DUB
DESIGN
USE
BUILD

W PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

CSE 440: Introduction to HCI

User Interface Design, Prototyping, and Evaluation!

Lecture 13: Interface Evaluation

Instructor: Amy Zhang, 11/16/2021

Today's Topics

- UI Hall of Fame and Shame

- Where are we now?

- Methods for interface evaluation

  - UI Inspection

  - Usability Testing

  - Formal User Testing

- Ethics in human subjects research

- Designing tasks in usability testing

- Groups pilot usability testing of their paper prototype!

UI Hall of Fame and Shame
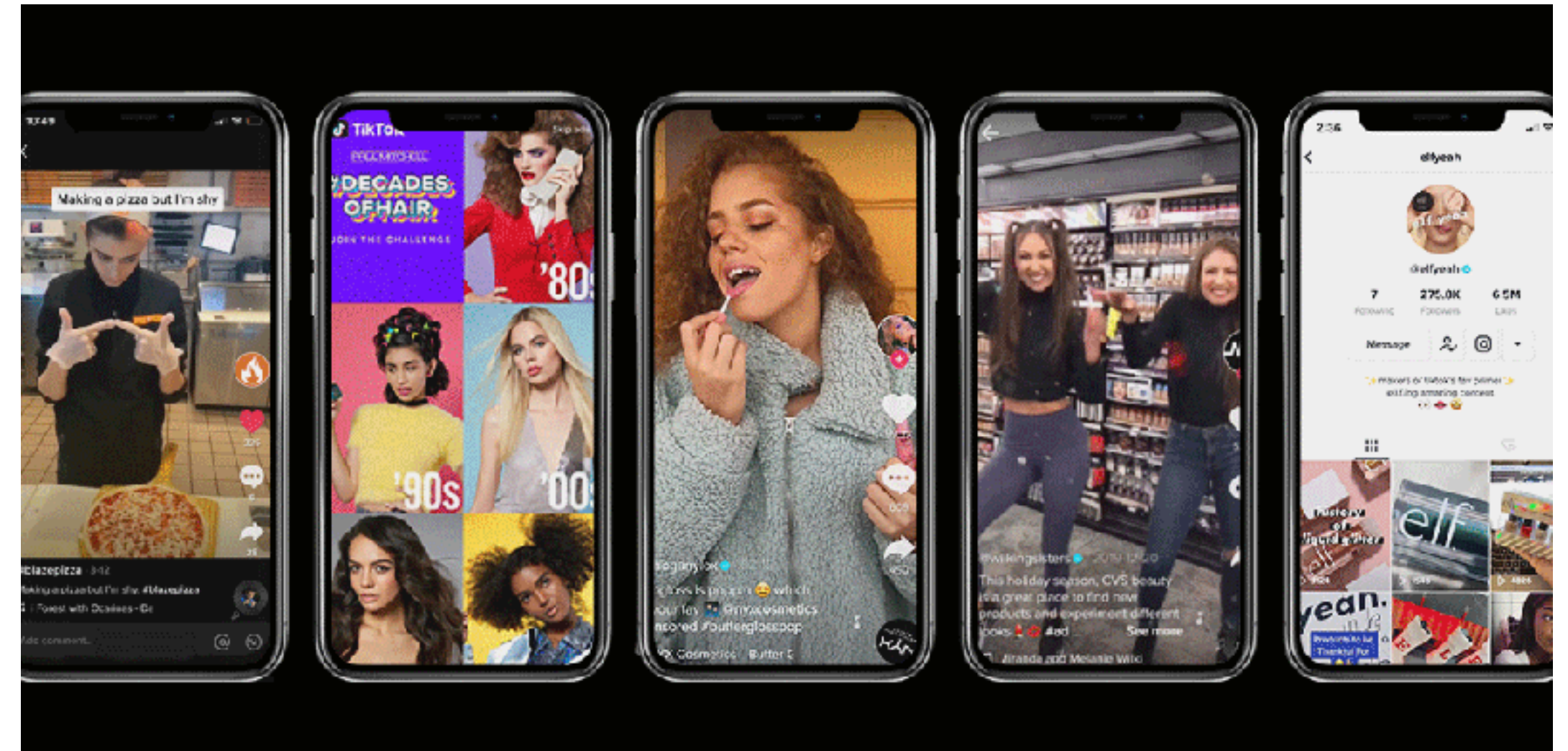
Infinite Scrolling

Pros and Cons?

