University of Colorado
Boulder

Human-Robot Interaction

Experimental Design Fundamentals

Professor **Dan Szafir**

Computer Science & ATLAS Institute
University of Colorado Boulder

Previously…

Methodology

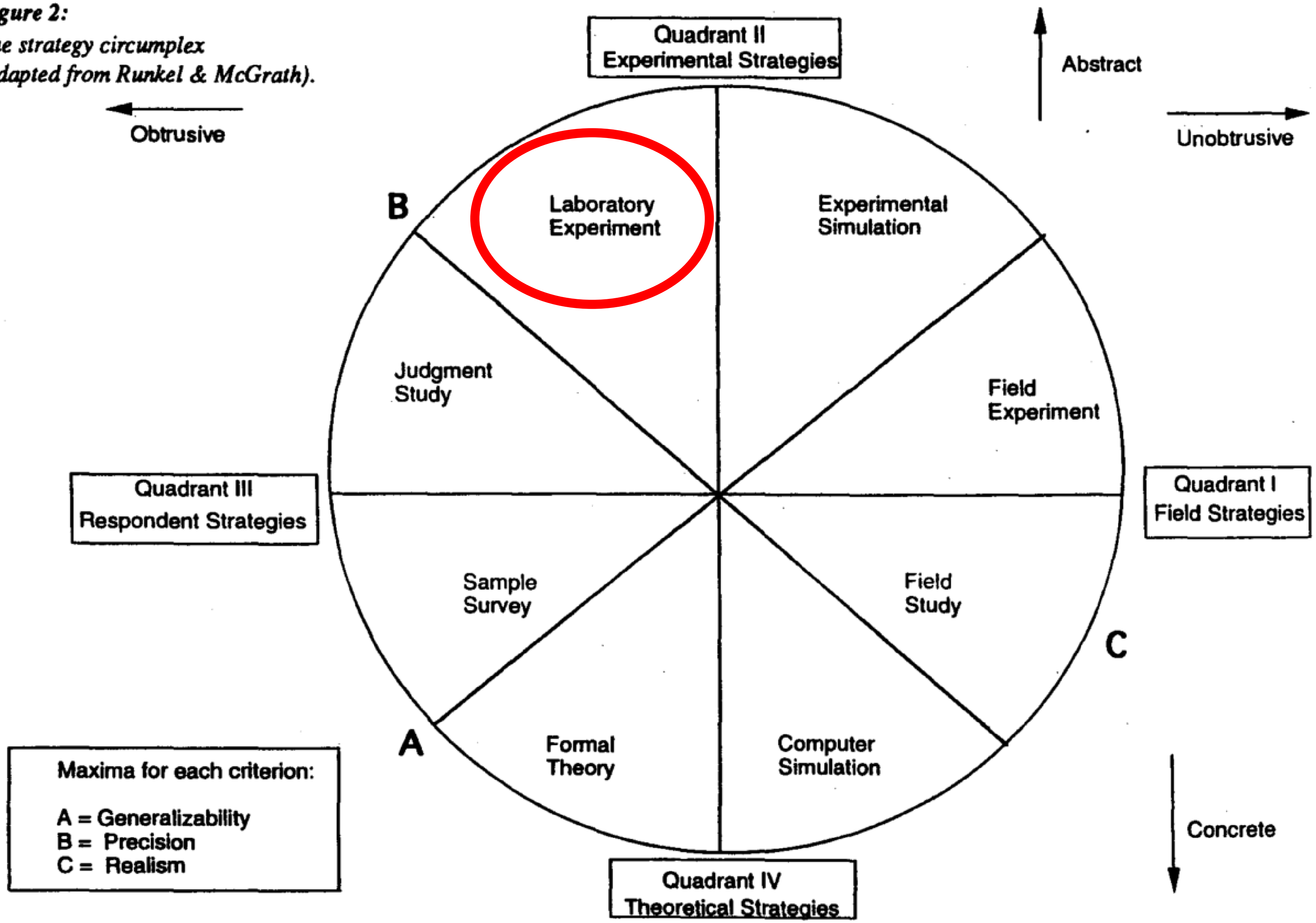
Research strategies

- Field
- Experimental
- Respondent
- Theoretical

Nascent, Intermediate, and Mature research

Examples of poor fit

Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



Fundamentals of Experimental Design

What is a **hypothesis**?

Hypothesis

hypothesis |hīˈpäθəsis|

noun (pl. **-ses** |-ˌsēz|)

a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation : professional astronomers attacked him for popularizing an unconfirmed hypothesis.

- *Philosophy* a proposition made as a basis for reasoning, without any assumption of its truth.

ORIGIN late 16th cent.: via late Latin from Greek **hupothesis** **'foundation**,' from **hupo 'under' + thesis 'placing.**'

Hypothesis

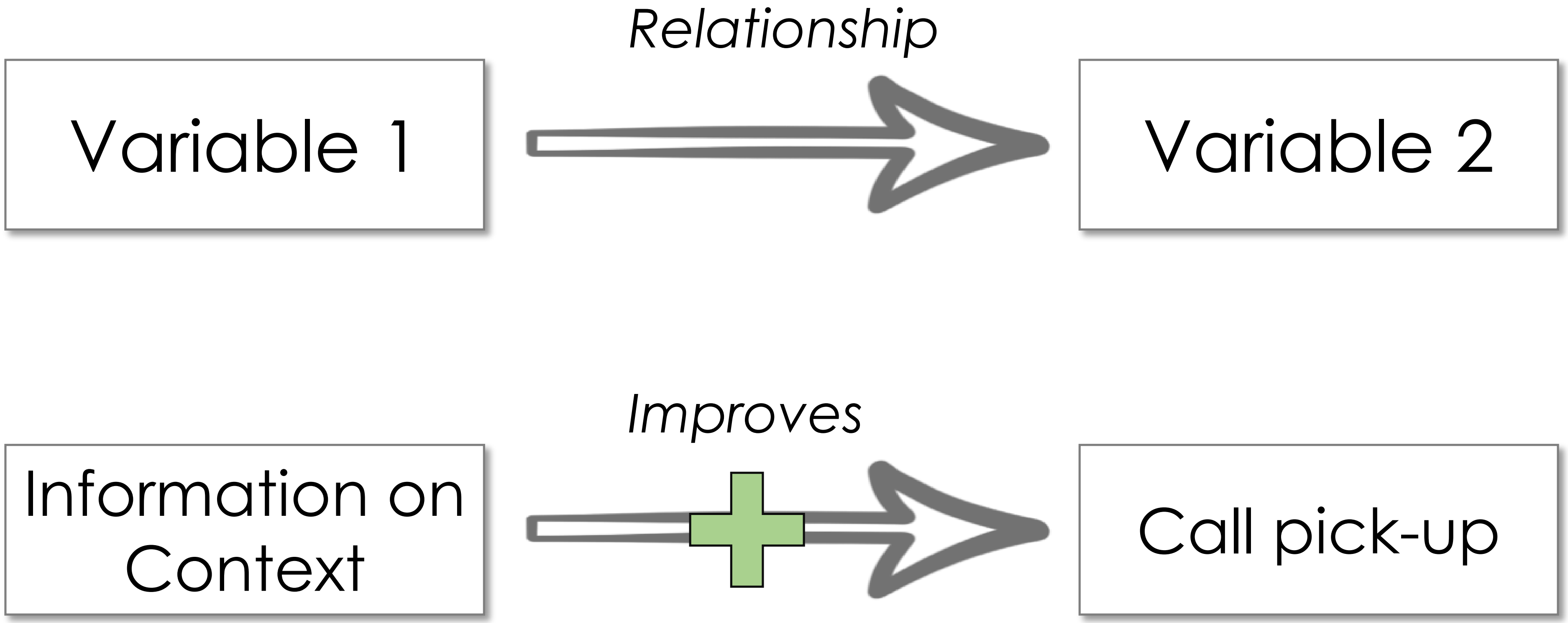
A statement of the predicted or expected relationship between at least 2 variables

- A provisional answer to a research question
- Has to define the variables involved
- Has to define a relationship

Example:

- **Question:** How does having information on the context of a caller affect whether the receiver picks up the call?
- **Hypothesis:** Receivers will be more likely to pick up a call when they have information of their callers' context than they will be when they do not

Hypothesis


Relationship
Variable 1
Variable 2
Improves
Information on Context
Call pick-up

Good Hypothesis Formation

Testable: The means for manipulating the variables and/or measuring the outcome variable must potentially exist

Falsifiable: Must be able to disprove the hypothesis with data

Parsimonious: Should be stated in simplest adequate form

Precise: Should be specific (operationalized)

Useful:

- Relate to existing theories and/or "point" toward new theories
- Lead to studies beyond the present one (may be hard to determine in advance)

Variables

Independent variable
What is manipulated

Dependent variable
What is measured

Relationship

Variable 1 → Variable 2

Independent variable

Dependent Variable

Variables

Control variables

- What is held constant

Random variables

- What is allowed to vary randomly

Confounding variable

- What influences the independent + dependent variable(s)

What is the difference between **causality** and **correlation**?

