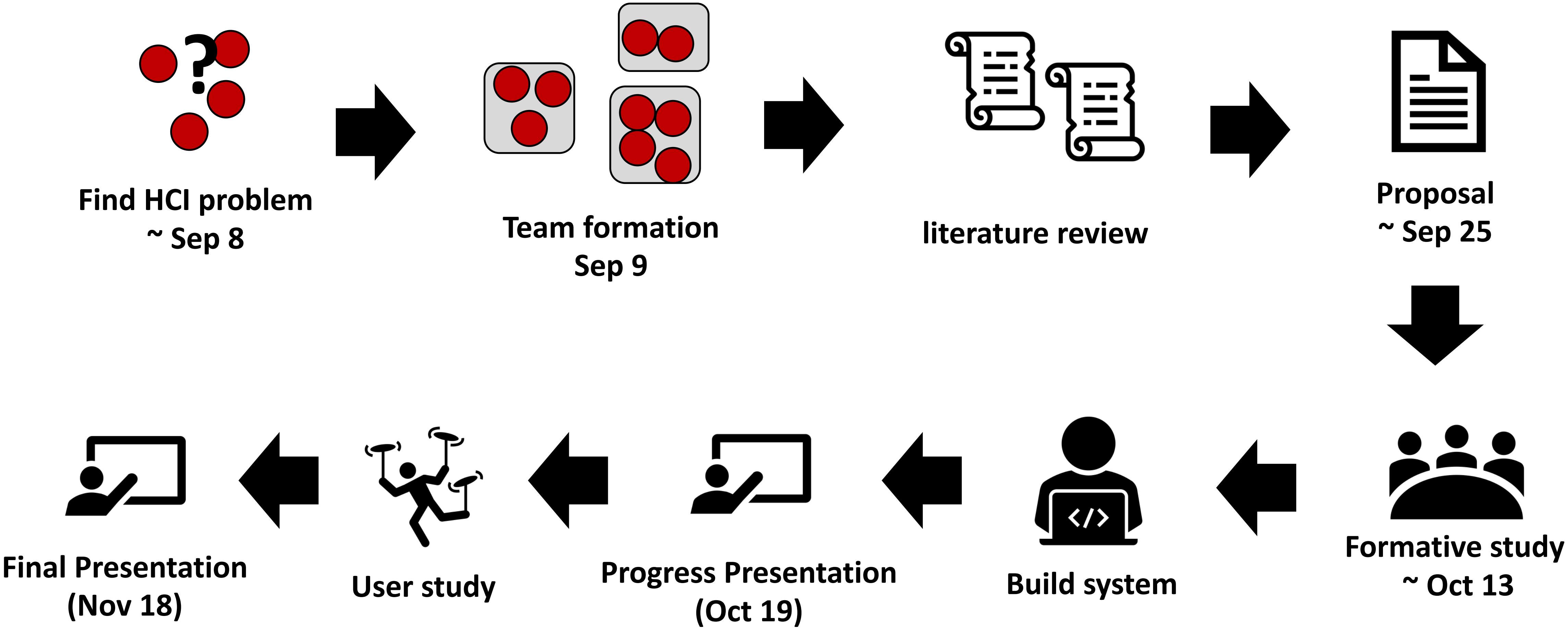
Course Project:
Make people's lives better in this pandemic

What do we miss, what do we struggle with, or what do we like?

