Review and overview

Research Methods for Human-Centered Computing

Review and overview

Today's goal:

Review the materials covered so far, and give an overview of the non-experimental research methods

Outline:

- Review of part 1 (RQs, ethics, literature, theory)
- Review of part 2 (experiments)
- Overview of part 3 (other research methods)

Review of part 1

RQs, ethics, literature, theory

HCC as a science

Why do we need the scientific method in HCC?

- Practitioners rely on personal experience and authority
- Without a scientific foundation, these can introduce **bias**

The science of HCC fills this gap by building an organized body of knowledge

Research questions

A good research question has the potential to expand this body of knowledge

- It must be grounded in the existing knowledge base
  - Find gaps in related work
- It must be researchable
  - Be ready to operationalize them in hypotheses
- It must be important
  - What is the intellectual merit?
  - What are the broader impacts?

Types of questions

Descriptive: describe a certain phenomenon

Univariate: questions pertaining to a single variable

Multivariate/correlative: questions pertaining relationships between multiple variables

Causal: how one variable influences the other

PICO

A good research question identifies a:

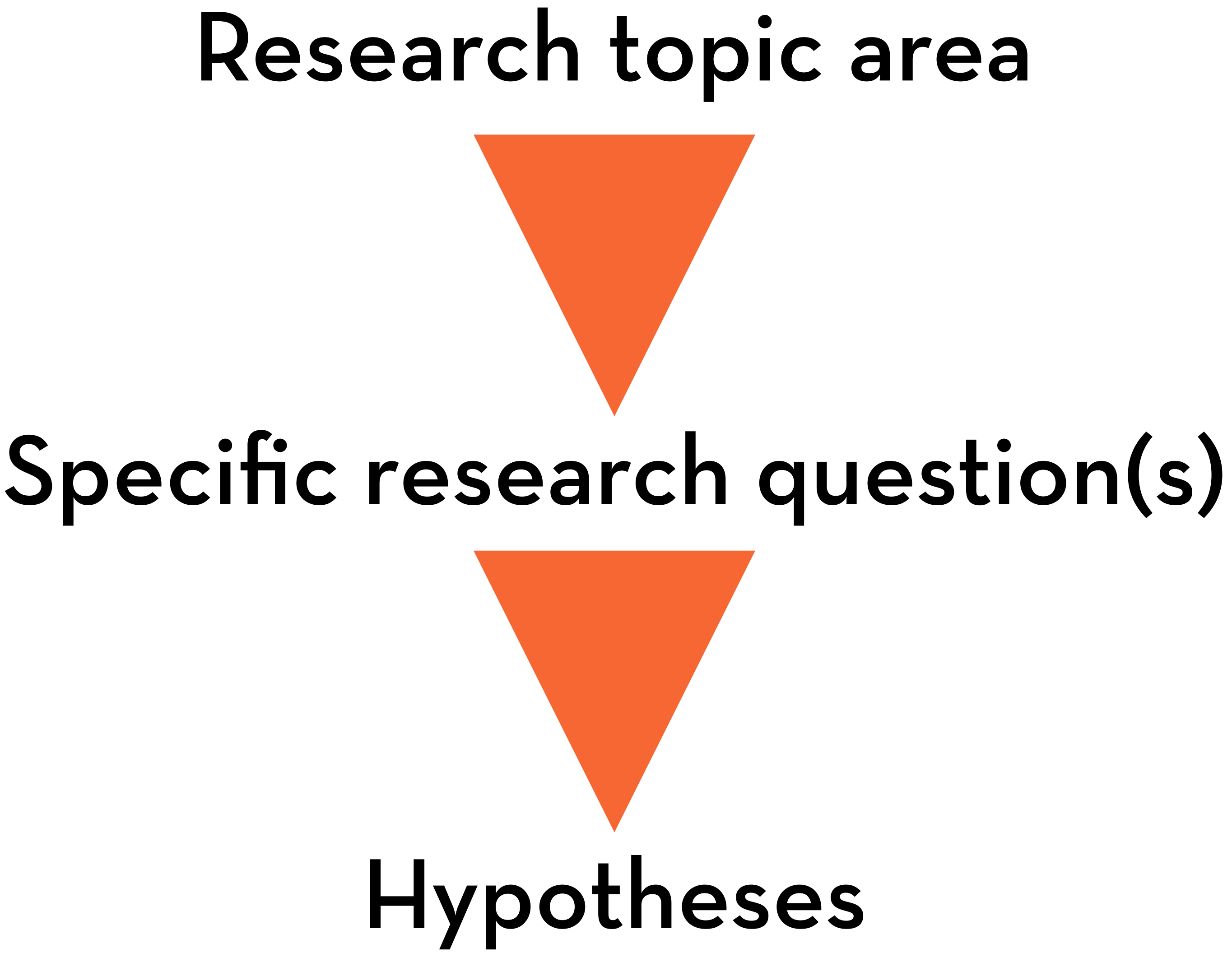
- Population
- Intervention (causal) or exposure (multivariate)
- Comparative / control (causal and multivariate)
- Outcome