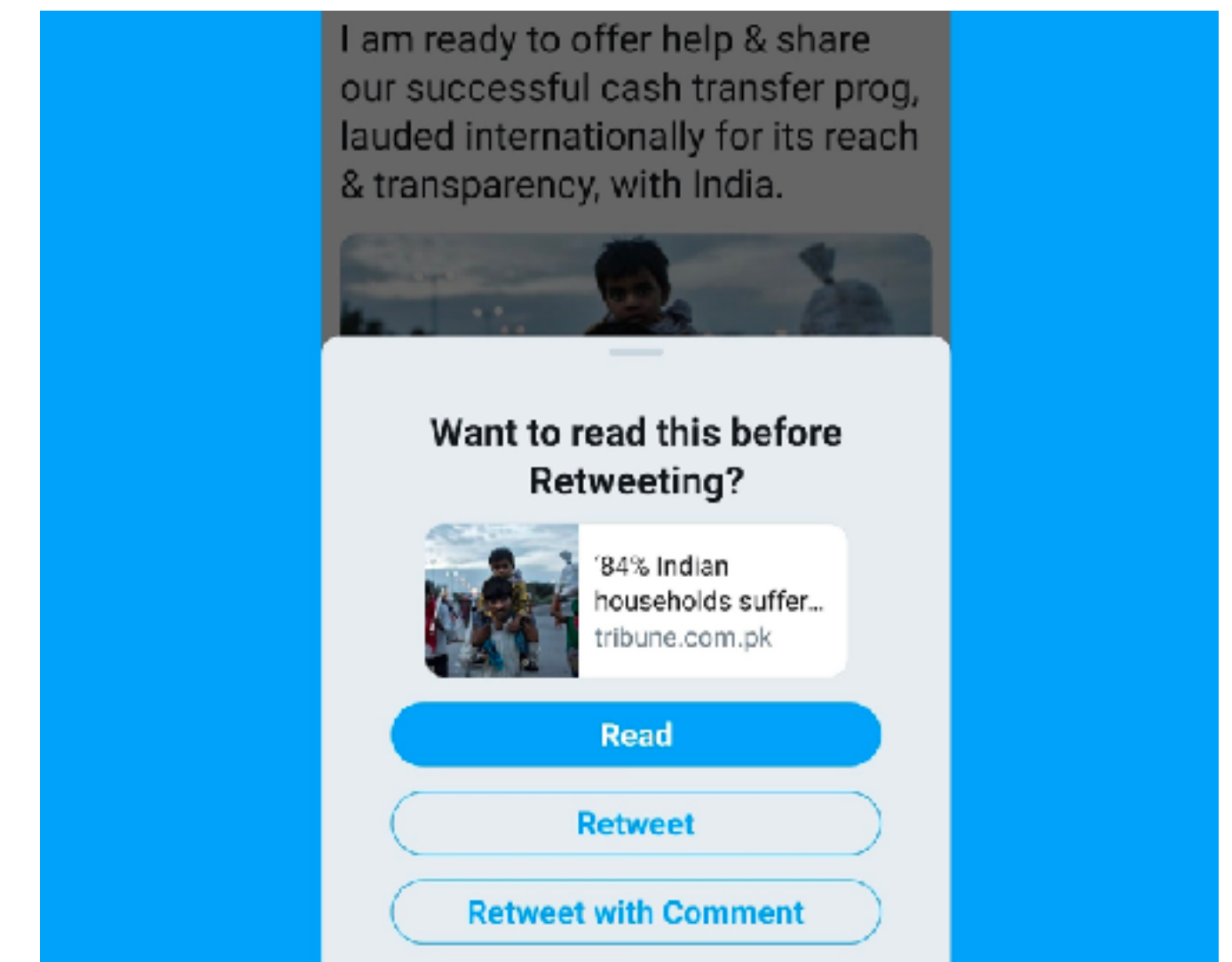# ...vs. Pagination

Pros and Cons?

