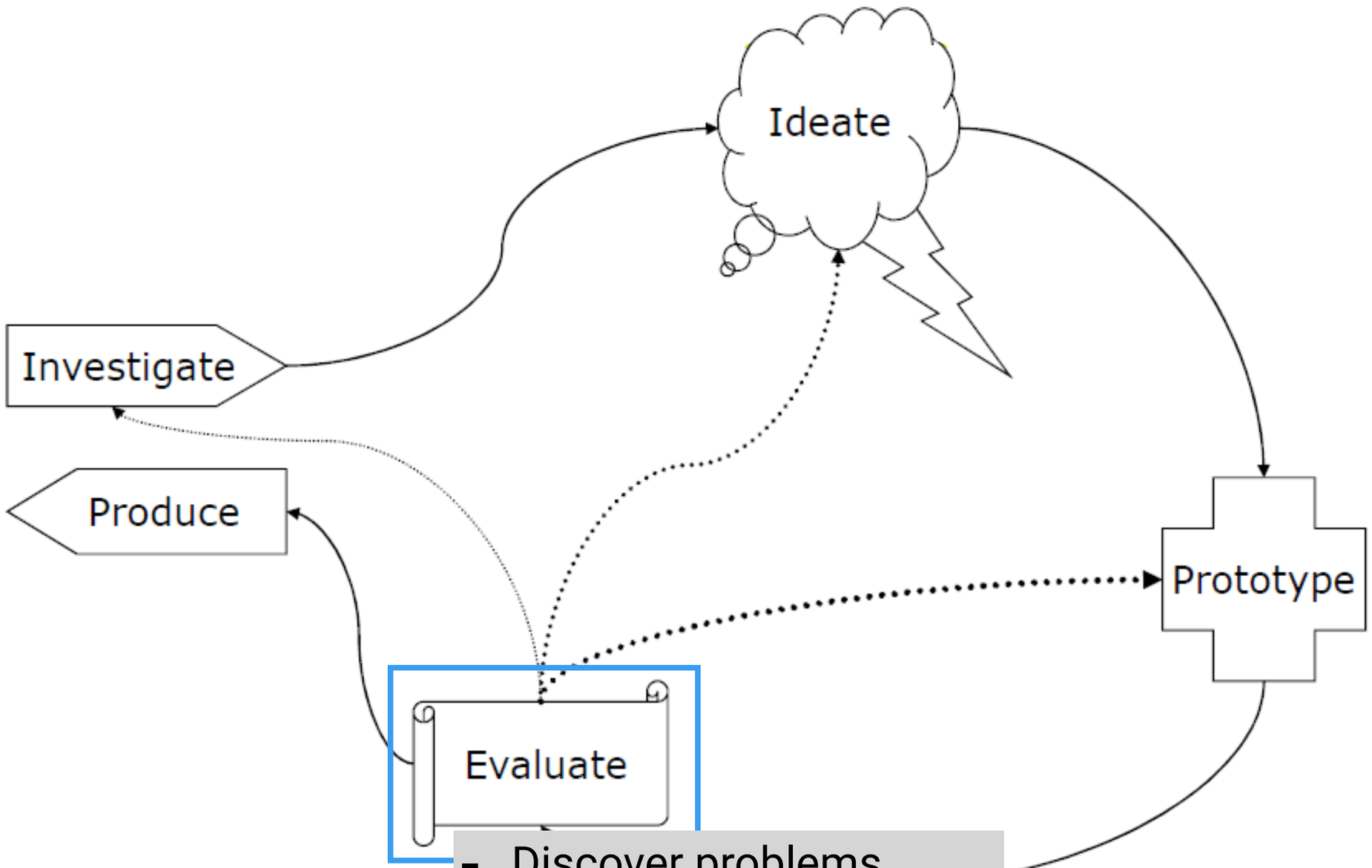
Human-Computer Interaction

CPSC 481 - Winter 2019

Usability Evaluation

With acknowledgements to Tony Tang

Evaluate



- Discover problems
- Assess progress
- Determine next steps

Why evaluation?

- Automated processes can find bugs, but not usability issues
- Evaluation gives you a way to move forward
  - What needs to be fixed, added, removed?
- Answers to two questions:
  - Did we build the right thing?
  - Did we build the thing right?