Example: Does the new Facebook app (I) increase usage (O) among existing Facebook users (P) compared to the old app (C)?

versus hypotheses

Research questions:	**Hypotheses:**
Questions	Statements
Usually open-ended	Supported or rejected
General (each paper usually has 1–3 research questions)	Specific (each research question can lead to multiple hypotheses)
Describes broad relationships	Describes relationships between specific variables

Source



Research topic area

Specific research question(s)

Hypotheses

Example

Research topic area:

Preference elicitation in recommender systems
(i.e., the way such systems collect information about your preferences)

Research question:

Is there a relationship between domain knowledge and preference elicitation (PE) method in terms of the optimal user experience of a recommender system?

Example

Hypotheses:

Novices have a higher satisfaction when they use the case-based PE method (compared to the attribute-based PE method)

Experts have a higher satisfaction when they use the attribute-based PE method

Novices perceive the system with the case-based PE method as more useful

Experts perceive the system with the attribute-based PE method as more useful

Plagiarism

Copying text or ideas without giving proper credit

Not just copying others!

- Self-plagiarism: when you submit the same work twice

Not just copying paragraphs of text!

- Also when you summarize or paraphrase things **without** giving credit

Citing the source elsewhere in the paper is not “giving credit”

When you copy verbatim, use quotation marks

Plagiarism

Examples:

- Submitting a non-archival paper to a conference?
- Submitting the same paper to two different conferences?
- Adding a few paragraphs to a conference paper and submitting it to a journal?
- Combining the results of two studies and adding a substantial reflection section?
- Paraphrasing a wikipedia definition without citing the source?
- Using measurement scales from other researchers?

Research fraud

General principle: Be honest about your research!

Illegal practices:

- Fake studies / fudged data
- Selective data

Bad practices:

- p-hacking
- selective reporting

IRB

Minimize physical and/or psychological harm

Informed consent

- Make sure to give sufficient information!
- Signed consent generally not necessary (use a checkbox)

Deception means no fully informed consent

- Only allowed in certain cases
- Usually subject to more extensive review
- Requires debriefing

Literature review

Purposes of reviewing the literature:

- Provide a context for your research
- Avoid duplication effort
- Argue relevance of your work
- Find relevant theories
- Identify potential problems in conducting the research
- Identify "acceptable" practices of the field

Literature review

Types of papers you may want to look for:

- Overview(s) of the topic(s)
- Factoids
- Theory/theories
- Works whose "gap" you will attempt to fill
- Work that supports (an) assumption(s)
- Work that contains support for your hypotheses
- Work with similar methods
- Work with aligned conclusions

Cite correctly

Don’t cite something that is cited in another paper, cite the original work instead!

Don’t use string citations:

Example: “Multiple researchers note that privacy is an undying problem in social media [1, 6, 8, 12, 24, 46, 51, 56].”

Are you citing too many papers?

Usually symptom of not covering each work deep enough

- What does each of these works contribute?
- How is your work different from each of these works?