Infinite scrolling
Auto-queue next item
Auto-play video
Auto-looping video
Volume on by default
One-click share

...

Slow Design
Design Friction
Seamful Design (as opposed to seamless)

Where are we now?

Where are we now?

1a: Three Project Proposals

Frame the Problem

Explore the Solution Space

1b, 2b, 2c
User Research

User research
Competitive Analysis

2a Project
Ideation

Brainstorming
Ideation through Sketching

2d Task analysis

Scenarios
Storyboards
Personas
Design Rationale

Find a Good Solution

2e Sketches
2f Storyboard

3a Paper Prototype

Refine the Solution

Lo-fi Prototypes
Early Evaluations
Mid-fi prototypes
Additional Evaluations

3b Heuristic Evaluation

3c Usability Testing

Looking ahead...

- This week:
  - Thursday: 3b: Heuristic Evaluation in class
- Next week (week 9):
  - work on 3c: Usability Testing
  - Thanksgiving
- The following week (week 10):
  - work on 3d: Digital Mockup
  - Take-home exam Thurs-Fri
- The following week (week 11):
  - work on 4a and 4b (initial and final poster designs)
  - Poster session on Thursday
  - Submit 3e (final writeup of Assignment 3 as a blog post) later in the week
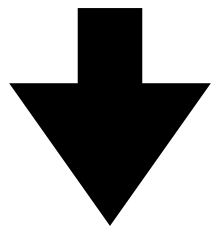
Methods for Interface Evaluation

Interface Evaluation Methods

- UI Inspection

  ⬇

- Usability Testing

  ⬇
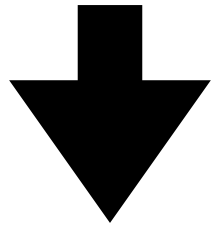
- Formal user testing (beyond paper prototypes)

- *Increasingly sophisticated, laborious, and rigorous!*

UI Inspection

What is UI Inspection?

- So far, you've been practicing giving feedback mainly on **project ideas** that have been rapidly explored and iterated on in parallel.

- As we move into serial prototyping and slightly more involved prototypes, it's time to get a little more rigorous in how we give feedback on the **specific interfaces** that you come up with.

- **UI inspection methods involve systematic approaches for experts to evaluate a design.**

- You've already started to practice this with our UI HOFS!

Why UI Inspection?

- It takes time, money, and effort to find good study participants to try out your prototype.

- Ideally, you "save" study participants for uncovering less obvious design issues, and do UI inspection first with the design team to identify obvious problems.

- Unlike user tests, UI inspection is **rational**, as opposed to **empirical**. Incorporating different kinds of evaluation can paint a more well-rounded picture.

Types of UI Inspection

1) Heuristic Evaluation

  • We'll go into this in more detail on Thursday (and then you'll do one!)

2) Cognitive Walkthrough

1) Heuristic Evaluation

- Developed by Jakob Nielsen

- Helps find usability problems in a design

- Method:

  - A small set of 3–5 evaluators examine the interface

  - They independently check compliance with a set of design principles (e.g., how easy is it to prevent errors?).

  - Details Thursday!

2) Cognitive Walkthrough

- Heuristic evaluations focus on the **product as a whole**, while cognitive walkthroughs focus on **specific tasks**.

