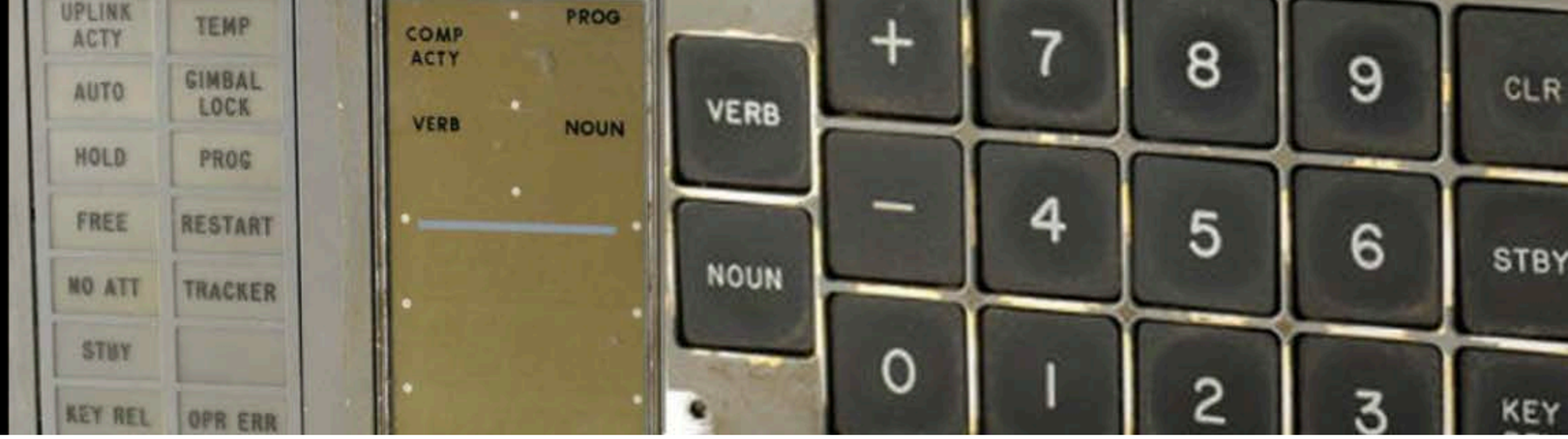
CS160

USER INTERFACE DESIGN

FALL 2020



USABILITY TESTING

19 OCT 2020

ERIC PAULOS
www.paulos.net

UNIVERSITY OF CALIFORNIA Berkeley

ANNOUNCEMENTS

Thank you for feedback

PROG 02

VOTE!

Teams

DESIGN 06: Understanding Uses and Competitive Analysis

DESIGN 07: Low-Fidelity Prototype

DESIGN 08: Low-Fidelity Feedback

DESIGN 09: Coding Towards High-Fidelity

berkeley certificate in design innovation

bcdi.berkeley.edu

College of Engineering • College of Environmental Design • Haas School of Business • College of Letters and Science – Arts and Humanities Division

The Berkeley Certificate in Design Innovation introduces design as the creative and critical means to innovate — to define, to imagine, and to advance a globally just future.

USABILITY TESTING METHODS

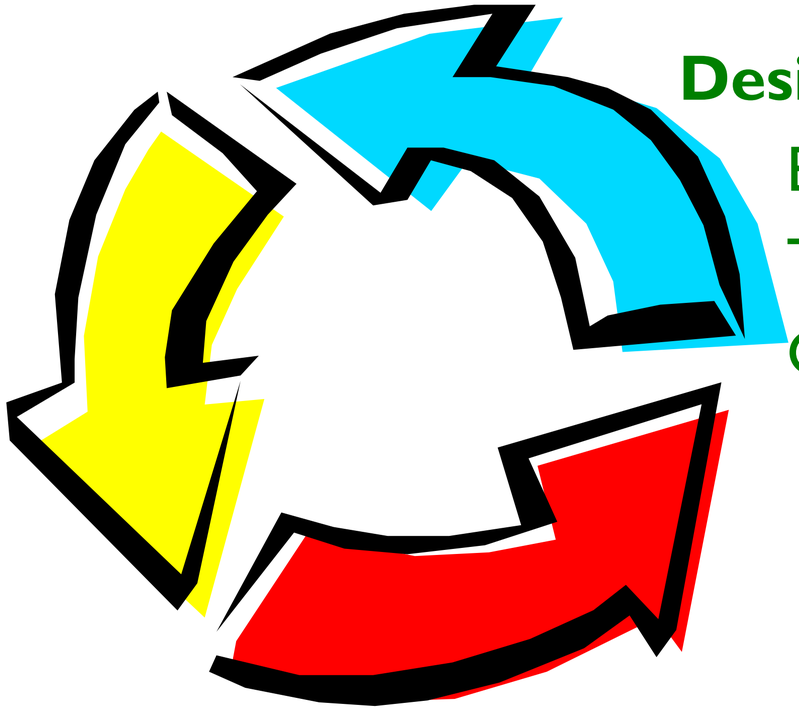
ITERATIVE DESIGN

Design

Brainstorming

Task analysis

Contextual inquiry



Prototype

Low-fi, paper

Evaluate

Low-fi testing,

Qualitative eval

Quantitative eval

GENRES OF ASSESSMENT

Automated	*Usability measures computed by software*
Inspection	*Based on skills, and experience of evaluators*
Formal	*Models and formulas to calculate measures*
Empirical	*Usability assessed by testing with real users*

EMPIRICAL TESTING IS COSTLY

User studies are very expensive – you need to schedule (and normally pay) many subjects.

User studies may take many hours of the evaluation team's time.

A user test can easily cost $10k's

“DISCOUNT USABILITY” TECHNIQUES

Cheap

No special labs or equipment needed

The more careful you are, the better it gets

Fast

On order of 1 day to apply

(Standard usability testing may take a week)

Easy to use

Can be taught in 2–4 hours

“DISCOUNT USABILITY” TECHNIQUES

Heuristic Evaluation

Assess interface based on a predetermined list of criteria

Cognitive Walkthroughs

Put yourself in the shoes of a user

Like a code walkthrough

Other, non-inspection techniques are on the rise

e.g., online remote experiments with Mechanical Turk

COGNITIVE WALKTHROUGH

COGNITIVE WALKTHROUGH

Formalized technique for imagining user's thoughts and actions when using an interface:

"*Cognitive walkthroughs involve simulating a user's problem-solving process at each step in the human-computer dialog, checking to see if the user's goals and memory for actions can be assumed to lead to the next correct action*." (Nielsen, 1992)

COGNITIVE WALKTHROUGH

Given an interface prototype or specification, need:

- A detailed task with a concrete goal, ideally motivated by a scenario
- Action sequences for user to complete the task

Ask the following questions for each step:

- Will the users know what to do?
- Will the user notice that the correct action is available?
- Will the user interpret the application feedback correctly?

Record: What would cause problems, and why?

COGNITIVE WALKTHROUGH EXAMPLE

Task: Find the call number and location of the latest edition of the book “Interaction Design” by Preece, Rogers & Sharp in the Berkeley library

Typical users: Students who are familiar with the web, but not necessarily with the library or its website

