Week 13-1 **Experiment**

SFWRENG 4HC3/6HC3 Human Computer Interfaces

** Slides adapted from previous and current instructors of COMPSCI/SFWRENG 4HC3/6HC3*

McMaster University

Week 13 Overview

- **Monday**
  - **Evaluation with Users: Experiment**
- **Wednesday**
  - Evaluation with Users: Experiment continue
- **Friday**
  - Additional Design Topics

Usability Test vs. Experiment

Usability Testing

- Goal: Check that the system being developed is usable by the intended user population for their tasks

Experiments

- Goal: Test hypotheses to discover new knowledge by investigating the relationship between two or more things
- Involves things like conditions, variables, etc.

Usability Test vs. Experiment

Usability Testing

- Few participants
- Formative results for design
- Not perfectly replicable
- Controlled conditions
- Procedure planned

Experiments

- Many participants
- Results validated statistically
- Replicable
- Strongly controlled conditions
- Experimental design

Experiment: Process

1. **Formulate a hypothesis**
2. Identify **independent, dependent variables**
3. Design a **controlled experiment**
4. **Check for:** Confounds, Validity, Reliability
5. Select **representative participants**
6. **Randomly assign** to conditions
7. Run experiment, collect data
8. Analyze results

Experiment: Hypothesis

A **suggested explanation** of a phenomenon

- "If I change A, then B will change in this manner…"

In experimentation, want a hypothesis to be **as specific as possible**

- Makes it easier to test

Experiment: Hypothesis

To test hypothesis, **must identify** what variables we think will lead to expected outcome

- Must identify how **manipulating these variables** will **result in expected outcome**
  - “If I provide keyboard shortcuts, users will be able to complete the tasks faster than with just menus”
  - “If I use pie menus rather than vertical context menus, users will be able to select items faster”
- **Clearly identify** which variables will influence what outcomes, and how

Experiment: Independent Variables

Independent variables are those directly manipulated as part of the experiment

Examples:

- Keyboard shortcuts: available, not available
- Menu type: pie or vertical context

- Everything else should be kept constant
- They are **not dependent** on anything in the experiment, **you change them** on purpose

Experiment: Dependent Variables

Dependent variables are those that change **in response to** the independent variables

Examples

- Completion time
- Error rate
- User preference
- Quality of user response

- Their value is **dependent on** the changes you made to the **independent variables**

Experiment: Relationships

Independent variables are assumed to **produce an effect on dependent variables' values** when manipulated