Relationships

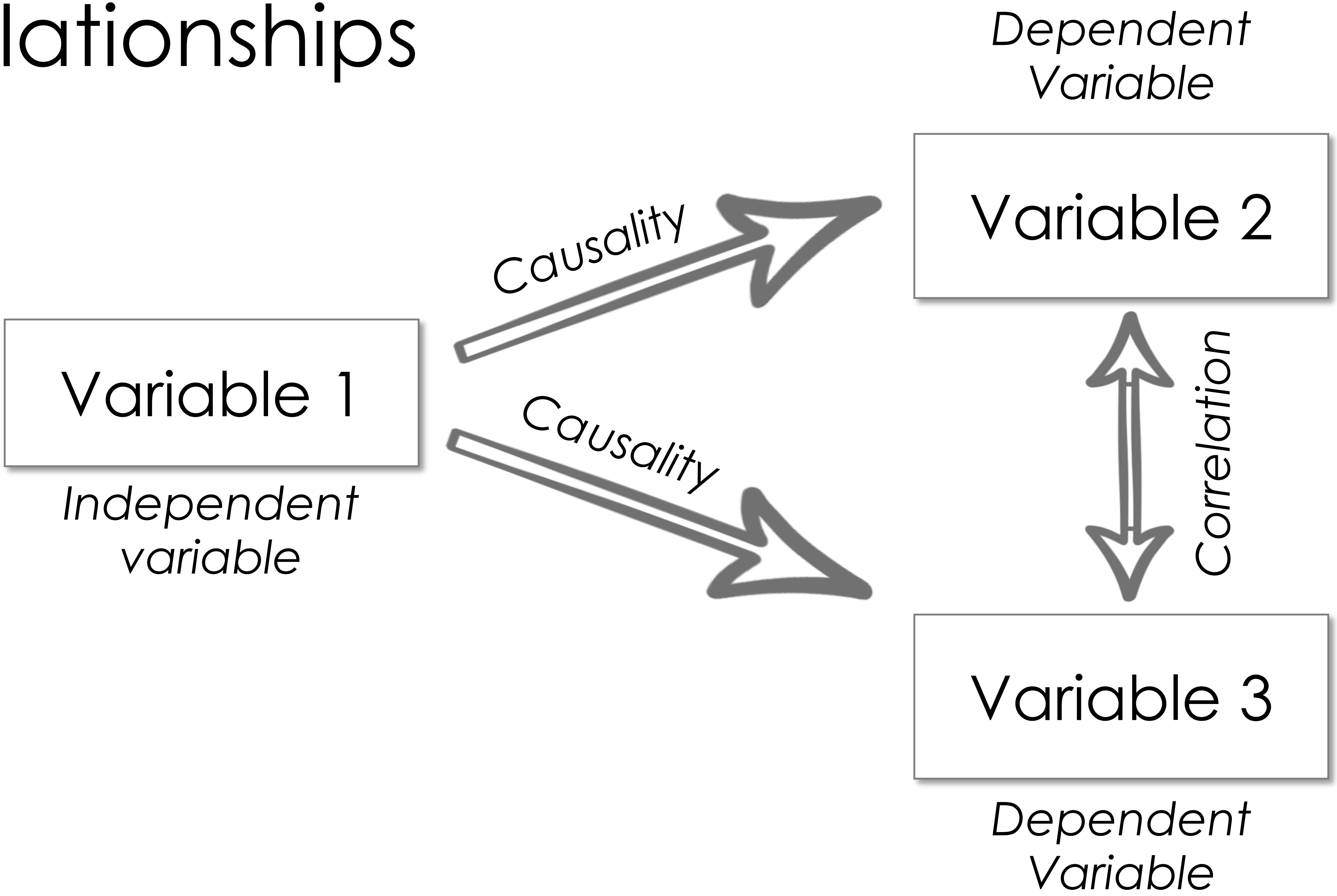
Causal

One variable depends on and is affected by the other

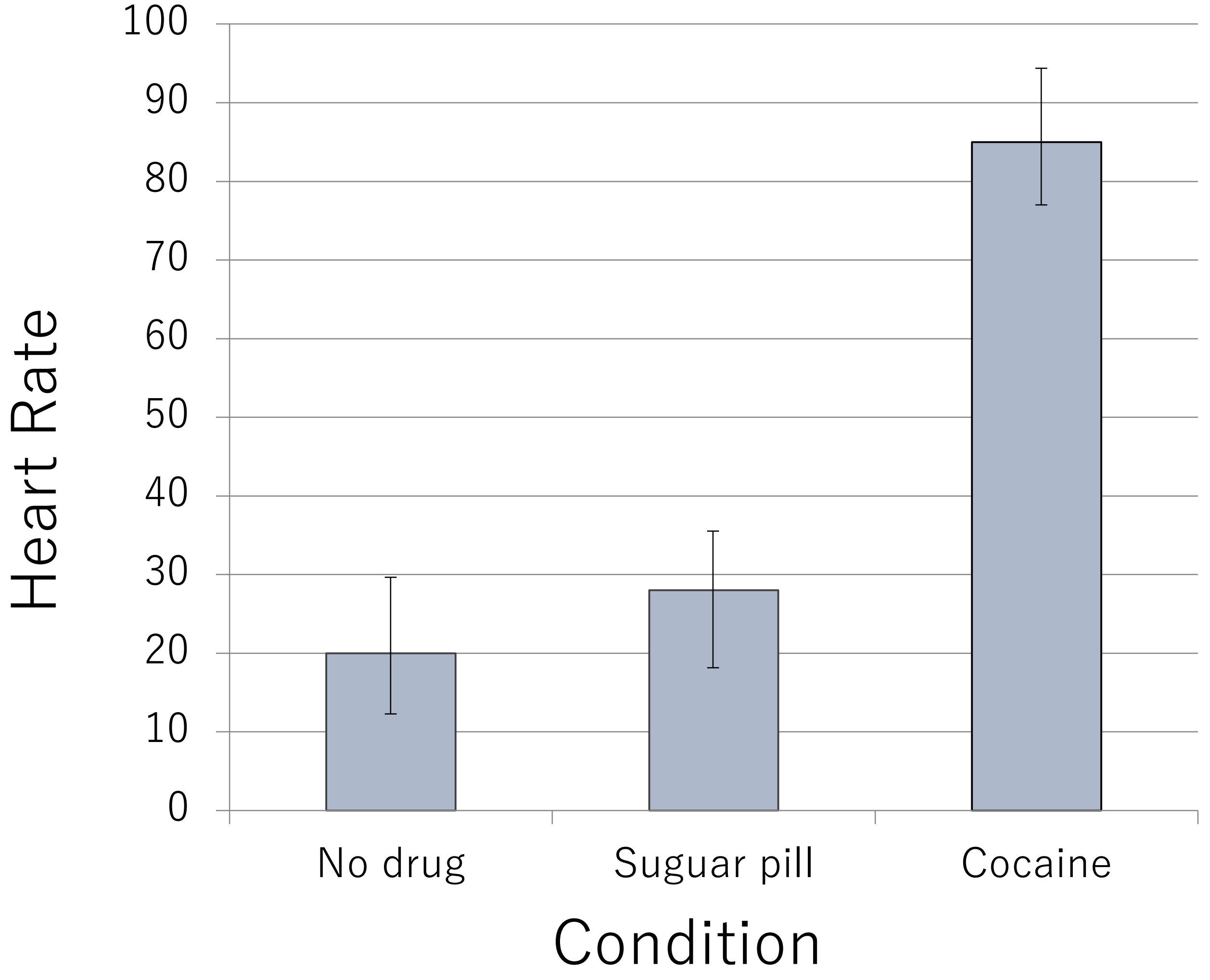
Correlational

Two variables are affected by a third variable in the same direction

Relationships

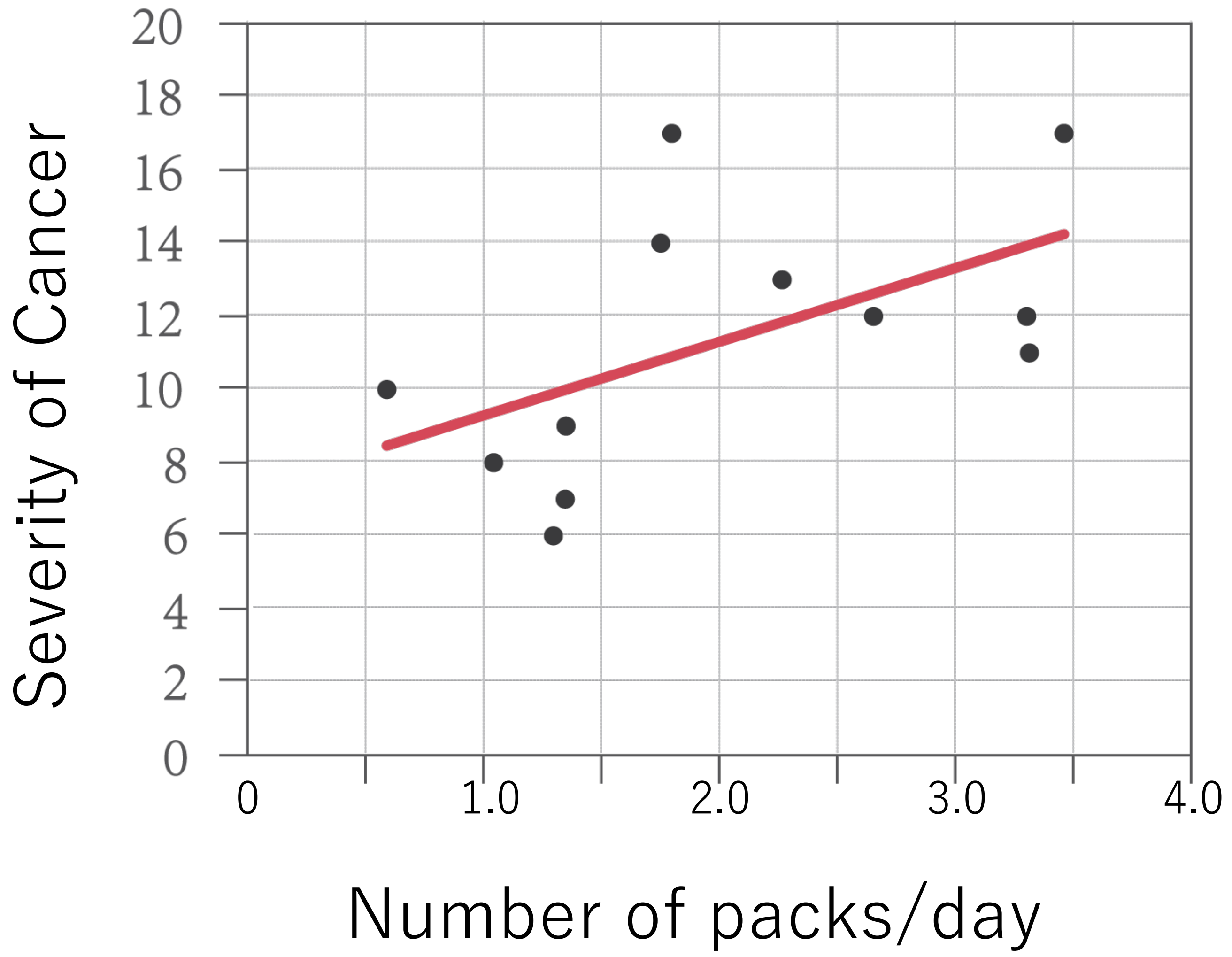


Causal



100
90
80
70
60
50
40
30
20
10
0
Heart Rate
No drug
Suguar pill
Cocaine
Condition

Correlational



How do we calculate correlation?

Different ways (different interpretations of meaning of *correlation*)

Commonly: correlation = do two variables have a linear (or close to linear) relationship?

To determine, calculate Pearson's correlation coefficient *r*

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- r = +1 indicates a perfect positive linear fit
- r = 0 indicates no linear dependency
- r = -1 indicates a perfect negative linear fit
- Often we care more about the coefficient of determination (r^2) – we will discuss more later

Determining Causality

- Fundamental Problem of Causal Inference
  - It is impossible to directly observe causal effects