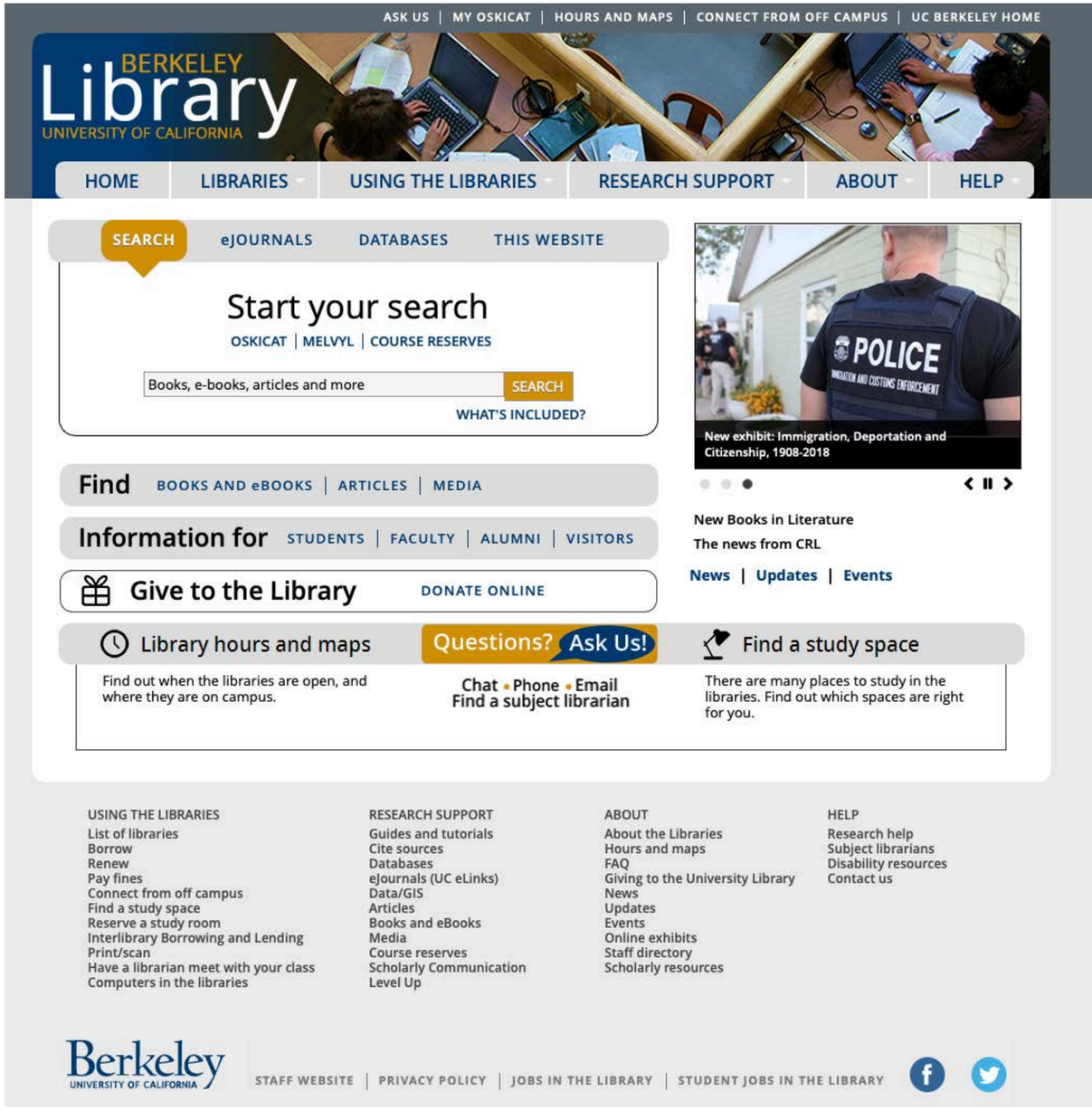
COGNITIVE WALKTHROUGH EXAMPLE

Step1: Select library catalog.

Will the user know what to do?

Will user notice that action is available?

Will user interpret feedback correctly?



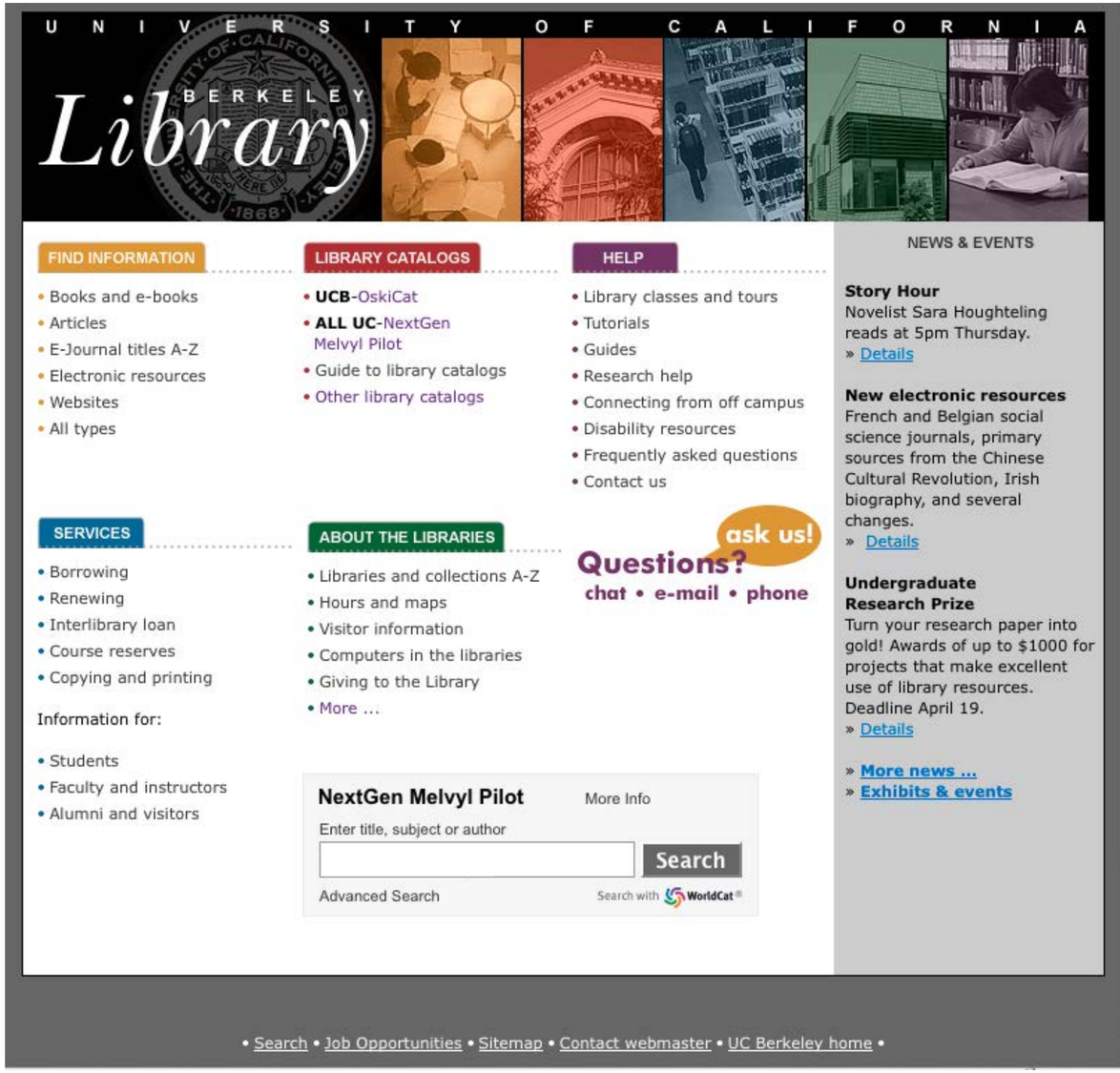
COGNITIVE WALKTHROUGH EXAMPLE

Step1: Select library catalog.

Will the user know what to do?

Will user notice that action is available?

Will user interpret feedback correctly?