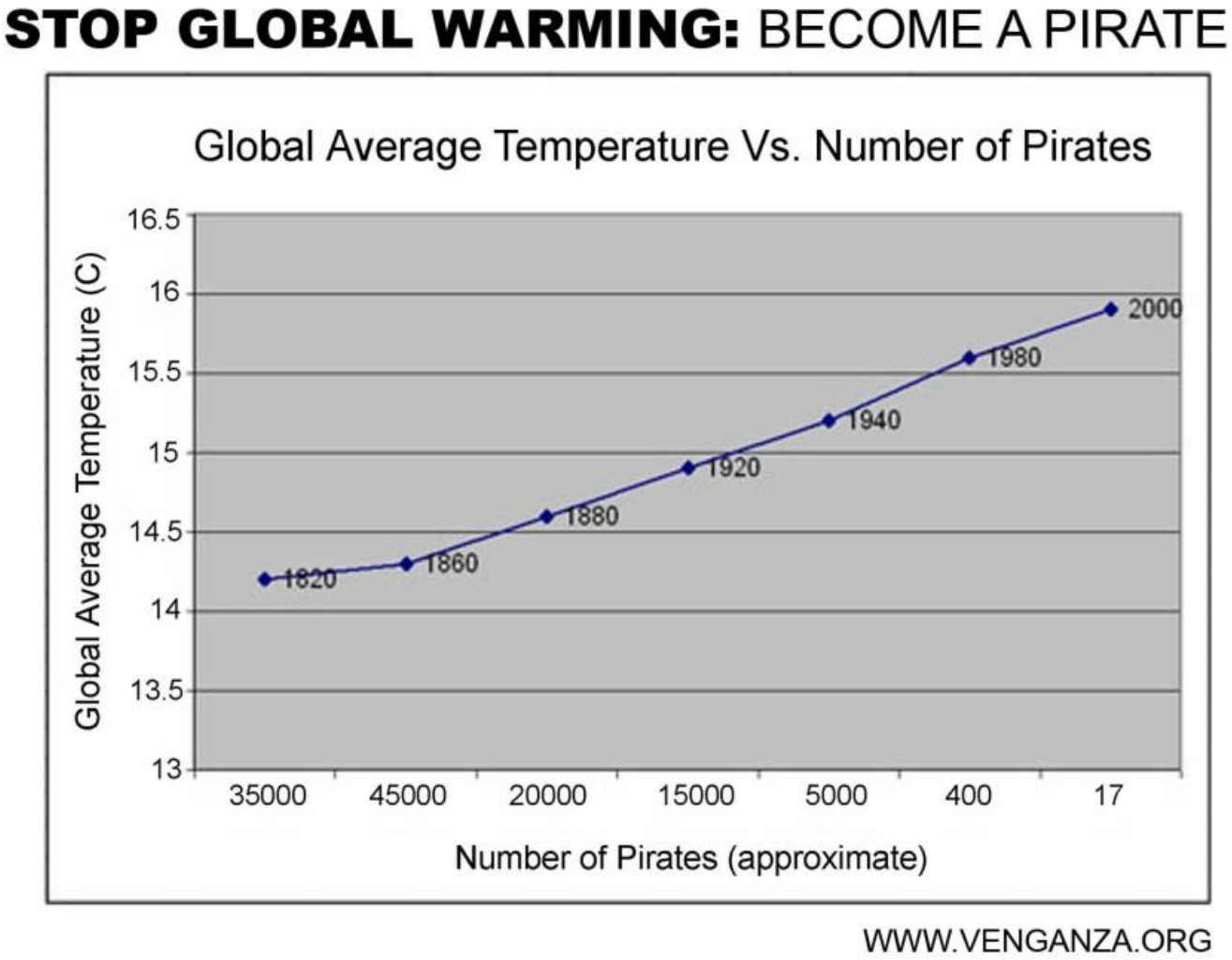
- "If I use pie menus rather than context menus, users will be able to select items faster"
- Pie menus vs. context menus (independent variables)
- Item selection speed (dependent variable)

Only manipulating independent variables increases our confidence that any observed changes in dependent variables due to changes in independent variables

Experiment: Causation vs. Correlation

- You notice that people seem to **be faster with your interface when they use a mouse, and slower with hotkeys**
  - Therefore, the mouse is more effective for your interface than hotkeys?
  - But, what if it was luck? what if there were other variables that you missed? What if you just noticed what you wanted? What if some other reason is why you have mouse vs hotkey users?
- **This is a correlation.** You noticed that two things seem to be linked. This does not mean that one caused the other
- You can **use experimental design to**, on purpose, change one variable to see if the other is impacted (test causation)

Experiment: Causation vs. Correlation



As the number of pirates decreases, global warming increases!

Experiment: Hypothesis Testing

In testing hypothesis, we are seeking to **reject the null hypothesis**

Null hypothesis

- There exists no relationship between manipulating the independent variables and the resultant changes in the dependent variables
- Example:
  - “There is no difference in selection speed between pie-menus and vertical context menus”

Experiment: Example

You are designing a **color scheme** for your interface, and recruit participants over an entire day. For morning participants, you **use interface A**; for afternoon participants, you **use interface B**.

- Participants for interface A take **a lot more time** to complete the task
- Therefore, you conclude that the color scheme impacts the task completion time

Experiment: Nuisance Variables

Any other factors that can affect the dependent variables?

Examples

- Time of day
- Handedness

- Goal is to have as few of these as possible
- Use techniques to mitigate the effect (e.g., **counterbalancing**)

Experiment Design

Need **at least** two conditions

- **Control condition**
  - No experimental manipulations performed
- **Experimental condition**
  - Experimental variable is manipulated

Results are **compared between two conditions**

- Statistics tell you whether the differences are expected randomness or not

Experiment Design: Considerations

Validity

- Are we measuring what we say we are measuring?