Norman's Theory

The action cycle and gulfs of execution/evaluation

- Explains how people use interfaces, and why they sometimes fail

Designer image, system image, use image

- Explains what causes some systems to be less usable than others

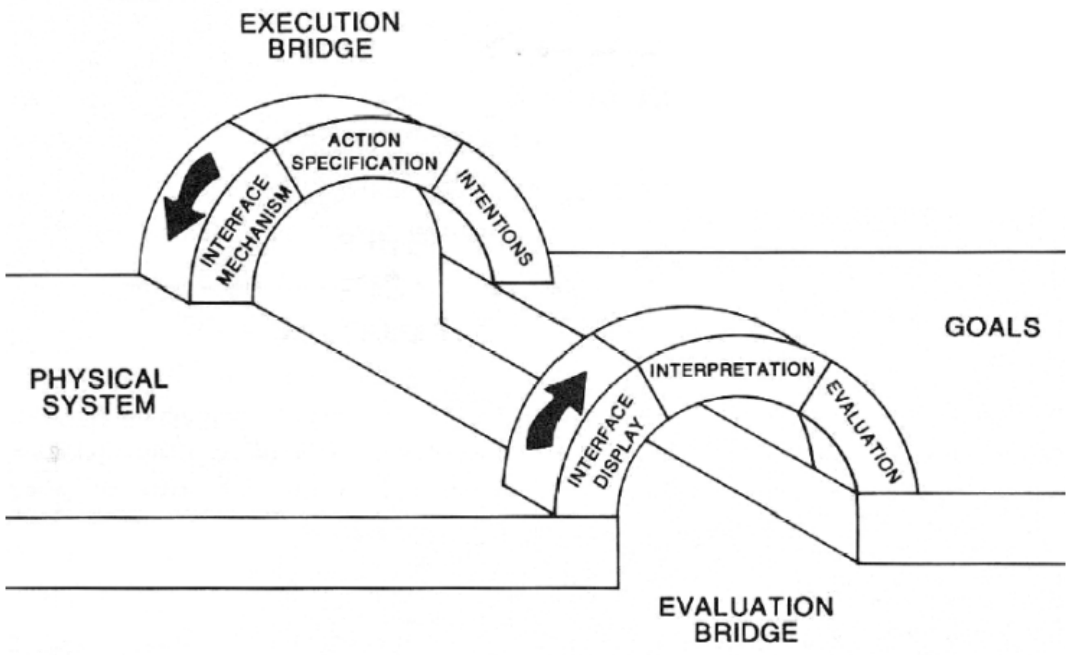
Constraints, signifiers, and feedback

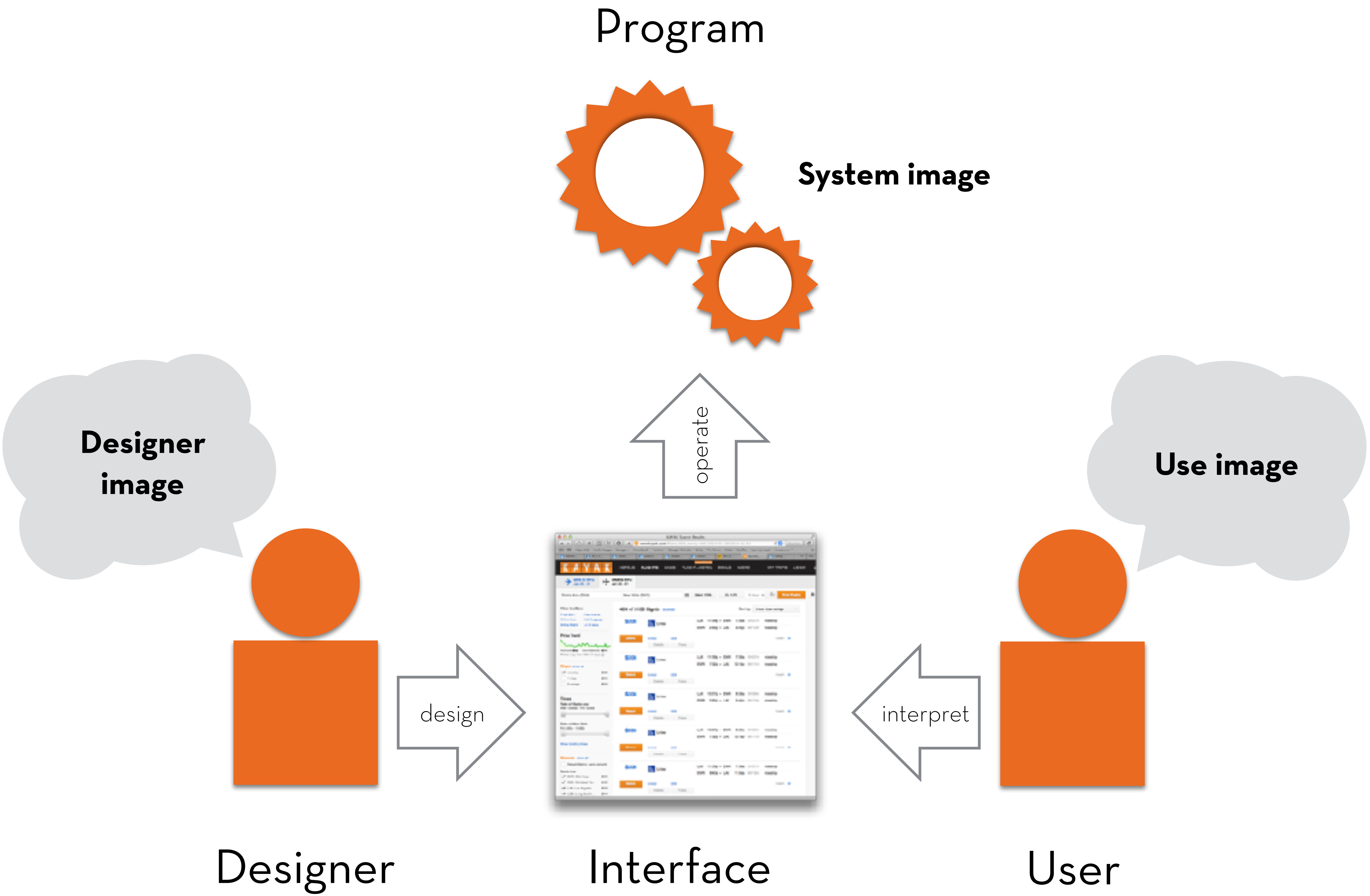
- Explains how you can increase the usability of interfaces