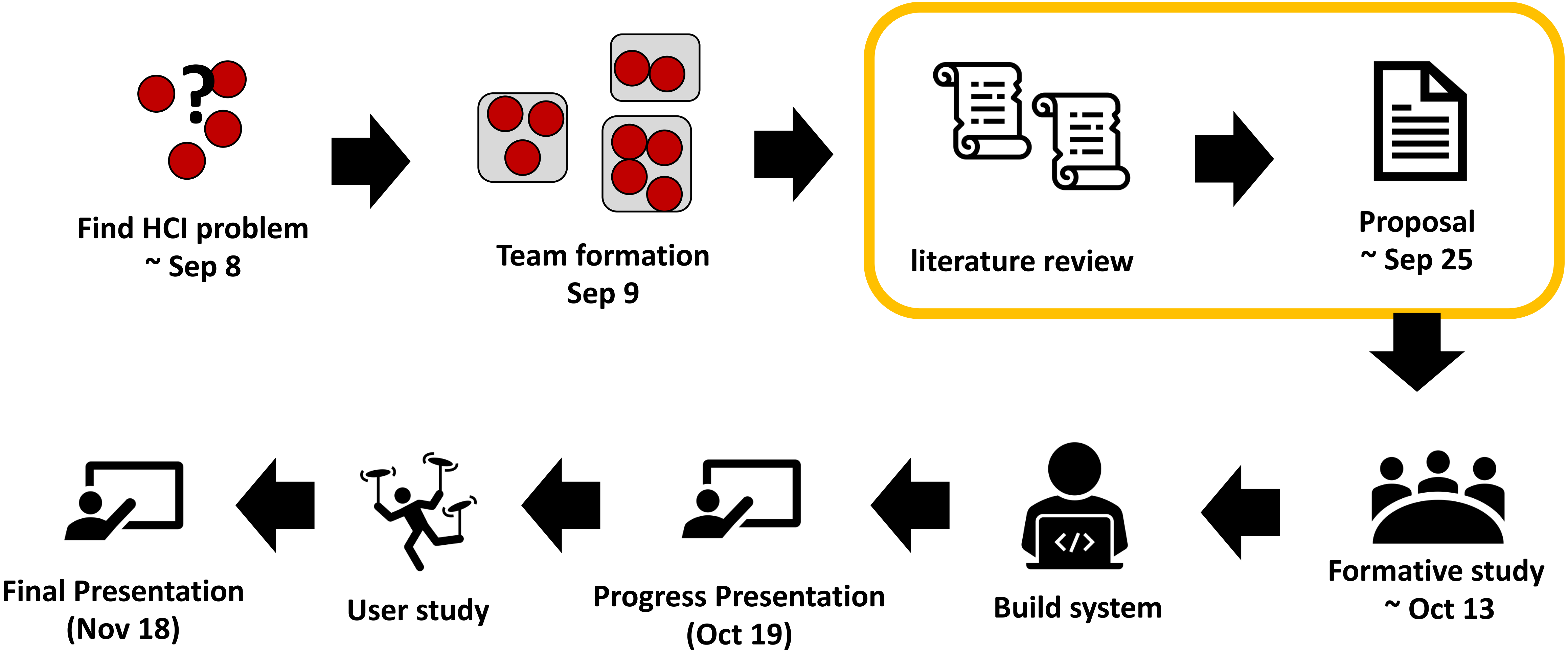
Course Project: Timeline



Course Project: Team Formation Report

- One report per team
- Should include
    - Team Name
    - Team Member Names
    - HCI Problem
- Team Formation Report is Due Next Tuesday (11:59 pm)

Course Project: Timeline



Course Project: Team Formation Report

- One report per team
- Should include
    - Team Name
    - Team Member Names
    - HCI Problem
- Team Formation Report is Due Next Tuesday (11:59 pm)

Problems to Research Questions

- You now have teammates interested in a similar problem 
  e.g., my home coffee sucks, because of the wrong tamping pressure
- You probably have an idea or two to solve the problem
  e.g., showing pressure level on the portafilter will help
- Define a **research question** you want to answer by conducting this research project

Research

- Research is not about products, but about questions
- Research is narrowly focused, research questions are small in scope
- Research is incremental, not monumental
- Research ideas build on previous research ideas
- Good ideas are refined, advanced (into new ideas)
- Bad ideas are discarded, modified

Example Research Questions

- *Can changing portafilter color make users apply adequate amount of pressure?*
- *Can using pressure-sensitive steel be used to create a color-changing portafilter?*
- *Which type of visual feedback would be the most effective in helping users apply adequate amount of pressure?*
- *How can we let people know how to read the pressure level on portafilter?*
- *How much information do we need to visualize to persuade users to change their tamping habits?*
- *When exactly people struggle to do the tamping right?*

Project Proposal (Presentation + Document)

- You will need
  - HCI Problem
    - The problem you want to solve
  - Related work
    - What others have done
  - Research Question
    - What you want to know by conducting this research
  - Method
    - What you suggest or design

Project Proposal (Presentation + Document)

- **Project Proposal (Due Sep 25)**
  - Similar to Introduction of a CHI paper
  - Three sections
    - Motivation & Background
    - **Related Work**
    - Research Question and Method

- 3+ papers that are relevant to the problem
- 2+ papers that are relevant and aligned with the solution you are suggesting
- (optional) suggesting a similar solution but used for other problems
- (optional) study papers that help understanding the problem

Project Proposal (Presentation + Document)

