CMPT 363: User Interface Design
Summer 2021

Week 3: Design Heuristics, Usability Testing

Instructor: Victor Cheung, PhD

School of Computing Science, Simon Fraser University

Recap from Previous Part

- Evaluating Interfaces
  - Understand the users, discover flaws early & fix them
  - At all levels, all parts, and all attributes
  - In laboratories with controllable environments, natural settings
  - Anytime during the design and development process
- Be aware of possible factors that could impact validity of evaluation
- Heuristic Evaluation – a type of analytical evaluation where experts are invited instead of users
  - Design Heuristics – applying common usability principles to discover design issues

Some Questions in the Previous Part

- Is A-B testing/beta testing done in natural settings?
  - Not always. They are independent of where the testing is done and can both be done in a controlled or natural setting
- What is the relationship between Heuristic Evaluation and Analytical Evaluation?
  - Heuristic Evaluation is a type of Analytical Evaluation, which we use in the course to cover all evaluations that don't directly involve end users. We'll learn other types of Analytical Evaluation later in the course
- Aren't Match between system + real world (HE#2) the same as Recognition > Recall (HE#6)?
  - They can be achieved using similar design (e.g., familiar icons/wordings), but HE#2 aims at making the interface more familiar and easier to understand, whereas HE#6 aims at reducing memory load
- Are user experience & satisfaction always the goals for designing an interface?
  - Typically yes, at least that's the idea scenario. However, we do want the interface to at minimal supports the tasks it is supposed to do. In some cases having redundancy and not-so-simple interfaces could take priority

Evaluating Interfaces with Users

Time to put things in the hands of the actual users

Evaluating Interfaces with Users

- Why: helps check that people can use the product and that they like it (or why not)
- What: can be done on conceptual models, early prototypes, and more complete ones
- Where: can be in laboratory or natural settings
- When: should occur throughout the design, finished products can also be evaluated to collect information to inform new versions/products

Typical Arguments against Evaluation

- "Evaluation takes up too much time and money"
  - The time and money one saves from fixing problem is worth it
- "There's nothing to test… we'll finish the system first and then have something to evaluate"
  - The more complex (finished) a system is, the harder it is to evaluate and make changes
- "The interface is very trivial and everybody should understand it"
  - Even a simple design can have issues, for example, hard to find help, relies too much on recall
- "We have been developing this kind of system for years, we know our stuff"
  - Knowing too much about the system might lose sight of how a novice user approaches it