The action cycle

Norman created an abstract representation (a model) of how users perform tasks:

- How they turn their goals into actions (system input)
- How they evaluate the resulting system output





Program

System image

operate

Designer image

Use image

design

interpret

Designer

Interface

User

User interfaces

Norman argued that certain aspects of a user interface can help align the use image and system image:

- Constraints (making it impossible to do the wrong thing)
- Signifiers (demonstrating the right way to use a thing)
- Feedback (letting the user know what happened)

Careful use of constraints, signifiers, and feedback help reduce the mismatch between system image and use image

Cognitive Modeling

Cognitive architectures

- Abstractions of the mind, useful for reasoning
- Examples: model-human processor, ACT-R

Cognitive modeling

- A usability analysis based on how the brain works
- Examples: GOMS models, CogTool

Distributed cognition

Combination of people, systems, and artifacts is a cognitive system

- The system (a combination of subjects and artifacts that together perform a task) provides the goal
- Artifacts are pulled to the human side, and assigned cognitive capabilities
- Generalizations result from analyzing the collective manipulation of artifacts / representations
- Provides a formal analysis of artifacts and how they are used, and produces comparative data across settings

Situated Action

Examine the social context in which HCI occurs

- Goals are retrospective reconstructions of what happened; the situation is the driving factor
- Humans are pulled to the artifact side; they are reactive ciphers that react to stimuli in a behaviorist manner (controlled by the situation)
- Generalizations do not happen, due to the idea of moment-by-moment analysis
- Acknowledges the fluidity of goals and plans, but the exclusive focus on the situation may reduce its usefulness

Activity Theory

Treat plans as anticipatory reflections of recurring activity, transforming and transformed by culture an society

- Goals exist at several levels (activity: motives, actions: goals, operations: orienting basis), but originate from the subject's intentionality
- Humans control their activities; artifacts are mediators
- Generalizations can occur by looking at the historical development of activities and the artifacts that exist as mediators between subject and activity
- Treats consciousness at the individual level; situation influences but does not determine the actions

Review of part 2

User Experiments

Core components

A user experiment systematically tests how different system aspects (manipulations) influence the users' experience and behavior (observations)

- Manipulations: independent variables, system aspects, experimental conditions... ceteris paribus!
- Observations: dependent variables, objective or subjective outcomes of your manipulations
- Also consider covariates: things that change/moderate the effect(s) of the manipulation(s)

Hypotheses

Hypotheses are predictions regarding the influence of your independent variables (manipulations) on your dependent variables (outcomes)

Example: compared to text, comic-based policies increase privacy knowledge

Calculate the means in each condition. Do they differ a lot?