- **Project Presentation (Sep 28)**
    - Each team will have up to 10 minutes
    - Should include all four components
    - We will have a shared document for feedback and questions (and this is where participation counts)
    - If you can’t join live, you can record and send the video

Video Break

153854
CHI Full Paper
Slow Robots for Unobtrusive Posture Correction
Joon-Gi Shin, Eiji Onchi, Maria Jose Reyes, Junbong Song, Uichin Lee, Seung-Hee Lee & Daniel Saakes

Quantitative Evaluation 1

CS6501: Human-Computer Interaction

Seongkook Heo

Fall 2020, Department of Computer Science

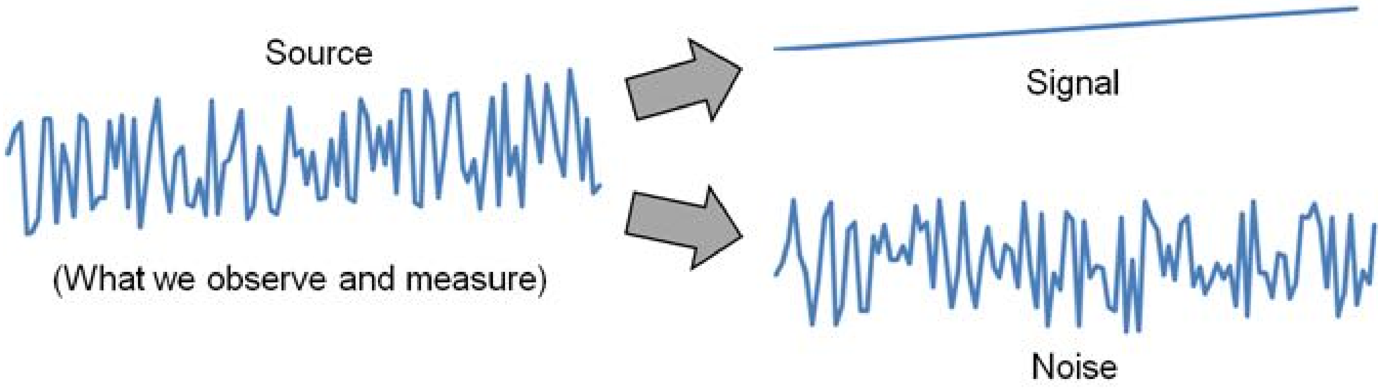
University of Virginia | School of Engineering & Applied Science

Research Methods

- Observational Method
  - Observe humans interacting with computers in a natural setting.
  - Using interviews, field investigations, case studies, focus groups, etc.
  - Tends to be qualitative.
  - High relevance, but sacrifices precision.
- **Experimental Method**
  - Acquire knowledge through controlled lab experiments.
  - Tests if changes to a manipulated variable result in changes to a response variable.
  - High precision, low relevance.

Signal and Noise Metaphor

- Signal and noise metaphor for experiment design:



- Signal → a variable of interest,
  e.g., task completion time, error rate, learning rate, fatigue, etc.
- Noise → everything else (random influences)

Experimental, Quantitative Evaluation

- What task to evaluate?
  - Depends on application
  - Attempt to find canonical task(s)

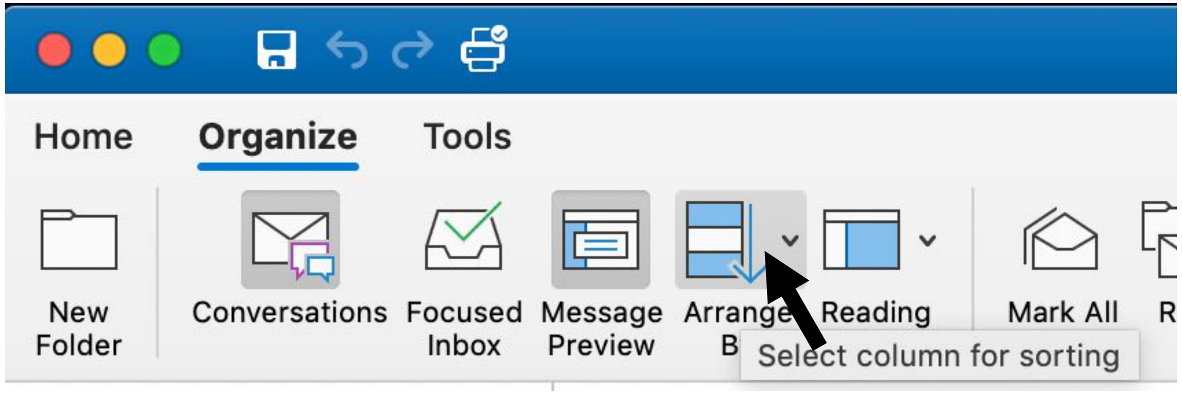
- Common measures
  - Task completion time
  - Error rate
  - Learning rate (novice -> expert transition)
  - Fatigue, comfort?
  - etc.