Usability Test: Essentially

- Bring in real users
- Have them complete tasks with your design, while you watch **with your entire team**
- Use a think-aloud protocol, so you can “hear what they are thinking”
- Measure
  - Task completion, task time
  - Satisfaction, problem points, etc.
- Identify problems (major ones | minor ones)
- Provide design suggestions to design/engineering team
- Iterate on the design, repeat

Corel Paper Prototype Test

- http://www.youtube.com/watch?v=ppnRQD06ggY

Three Basic Usability Test Protocols

- Think-Aloud Protocol

- Co-Discovery Protocol

- Conceptual Model Extraction

Think-aloud protocol

- As participants complete a task, you ask them to report
  - what they are thinking
  - what they are feeling
  - rationale for their actions and decisions
- **Idea:** rather than interpret their actions/lack of action, you can actually understand why they are doing what they are doing

Think-aloud protocol

- *What's weird*:

- People are not normally used to saying things out loud as they work.

- They may also be embarrassed to say things out loud.

Co-discovery Learning protocol

- **Main idea**: remove the awkwardness of think-aloud

- Two people sit down to complete tasks
- Only one person is allowed to touch the interface
- Monitor their conversation

- **Variation**: use a semi-knowledgable “coach” and a novice (only the novice gets to touch the design)

Conceptual Model Extraction

- Show the design, but don’t say how it works

- Ask the user to explain
  - function of each element
  - how they would perform a particular task

Conceptual Model Extraction

- Initial conceptual model (before they use it)
- Formative conceptual model (after they've used it)
- *Good for*: eliciting a user's understanding before and after use
- *Bad for*: understanding exploration and learning

CPSC 481 - Winter 2019 - Tiago Silva da Silva

Learning Objectives

- By the end of this lecture, you should be able to:
  - Identify and label several types of biases as they relate to internal validity in experiments
  - Describe several methods of addressing internal validity by changing how a study is run
  - Understand how this usability testing framework can be applied in research

Experimental Validity

Experimental Validity

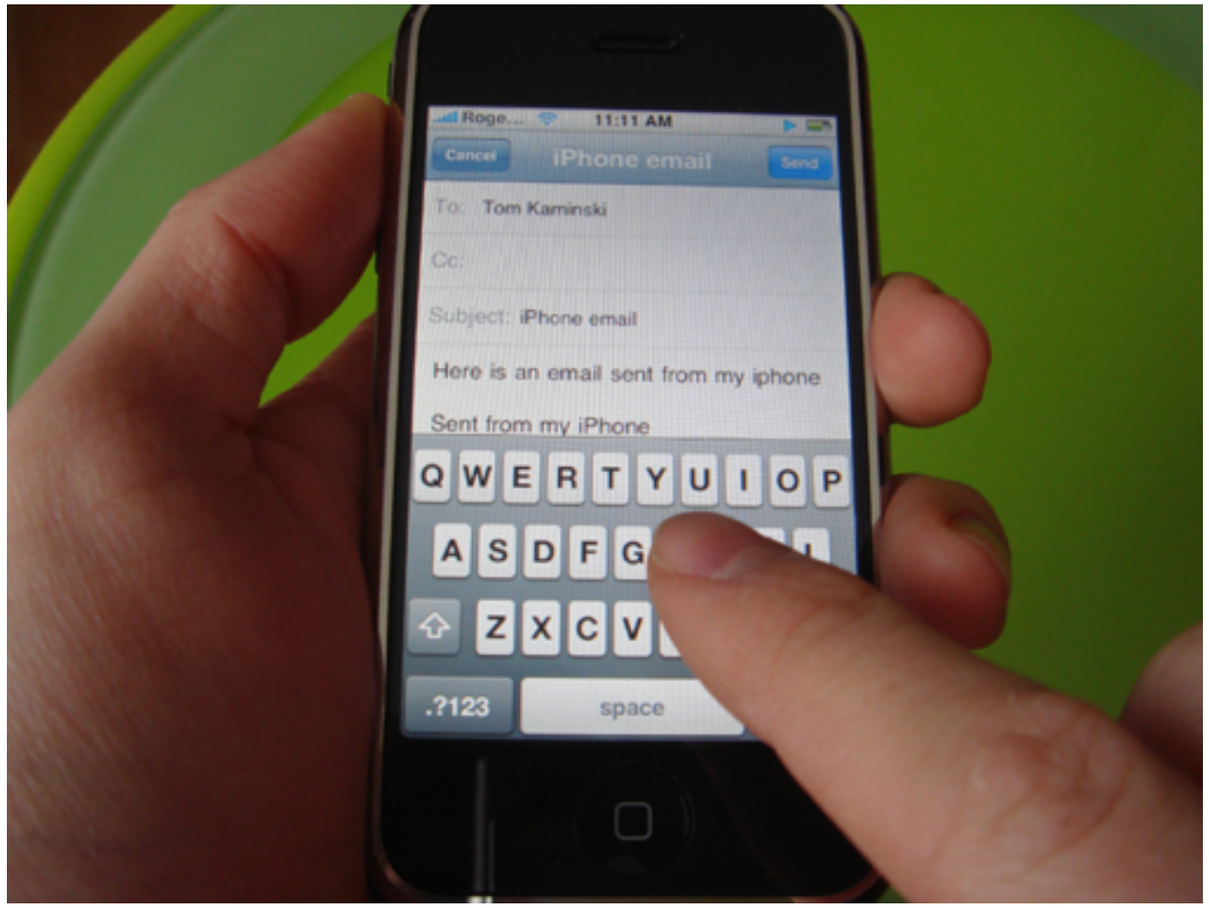
- **External validity » realism**
  - Across situations
  - Across people