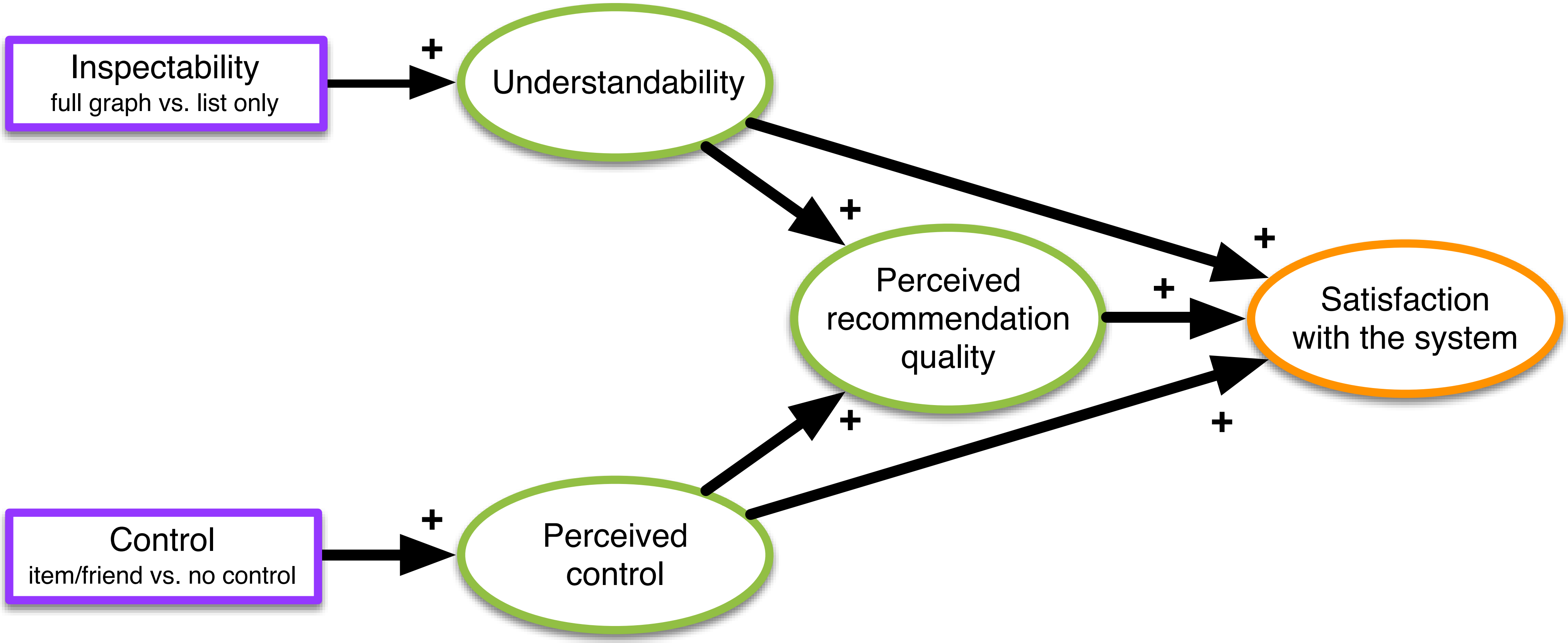
Given no effect, we expect the means to be roughly equal

H0: Mcomic = Mtext

To test H1, we try to **reject H0**

Hypotheses

Build a model of hypotheses!


Inspectability
full graph vs. list only
+
Understandability
+
+
Perceived recommendation quality
+
Satisfaction with the system
+
+
Control
item/friend vs. no control
+
Perceived control

Random assignment

Randomization neutralizes (but doesn't eliminate) participant variation

- Between-subjects: realistic, manipulations hidden, but many participants needed
- Within-subjects: removes subject variability, but potential demand characteristics and spill-over effects
- Simultaneous within: allows for comparison, not realistic

Random assignment

Generally, use between-subjects for UX research, within-subjects for psych research

- In factorial designs you can do both!

Multi-level models: multiple hospitals, multiple patients per hospital, multiple doctor visits per patient

- Variables exist on multiple levels

Participants

You should sample from your **target population**

- An unbiased sample of users of your system
- Avoid limiting your scope to a specific set of users
- Go beyond WEIRD participants

What if your sample is skewed?

- Identify the issue, and then either use the skewed parameter as a covariate, or stratify your sample

Study location

In the lab:

- More control, advanced measurement and instruments
- But more difficult to recruit, often skewed samples

Online:

- Fast recruitment, cheaper, anonymous
- But limited control, measurement and manipulation

On location:

- Convenient and contextual (mostly for qualitative tho)

Sample size

Small studies (N << 100) may find medium or large effects that are not significant

Large studies (N >> 100) may find very small effects that are significant

Do a power analysis!

- Doing an actual power analysis is not part of the test, but make sure you understand the principles

Power

	There is a real effect	There is no real effect
Found an effect	**Power**	alpha (false positive)
Found no effect	beta (false negative)	1–alpha (true negative)

Power

1-beta = power

The probability of finding an effect that is really there

How high is our power? Power depends on...