- Is the relationship conceptually valid?
  - Can you rule out reverse causality?
    - E.g., Windmill rotation causes wind
  - Is the relationship temporally correct as effects cannot precede causes?
- Can you remove or control for potential confounding variables?
- Can you rule out correlation by coincidence?
  - Use large sample sizes
  - Perform cross validation
  - Are your findings repeatable?
- How important is causality?
  - Correlational models can often provide useful predictions

Questions?

Experimental Design

Research Designs

Correlational research

Quasi-experimental research

True Experimental research

Correlational Design

For studies examining the relationships between variables such as personality traits, work habits, gender, etc., the hypothesis is a specific statement about relationships

When we observe an increase in X, then we will also observe an increase (or decrease) in Y

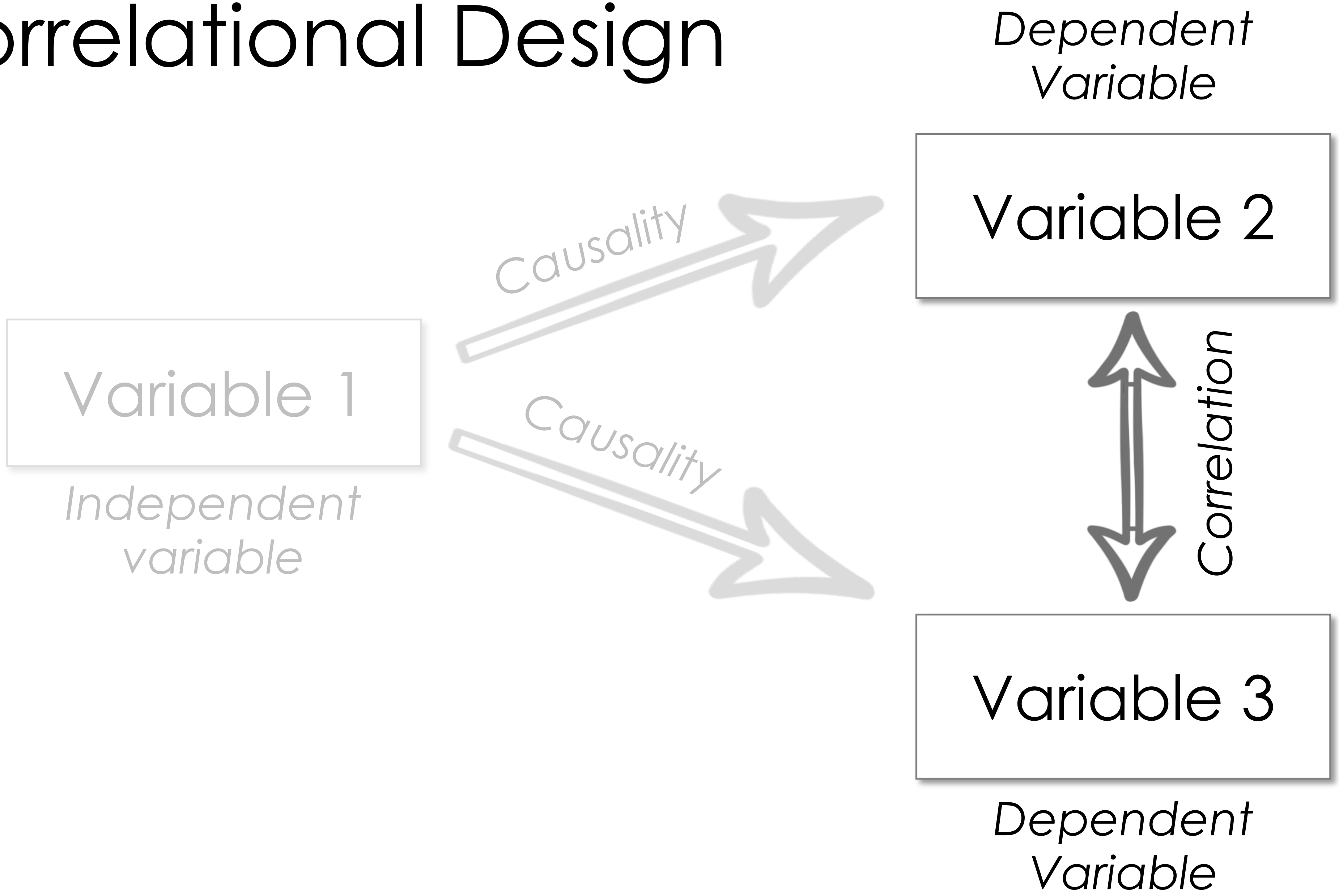
Example questions:

Is there a relationship between smoking and lung cancer?