Experimental Validity

- **External validity » realism**
  - Across situations
  - Across people
- **Internal validity » integrity**
  - Confound
  - Selection bias
  - Learning effects
  - Experimenter bias
  - Demand characteristics

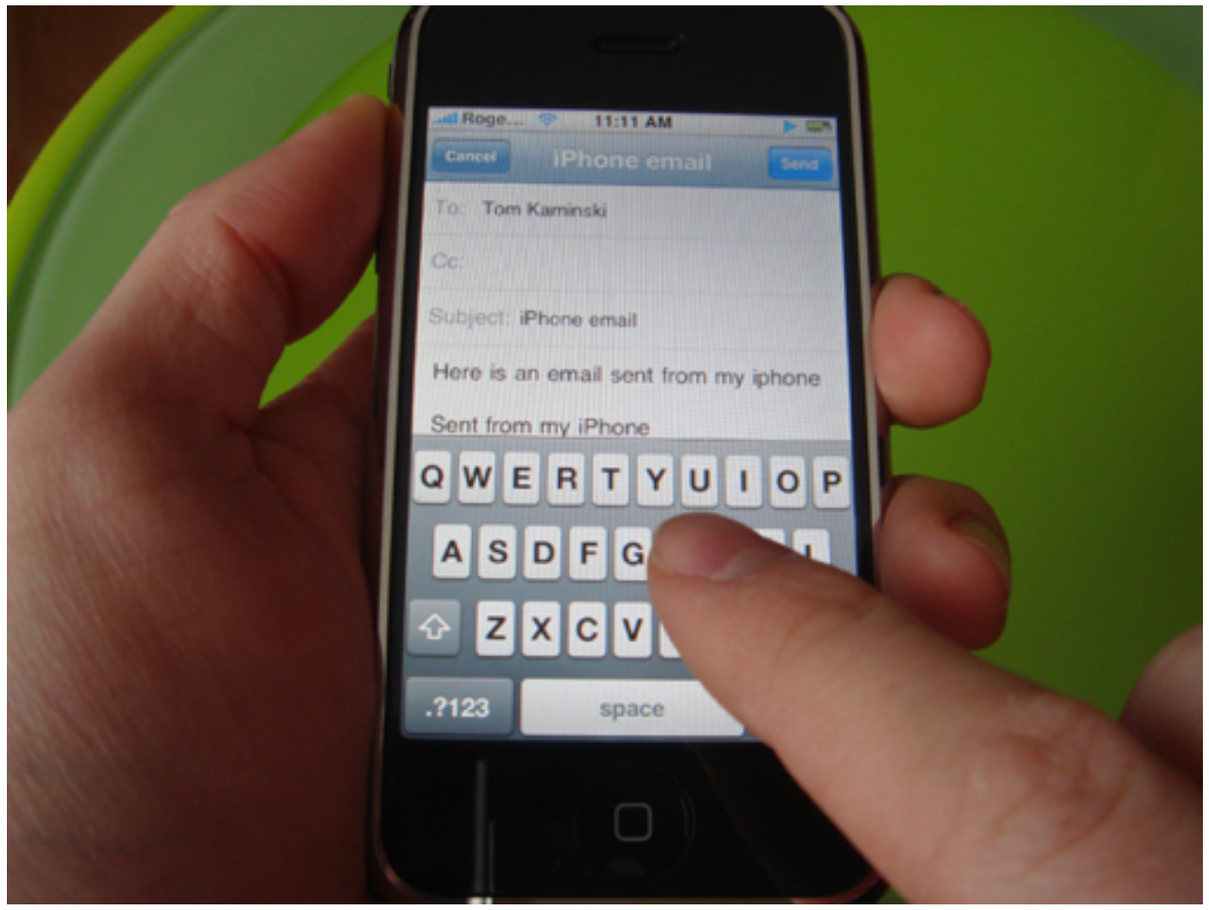
Imagine this test...