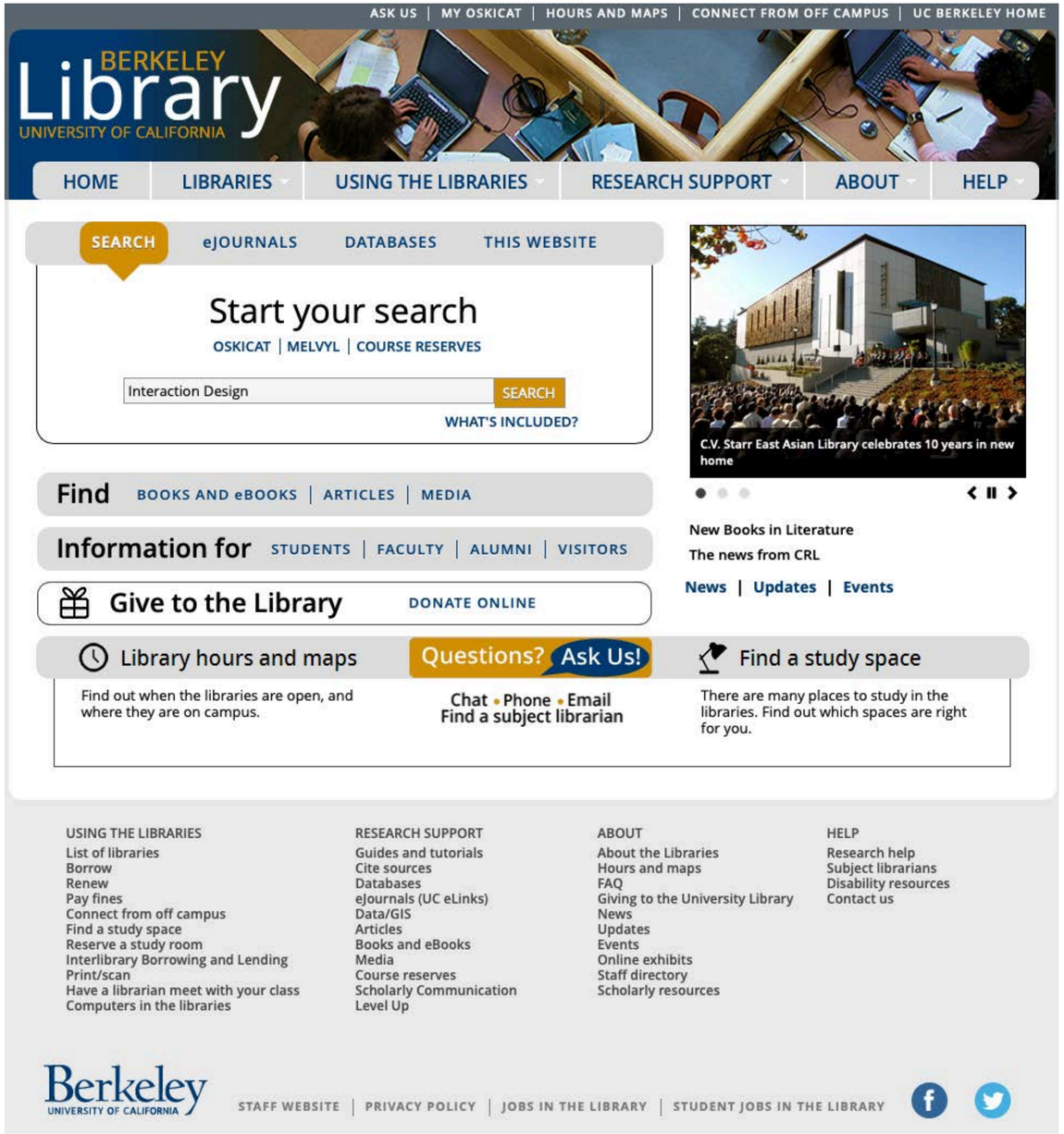
COGNITIVE WALKTHROUGH EXAMPLE

Step 2: Complete the search form

Will the user know what to do?

Will user notice that action is available?

Will user interpret feedback correctly?



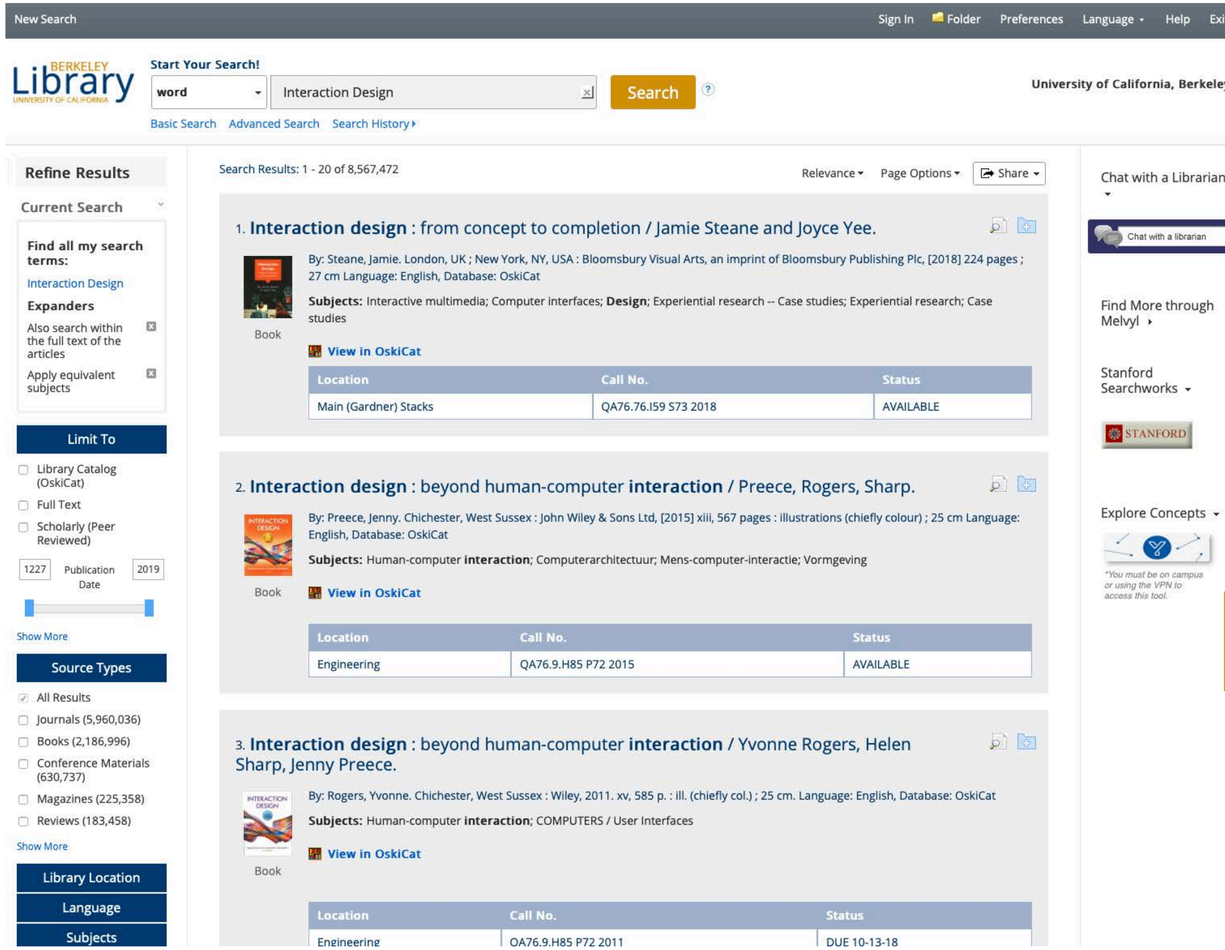
COGNITIVE WALKTHROUGH EXAMPLE

Step 3: Locate the right edition, click to detail screen

Will the user know what to do?

Will user notice that action is available?

Will user interpret feedback correctly?



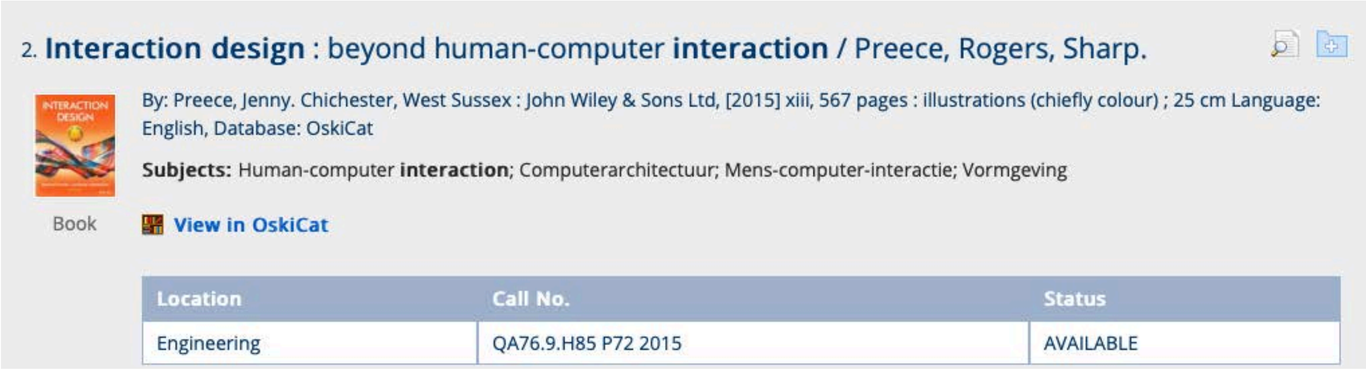
COGNITIVE WALKTHROUGH EXAMPLE

Step 4: Locate call number and library location

Will the user know what to do?