Reliability

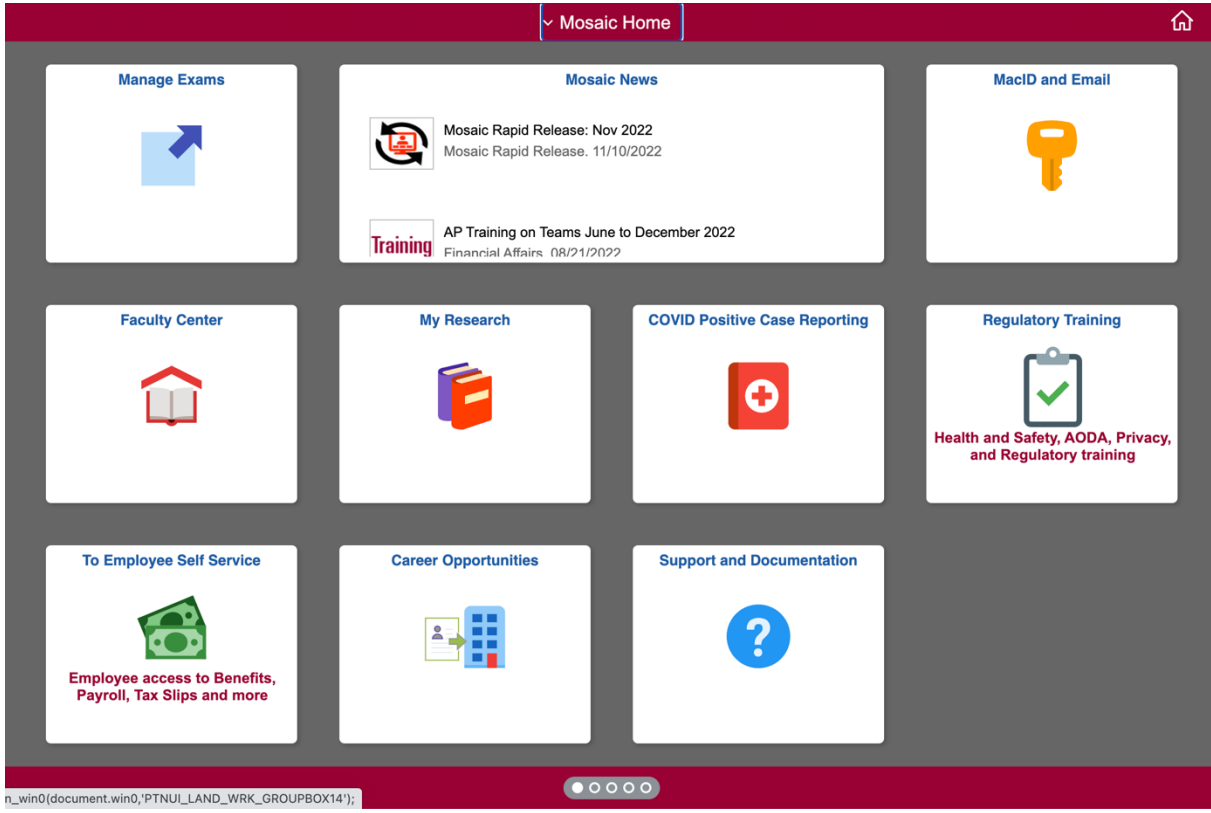
- If we run the experiment multiple times, will we get the same results?

Confounds

- Are there variables we didn't control for which may be influencing the results we're obtaining?

Experiment Design: Example

- You have designed a new interface for Mosaic
- You recruit students to test your interface
- For each participant, you give them Mosaic first to do the task, and then they use your new interface second
- You find your new interface is faster



What is the problem here?

Experiment Design: Effects

- Experience gained from using the first interface **affects** how they think about and use the second interface
  - Validity problem

- Too much testing means participants **get tired of testing**
  - Mix it up: for some participants, A then B; for others, B then A

Experiment Design: Validity

Internal validity

- The changes in the dependent variables are caused by the independent variables

External validity

- Results can be generalized to other settings, populations, tasks, etc.
- There is often **a trade-off** between the two
    - The more tightly you control the experiment (to increase internal validity), the less generalizable the results

Ecological validity

- To what extent do the study conditions mimic those in the real world
- Related to external validity, but not the same

Experiment Design: Example

- You design two computer games for children, and bring it to a school to test
- The first 10 students that complete their homework are sent to your testing office for the first game
- The next 10 students that complete their homework are sent to play the second game
- You find that the first game is much easier than the second

What is the problem here?

