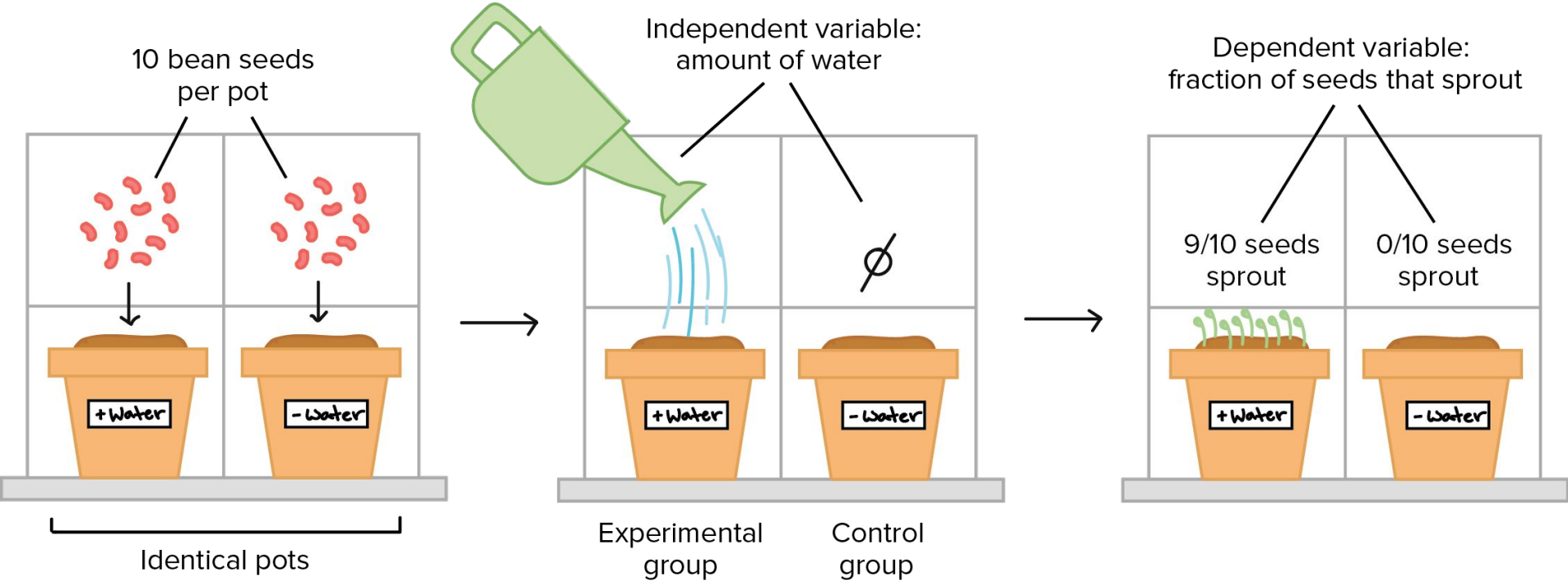
Experiment Design

Kyungsik Han

Experiment Design

- Designing good controlled experiment is an art



How to design a study well

Experiment Design

- Why bother with experiment design?

- To establish strong evidence linking manipulated treatments to changes in one or more outcomes
- To determine causation
  - Changes to x cause changes to y in this measurable way

Types of HCI Studies

- Descriptive investigations focus on constructing an accurate description of what is happening
- Relational investigations enable the researchers to identify relations (correlations) between multiple factors
- However, relational studies can **rarely** determine the causal relationship between multiple factors

Experimental research allows the establishment of a **causal relationship**. Usually these are controlled experiments

Types of HCI Studies

Types of research	Focus	General Claims	Typical Methods
Descriptive	Describe a situation or a set of events	X is happening	Observations, field studies, focus groups, interviews
Relational	Identify relations between multiple variables	X is related to Y	Observations, field studies, surveys
Experimental	Identify causes of a situation or a set of events	X is responsible for Y	Controlled experiments

Hypotheses

- An experiment normally starts with a hypothesis
- A hypothesis is a precise problem statement that can be directly tested through an empirical investigation
- Compared with a theory, a hypothesis is a smaller, more focused statement that can be examined by a single experiment

Example: *The iOS virtual keyboard is faster and more accurate than the Android virtual keyboard*

Null Hypotheses

- Null hypothesis: typically states that there is no difference between experimental treatments
  - Example: *There are no detectable differences in the speed or accuracy of the iOS virtual keyboard than the Android virtual keyboard*

- The goal of an experiment is to find statistical evidence to confirm or reject null hypothesis in a reliable fashion
- A hypothesis should specify the independent variables and dependent variables