- Design a typing interface for use while driving cars.



Imagine this test...

- Design a typing interface for use while driving cars.
- Bring people into the lab, put them at a desk.
- Ask them to write an email, and time how long it takes.



External Validity >> across situations

- Does the test situation match the situation that the design will be used in?

External Validity >> across situations

- Does the test situation match the situation that the design will be used in?

- Does it match at least in critical ways?

External Validity >> across situations

- Does the test situation match the situation that the design will be used in?

- Does it match at least in critical ways?

- What are aspects that are different?

External Validity >> across situations

- Does the test situation match the situation that the design will be used in?

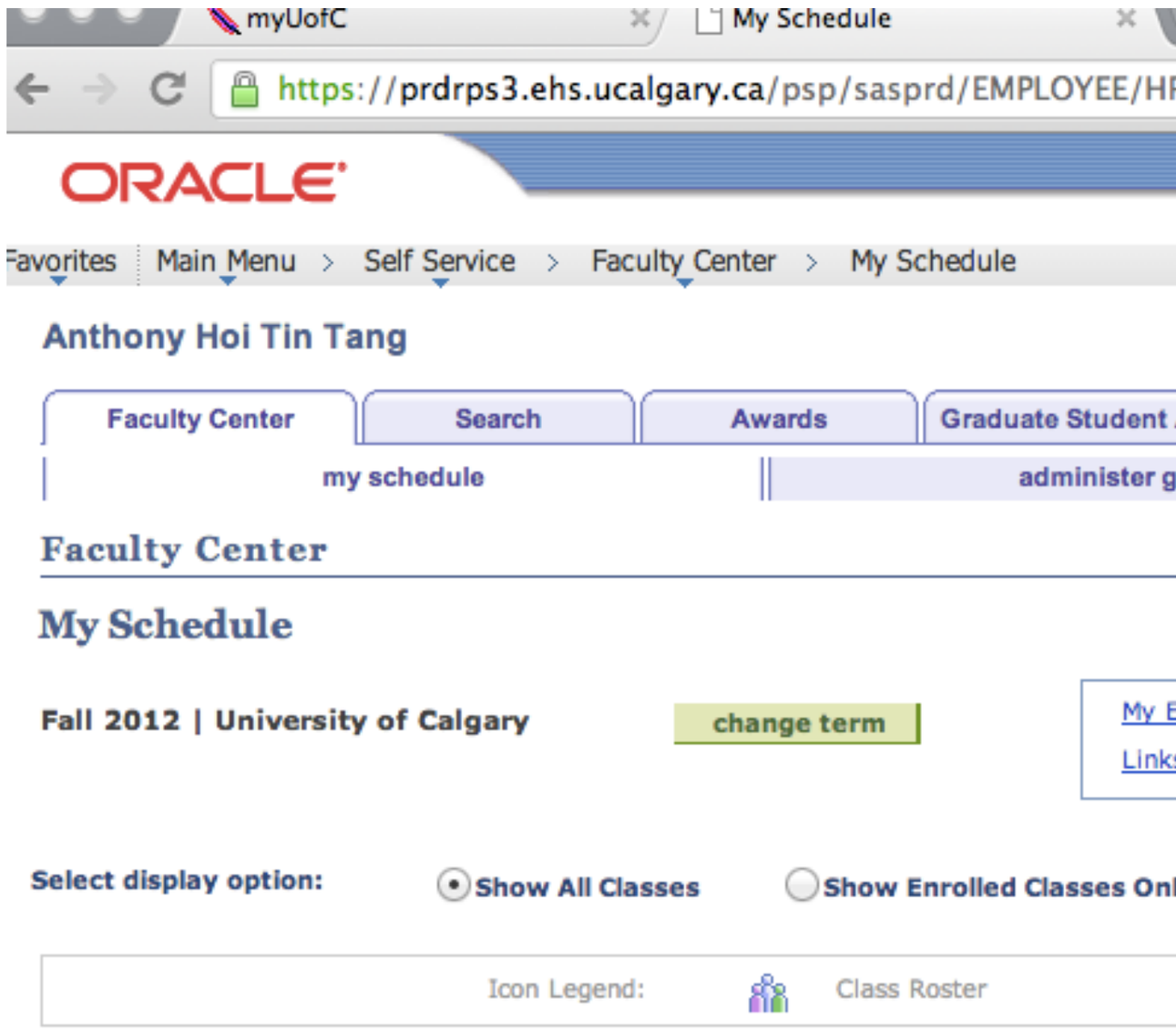
- Does it match at least in critical ways?