Will user notice that action is available?

Will user interpret feedback correctly?



2. **Interaction design** : beyond human-computer **interaction** / Preece, Rogers, Sharp.

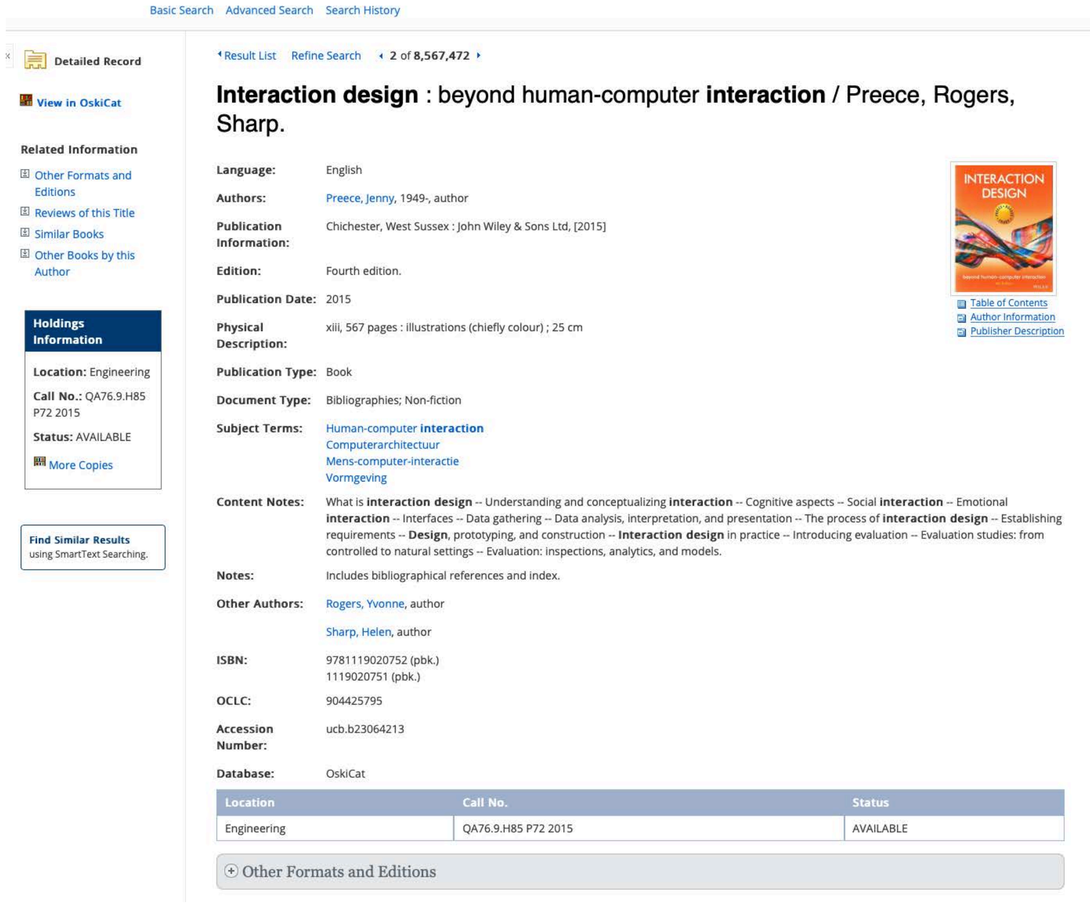
By: Preece, Jenny. Chichester, West Sussex : John Wiley & Sons Ltd, [2015] xiii, 567 pages : illustrations (chiefly colour) ; 25 cm Language: English, Database: OskiCat

Subjects: Human-computer **interaction**; Computerarchitectuur; Mens-computer-interactie; Vormgeving

Book

View in OskiCat

Location	Call No.	Status
Engineering	QA76.9.H85 P72 2015	AVAILABLE



Basic Search Advanced Search Search History

Detailed Record

View in OskiCat

Related Information

- Other Formats and Editions
- Reviews of this Title
- Similar Books
- Other Books by this Author

Holdings Information

Location: Engineering

Call No.: QA76.9.H85 P72 2015

Status: AVAILABLE

More Copies

Find Similar Results using SmartText Searching.

Result List Refine Search 2 of 8,567,472

Interaction design : beyond human-computer **interaction** / Preece, Rogers, Sharp.

Language: English

Authors: Preece, Jenny, 1949-, author

Publication Information: Chichester, West Sussex : John Wiley & Sons Ltd, [2015]

Edition: Fourth edition.

Publication Date: 2015

Physical Description: xiii, 567 pages : illustrations (chiefly colour) ; 25 cm

Table of Contents
Author Information
Publisher Description

Publication Type: Book

Document Type: Bibliographies; Non-fiction

Subject Terms: Human-computer **interaction**
Computerarchitectuur
Mens-computer-interactie
Vormgeving

Content Notes: What is **interaction design** -- Understanding and conceptualizing **interaction** -- Cognitive aspects -- Social **interaction** -- Emotional **interaction** -- Interfaces -- Data gathering -- Data analysis, interpretation, and presentation -- The process of **interaction design** -- Establishing requirements -- **Design**, prototyping, and construction -- **Interaction design** in practice -- Introducing evaluation -- Evaluation studies: from controlled to natural settings -- Evaluation: inspections, analytics, and models.

Notes: Includes bibliographical references and index.

Other Authors: Rogers, Yvonne, author

Sharp, Helen, author

ISBN: 9781119020752 (pbk.)
1119020751 (pbk.)

OCLC: 904425795

Accession Number: ucb.b23064213

Database: OskiCat

Location	Call No.	Status
Engineering	QA76.9.H85 P72 2015	AVAILABLE

Other Formats and Editions

EMPIRICAL ASSESSMENT: QUALITATIVE

Qualitative: What we've been doing so far

Contextual Inquiry: try to understand user's tasks and conceptual model

Usability Studies: look for critical incidents in interface

Qualitative methods help us:

Understand what is going on

Look for problems

Roughly evaluate usability of interface

EMPIRICAL: QUANTITATIVE STUDIES

Quantitative

Use to reliably measure some aspect of an interface

Compare two or more designs on a measurable aspect

Contribute to theory of Human-Computer Interaction

Approaches

Collect and analyze user events that occur in natural use

Controlled experiments

Examples of measures

Time to complete a task, Average number of errors on a task, Users' ratings of an interface*

** You could argue that users' perception of speed, error rates etc is more important than their actual values*

COMPARISON

Qualitative studies

Faster, less expensive → especially useful in early stages of design cycle

Quantitative studies

Reliable, repeatable result → scientific method

Best studies produce generalizable results

DESIGNING CONTROLLED EXPERIMENTS

STEPS IN DESIGNING AN EXPERIMENT

1. State a lucid, testable hypothesis
2. Identify variables
   (independent, dependent, control, random)
3. Design the experimental protocol
4. Choose user population
5. Apply for human subjects protocol review
6. Run pilot studies
7. Run the experiment
8. Perform statistical analysis
9. Draw conclusions

EXPERIMENT DESIGN

Testable hypothesis

Precise statement of expected outcome

Independent variables (factors)

Attributes we manipulate/vary in each condition

Levels – values for independent variables

Dependent variables (response variables)

Outcome of experiment (measurements)

Usually measure user performance

EXPERIMENT DESIGN

Control variables

Attributes that will be fixed throughout experiment

Confound – attribute that varied and was not accounted for

Problem: Confound rather than independent variables could have caused change in dependent variables

Confounds make it difficult/impossible to draw conclusions

Random variables

Attributes that are randomly sampled

Increases generalizability

VARIABLE TYPES

Nominal: categories with labels, no order

Ordinal: categories with rank order

Continuous: interval (w/o zero point), ratio (w/ zero point)

COMMON METRICS IN HCI

Performance metrics:

- Task success (binary or multi-level)
- Task completion time
- Errors (slips, mistakes) per task
- Efficiency (cognitive & physical effort)
- Learnability

Satisfaction metrics:

- Self-report on ease of use, frustration, etc.