Independent Variables

- Independent variables (IV): the factors that the researchers are interested in studying or the possible “cause” of the change
- IV is independent of a participant’s behavior
- IV is usually the treatments or conditions that the researchers can control

Independent variables are things the experimenter manipulates

Typical Independent Variables in HCI

- Those that relate to technology
  - Types of technology or device (e.g., keyboard type)
  - Types of design (e.g., design A vs design B)
- Those that relate to users
  - age, gender, computer experience, professional domain, education, culture, motivation, mood, disabilities, etc.
- Those that relate to context of use
  - Physical status
  - User status
  - Social status

Dependent Variables

- Dependent variables (DV) refers to the outcome or effect that the researchers are interested in
- DV is dependent on a participant's behavior or the changes in the IVs
- DV is usually the outcomes that the researchers need to measure

Dependent variables are things the experimenter measures

Typical Dependent Variables in HCI

- Efficiency
  - e.g., task completion time, speed
- Accuracy
  - e.g., error rate
- Subjective satisfaction
  - e.g., Likert scale ratings
- Ease of learning and retention rate
- Physical or cognitive demand
  - e.g., NASA task load index
- and more…

Factors

- Same as independent variables
- An experiment with a control group and a treatment (or experimental) group is a single-factor (or one-way) experiment

- Example
  - Two groups: treatment get broccoli every morning, control does not
  - The factor or independent variable might be called food
  - The measure or dependent variable number of pushups at 11am

Levels

- Levels are values a factor can assume (i.e., groups)
- Examples
  - Factor food has two levels: broccoli, no-broccoli
  - Factor keyboard has two levels: iOS, Android
  - Factor posture has three levels: sitting, standing, walking

Finding differences among levels is
what an experiment is all about (very important)

Between-Subjects Design

- Each participant (subject) experiences only one level of a factor
  - Requires more participants
  - But avoids possible confounds
  - Easier to analyze statistically

- Example
  - Participants type using either iOS keyboard OR Android keyboard, but not both

Within-Subjects Design

- Each participant (subject) experiences all levels of a factor
  - Much more powerful statistically
  - But can introduce confounds

- Example
  - Participants complete typing tasks using both an iOS keyboard AND Android keyboard

When to use between-subjects vs within-subjects?

Carryover Effects

- The effect of one condition “carries over” into the next condition
- Common in within-subjects designs
  - e.g., learning from one condition to the next
- Neutralize carryover effects with counterbalancing

Counterbalancing

- Choose an order of presentation to neutralize any carryover effects

Example (two conditions):
p1: iOS, Android
p2: Android, iOS
p3: iOS, Android
p4: Android, iOS
…

Example (three conditions):
p1: A, B, C
p2: A, C, B
p3: B, A, C
p4: B, C, A
p5: C, A, B
p6: C, B, A

Mixed Factorial Design

- Contains at least one between-subjects factor and one within-subjects factor
- Also called split-plot designs

- Example: Do males and females perform differently with different mobile keyboards?
  - Between subjects factor sex with two levels: male, female
  - Within subjects factor keyboard with two levels: iOS and Android

Confounds

- Any unaccounted for factors that could explain your results
- Serious confounds ruin experiments

Examples

- Unequal treatments or procedures
  - e.g., participants typed 5 phrases with iOS and 20 with Android
- Sources of non-random variation
  - e.g., all participants who used iOS were teenage boys
- Systematic measurement error
  - e.g., task start time was different from Android than for iOS
- Various other biases

Avoiding Confounds

- Remove or exclude
  - simply make the confound not exist
- Spread equally
  - randomize such that the confound is ‘noise’
- Manipulate as a factor
  - systematically control a confound’s influence
- Record as a covariate
  - we can test whether it had an effect

Randomization

- Randomization: the random assignment of treatments to the experimental units or participants

- In a totally randomized experiment, no one, including the investigators themselves, is able to predict the condition to which a participant is going to be assigned