- What are aspects that are different?

- **Artificiality**

Imagine this test…

- Recruiting developers of PeopleSoft, ask them to register for courses.
- Because they can register for their courses within 5 minutes, the interface is deemed usable.



External Validity >> across people

- Are test subjects representative of the target user population?

External Validity >> across people

- Are test subjects representative of the target user population?

- Is it a randomly selected group, or are there constraints on how the group is selected that may affect test results?

External Validity >> across people

- Are test subjects representative of the target user population?

- Is it a randomly selected group, or are there constraints on how the group is selected that may affect test results?

- **Generalizability across a population.**

Experimental Validity

- **External validity » realism**
  - *Across situations*
  - *Across people*
- **Internal validity » integrity**
  - Confound
  - Selection bias
  - Learning effects
  - Experimenter bias
  - Demand characteristics

Imagine this test...

- You are designing a *colour scheme* for your interface, and recruit participants for the entire day.

  - For morning participants, you use interface **A**; for afternoon participants, you use interface **B**.

Imagine this test...

- You are designing a *colour scheme* for your interface, and recruit participants for the entire day.
    - For morning participants, you use interface **A**; for afternoon participants, you use interface **B**.
- Morning participants seem to have no problems with the interface, but participants take a lot more time to complete the task.

Internal Validity >> confound

- When you are testing something, and changing *one* aspect of the test (i.e. a variable), if something else changes along with that variable, then you have a confound.

- This means that you cannot tell what is causing the difference.

Imagine this test...

- 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.

Imagine this test...

- 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.

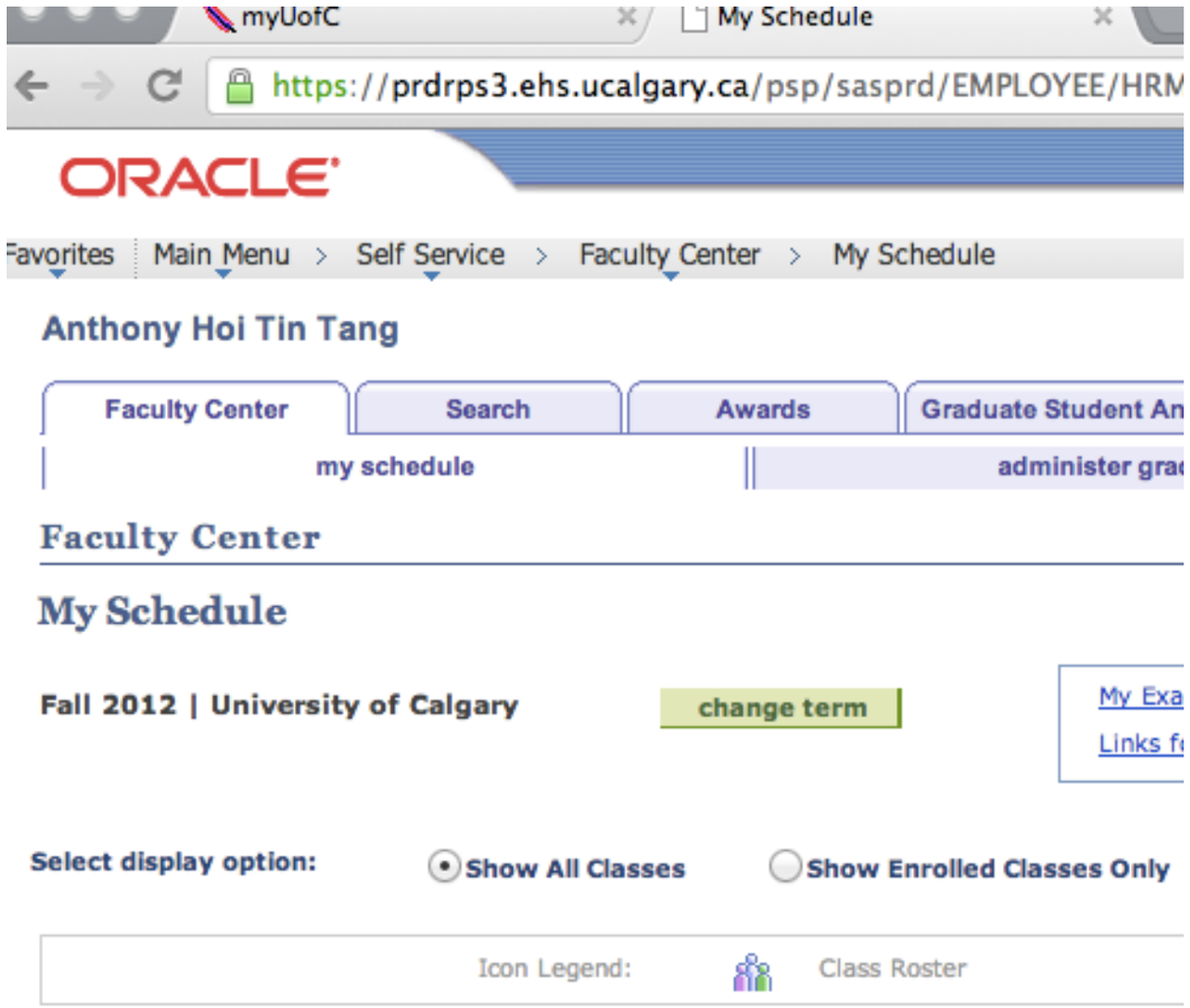
- They find both games very easy to play.