Usability Testing

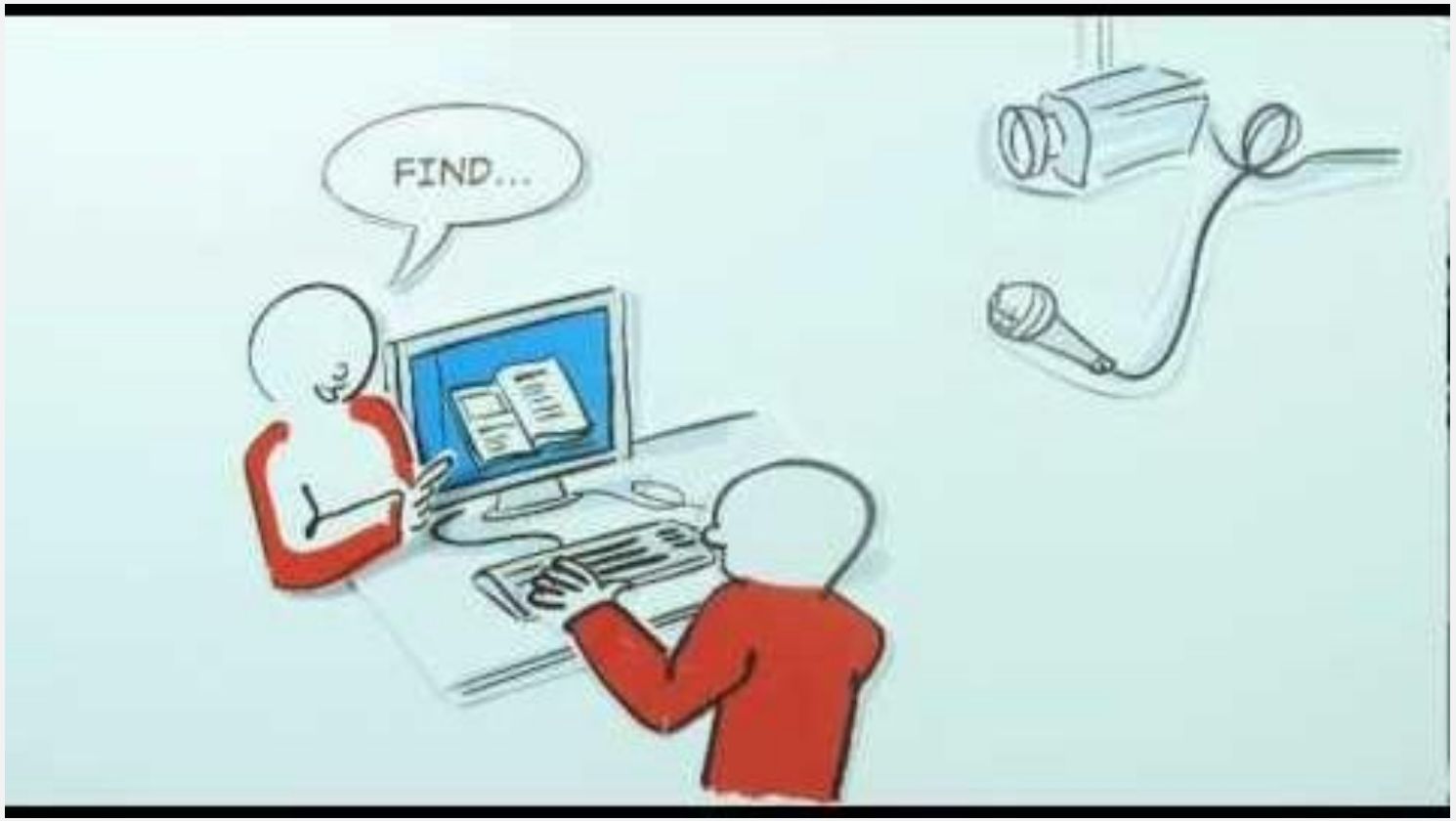
- A form of evaluation that involves measuring performance and satisfaction with the system, in a laboratory setting (here we loosen this up to include natural settings)

What Do You Get out of a Usability Test?

- Identify problems in the design of the product or service
  - Users might find issues that the designers/developers haven't thought of
- Uncover opportunities to improve
  - The way users use/adapt the design might provide insights on how to make things better
- Learn about the target users' behaviour and preferences
  - Understanding how they do things helps designing products or services closer to their expectations

- Fun read on don't rely on just a "genius designer"
  https://www.nngroup.com/articles/the-myth-of-the-genius-designer/

Usability Testing in Action



https://www.youtube.com/watch?v=BrVnBdW6_rE

A Typical Usability Lab

http://iat.ubalt.edu/usability_lab/

https://www.nngroup.com/articles/usability-testing-101/

A Typical Usability Test – The Core Elements

Gives instructions, answers participant's questions, observe and ask follow up questions.

Someone who is already a user, or has a similar background and needs as a target user.

Facilitator

Guides the participant through the test process

Tasks

Realistic activities that the participant might actually perform in real life

Participant

Realistic user of the product or service being studied

NNGROUP.COM NN/g

Source: https://www.nngroup.com/articles/usability-testing-101/

A Typical Usability Test – The Flow

- The facilitator gives instructions and task scenarios to the participant, who provides behavioural and verbal feedback about the interface while he/she performs those tasks
- Some performance parameters (e.g., time for completion, error rates, hot spots) might also be measured

Facilitator:
- observers & interviews Participant
- administers Tasks

Participant:
- gives feedback to Facilitator
- performs Tasks

The flow of information in a Usability test.
Source: https://www.nngroup.com/articles/usability-testing-101/