SLIPS VS MISTAKES

Mistake — when a person plans to do the wrong thing and is successful (conscious)

Example: Sitting at a table when you have to order from a bar or trying to use an old Xbox game controller like a motion-sensitive Wiimote and gesturing with it in the air when you need to press the buttons , misreading low oil-pressure light as low tire pressure

Slip — when a person plans to do one thing, but then inadvertently does something else (unconscious)

Example: Leaving your change in a vending machine or forgetting to replace the gas cap after filling up your car with fuel. Or even accidentally typing a wrong word when you're writing a text, even though you know how to spell it, sometimes you'll still type it wrnogly, or putting liquid hand-soap on toothbrush

SLIPS VS MISTAKES

The difference between **slips** and **mistakes** is important.

If someone makes a **mistake** because they don't know what to do, we can train them to improve their performance (informing a person that they have to order at the bar). They have the **wrong mental model.**

But training doesn't make us any less likely to slip up.

A **slip** is when the user has the **correct mental model** of the interaction **yet makes an error** on accident.

PERFORMANCE METRIC: LOSTNESS

Calculated using the ratio of visited and optimal node counts as shown below:

N = total number of different pages visited (including revisits)

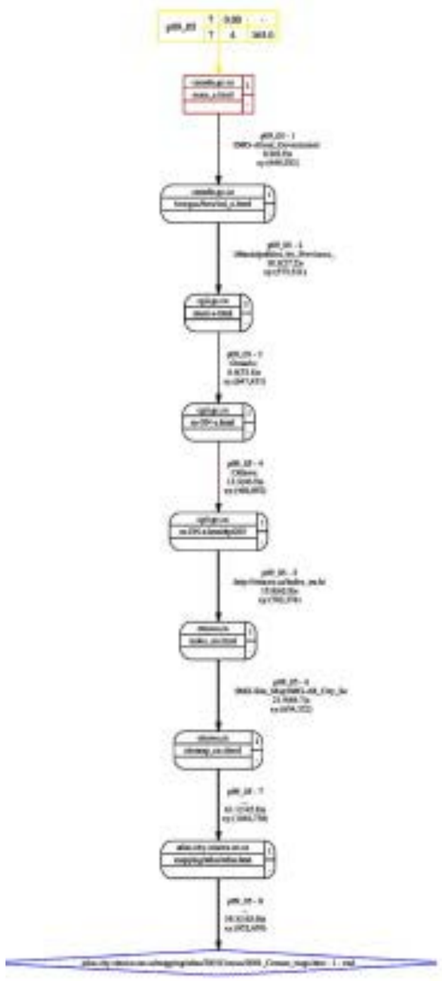
U = total number of unique pages visited

O = minimum (optimal) number of pages to accomplish task

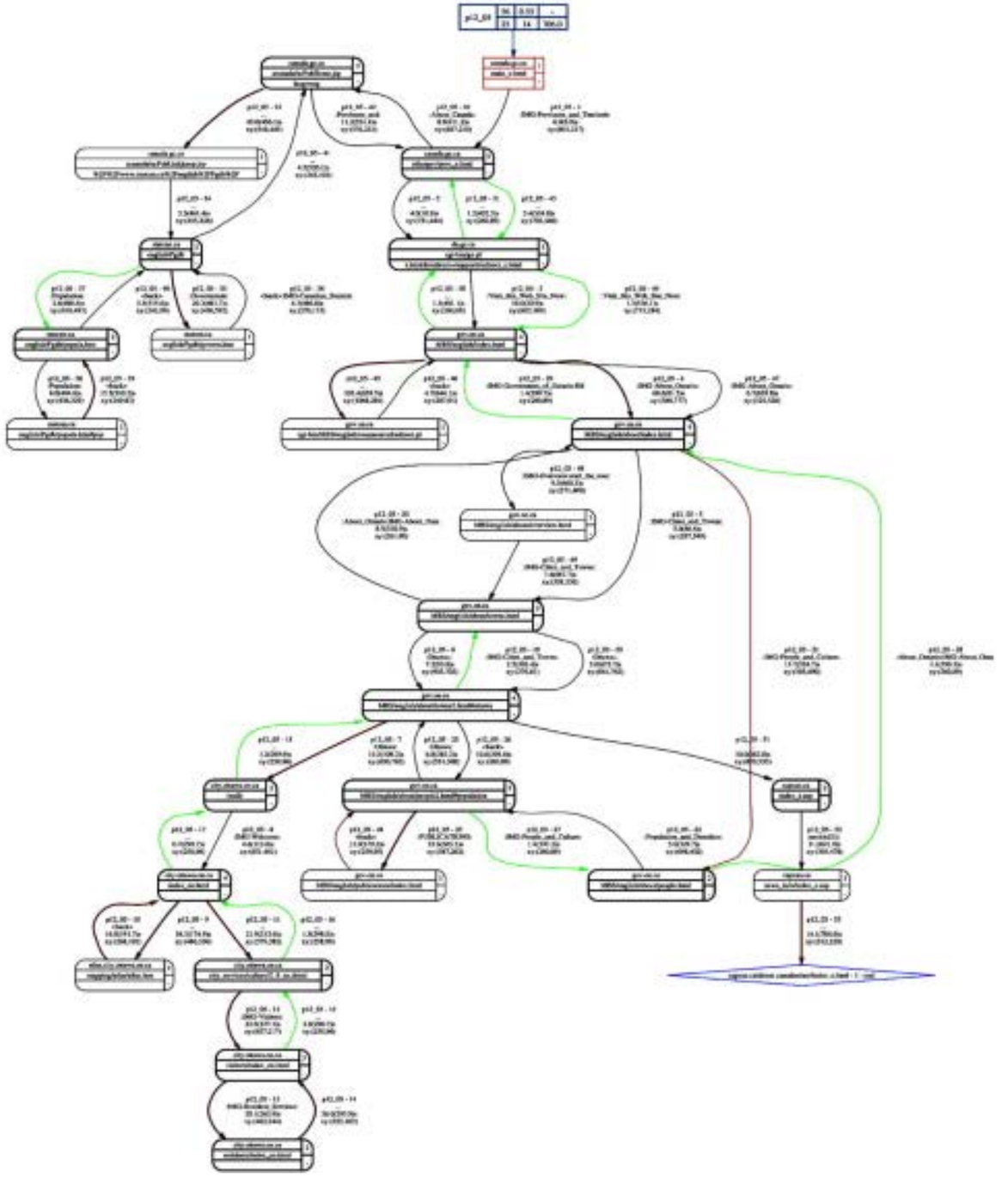
$$\texttt{Revisits} = 1 - \texttt{U}/\texttt{N},$$

$$\texttt{Lostness} = \sqrt{(\texttt{U}/\texttt{N} - 1)^2 + (\texttt{O}/\texttt{U} - 1)^2}.$$

Lostness_R = 1.3



Lostness_R = 2.8



SATISFACTION METRIC: LIKERT SCALES

Respondents rate their level of agreement to a statement

Likert data is ordinal, not continuous (matters for analysis)!

"Overall, I am satisfied with the ease of completing the tasks in this scenario"