Internal Validity >> selection bias

- Systematic, non-random sampling of the population distorts your ability to generalize from the results.

Imagine this test...

- You have designed two new interfaces for PeopleSoft. You recruit students to test your interface.
- For each participant, you give them your least favourite interface first to complete the task, and then you give them your favourite interface second.



Internal Validity >> learning | fatigue effects

- Experience gained from using the first interface (to conceptual model) affects how they think about and use the second interface.

Internal Validity >> learning | fatigue effects

- Experience gained from using the first interface (to conceptual model) affects how they think about and use the second interface.

- Too much testing means participants get tired of testing.

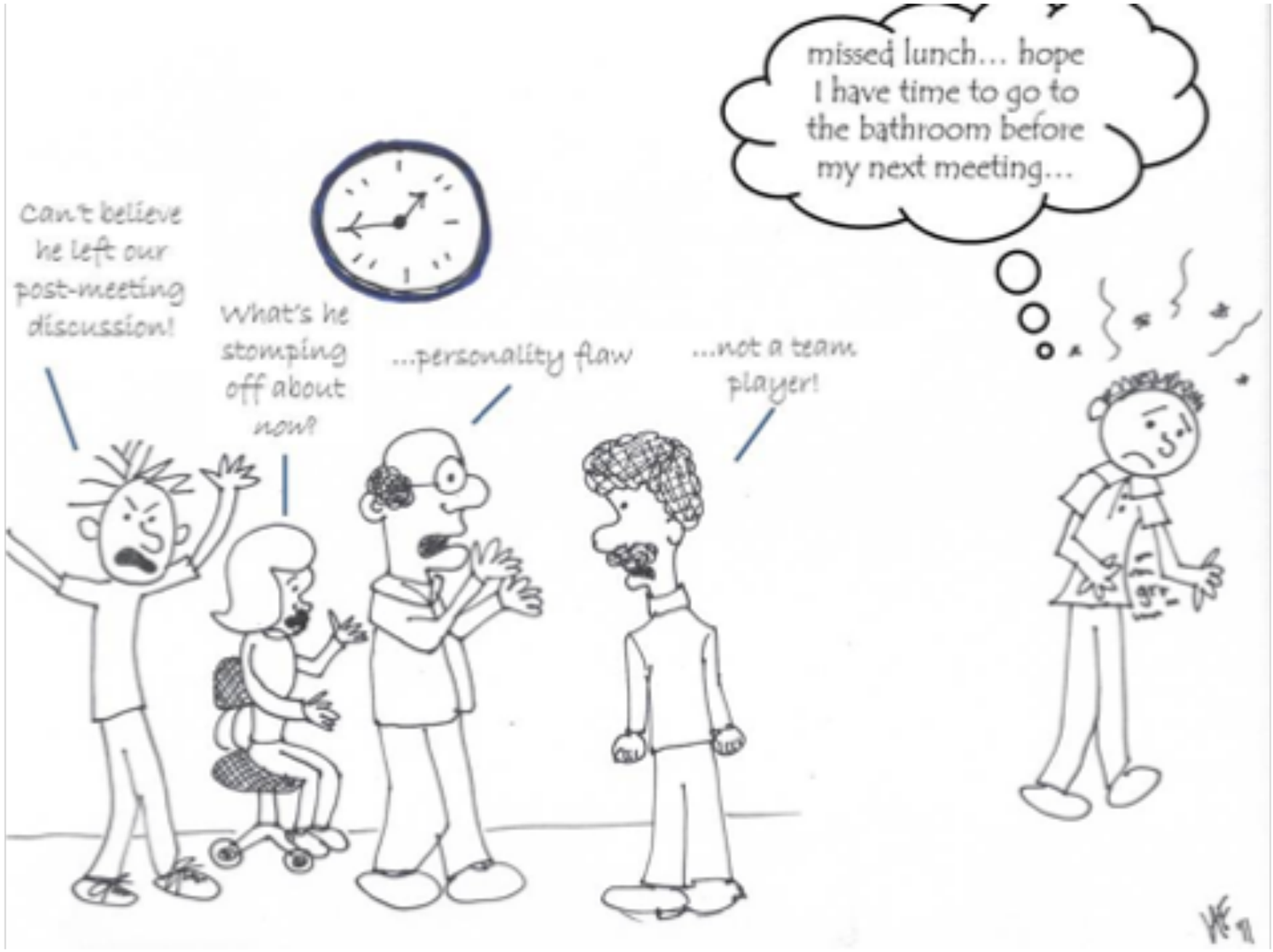
Internal Validity >> learning | fatigue effects