Steps in Conducting A Usability Test

Prepare tasks, recruit participants, setup the test materials

Invite participants, observe and ask questions during the session, thank them when done

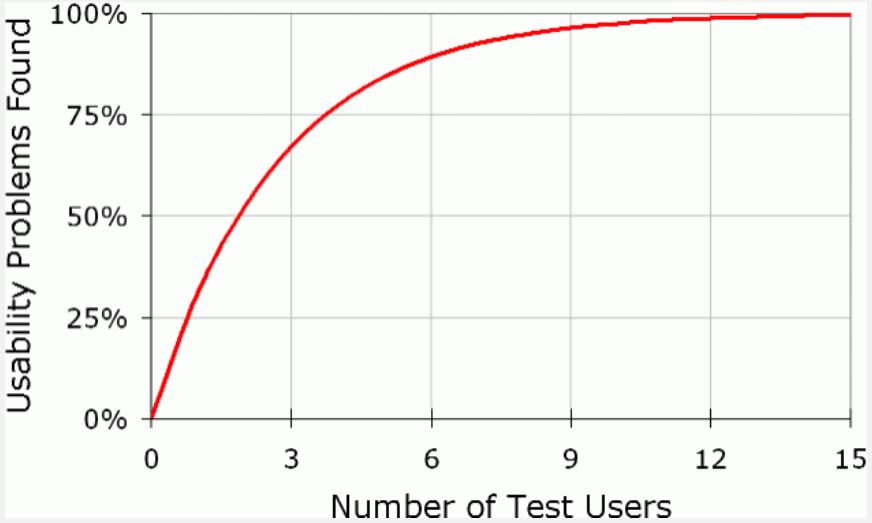
Analyze data collected, pay attention to problems, summarize results, make recommendations

Creating the Tasks

- Snippets of activities that the target user would perform in real-life with the design
- Set the stage with a short paragraph, for example,
  - "We are interested in improving the mobile interface for people to save, update, and use contacts"
- Use representative tasks, e.g., "you are writing an email" for email clients, "you want to go to X" for maps
  - Can be specific like "write the email on this topic to your course instructor", or "navigate around the landing page"
- Wording is very important as it might confuse participants or steer them to a certain direction
  - Read 10 mistakes to avoid by Amy Schade: https://www.nngroup.com/articles/better-usability-tasks/

How Many Participants Do You Need for Usability Testing?

- About 5 for most cases (qualitative measurements)
  - Less than that might not find many problems, or biased by a single person
  - More is better, but benefit-cost ratio is less
- For other types of measurements
  - Quantitative: at least 20 (for statistical power)
  - Card sorting: at least 15 (to generate useful concepts)
  - Eyetracking: at least 39 (to have stable heatmaps)
- Read more about the numbers
  https://www.nngroup.com/articles/how-many-test-users/



Source: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Observing Participants & Collecting Data

- Typically done with video & audio recording (acquire consent from participants first!), can also supplement with screen captures and questionnaires
  - Your notes are also very useful (jog down special moments such as breakdowns or delights)
- Encourage participants to verbalize what they are doing and how they feel at the moment
  - Known as the "Think-aloud" method, referred to as the "single most valuable usability engineering method" by Nielsen
  - Great way to find out why users do a certain thing, why they like/dislike something, why are they stuck/frustrated

Think-Aloud Protocol

- Main idea is to let users continuously verbalize their thoughts as they move through the design
  - Recruit representative users
  - Give them representative tasks to perform
  - Shut up and let them do the talking
- Benefits: cheap, robust, flexible, convincing, easy to learn
- Downsides: unnatural situation, filtered statements, potential bias from facilitator
- Learn more about the protocol
  https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

Tips to Do Think-Aloud Well

- Be encouraging (there is no "wrong" answer)
- Coach participants with general questions
  - "what do you see?", "what did you expect?", "what are you thinking now?"
  - Instead of "what do think this button is for?", "why didn't you click here?" – potentially giving away or bias the participant
- Have more than one participants to encourage discussion
  - Might not be as easy to schedule more than one at the same time
  - There is a potential for them to bias each other