Example: Pointing Device Evaluation

- Which device is better?

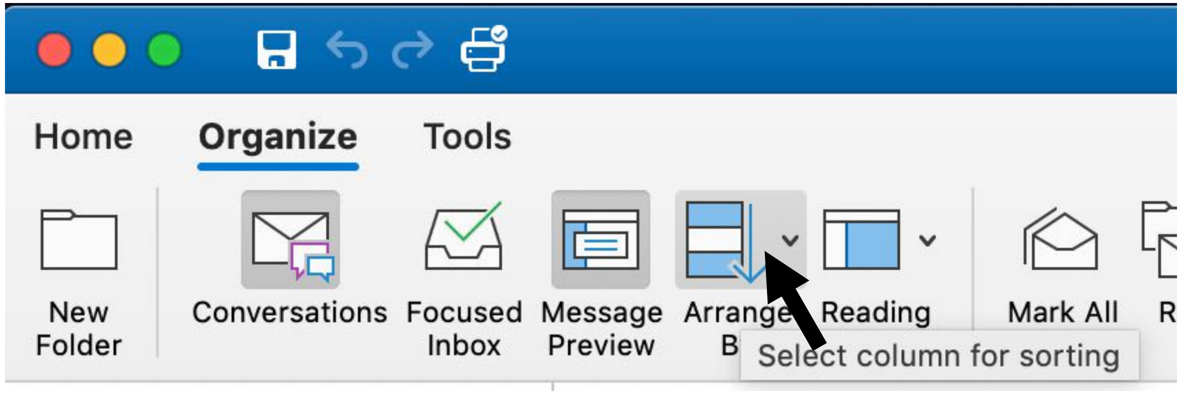
Example: Pointing Device Evaluation

- Real task: interacting with GUI's
    - Pointing is fundamental

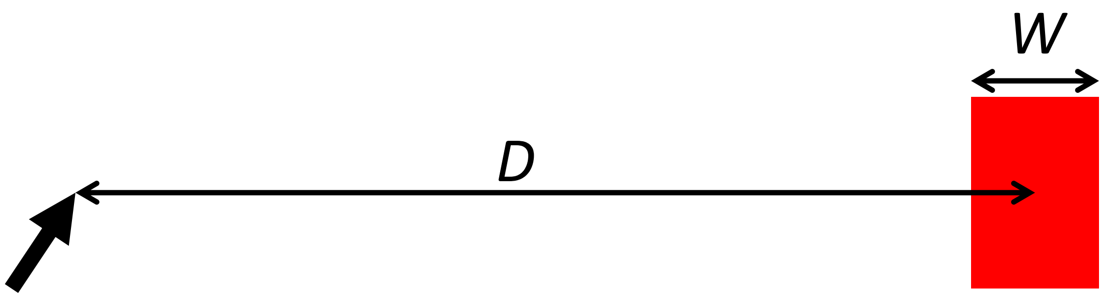


Example: Pointing Device Evaluation

- Real task: interacting with GUI's
  - Pointing is fundamental



- Experimental task: target acquisition
  - Abstract, elementary, essential



What Variables to Manipulate/Measure?

- Independent variables
  - Factors that are manipulated in the experiment
- Dependent variables
  - Factors which are measured
- Chosen based on task
  - Seek external validity

Independent Variables

- Factors that are manipulated in the experiment
    - e.g., W, D in pointing task
- Operational definitions
    - Specific definition of a conceptual variable
    - E.g. Level of experience -> Number of hours of system use
- Pick the most important for the task
- Choose appropriate ranges
    - Realistic range for task
    - A range that shows an effect and trend (but realistic)
    - E.g. Pointing experiment:
        - W 's range from character size (10) to icons (40) pixels
        - D 's from short (50) to large (screen size ~800 pixels)

How Many IVs?