- Experience gained from using the first interface (to conceptual model) affects how they think about and use the second interface.

- Too much testing means participants get tired of testing.

- Mix it up: for some participants, A then B; for others, B then A.

Internal Validity >> experimenter bias



Internal Validity >> demand characteristics

- If participants know what your hypothesis is, they will actively try to be “good participants” and help you out.

Experimental Validity

- **External validity » realism**
  - Across situations
  - Across people
- **Internal validity » integrity**
  - *Confound*
  - *Selection bias*
  - *Learning effects*
  - *Experimenter bias*
  - *Demand characteristics*

Ways to improve validity

Ways to improve validity

- “Double-blind” experiment
    - neither participant nor experimenter know the hypothesis

Ways to improve validity

- “Double-blind” experiment
    - neither participant nor experimenter know the hypothesis
- Active deception
    - tell participants you’re expecting the opposite of what you expect

Ways to improve validity

- Randomized assignment to conditions
  - reduces systematic assignment biases

Ways to improve validity

- Randomized assignment to conditions
  - reduces systematic assignment biases
- Randomized ordering of conditions
  - normalizes the effect of order/learning/fatigue

Ways to improve validity

- Randomized assignment to conditions
    - reduces systematic assignment biases
- Randomized ordering of conditions
    - normalizes the effect of order/learning/fatigue
- Large sample size
    - reduces effect of “randomness”

Experimental Validity Summary

- **External validity » realism**
  - confidence that results applies to real situations
- **Internal validity » integrity**
  - confidence in our explanation of experimental results