- Watch this example of a Think-aloud recoding: https://www.youtube.com/watch?v=nJ2udLjdsx4

Designing Questionnaires

- Start with general demographic information that may be relevant
  - Age (use range when possible)
  - Gender (include "prefer not to say" or "self-identify as" when possible)
  - Experience (can ask for specific examples for clarity, useful to gauge level of expertise)
- Use carefully worded questions to evaluate feelings or perception towards the design
  - No leading questions (e.g., "why do you think this is good?", "users tend to think this is bad, do you?)
  - Concise use of terms (provide a short example if needed, but be careful about drawing too much attention)
  - Use open-ended questions instead of yes/no questions to elicit more information
- Pilot test your questions with others

Types of Questions

- Likert Scale

	Strongly Agree … Agree…Neutral…Disagree…Strongly Disagree
The system is easy to use	1……..2……3……..4…….5

  - Use odd number of options

- Semantic Scale

	Difficult…….Easy
Finding the right information was	-2……..-1……...……0……...…..1…….2

  - If multiple questions, can mix the polarity for good/bad

Some Standard Questionnaires for Usability Testing

- Questionnaire for User Interaction Satisfaction (QUIS) http://www.cs.umd.edu/hcil/quis/
- System Usability Scale (SUS) https://www.usabilitest.com/sus-pdf-generator
- User Experience Questionnaire (UEQ) https://www.ueq-online.org/
- NASA-TRX https://humansystems.arc.nasa.gov/groups/TLX/index.php

Some Issues about Usability Testing to Consider

- Reliability and validity – Can the results be replicable? Can the results be generalized to others?
- Ethics – Is there any harm done to the users? Was the testing stressful? Is their identity protected?
  - Universities have their "institutional review board" (IRB) to monitor that
  - Treat your participants well, even if you are paying them

Some Useful/Helpful Things to Say

- Set the scene
  - "We are interested in learning about X and that's where you come in!"
  - "Imagine that you are doing Y and want to do Z."
- Set the mood
  - "We are testing the design here, not you."
  - "It is important that you say it out what it is in your mind at the moment. I may even prompt you if you stop talking."
  - "I will not be able to answer your questions about the interface to avoid biasing you."
- Close the deal
  - "Thank you very much for helping us to understand how you use the design!"

During The Usability Test

- Do's
  - Give the scenario and first task verbally & in text
  - Give subsequent tasks in text, one at a time
  - Take short breaks if the tasks are long
  - Observe task time, errors, breakdowns, workarounds, confusions, and success/failure
  - Take notes, video- & audio-record
- Don't's
  - "Correct" the participants if they make a mistake (unless the whole system crashes)
  - Answer questions to explain or clarify something (materials for improvements)
  - Rush the participants or show annoyance

After The Usability Test

- Debrief them by
  - Providing more information about what you were interested in finding (at the start you don't want to tell them everything)
  - Answering any questions they have (can even show them the "expected")
- Thank them by
  - Saying thank you and acknowledging them for helping you out
  - Pay them if you have promised monetary reward
  - Ask for suggestions

Take A Break While Watching This ☺

https://www.youtube.com/watch?v=3Qg80qTfzgU

Summary

- Usability Testing
  - Why?What?Where?When?
  - Expected outcome
  - Typical setup, core components, & flow
  - Ways to observe & collect data
  - Steps involved in a usability testing

Post-Lecture Activity

- Read/watch these (and those in the slides)
  - Chapters 14 & 15 of ID-Book: Introducing Evaluation & Evaluation Studies
  - Thinking Aloud: The #1 Usability Tool
    https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/
  - Usability evaluation and analysis
    https://www.hotjar.com/usability-testing/evaluation-analysis/

- Imagine you are helping Canvas SFU to design the interface for new SFU students
  - How would you propose a heuristic evaluation and a usability testing?
  - Who would you invite to participate for each activity? Why?