...alpha (if we use $p < .01$, our power is lower)

...effect size (if the effect is smaller, power is lower)

...N (if we use a larger sample, we increase our power)

Given alpha = 0.05, and a certain expected effect size, how large should our N be to find a true effect 80% of the time?

Measuring data

Levels of measurement

Categorical: Nominal - you can do counts

Categorical: Ordinal - you can order them, but differences not equal

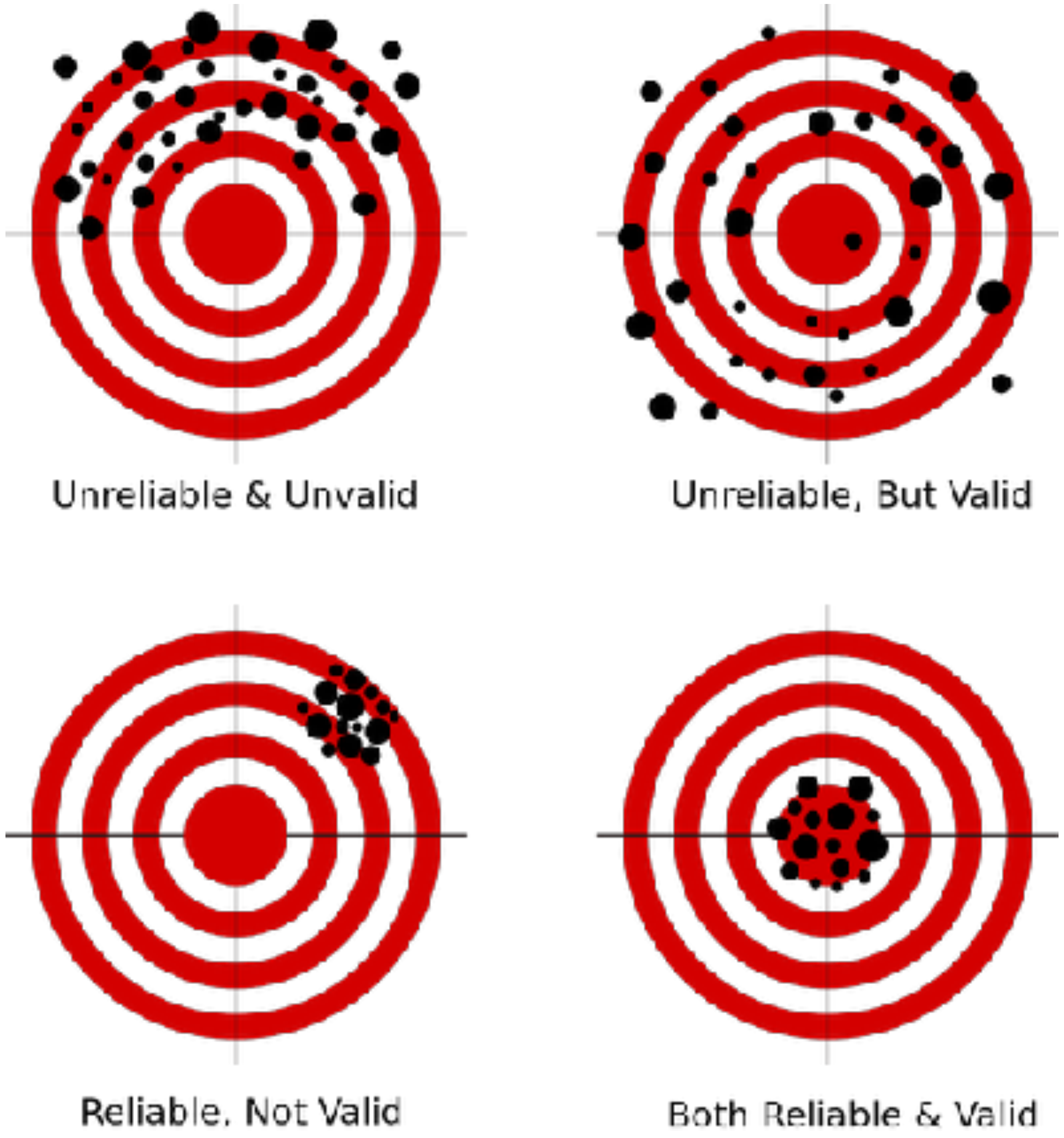
Continuous: Interval - you can do addition, averaging, but no meaningful zero point

Continuous: Ratio - you can multiply

Validity and error

Validity: does it measure what you intend to measure?