Recording Observations

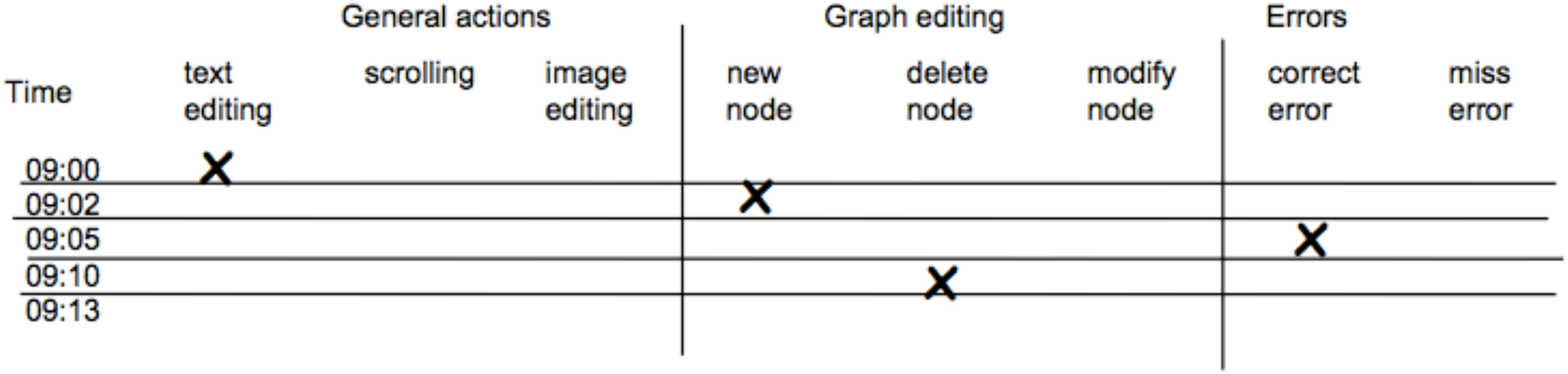
- We risk forgetting, missing, or misinterpreting events if we do not record user actions for later analysis.

Recording Observations

- We risk forgetting, missing, or misinterpreting events if we do not record user actions for later analysis.

- Lots of different methods for doing this with various pros and cons:
  - paper and pencil (2nd observer helps)
  - audio recording
  - video recording
  - system recording (user actions w/ system, user actions)

Coding sheet example

- 

Time	General actions: text editing	General actions: scrolling	General actions: image editing	Graph editing: new node	Graph editing: delete node	Graph editing: modify node	Errors: correct error	Errors: miss error
09:00	X							
09:02				X				
09:05							X	
09:10					X			
09:13								

Coding sheet example

- Tracking a person's use of an editor…

A	B	C	D	E	F	G
	P1	P2	P3	P4	P5	P6
User is annoyed by the slow login process						
Uses the "Calculate" button						
Says something positive about the calculator feature unprompted						
Asks for help with finding the "Management" page						
Expected the process of adding a new contact to be much quicker						
Discovered the voice feature						
Understands the meaning of submitting a lead request						
Surprised (in a bad way) by the results page of the lead search						
The user experiences flow						
Double-clicks icons (rather than single-clicks)						
Confused by the last paragraph in the Lead page						
Says that the Reports page has a clean look						
Keep clicking the "visit now" text in the home page						
Attempted submitting a lead request without filling in 2 mandatory fields						
Thinks the product is relatively slow						
Didn't notice the message bar						
Complains about font size						

The "Observation" sheet: repeated observations are highlighted in different colors. (Large preview)

Usability Tests: HowTo

- Determine goals of usability test
- Determine testing timeframe
- Determine target audience & recruitment plan
- Develop testing plan
    - What are the most important things you want to know?
    - Conceptual model extraction
    - Provide non-leading questions
    - Simple/realistic scenarios
    - Prepare any written materials (audience-specific, if necessary)
- Run a **pilot study**
- Run your test with real participants

When should I use Usability Test?

When should I use Usability Test?

- **Any time!**

When should I use Usability Test?

- **Any time!**
- **Early:**
  - Exploring potential possible designs

When should I use Usability Test?

- **Any time!**
- **Early:**
  - Exploring potential possible designs
- **Late:**
  - Close to end stage to determine possible showstoppers

When should I use Usability Test?

- **Any time!**
- **Early:**
    - Exploring potential possible designs
- **Late:**
    - Close to end stage to determine possible showstoppers
- **After:**
    - Investigate reported problems

Acknowledgements

- Tony Tang
- Lora Oehlberg
- Ehud Sharlin
- Frank Maurer
- Saul Greenberg

Course information

- Website
  - GitHub Pages https://silvadasilva.github.io/CPSC481-2019W/en/#!index.md
- Communications
  - Slack https://cpsc481-2019w.slack.com/
- Readings and Slides
  - Posted online at the main website

HowTo

- Determine goals of usability test
  - What are you trying to find out?
  - What are the problems you expect to see?
  - How much of the interface you expect people to use?
- Determine testing timeframe
- Determine target audience & recruitment plan
- Develop testing plan
  - What are the most important things you want to know?
  - Conceptual model extraction
  - Provide non-leading questions
  - Simple/realistic scenarios
  - Prepare any written materials (audience-specific, if necessary)