Is there a relationship between anxiety and test-taking performance?

Correlation does NOT imply causation

Correlational Design



Quasi-experimental Design

Used when randomization is impossible and/or impractical

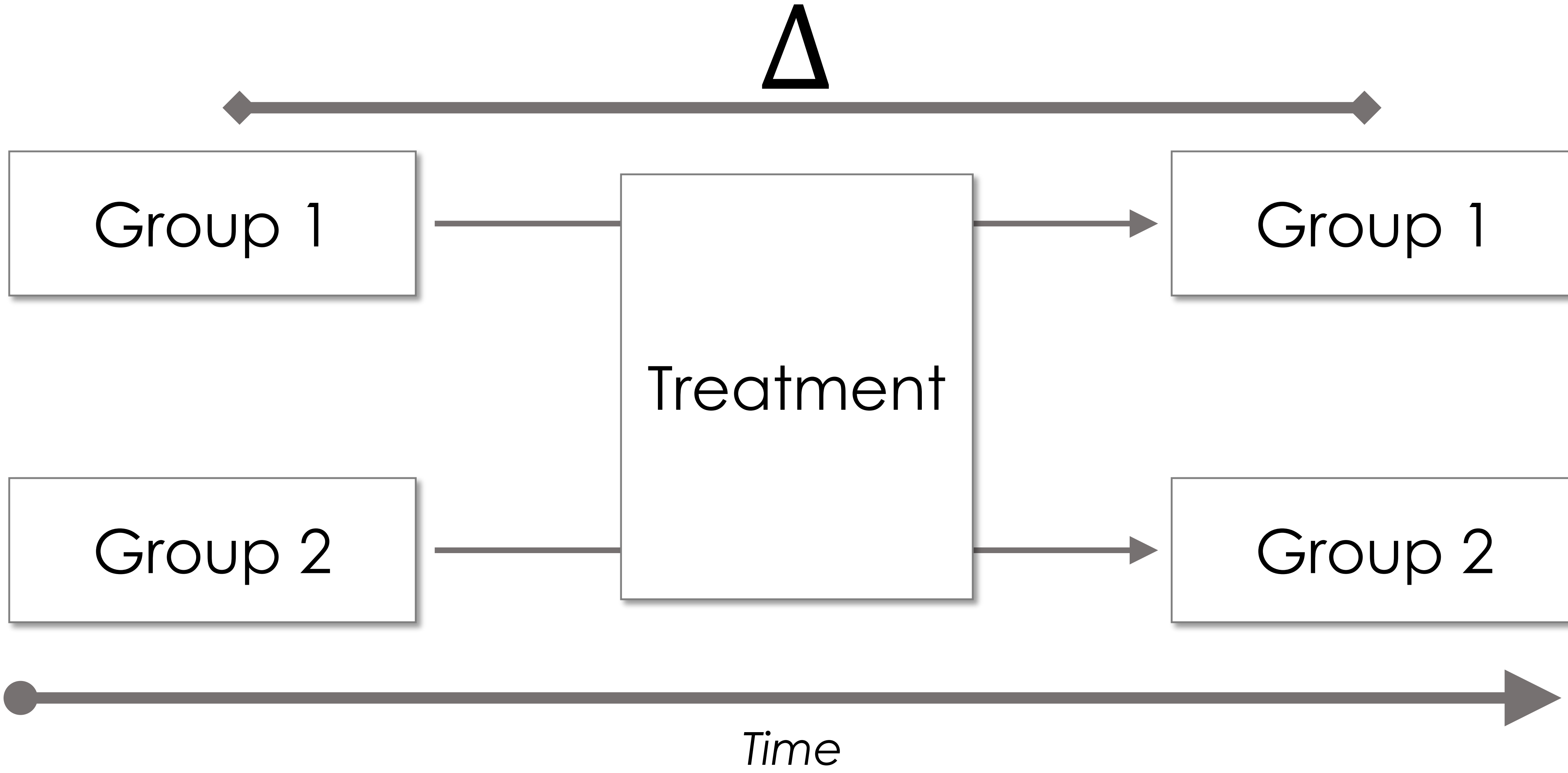
Separate participants based on some characteristic

- No random assignment
- E.g., Gender, occupation, verbal ability (VSAT)

Example questions

- Do people with high verbal ability learn new languages faster?

Quasi-experimental Design



True Experimental Design

Studies in which variables are manipulated and outcomes measured, the hypothesis is a cause and effect statement

Y will occur, when X is manipulated

Examples:

Students will remember more items from a word list if they learn the list in quiet, rather than in the presence of intense music

Reading speed (words/minute) will change when font size is manipulated, such that reading speed will increase as font size is increased from 4pt to 20pt, but reading speed will decrease as font size is increased above 20pt

Number of Variables

Single Variable

- Only one independent variable
- Cannot look at interactions

Multiple Variables

- Two or more independent variables
- If use **factorial design**, can look at interactions
- Will require more participants (*between* design) or time (*within* design)

Within vs Between Designs

Comparisons between conditions **within participants**

- Demands time
- Statistical power with smaller number of participants
- Potential for *transfer effects*

Comparisons between conditions **across participants**

- Demands larger sample
- Avoids transfer effects
- Easier to avoid bias

Can also have **mixed design**

- Complicates statistical analysis

Factorial Design

Suppose we are interested in the effect of both how gaming and computer use affect perception of robots

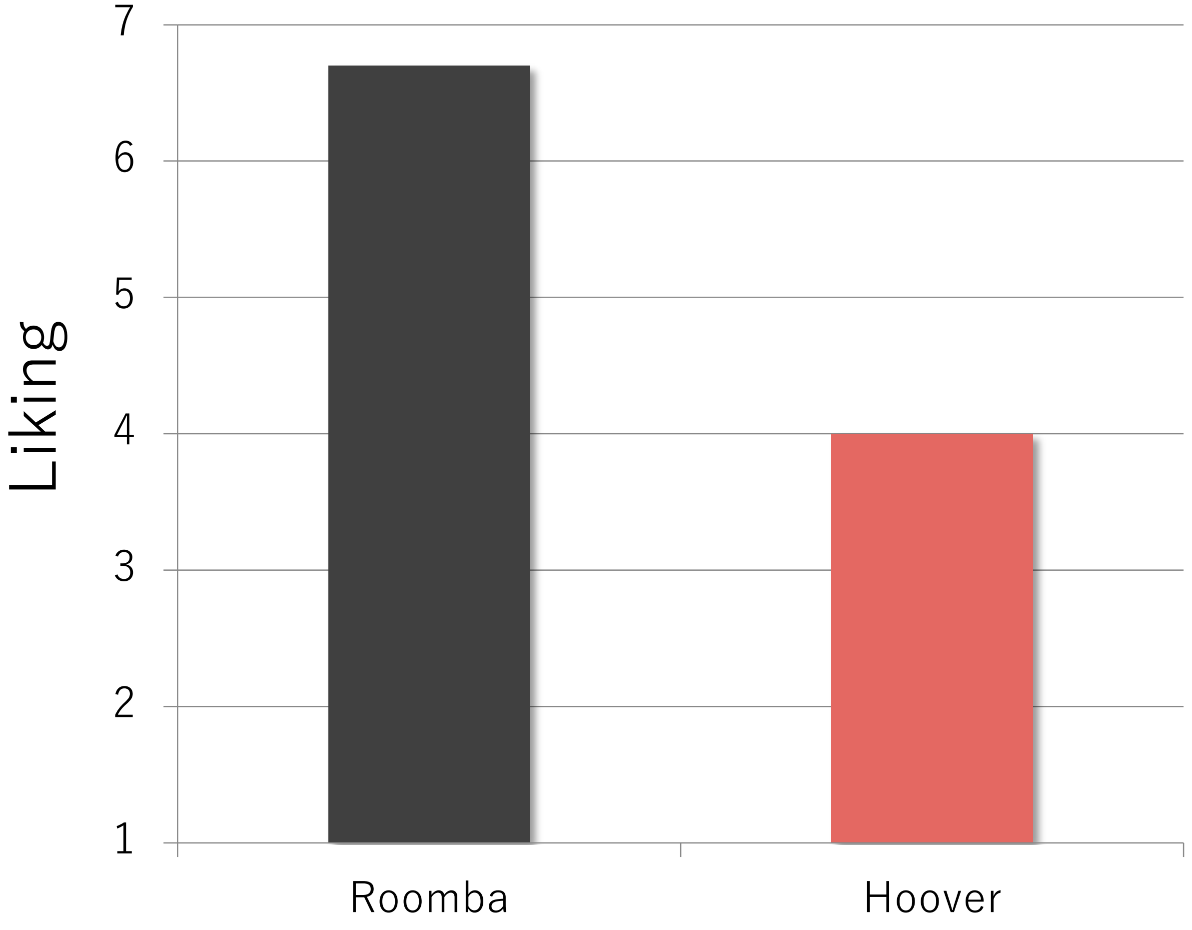
Ideally: look at all 4 populations in one experiment