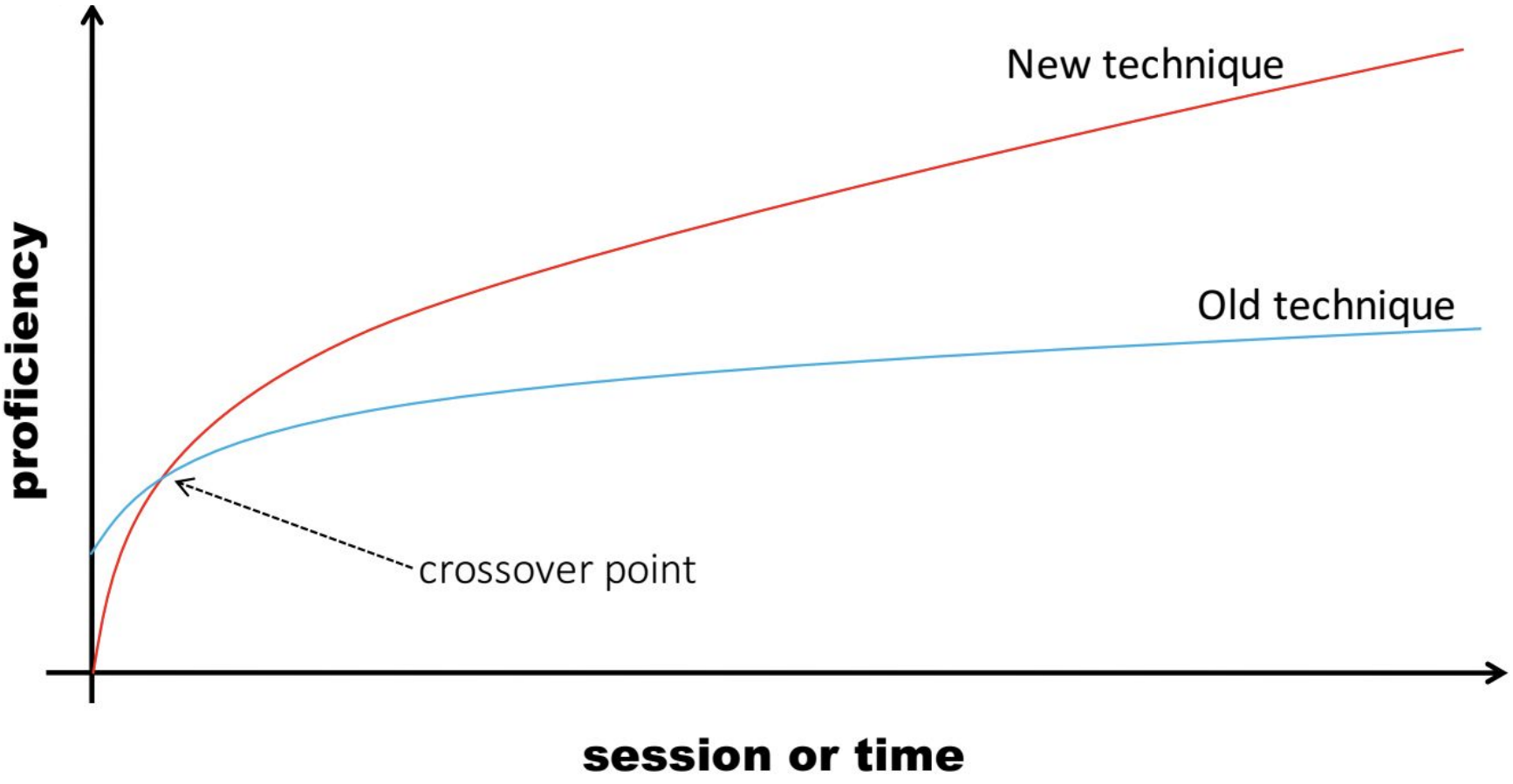
Practice?

- When comparing a ‘new thing’ to an ‘old thing,’ how can we make a fair comparison?
  - What is a fair comparison?
- How do we handle practice?
- Example
  - Typing on a familiar QWERTY keyboard versus a new, unfamiliar experimental keyboard

Handling Practice

- Recruit participants with equal (non-) familiarity with treatments
  - could we find people who have never used a QWERTY keyboard?
  - would testing them answer our research question?
- Give fixed amount of practice
  - can be fixed amount of time, fixed number of trials, etc.
- Practice until a certain proficiency is reached
  - requires real-time feedback, go until performance is equal, report time taken to that point, then study further
- Run a longitudinal study
  - test over multiple sessions and construct learning curves

Learning Curves



Example

- HCI researchers wanted to determine if the size of a device's screen affects how quickly people are able to read news articles. They created an experiment in which they asked 40 participants to read a news article on either a smartwatch, smartphone, tablet, or desktop. They measured how long it took each participant to read the article.
- Factor(s)/Independent variable(s)?
  - Within or between subjects?
  - Level(s)?
- Dependent variable(s)?
- Possible issues/confounds to think about?

Limitations of Experimental Research

- Experimental research requires well-defined, testable hypotheses that consist of a limited number of dependent and independent variables
- Experimental research requires strict control of factors that may influence the dependent variables
- Lab-based experiments may not be a good representation of users' typical interaction behavior
- Experiments done "in-the-wild" are more difficult to control