- Method:

  - Evaluation is performed by an expert, such as someone on the design team

  - The expert identifies a user goal and list of tasks to accomplish that goal.

  - They work through the tasks, asking themselves 4 questions from the perspective of the user

Cognitive Walkthrough Questions

- At each step:

    - Will the user try and achieve the right outcome?

    - Will the user notice that the correct action is available to them?

    - Will the user associate the correct action with the outcome they expect to achieve?

    - If the correct action is performed; will the user see that progress is being made towards their intended outcome?

Usability Testing

What is a Usability Test?

- Review from last Thursday!

- We give participants tasks and watch them as they try to perform them with our prototype.

  - Observe and take notes when they succeed/fail at the task. Note **critical incidents**, or moments while using an interface that strongly affect user performance or satisfaction.

  - Have them speak aloud their thoughts.

  - Ask questions of their expectations and thought process.

  - Results are observational and primarily qualitative.

Why Usability Test?

- Unlike UI inspection, we are now collecting **empirical data from non-experts**.

    - UI inspection is faster but possibly more biased and misses problems (since it's just from the expert's perspective)

    - Usability testing is more accurate and realistic but takes more time and you have to interpret what the user says and does

    - One approach is to alternate between them

Activity 1

Watch for the **critical incidents** in this usability test.

Write your observations in: https://tinyurl.com/xvcexm8e

Keep your form open and don't submit yet.

Formal User Testing

What is a formal user test?

- Approach evaluating new design ideas from the frame of a **scientific experiment** to prove a hypothesis or answer a research question rigorously

  - Set up the environment and task to be uniform as possible across participants

  - Collect summary or statistical data while or after users are performing the task. Quantitative measures (i.e., time taken, eye gaze, number of errors made) are common.

  - Potentially have a baseline/alternative prototypes as an explicit comparison

Why formal user test?

- Unlike a usability test, this is a scientific experiment. You are trying to **prove rigorously with empirical data** that a design is better than others.

- Usually follows usability tests

- Slower to do than usability tests

- Can do explicit comparisons

- Can do tests for statistical significance

Within-subjects design
The same participant tests all conditions corresponding to a variable.

Between-subjects design
Different participants are assigned to different conditions corresponding to a variable.

Examples of User Testing Methods

- Interviews or Focus Groups - Qualitative & Moderated

- Survey - Qualitative or Quantitative & Unmoderated

- Clickstream Analysis - Quantitative & Unmoderated

- Eyetracking Study - Quantitative & Unmoderated

- Lab Study - Qualitative or Quantitative & Moderated

- Remote Usability or User Tests - Qualitative or Quantitative & Unmoderated or Moderated