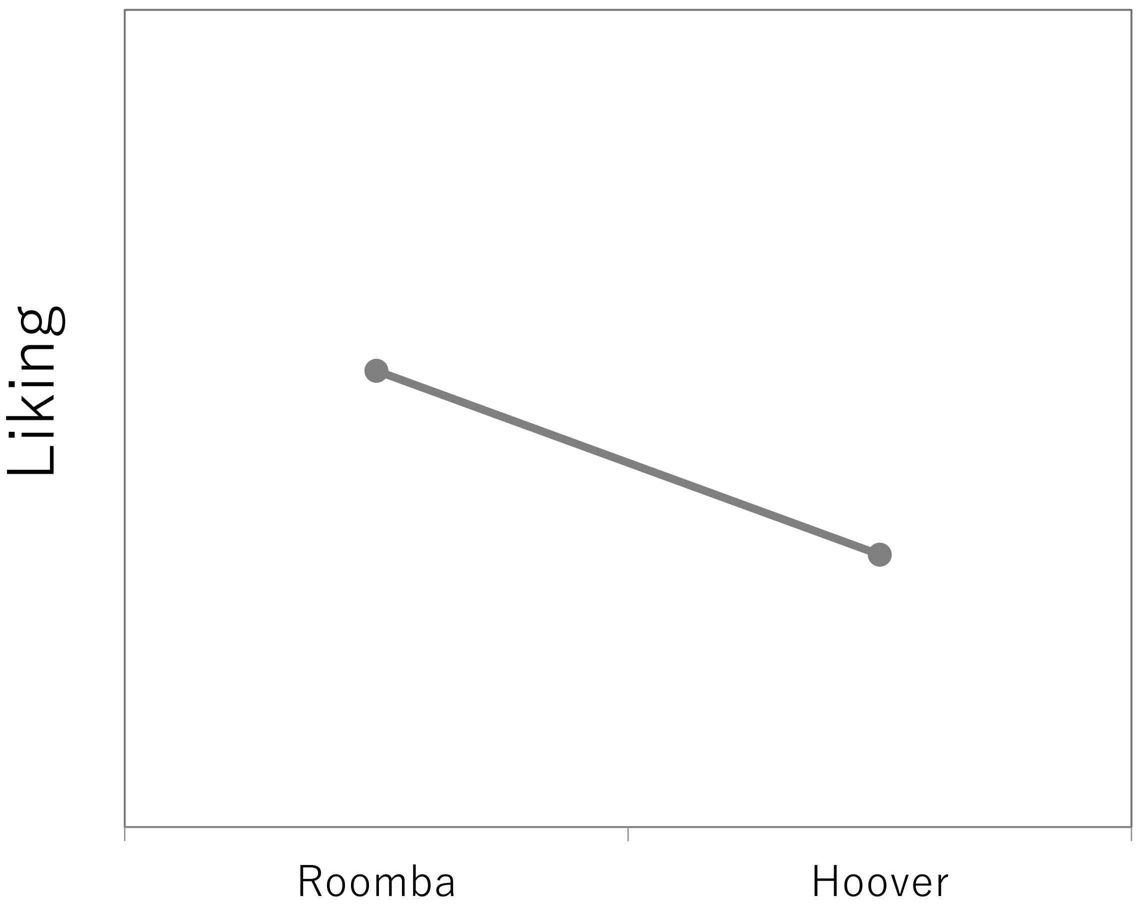
		Factor 1: Computer use	
		Level 1: High computer use	*Level 2:* Low computer use
Factor 2: Gaming	*Level 1:* High gaming	Population 1	Population 2
	Level 2: Low gaming	Population 3	Population 4

Why?

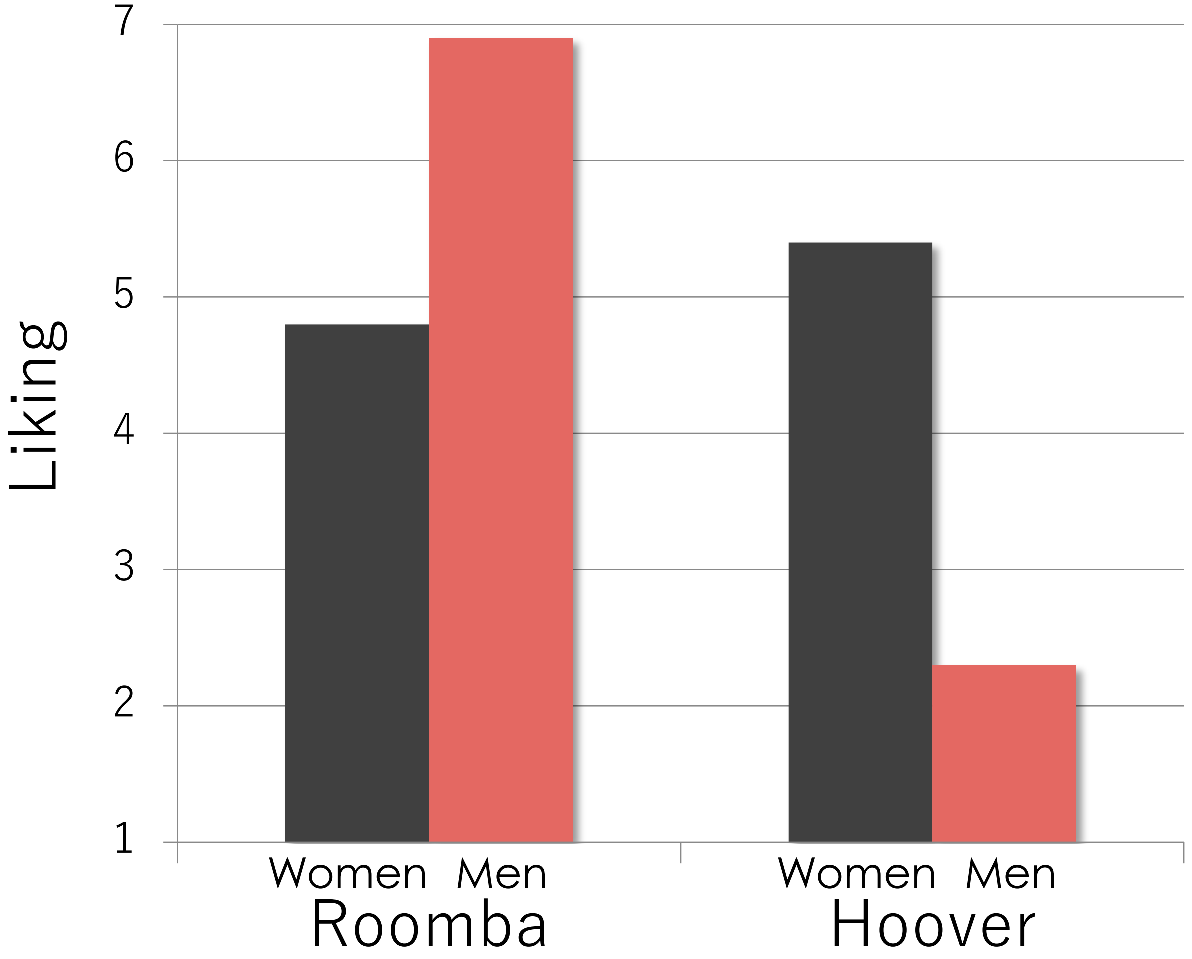
- We can learn more
- More efficient than doing numerous single-factor experiments