Experiment Design

Participant pool

- Are the study participants representative of the intended user population?

Modify **only one** thing between conditions

Bias

- Ensure experimenter is not biasing subject
- Blind vs. double-blind studies

Experiment Design: Considerations

How will participants be assigned to conditions?

Two options:

- Between-subjects
- Within-subjects

Experiment Design: Between-Subjects

- Each participant does **one** of the experimental conditions
- Doesn't account for **individual variability** (potentially bad)
- Need **more** participants
- **No learning effects** (good)

Experiment Design: Within-Subjects

- Each participant completes **all** experimental conditions
- Better - able to account for individual differences
- Requires **fewer** participants
- Allows participants to **make direct comparative statements**
- **Learning effects** are possible
  - To account for these, order of conditions are usually counterbalanced

Number of Participants

The more participants the better

- Increases confidence of findings

For large effects, fewer participants needed to achieve statistical significance

- The smaller the effect, the more participants needed

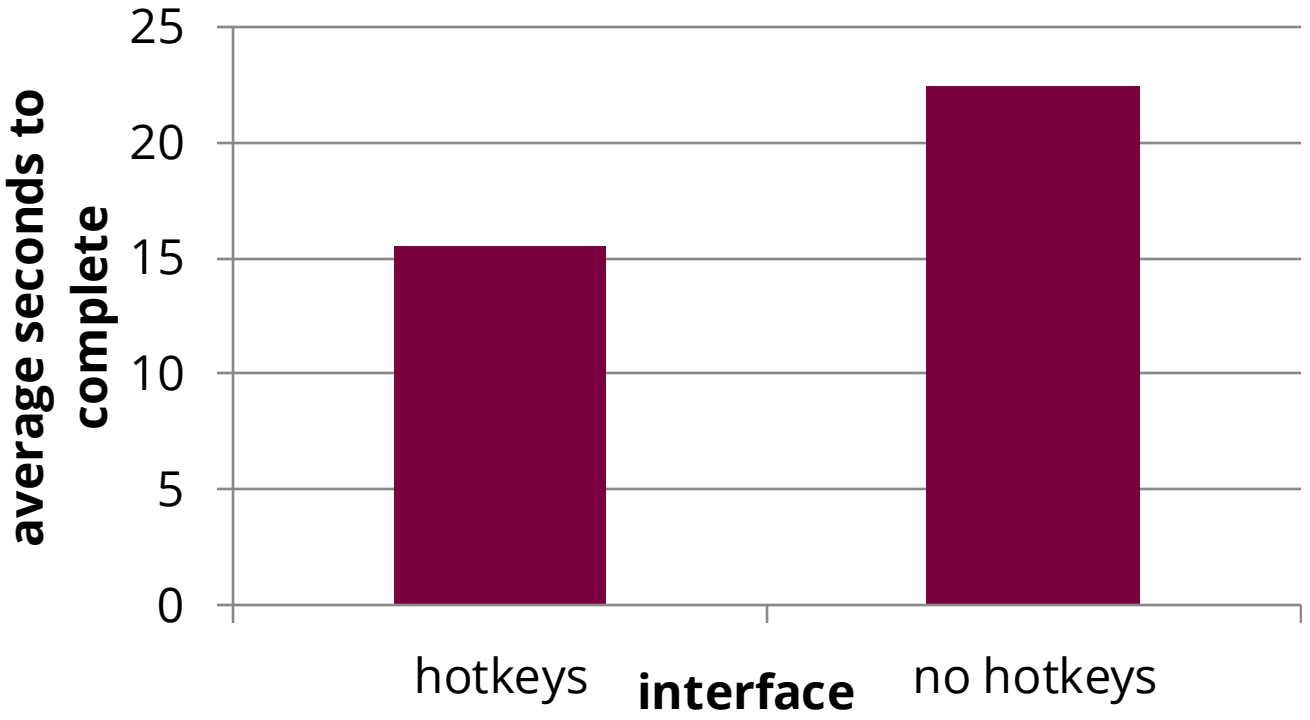
Experiment Design: Analysis

Which is faster?

We don’t know with what’s shown here..