- An experiment must have at least one independent variable
- Possible to have 2, 3, or more IVs
- But the number of “effects” increases rapidly with the size of the experiment:

Independent Variables	Effects					Total
	Main	2-way	3-way	4-way	5-way	
1	1	-	-	-	-	1
2	2	1	-	-	-	3
3	3	3	1	-	-	7
4	4	6	3	1	-	14
5	5	10	6	3	1	25

- Advice: Keep it simple (1 or 2 IVs, 3 at the most)

Dependent Variables

- Factors which are measured
  - E.g., Trial completion time, error rate, in a pointing task
  - May depend on independent variables
- May require operational definitions
  - E.g. Time between the first and last click

Dependent Variables

- Single dependent variable
  - Usually hard to find one variable that is indicative of task
- Multiple dependent variables
  - Total time, reaction time, physical movement, eye movement, accuracy, etc.
  - There is often a speed-accuracy tradeoff
- Composite dependent variables
  - Indication of overall performance

Dependent Variables

- Unique dependent variables are also possible:
  any observable, measurable behavior is a legitimate DV
- E.g. Negative facial expressions[1]
  - Application: user difficulty with mobile games
  - Events logged included frowns, head shaking

[1] Duh, H. B.-L., Chen, V. H. H., & Tan, C. B. (2008). Playing different games on different phones: An empirical study on mobile gaming. *Proceedings of MobileHCI 2008*, 391-394, New York: ACM.

Data Collection

- Obviously, the data for dependent variables must be collected in some manner
- Ideally, engage the experiment software to log timestamps, key presses, button clicks, etc.
- Planning and pilot testing important
- Ensure conditions are identified, either in the filenames or in the data columns

Data Collection

```
min_keystrokes,keystrokes,presented_characters,transcribed_characters, ...
55, 59, 23, 23, 29.45, 0, 9.37, 0.0, 2.5652173913043477, 93.22033898305085
61, 65, 26, 26, 30.28, 0, 10.3, 0.0, 2.5, 93.84615384615384
85, 85, 33, 33, 48.59, 0, 8.15, 0.0, 2.5757575757575757, 100.0
67, 71, 28, 28, 33.92, 0, 9.91, 0.0, 2.5357142857142856, 94.36619718309859
61, 70, 24, 24, 39.44, 0, 7.3, 0.0, 2.9166666666666665, 87.14285714285714
```

Other Variables

- Control variable
  - Variables with constant value
    - e.g. Screen background color in pointing task
- Random variable
  - Variable which takes on a random value
    - e.g. Location of target in a pointing task
  - Usually randomized with constraints
    - e.g. Each location appears same number of times

Participants

- Researchers want experimental results to apply to people not actually tested – a population
- Population examples:
  - Computer-literate adults, teenagers, children, people with certain disabilities, left-handed people, engineers, musicians, etc.
- For results to apply generally to a population, the participants used in the experiment must be
  - Members of the desired population
  - Selected at random from the population

How Many Participants?

- Too few → experimental effects fail to achieve statistical significance
- Too many → statistical significance for effects of no practical value
- The correct number:
  - Use the same number of participants as used in similar research[1]

[1] Martin, D. W. (2004). *Doing psychology experiments* (6th ed.). Pacific Grove, CA. Belmont, CA: Wadsworth.

Within vs. Between Subjects Design

Within-subjects design:

- All subjects do all conditions
- Fewer participants needed
- Prone to learning transfer effects

Condition 1	Condition 2
Subject 1	Subject 1
Subject 2	Subject 2
.	.
Subject 10	Subject 10

Between-subjects design:

- Subjects only do one condition
- More participants needed
- No learning transfer effects
- Can train to high skill

Condition 1	Condition 2
Subject 1	Subject 11
Subject 2	Subject 12
.	.
Subject 10	Subject 20

Order Effects, Counterbalancing

- Only relevant for within-subjects factors
- The issue: *order effects* (aka *learning effects*, *practice effects*, *fatigue effects*, *sequence effects*)
- Order effects offset by *counterbalancing*:
    - Participants divided into groups
    - Test conditions are administered in a different order to each group
    - Order of administering test conditions uses a Latin square
    - Distinguishing property of a Latin square → each condition occurs precisely once in each row and column (next slide)

Counterbalancing

- Fully counterbalanced:
  - Combinatorial explosion when n>4
  - Needs lots of subjects

```
A B
B A
```

```
A B C
A C B
B A C
B C A
C A B
C B A
```

```
A B C D
A B D C
A C B D
A C D B
…
…
…
…
```