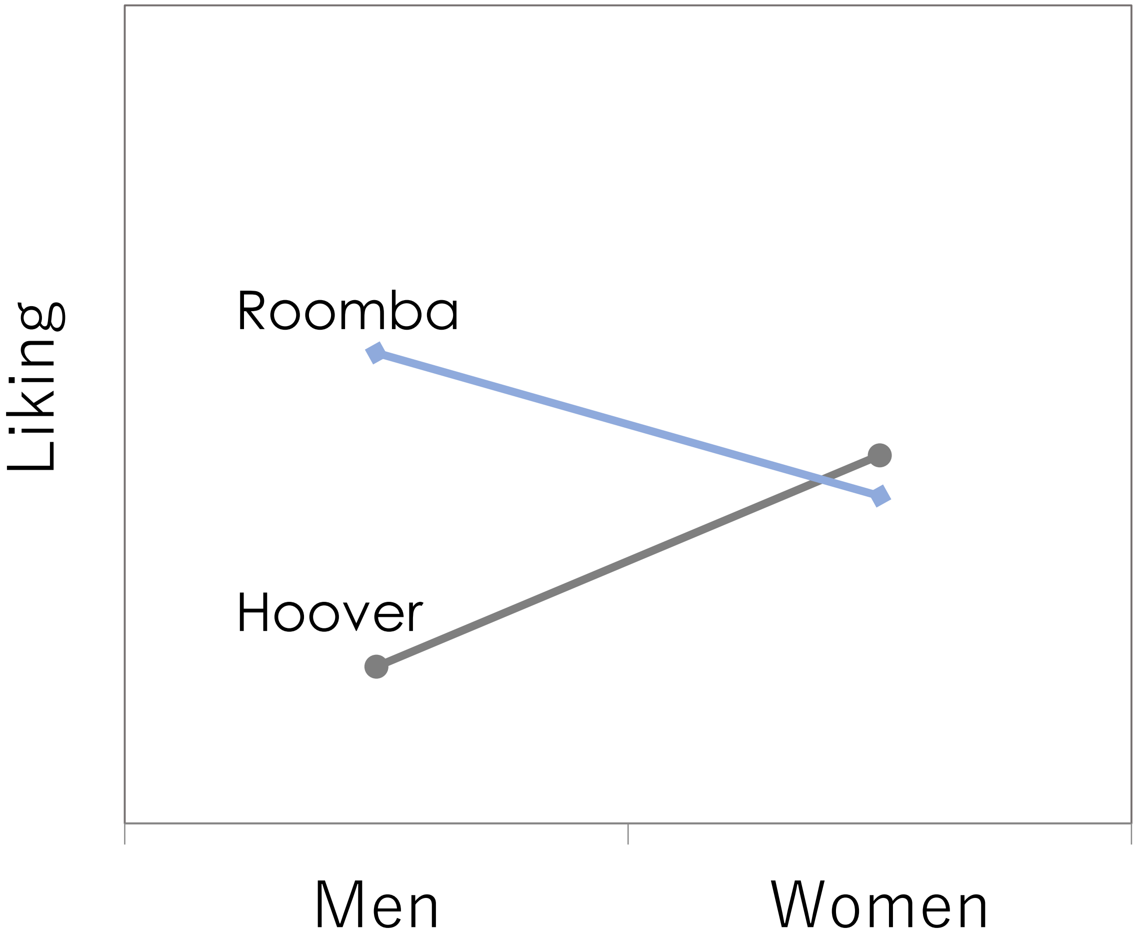
Main Effects





Interaction Effects





Random Sampling

- Random Sampling
  - Choosing participants randomly from the entire population
  - Allows generalization to the population
  - Randomization allows later use of probability theory and gives a solid foundation for statistical analysis
  - Avoid bias
    - The first six participants who come to the lab might be highly motivated
- Random Assignment
  - Random does not mean haphazard
  - One needs to explicitly randomize
    - Random assignment at arrival, counterbalancing, matching

Counterbalancing

Particularly important to **within** designs

Important because of **transfer effects**

- Taking part in earlier trials changes performance in later trials
- Examples: learning, fatigue, etc.

Makes within-subjects designs difficult to interpret

Counterbalancing

In within-subjects counterbalancing:

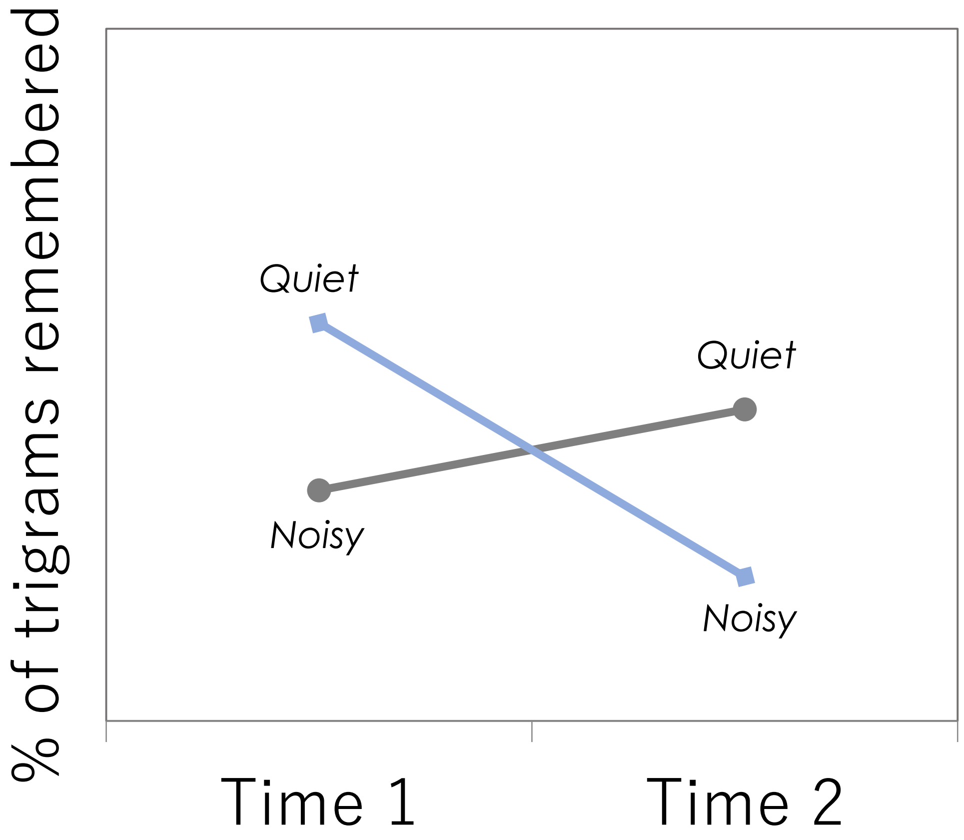
- Possible **linear transfer effects**
- Is the transfer from the 1st position to the 2nd position the same as the transfer from the 2nd to 3rd position?
  - E.g., sometimes the most important learning happens in 1st trials
- Always worry about asymmetrical transfer
  - Does A influence B more than B influences A?

Counterbalancing

Asymmetrical transfer

- Effect of noise depends on order
  - People stick with the strategy they pick first
  - Or mix strategies

Use a between-participants design



Counterbalancing

Full counterbalancing

Test all possible orderings:

Group 1:	A	B	C
Group 2:	A	C	B
Group 3:	B	A	C
Group 4:	B	C	A
Group 5:	C	A	B
Group 6:	C	B	A

May be infeasible

Counterbalancing

Partial counterbalancing

Latin Square

Every condition appears in every position equally:

Joe:	A	B	C
Mary:	B	C	A
John:	C	A	B

Transfer effects can still be a problem

Order is preserved (e.g., A always comes before B)

Solution: use a balanced Latin square

Counterbalancing

Balanced Latin Square

- Depends on even or odd number of conditions
- Even:
  - 1, 2, n, 3, n-1, 4, n-2, … where n = # of conditions
  - Subsequent rows: add one to the previous, returning to 1 after n rows
  - Example: 6 conditions:

Subject Group	1st Condition	2nd	3rd	4th	5th	6th
A	1	2	6	3	5	4
B	2	3	1	4	6	5
C	3	4	2	5	1	6
…						

Every condition follows every other condition once

Counterbalancing

Balanced Latin Square

- Depends on even or odd number of conditions
- Odd:
  - Create two Latin squares
  - First square: use same method (1, 2, n, 3, n-1, 4, n-2, … where n = # of conditions)
  - Second square: mirror image of first square
  - Example: 5 conditions:

1	2	5	3	4	4	3	5	2	1
2	3	1	4	5	5	4	1	3	2
3	4	2	5	1	1	5	2	4	3
4	5	3	1	2	2	1	3	5	4
5	1	4	2	3	3	2	4	1	5

Every condition follows every other condition twice

Matching

Try to reduce between-group differences

E.g., rank hearing as Good, Fair, Poor

Unmatched, could get:

Noisy: Poor1, Poor2, Fair1

Quiet: Good1, Good2, Fair2

Matched, get:

Noisy: Poor1, Fair2, Good1

Quiet: Poor2, Fair1, Good2

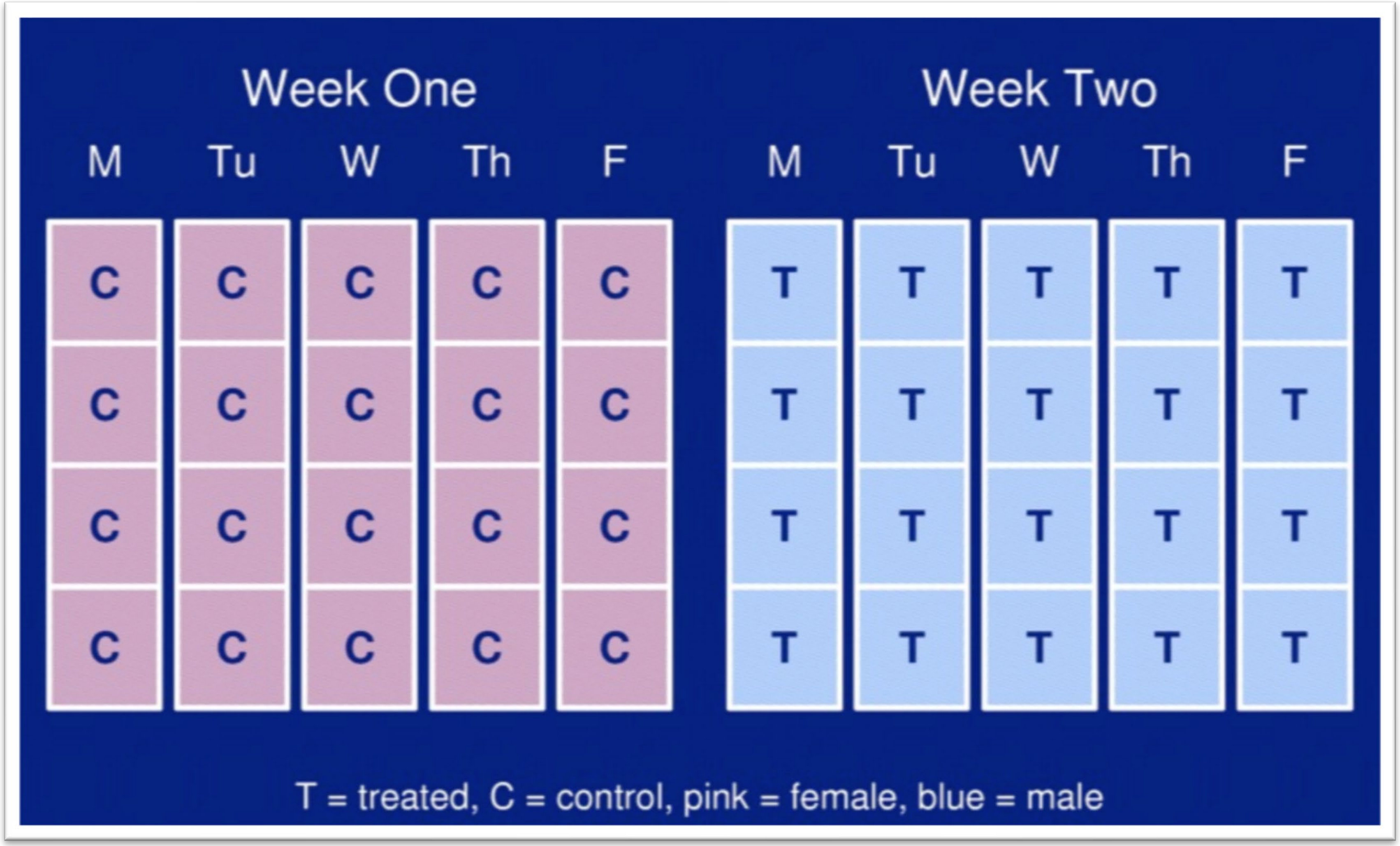
Stratification

Supposed that some social measurements will be made in the morning and some in the afternoon

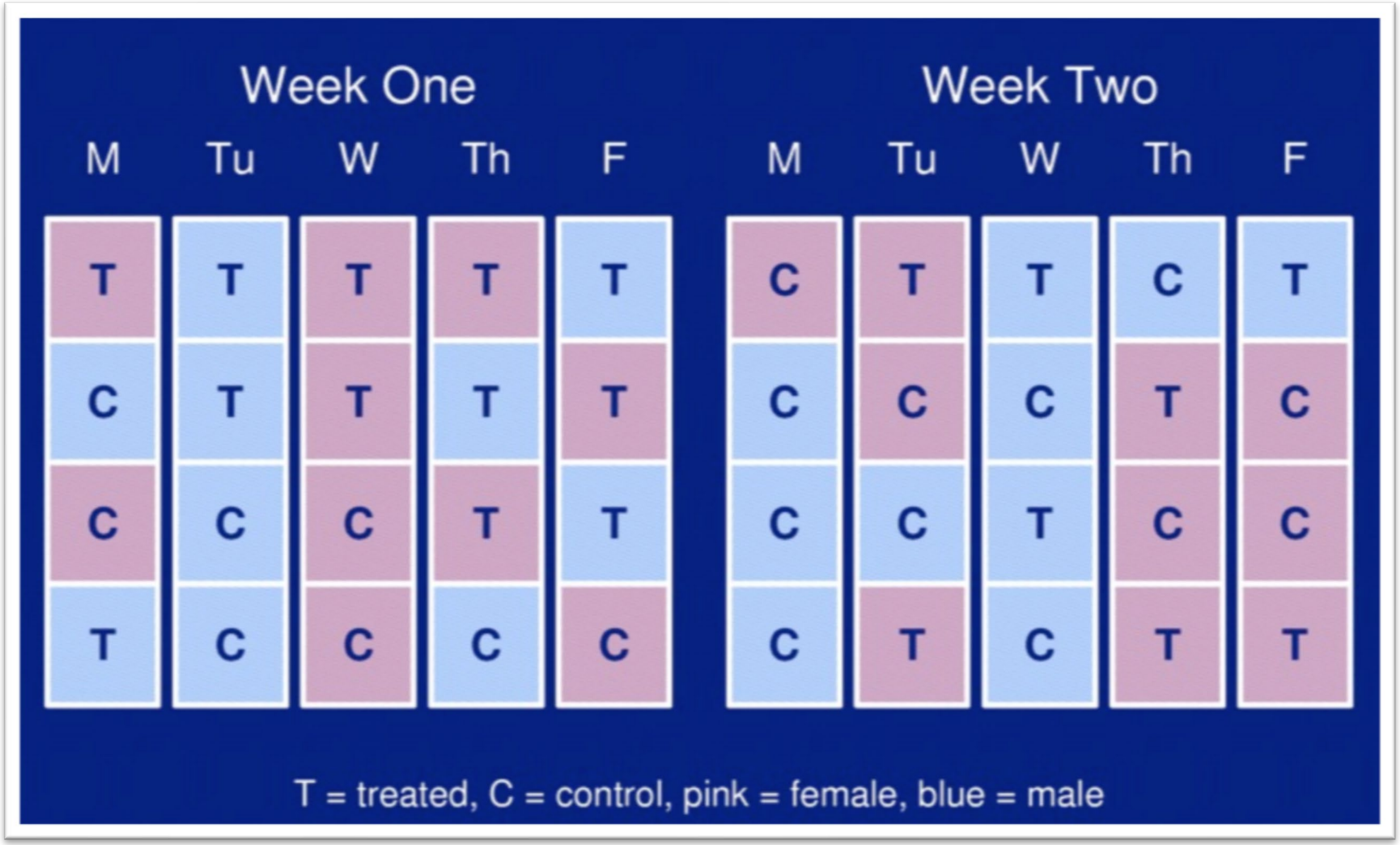
- If you anticipate a difference between morning and afternoon measurements:
  - Ensure that within each period, there are equal numbers of subjects in each treatment group
  - Take account of the difference between periods in your analysis (e.g., as covariate)

This is sometimes called "blocking"