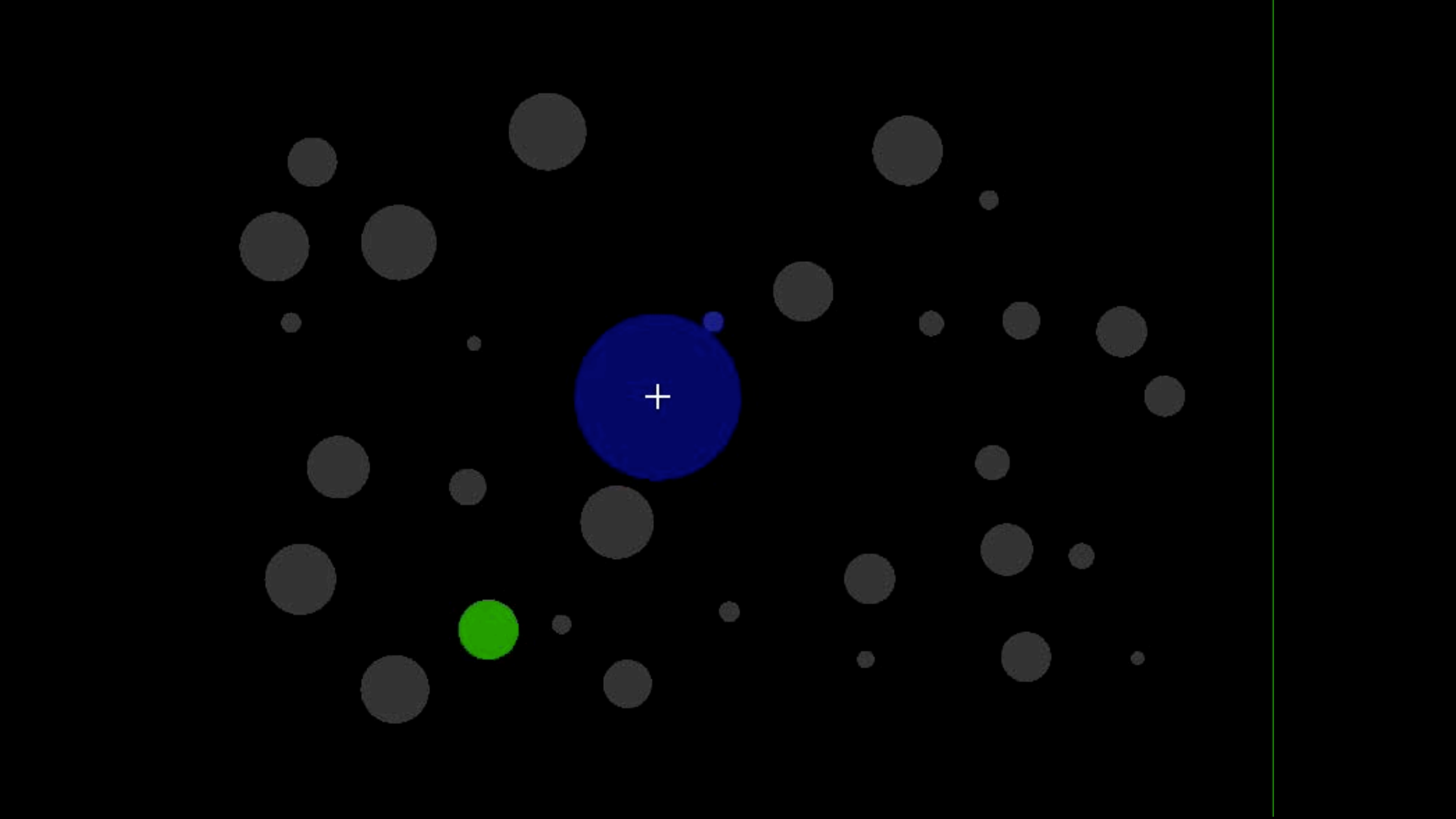
1: Strongly Disagree

2: Disagree

3: Neither agree nor disagree

4: Agree

5: Strongly agree



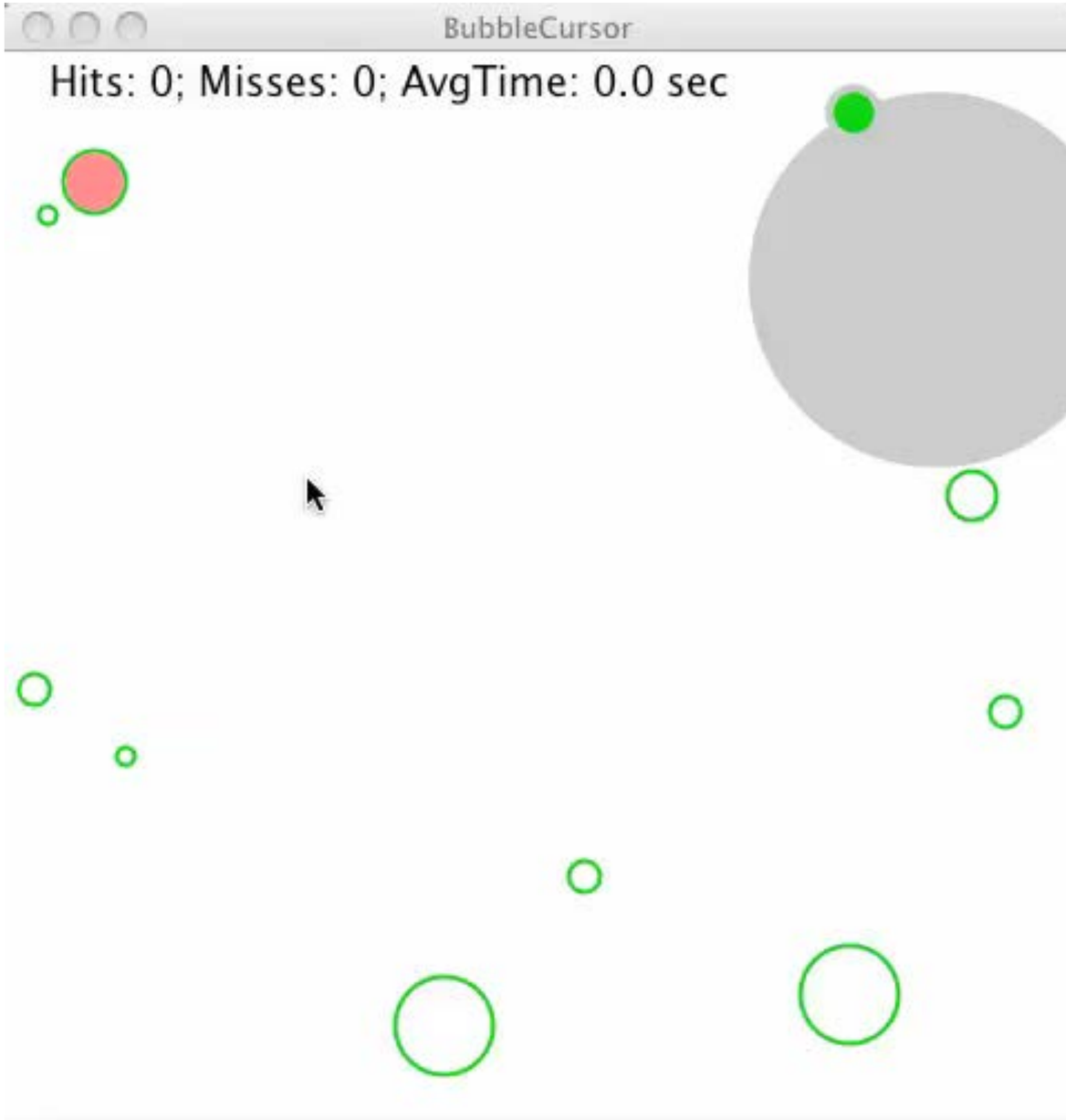
VARIABLES FOR THE BUBBLE CURSOR

Independent variables

Dependent variables

Control variables

Random variables



VARIABLES

Independent variables

Cursor type (bubble, normal, area?)
Target Distance
Target Width

Dependent variables

Movement Time
Error Rate
User Satisfaction

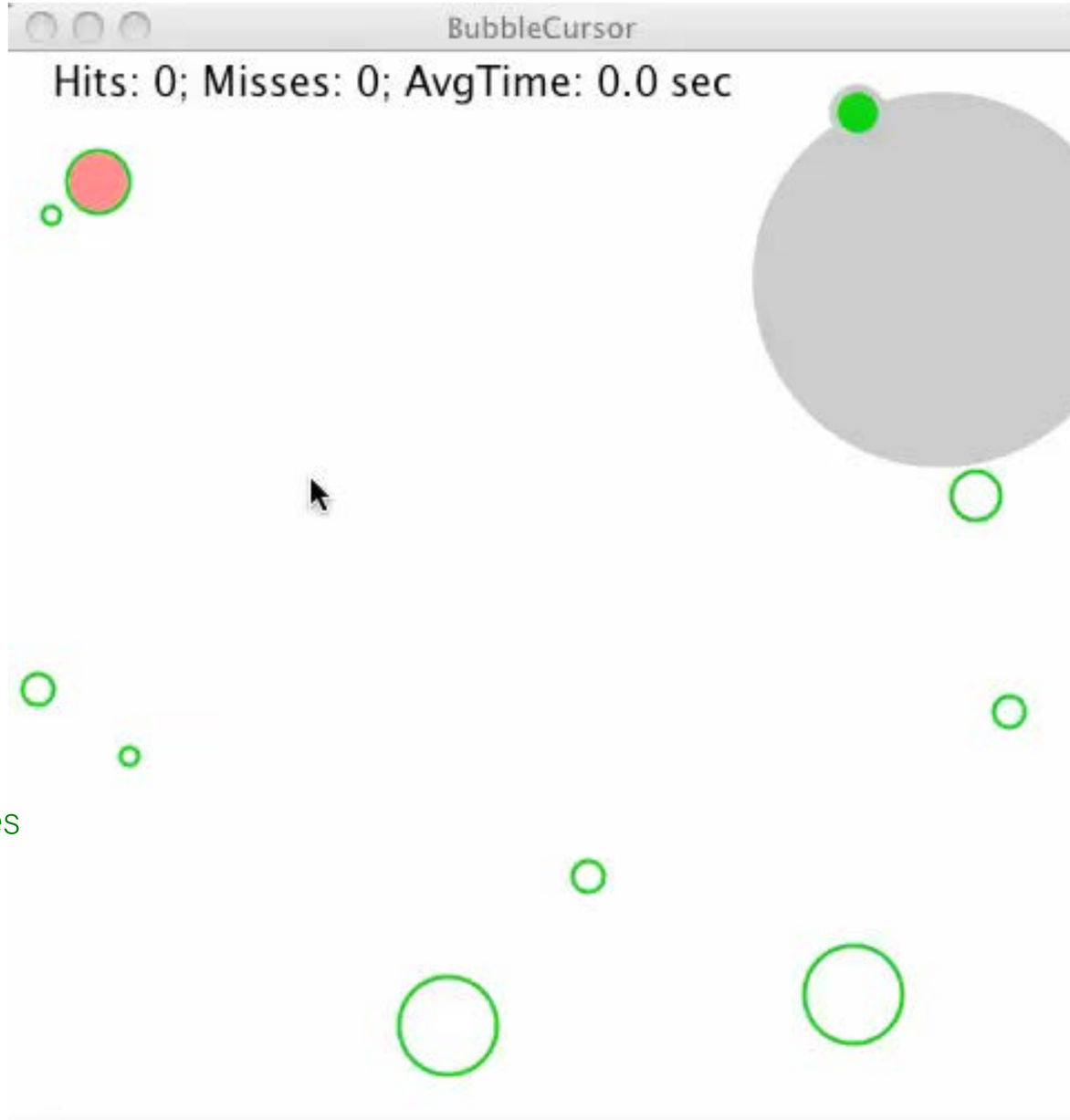
Control variables

Color scheme, input device, screen size

Random variables

Location, environment,
Attributes of subjects
Age, gender, handedness, ...

Conducting studies online
vs. in person strongly influences
which variables are controlled
and which are random



GOALS

Internal validity

Manipulation of **Independent Variable** is cause of change in **Dependent Variable**

Requires eliminating confounding variables (turn them into IVs or RVs)

Requires that experiment is replicable

External validity

Results are generalizable to other experimental settings

Ecological validity – results generalizable to real-world settings

Confidence in results

Statistics

EXPERIMENTAL PROTOCOL

What is the task? (must reflect hypothesis!)

What are all the combinations of conditions?

How often to repeat each combination of conditions?

Between subjects or within subjects

Avoid bias (instructions, ordering, ...)

NUMBER OF CONDITIONS

Consider all combinations to isolate effects of each Independent Variable (factorial design)

(3 cursor types) * (3 distances) * (3 widths) = 27 combinations

Adding levels or factors can yield lots of combinations!

REDUCING NUMBER OF CONDITIONS

Vary only one independent variable leaving others fixed

Problem: ?

REDUCING NUMBER OF CONDITIONS

Vary only one independent variable leaving others fixed

Problem: Will miss effects of interactions

OTHER REDUCTION STRATEGIES

Run a few independent variables at a time

If strong effect, include variable in future studies

Otherwise pick fixed control value for it

Fractional factorial design

Procedures for choosing subset of independent variables to vary in each experiment

CHOOSING SUBJECTS

Pick balanced sample reflecting intended user population

Novices verses experts

Age group

Gender

....

Example

12 non-colorblind right-handed adults (male & female)