Counterbalancing

- Partial counterbalancing. e.g., Latin square:
  - Ensures each level appears in every position in order equally often:

```
A B C D
B C D A
C D A B
D A B C
```

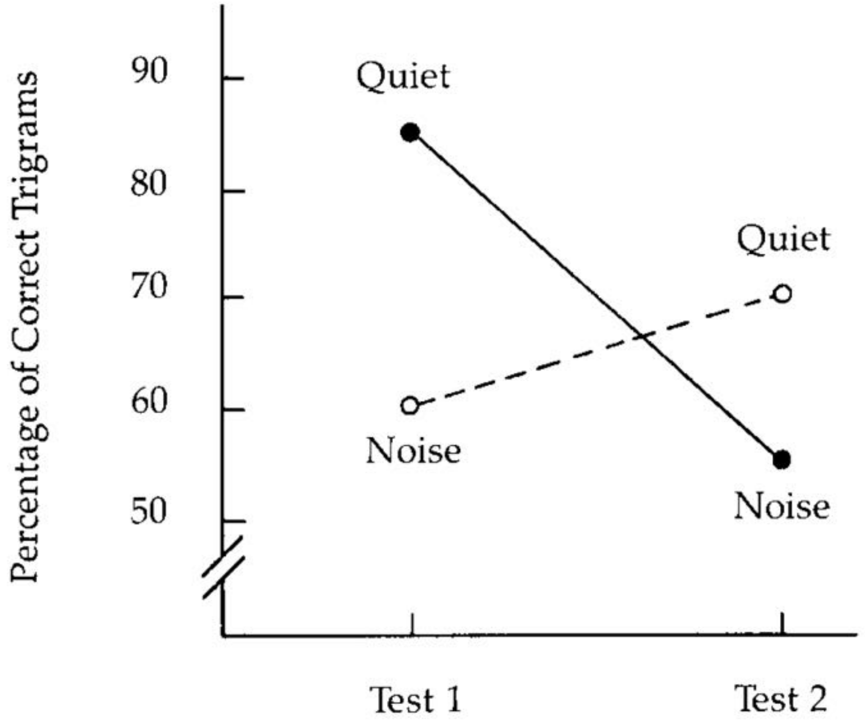
Counterbalancing

- Balanced Latin Square:
  - Each condition precedes and follows each of the other equally often:

A	**B**	**C**	**D**
B	**D**	**A**	**C**
D	**C**	**B**	**A**
C	**A**	**D**	**B**

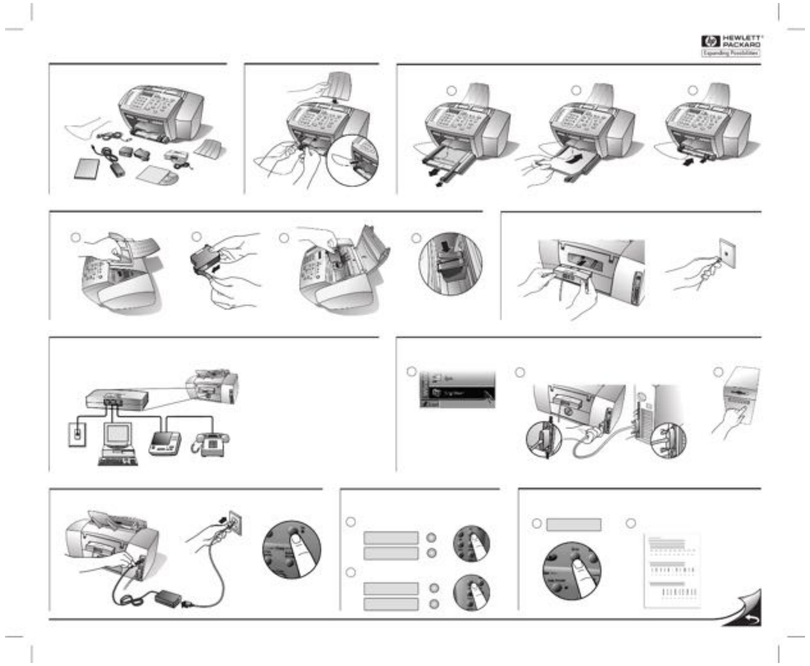
Experimental Design

- Problems with counter balancing
  - Assumes symmetric transfer effects



Experimental Design

- Problems with counter balancing
  - Assumes symmetric transfer effects
  - Initial conditions may invalidate subsequent tasks
    - E.g. Learning to do a task as a novice