- Controlled online A/B experiments - Quantitative & Unmoderated

Deep Dive: A/B Experiments
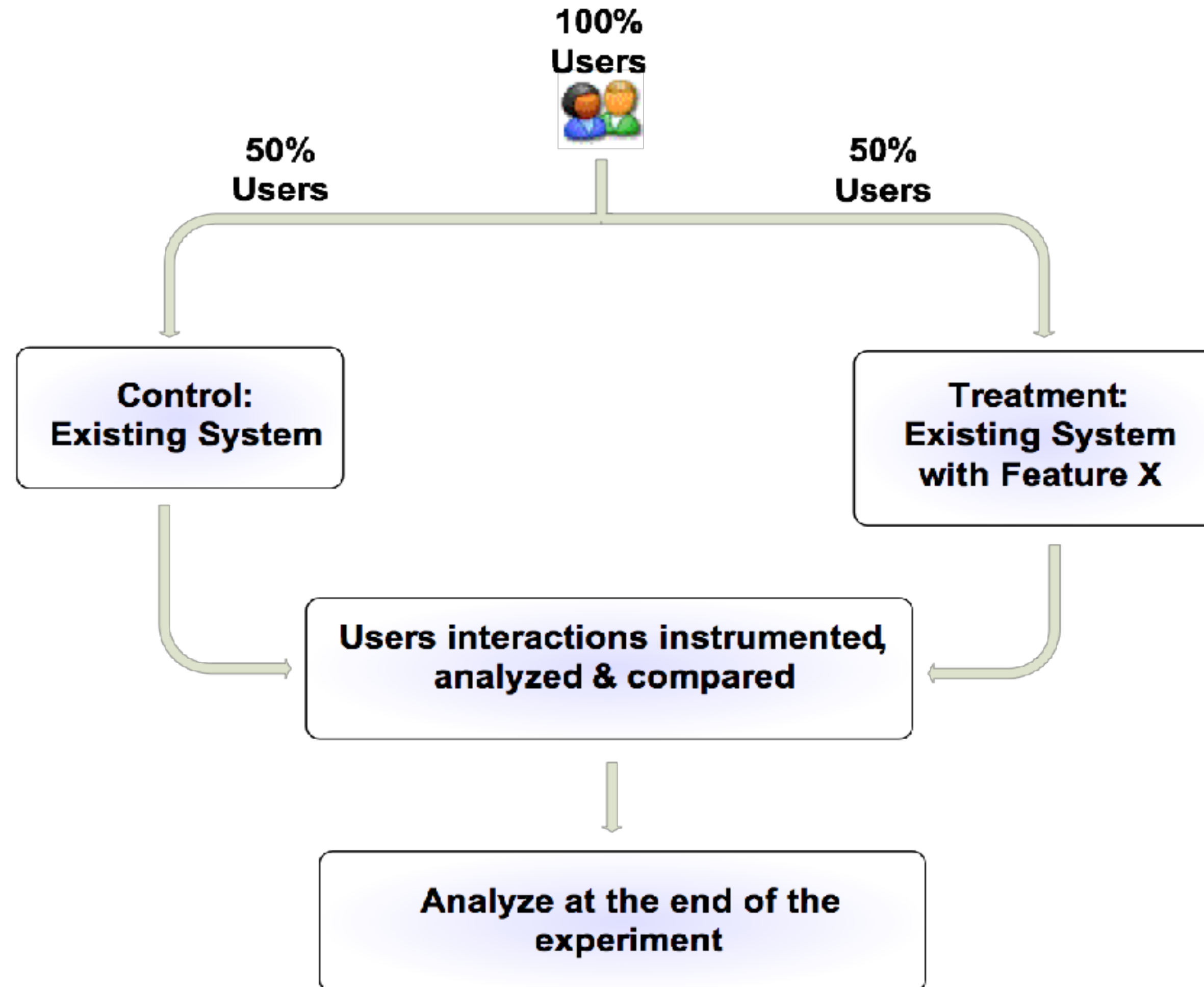
A/B Experiment Example

- Example: Amazon Shopping Cart Recommendations
  - Add an item to your shopping cart
  - Most sites just show the cart
- At Amazon, Greg Linden had the idea to also show recommendations of related items based on cart items

Example: Amazon Shopping Cart

- Evaluation

  - Pro: cross-sell more items

  - Con: distract people from checking out

- This con, if correct, could be a really expensive mistake!

- But after a simple A/B experiment on a subset of users, it was found to be wildly successful

Limitations of A/B Tests

- Drives hill-climbing, but not overall design

  - A design may perform better, but is it the best?

  - Marginal improvements

  - You can go overboard

  - No generalizable knowledge gained

- Impossible for novel designs to compete in an A/B test

- Difficult to scale for many comparisons

- Highly dependent on quantitative metrics that may lead you astray

~3 min stretch break

Ethics in Human Subjects Research

Ethical considerations in human subjects research

- **Testing is stressful and distressing**
  - People can leave in tears if they think the test is a reflection on them
  - People can reveal private information that you are now the steward of
- **You have responsibility to alleviate and mitigate immediate and potential harm**
  - Make voluntary, with informed consent, and avoid pressure to participate
  - Tell them they can stop at any time
  - Stress that you are testing the system, not them
  - Debrief people after the test is over, particularly if there was deception involved
  - Minimize immediate harm, such as physical or emotional discomfort
  - Respect people's time and fairly compensate users
  - Make collected data as anonymous as possible, store private data in a secure location

Activity 2

Point out as many **ethical issues** with this user test as you can.