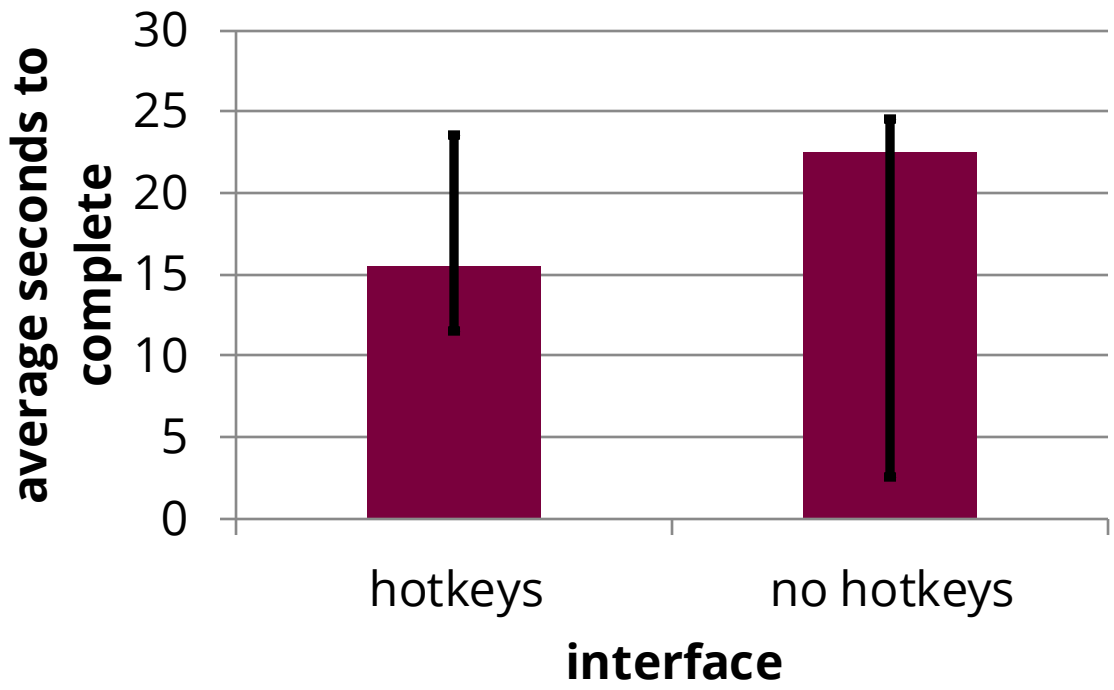
Statistics!

General idea: compare to random chance



Analysis: Statistics

- Statistics can use models of randomness to estimate **how likely this is by chance alone**
- This estimate is called a **p value**
- **p=0.05** means a 5% chance it matched expected randomness



Analysis: Statistics

Consult your friendly neighbourhood statistician

Different tests include:

- t-tests, ANOVA, CHI-squared, sign test
- Important to pick the right one for your data

Experiment: Pros and Cons

Advantages:

- High confidence in the phenomena that we observed

Disadvantages:

- Confidence applies to only narrow phenomena
- Can be difficult to create the right motivation for participants
- Can’t see things like workarounds that participants develop over time in the field
- Results can be statistically significant without being practically significant

Experiment: Takeaways

Proper study design is **very hard** – you need expertise, mentorship, iteration with trial and error, etc.

- Get a mentor, do a research project or get an advanced degree

Often, a study fails to find a result, but in retrospect **you can see confounds, nuisance variables**

- Re-design and re-conduct the study to fix it
- Negative results are not bad results (mixed-method)

Experiment: Example

- **Goal:** understand children's interactions with various speech interface design (personification, personalization)
- **Independent Variables:** age (child vs. adult), age groups (5-7, 8-12)
- **Dependent Variables:** Interaction strategies, task performances, design preferences
- **Setup:** Three conditions, within-subject design, counterbalanced order

Svetlana Yarosh, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, Ye Yuan, and A. J. Bernheim Brush. 2018. Children asking questions: speech interface reformulations and personification preferences. In Proceedings of the 17th ACM Conference on Interaction Design and Children (IDC '18).

Week 13 Overview

- **Monday**
  - ~~Evaluation with Users: Experiment~~
- **Wednesday**
  - **Evaluation with Users: Experiment continue**
- **Friday**
  - Additional Design Topics