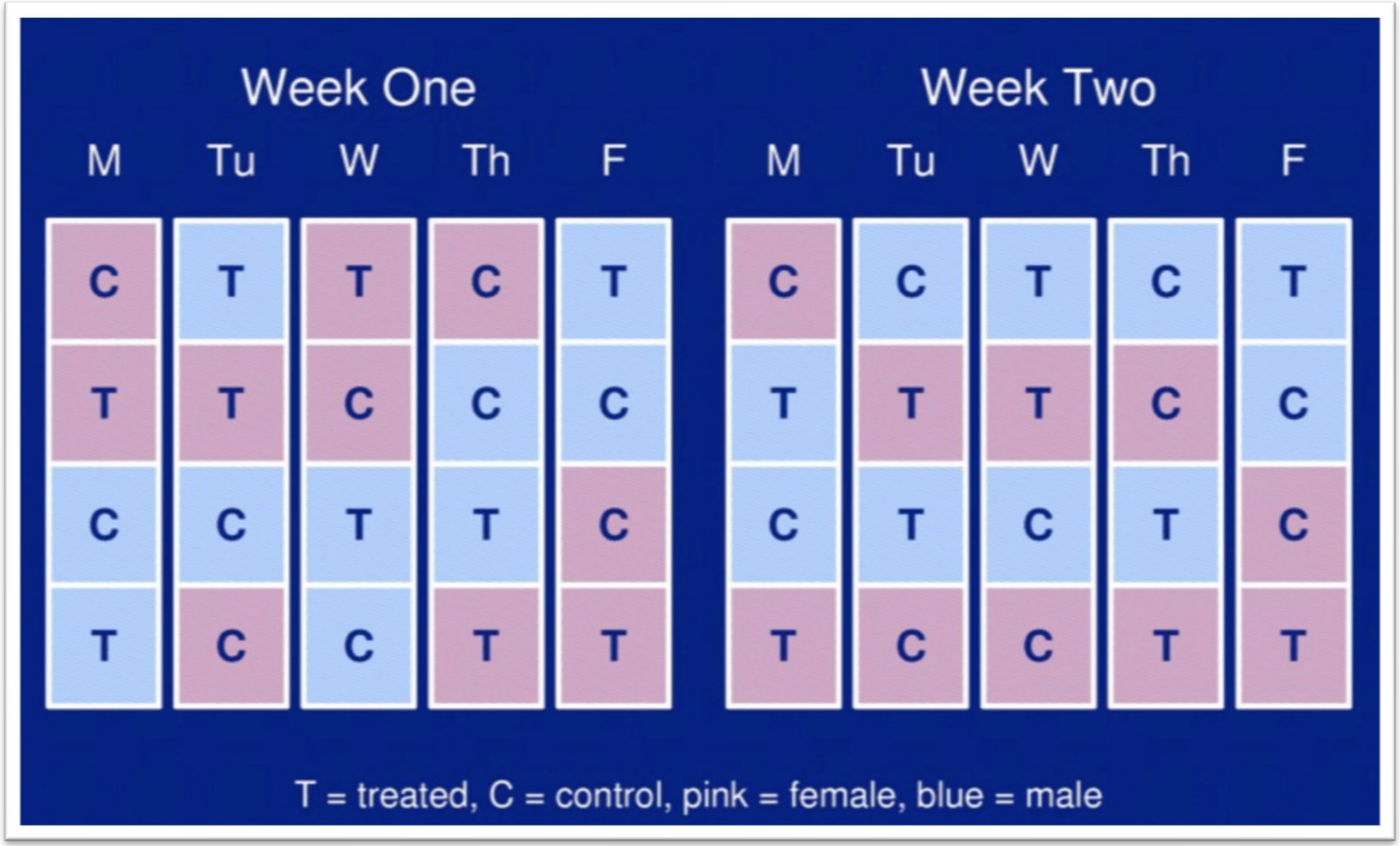
Very Bad Design



Randomized


Week One
M Tu W Th F
T T T T T
C T T T T
C C C T T
T C C C C
Week Two
M Tu W Th F
C T T C T
C C C T C
C C T C C
C T C T T
T = treated, C = control, pink = female, blue = male

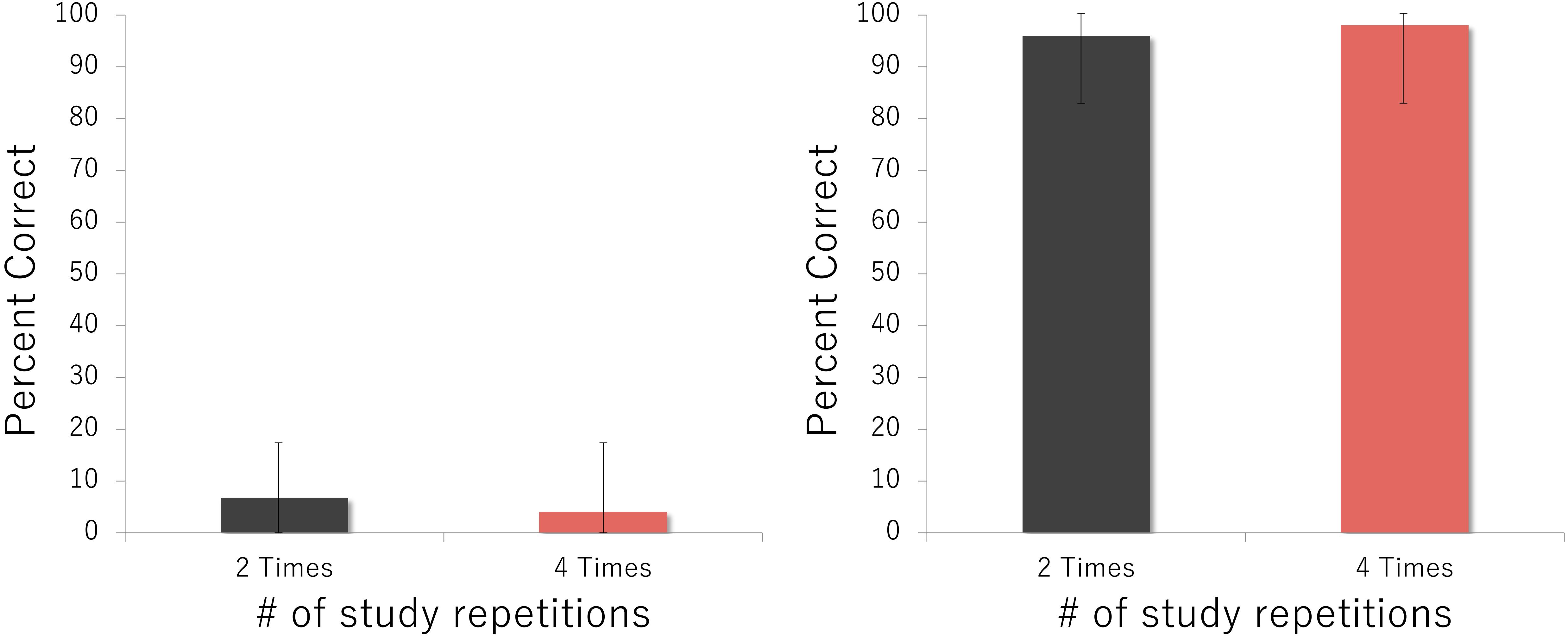
Stratified


Week One
Week Two
M Tu W Th F
M Tu W Th F
C T T C T
T T C C C
C C T T C
T C C T T
C C T C T
T T T C C
C T C T C
T C C T T
T = treated, C = control, pink = female, blue = male

All of these techniques make trade-offs between validity and practicality

Other considerations

Floor and Ceiling Effects



Blinding

Measurements made by people can be influenced by unconscious bias

Ideally, dissections and measurements should be made without knowledge of the treatment applied

Single vs. double-blind designs

Internal controls

- It can be useful to use the subjects themselves as their own controls (e.g., measuring the response after vs before treatment)
- Increased precision

Representativeness

Are the subjects you are studying really representative of the population you want to study?

Ideally, your study material is a random sample from the population of interest (true vs quasi experiment)

Summary:
Characteristics of Good Experiments

Unbiased
- Randomization
- Blinding

High precision
- Uniform material
- Replication
- Blocking

Simple
- Protect against mistakes

Wide range of applicability
- Deliberate variation
- Factorial designs

Able to estimate uncertainty
- Replication
- Randomness

Good methodological fit

Questions?

Next

Reading #5

Posted on Moodle
Need Pro/Con presenters for Wednesday 2/20

Assignment 2

Due Monday 3/4/19
See guidelines on Moodle

Project

Interim Presentation/Writeup Due Monday 3/11
See guidelines on Moodle

Guest Lecture next class (Monday 2/18)

University of Colorado Boulder

THANKS!

Professor **Dan Szafir**

Computer Science & ATLAS Institute
University of Colorado Boulder