Population group can also be an Independent Variable or a Controlled variable

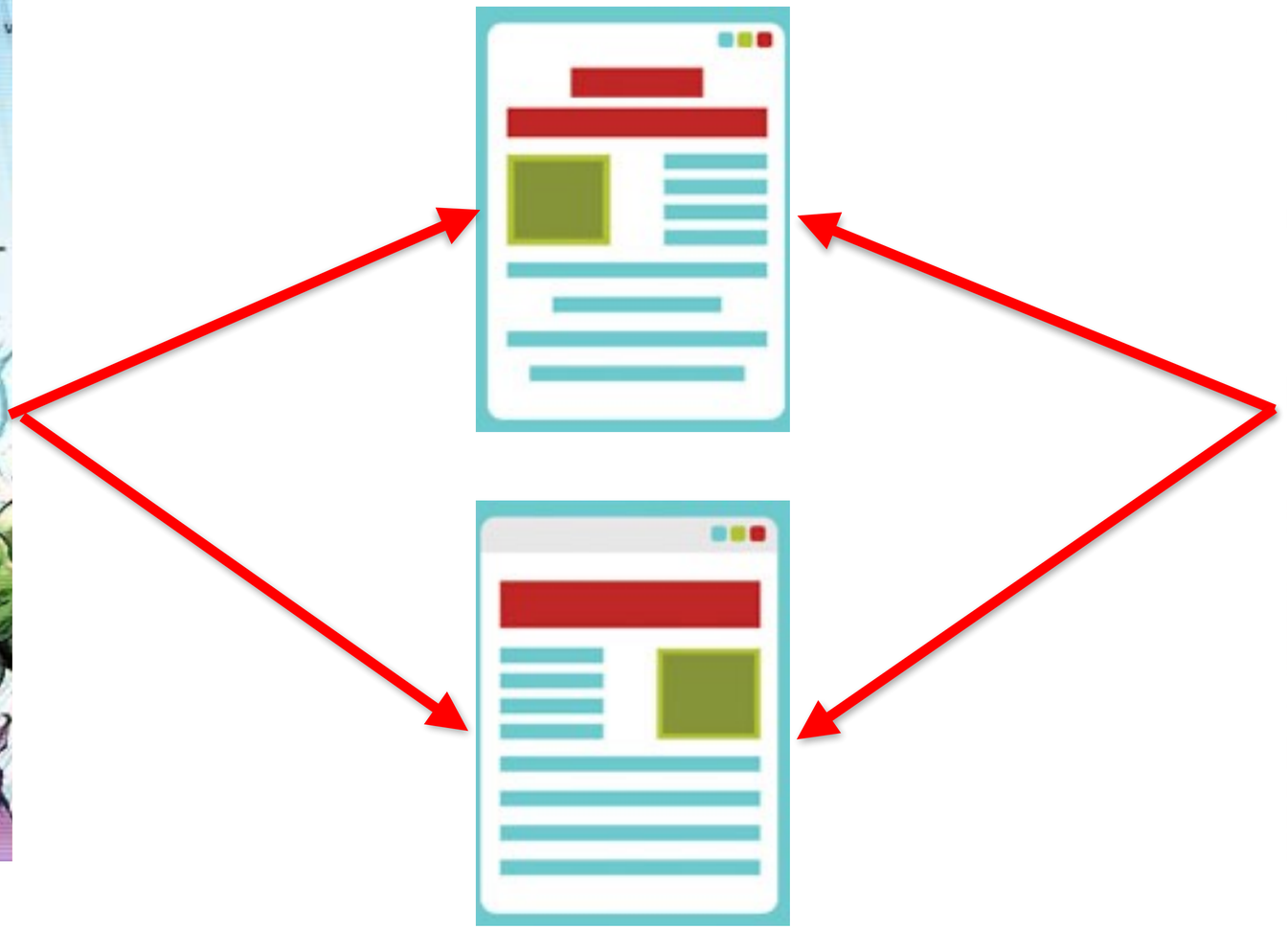
BETWEEN SUBJECTS DESIGN

Marvel Super Heroes use one interface

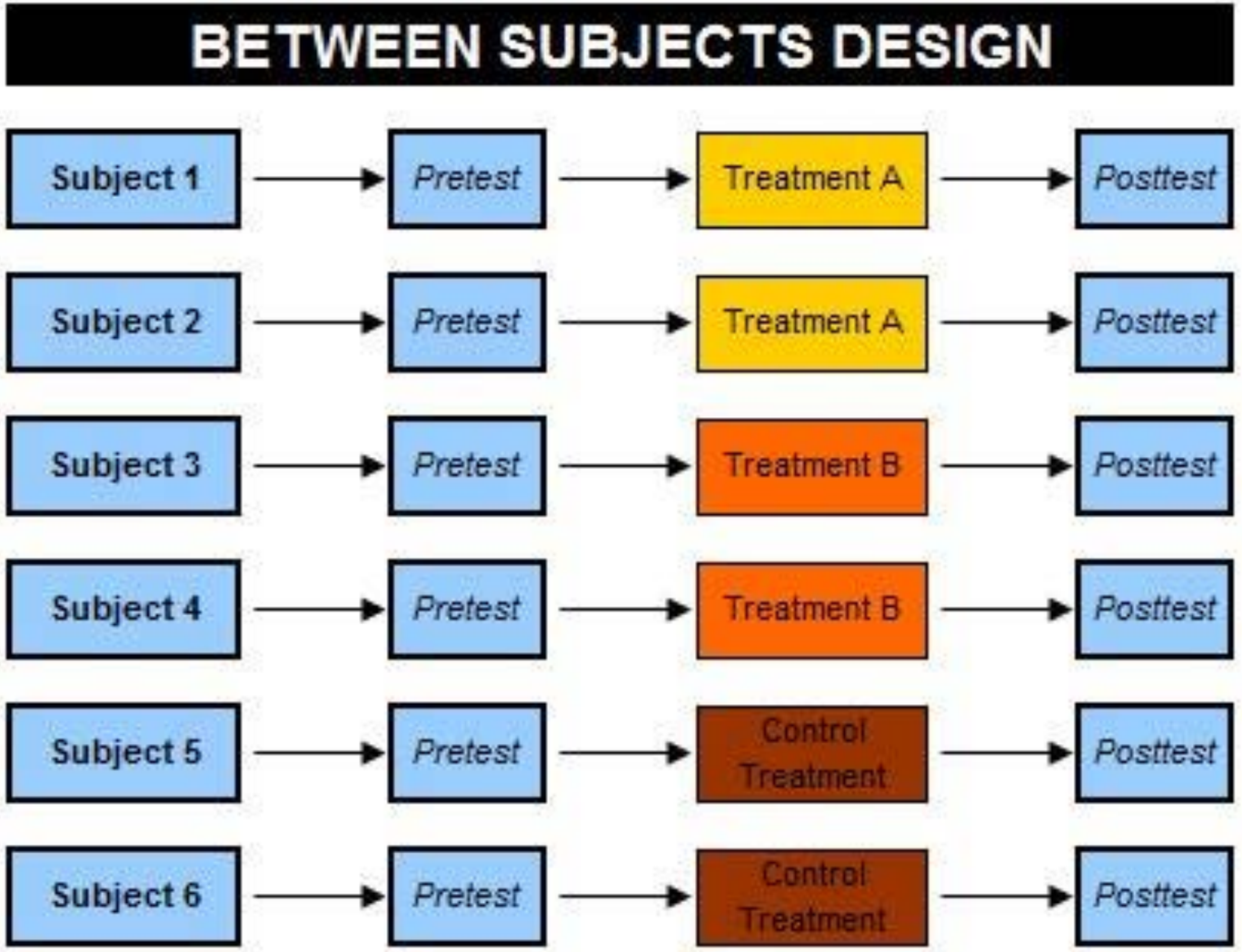
DC Heroes use one interface

WITHIN SUBJECTS DESIGN

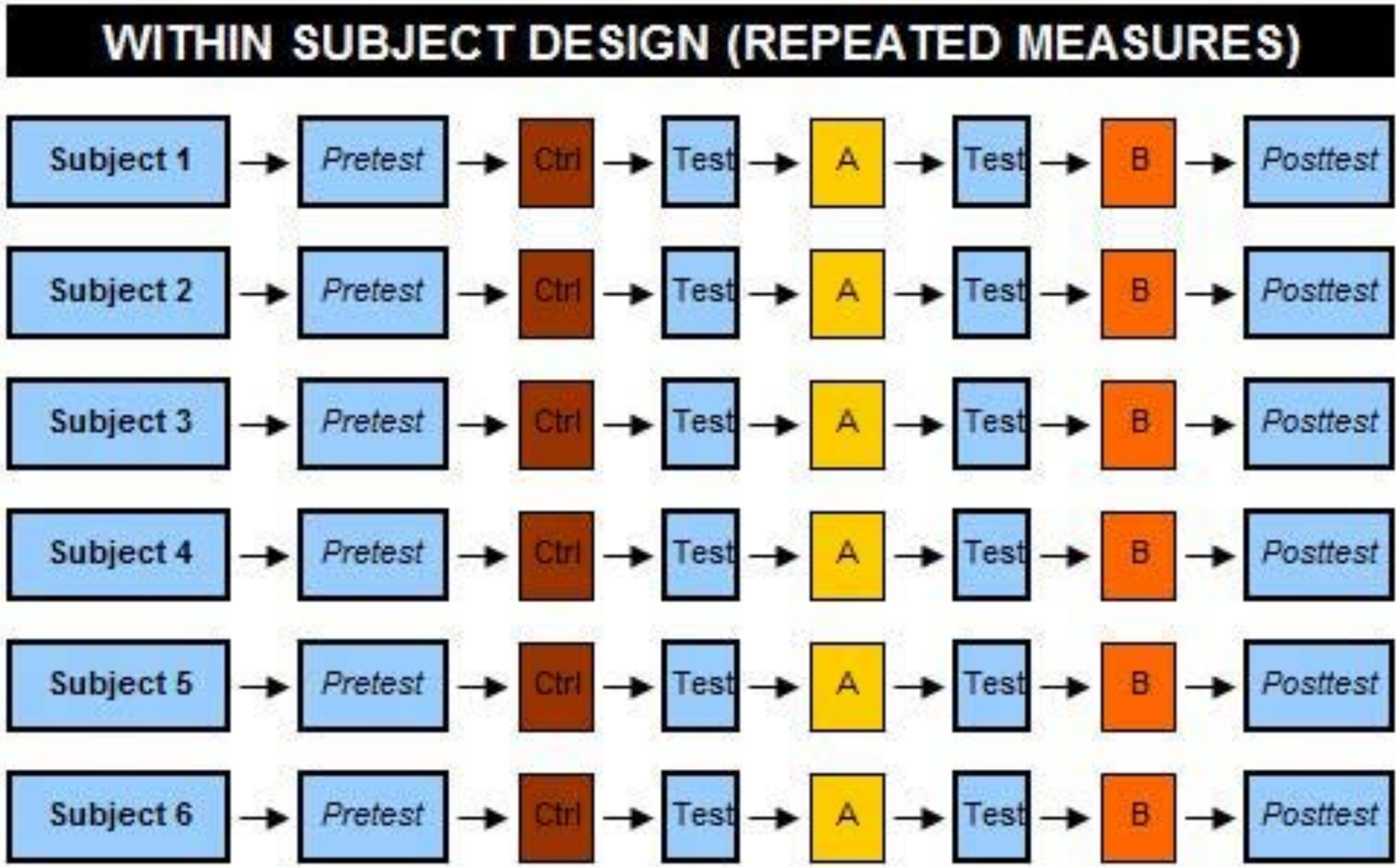
Everyone uses both interfaces



BETWEEN SUBJECTS DESIGN



WITHIN SUBJECTS DESIGN



Within Subjects

A group of people sees the test signs.

Between Subjects

One group of people sees one set of the test signs, and a different group sees another set.

BETWEEN VS. WITHIN SUBJECTS

Between subjects

Each participant uses one condition

+/- Participants cannot compare conditions

+ Can collect more data for a given condition

- Need more participants

Within subjects

All participants try all conditions

+ Compare one person across conditions to isolate effects of individual diffs

+ Requires fewer participants

- Fatigue effects

- Bias due to ordering/learning effects

WITHIN SUBJECTS: ORDERING EFFECTS

In within-subjects designs ordering of conditions is a variable that can confound results

Why?

Turn it into a random variable

Randomize order of conditions across subjects

Counterbalancing (ensure all orderings are covered)

Latin square (partial counterbalancing)

...

RUN THE EXPERIMENT

Always pilot it first!

Reveals unexpected problems

Can’t change experiment design after starting it

Always follow same steps – use a checklist

Get consent from subjects

Debrief subjects afterwards