Experimental Design

- Problems with counter balancing
  - Assumes symmetric transfer effects
  - Initial conditions may invalidate subsequent tasks
  - Range Effects: People may perform best in middle of range of values

Experimental Design

- Problems with counter balancing
  - Assumes symmetric transfer effects
  - Initial conditions may invalidate subsequent tasks
  - Range Effects: People may perform best in middle of range of values
- In these cases, no counterbalancing will help
  - Must use between-subject design

Summary

Within-Subjects Design

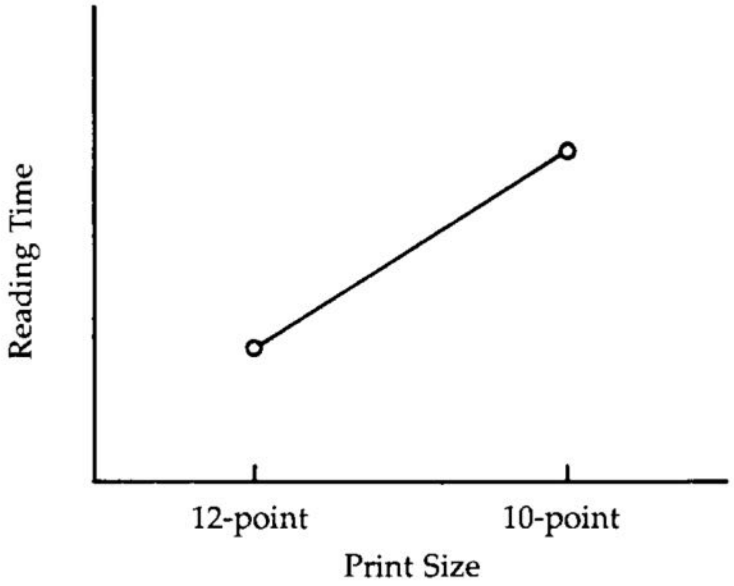
- Pros:
  - Fewer subjects
  - Smaller variability between groups
- Cons:
  - Transfer effects
  - Assumes symmetrical transfer

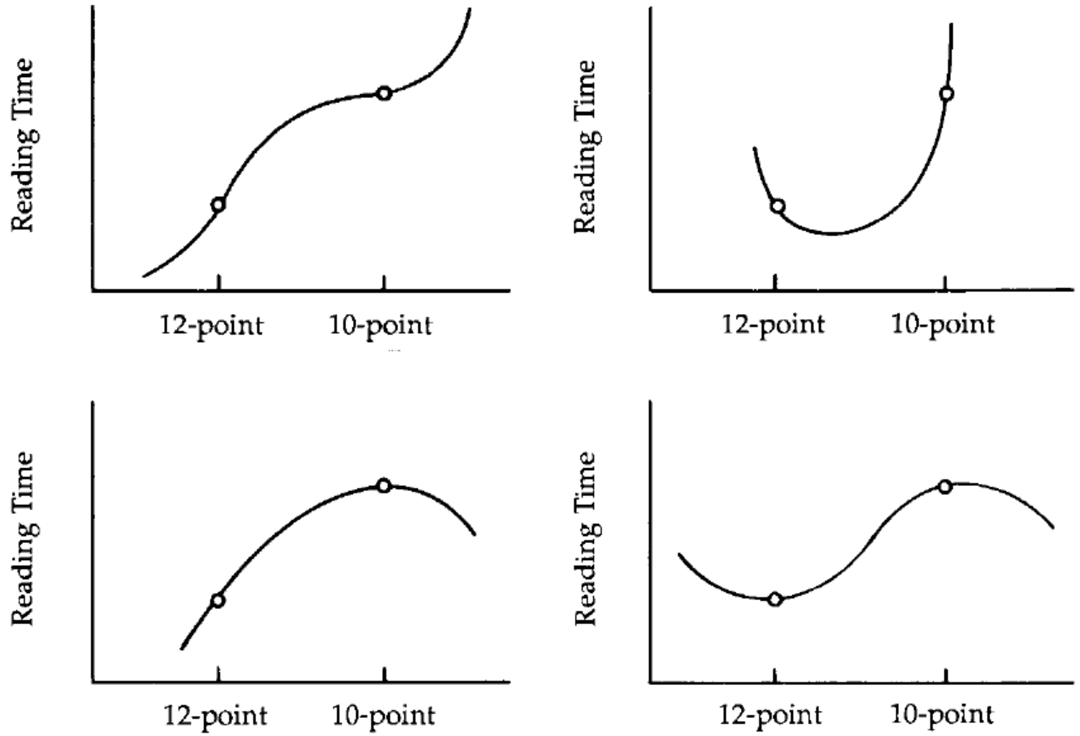
Between-Subjects Design

- Pros:
  - No transfer effects
  - No counterbalancing
- Cons:
  - Group differences
  - More subjects

