interaction exists, the main effect must be discussed with a **qualifier**

Example: Cell Phones and Driving

	Driving Conditions: Light Traffic	Driving Conditions: Heavy Traffic
No cell phone	No cell phone while driving in light traffic	No cell phone while driving in heavy traffic
Cell phone	Use cell phone while driving in light traffic	Use cell phone while driving in heavy traffic

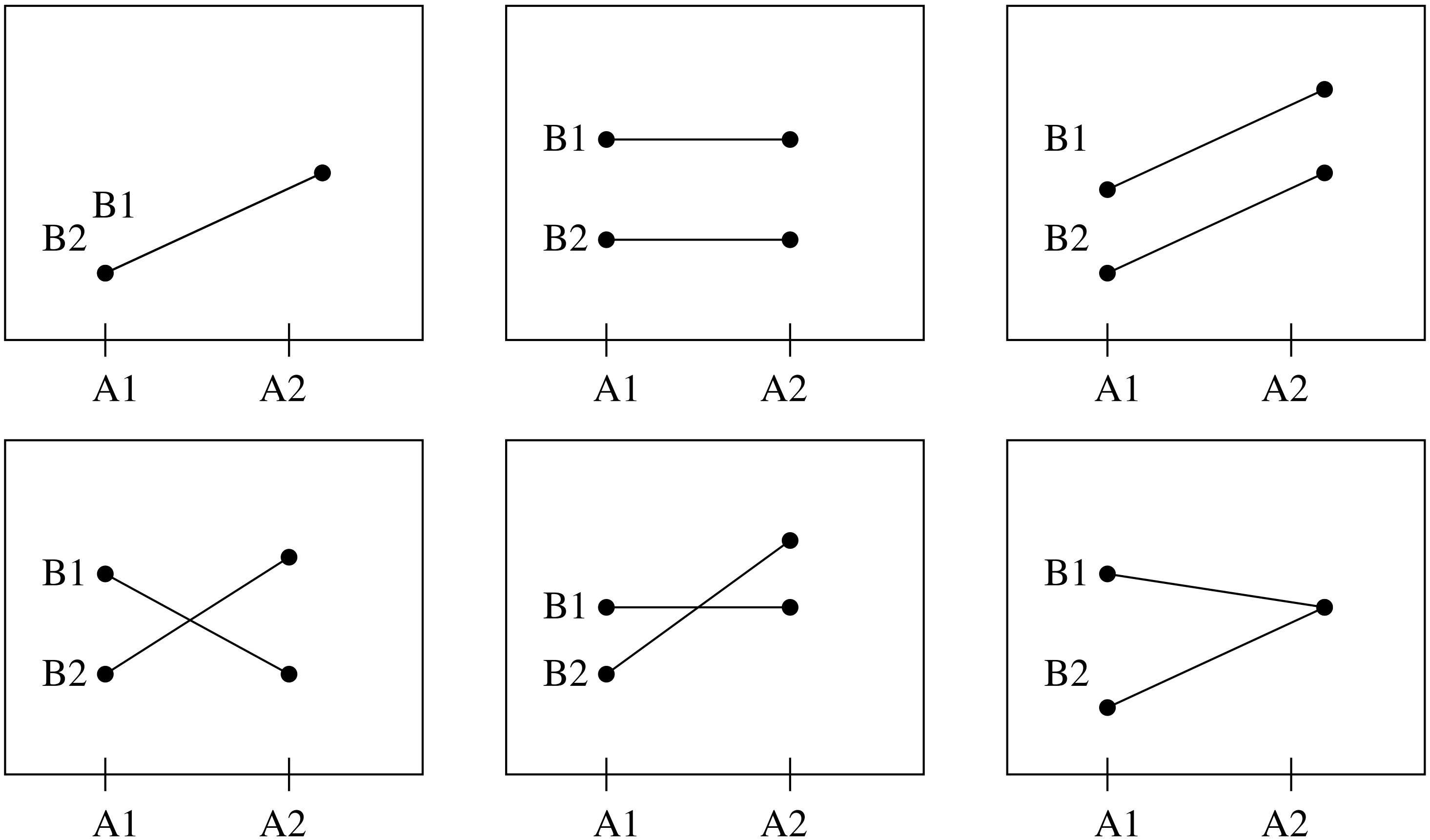
Four experimental conditions for a 2 x 2 factorial design

Example: Cell Phones and Driving



Interaction between cell phone use and driving conditions

Main and Interaction Effects



Validity and Control

- Validity: Establishing that the observed behavior is caused by the independent variable manipulated:
  - Construct validity
  - Internal validity
  - External validity
  - Face validity

Construct Validity

- **Construct validity** refers to the degree to which the researchers
  (a) manipulated the independent variable they wanted and
  (b) measured the dependent variable they wanted.

- Examples:

  Say, we are interested in the effects of fatigue in performance; how should fatigue be manipulated? How **not** to?

  What is the "performance" we are interested in? How should it be measured?

Internal Validity

- **Internal validity** refers to the situation or experiment where the causal or independent variable, **and no other extraneous variables**, caused the change in the dependent variables being measured

- It is often impossible to know about the existence of confounding variables and their influence on results.

- Examples?

External, or Ecological, Validity

- External validity refers to the degree to which the experimental results can be generalized to other situations, tasks, settings, and people.
- Threats to external validity:
  - Unrealistically simple experimental task
  - Unrepresentative subject population
- Examples?

Face Validity

- **Face validity** refers to the situation where the experimental setting or task has the same ‘look and feel’ as the real-word setting or task.

- Less important than the other types of validity (mostly!)

- Mainly influences subject’s acceptance of the experiment and later acceptance of the results

Hypothesis Testing

Null hypothesis (no effect)	Decision: True	Decision: False
True	Correct acceptance ($1-\alpha$)	Type I Error (α)
False	Type II Error (β)	Correct rejection ($1-\beta$)

The Type I Error or *p*-value

- More important than a "statistically significant" difference is
  - Which design is better?
  - By how much?
  - How much confidence can be placed on the difference
- Use common sense to extract meaning from data