Continue filling out the form you started at https://tinyurl.com/xvcexm8e

Submit your form!

Human Subjects Approval

- 1972 Tuskegee Study -> 1979 Belmont Report -> establishment of IRBs in all U.S. universities

- Published research now requires formal human subjects review of the process.

- You don't need it for your usability tests as you won't be formally publishing your work

- But understand why we do this. You can still check yourself informally.

- **Public trust in science depends on researchers treating subjects with dignity and respect.**

The New York Times

Syphilis Victims in U.S. Study Went Untreated for 40 Years

By JEAN HELLER
The Associated Press

Preparing your Usability Test

Tips for your task script

Bad: Artificial Subgoals

- People using the design "in the wild" may not necessarily form these same subgoals

- The task should give one top-level goal, while users form their own subgoals while pursuing it

Now you want to choose the type of paper you want to print your document on. Lets imagine that Bin "B" has the paper you want to print your paper on, please complete this task.

Now set the darkness of your copies to about 50% dark.
After setting the darkness, you decide you want to print 2 sides of copies on two sides of paper. Please complete this task.

Bad: Artificial Ordering of subgoals

- People might not proceed in this order.

- Order may be biased towards the layout of the interface, which would conceal any problems with finding the appropriate control

- Enter in 10 copies, with lightness set to 10%.
- Choose 1 sided to 2 sided, use paper source bin A.
- Cover sheet needed, using paper bin B for cover sheet.
- Set stapling feature on and collating on.
- Start printing.

Bad: Giving the Answers

- Don't tell the person what terminology the interface uses, which they might not otherwise know

You are a teacher and are trying to make 40 copies of a one-sided magazine article that is 10 pages long for your class tomorrow. Due to the large number of copies, you print the article double-sided, **in other words 10 page article would be printed on 5 sheets of paper.** Due to the high contrast of the article, you must lighten the copy, **in other words change the contrast.** You then want the copies to be **collated** and **stapled**.

Good: Give Context

- Giving realistic context through scenarios can reduce the artificiality of the task

It's your first day in the office, starting a new job. You would like to make some copies of several documents that your boss gave you to browse through. Your colleague in the next cubicle tells you that you need an access code to make copies. The code is 5150. You walk over to the copy machine at the end of the hall and realize that it is not the Xerox copier that you are accustomed to. Make 2 copies of the "Company Annual Report".

Consider: Under-Specified Tasks

- Many realistic goals are under-specified, as people only have a general idea what they want

- By under-specifying the task, you can elicit realistic confusion and decision-making

You just finished fixing up the old hot rod in the garage and now it's time to sell her. Make a couple copies of the pictures you took to send into the used car sales magazines. It's ok that they're in black and white but maybe you should lighten them up a bit. Your account billing code is 5150.

A cautionary tale: Clever Hans

- Beware of giving away clues to your participants with your actions.