Multiple Variables

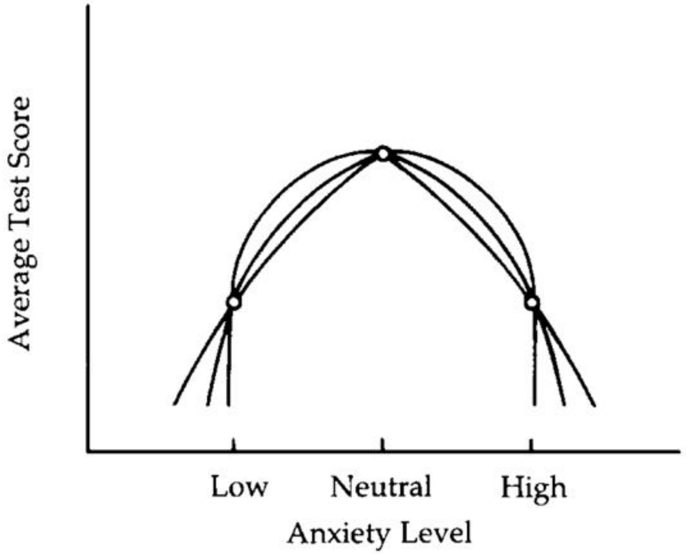
- Two level experiment
    - One variable with two possible values





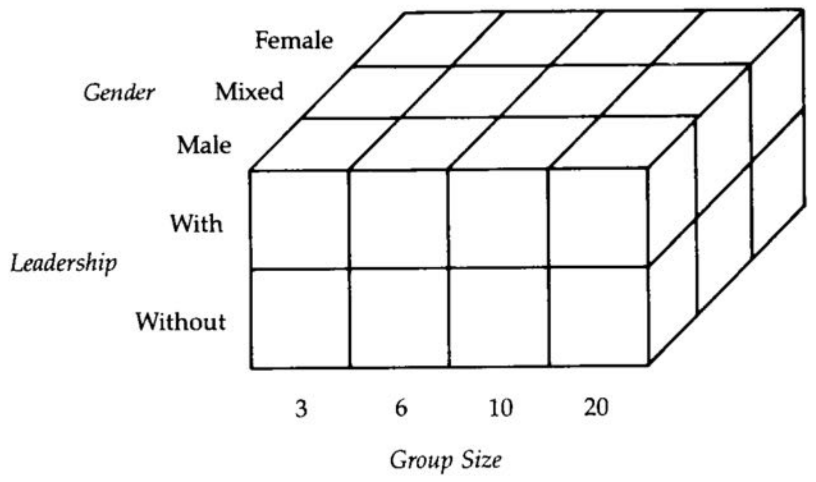
Multiple Variables

- Two level experiment
  - One variable with two possible values
- Multiple level experiment
  - One variable with three or more possible values



Multiple Variables

- Two level experiment
    - One variable with two possible values
- Multiple level experiment
    - One variable with three or more possible values
- Factorial design
    - Multiple variables, multiple levels



Multiple Variables

- Two level experiment
  - One variable with two possible values
- Multiple level experiment
  - One variable with three or more possible values
- Factorial design
  - Multiple variables, multiple levels
- Converging-Series design
  - Progressively close in on a solution

Factorial Design

- Multiple variables, multiple levels
- 2 x 4 = 8 cells/conditions
- 3 x 4 x 5 = 60 cells/conditions
- Advantages
  - Can analyze interactions
- Disadvantage
  - Experiment size can explode

Converging-Series Design

- Conduct a series of pilot experiments
- Determine set of independent variables
- Determine range of independent variables
- Must be careful of interactions

Other Definitions

- Block
  - A significant section of the experiment
  - Repeated to analyze learning
- Trial
  - An individual measurement for a single condition/cell
- Repetition
  - A trial which is repeated within a block
  - Increase number of data points, reliability
- Determining number of blocks/repetitions
  - Reasonable experiment duration
  - Enough data points for significant effects

Thank you!