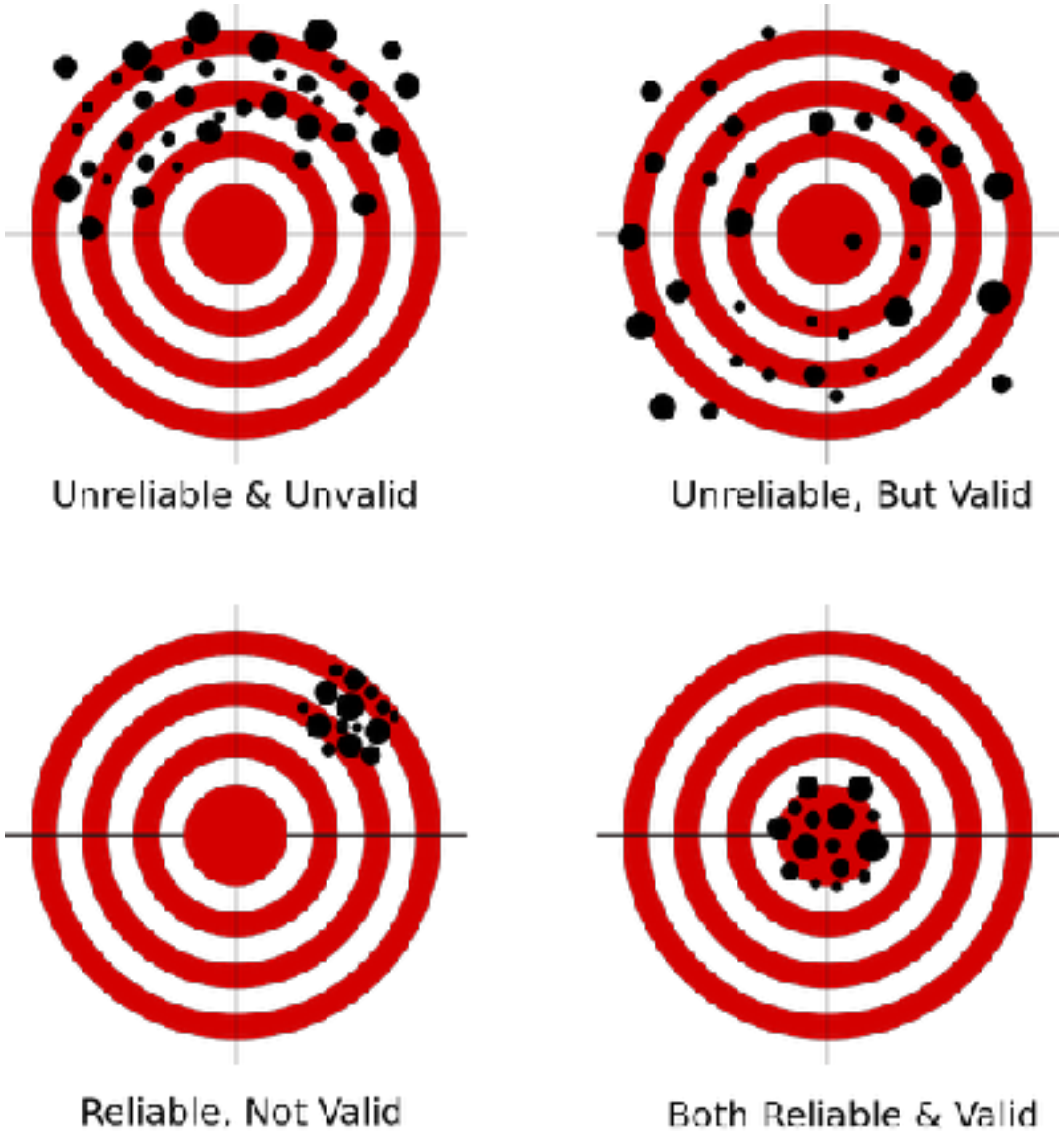
- Content validity
- Criterion validity (predictive, concurrent)
- Construct validity (discriminant, convergent)



Validity and error

Error: will you get the same value on repeated measurement?