- Beware of social desirability bias.
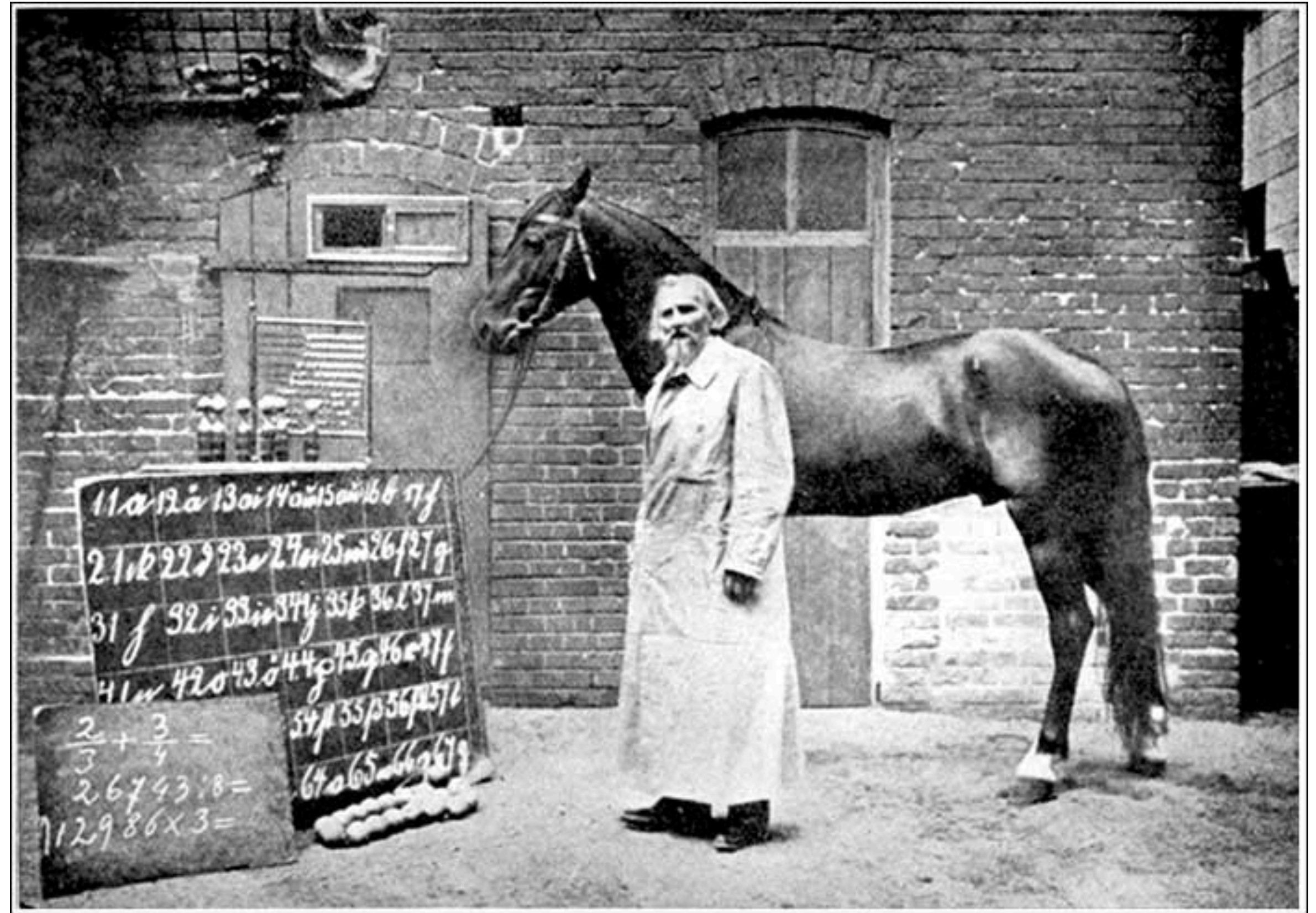
Pilot Usability Tests!

- Get into your group and find another group that is NOT in your section and sit across from each other.

- Within your group, decide on roles for this pilot usability test: computer, facilitator, note-taker

- [5 min] In your group, quickly write up a short script for the facilitator to say out-loud, where you give the user context and directions for their assigned task (just focus on 1 task for this pilot).

- [10 min] Next, conduct the pilot study with the other group.

- [10 min] Then swap!

- Use this pilot to "debug" your task script, people's roles, and the paper prototype so everything can run smoothly by Thursday.

- For Thursday, bring your prototype to class - you will be doing heuristic evaluations (3b) of them!