- Environment
- Participants
- Measurements



Psychometrics

Definition: The measurement of social and psychological concepts or traits

Scale: a collection of items, intended to reveal levels of a of a social/psychological concept

- Answers are an indirect observation on the concept/trait

Accurate measurement requires a **shared conceptual understanding** between all participants and researcher

- This can be accomplished by asking multiple questions per concept

Measurement scales

Use existing scales because:

- Constructing your own scale is a lot of work
- "Famous" scales have undergone extensive validity tests
- Ascertains that two related papers measure exactly the same thing

Create new / adapt existing scales when:

- Existing scales do not hold up
- Nobody has measured what you want to measure before
- Scale relates to the specific context of measurement

Creating a scale

1. Create a concept definition

    A careful explanation of what you want to measure

    Grounded in theory

2. Generate items

    Redundancy is good, but differentiate

    Use both positively and negatively phrased items

    Keep it clear and simple

Creating a scale

3. Determine the response format

    5- and 7-point scales are most common

    Usually "agreement", but also frequency, importance, etc.
    And semantic differential

4. Pre-Test the items

    Expert discussion, card sorting, think-aloud testing

5. Include validation items

    For concurrent validity

Creating a scale

6. Administer the scale to a development sample

   On a developmental sample

7. Evaluate the items

   M&E part 2

8. Optimize scale length

   Final scale should have least 3 (but preferably 5 or more) items per scale

Statistics

Covariance and correlation

Used to describe a relationship between two variables

Linear regression

How much each 1-point increase in X is associated with an increase in Y (assumes causality)

t-tests

Difference between two groups/conditions

Statistics

ANOVA

Differences between multiple groups (make sure to do an omnibus test, and to account for family-wise error)

factorial ANOVA

Differences related to multiple manipulations (to test for interaction effects)

Non-normal data

E.g. logistic regression to estimate probabilities of binary events

Statistics

Multi-level models

- To account for repeated measures, partially within-subjects designs, and grouped data
- Avoiding the violation of independent errors

Subjective data

- Measurement scales (see earlier); CFA and SEM (see M&E part 2)

Mediation analysis

- Does X -> Y because of some X -> M -> Y?

Overview of part 3

Non-experimental research methods

Other methods

Surveys

Think-aloud an other usability tests

Contextual inquiry and other field studies

Diary studies and experience sampling

Participatory design and design research

Grounded theory and other interview studies

Surveys

A study that measures (often subjective) aspects without manipulating anything

Purely correlational

- What causes what? Hard to determine without manipulations
- Third variable problem

No ceteris paribus

- Hard to get rid of confounding variables

Think-aloud

A means to study the usability of a system

- Conducted with the express purpose of improving the usability

Asks participants to verbalize (but not interpret!) their thought process

- Researcher takes notes, records errors/problems

Results of multiple participants are usually pooled and interpreted by the researcher to find solutions

Contextual inquiry

A means to investigate a work practice **before** a new system is introduced

- A user-centric form of requirements gathering

Observe and ask questions

- Apprentice-master relationship
- Observe breakdowns

Integrate findings (work models) and try to design a new system that is compatible with existing work practices

Diary studies / ESM

A way to get in-situ feedback about rare events

- Whenever observation is too time-consuming

Diary studies: log behavior and feedback at set intervals

Experience sampling: log behavior and feedback on demand

- Usually some sort of trigger

Often includes some sort of retrospective interview

Participatory design

Participatory design: Involve the user in the design and development of a product

Often paper-based

Designs are not used directly, but needs and desires are extracted and then designed for at a higher level

Design research: Design as part of research

Often: see how people react to a controversial design

Grounded theory

A means to study a field/phenomenon and develop theories about it

- Highly qualitative and inductive work

Requires deep, intensive interviews

- Sequential analysis, selective sampling

Transcription of interviews, thematic coding, and several rounds of synthesis

- Culminates in new theory

Comparison

Scientific studies usually result in papers

- E.g., experiments, most surveys, grounded theory, most diary studies

Practical studies aim to contribute to design/development

- Think-aloud testing, contextual inquiry, participatory design, some surveys, some diary studies

Comparison

Summative studies evaluate the quality of something

- E.g., most experiments, some surveys

Formative studies aim to improve something

- Think-aloud testing, contextual inquiry, participatory design, some surveys

Outside of this categorization:

- Grounded theory is neither, diary studies can be either

Comparison

Contextual inquiries are used **before** technical solutions are developed

Think-aloud tests and participatory design are used **during** the development (think-aloud late; participatory design early)

Experiments are often conducted **after** the development

Surveys can happen at any point; grounded theory falls outside of this categorization