Human-Computer Interaction

Experimental Research: An Introduction

Professor Bilge Mutlu

Questions

To ask questions during class:

- Go to slido.com and use code **#2938904** or direct link or scan QR code
- Anonymous
- I will monitor during class

Today's Agenda

- Topic overview: *Experimental Research: An Introduction*
- Hands-on Activity: *Reverse Engineering Experimental Design Choices*

What is a hypothesis?

Definition: An idea that proposes a tentative explanation about a phenomenon or a narrow set of phenomena observed in the natural world.

The two primary features of a scientific hypothesis are *falsifiability* and *testability* (proposed by Karl Popper).[1]

---

[1] Brittanica

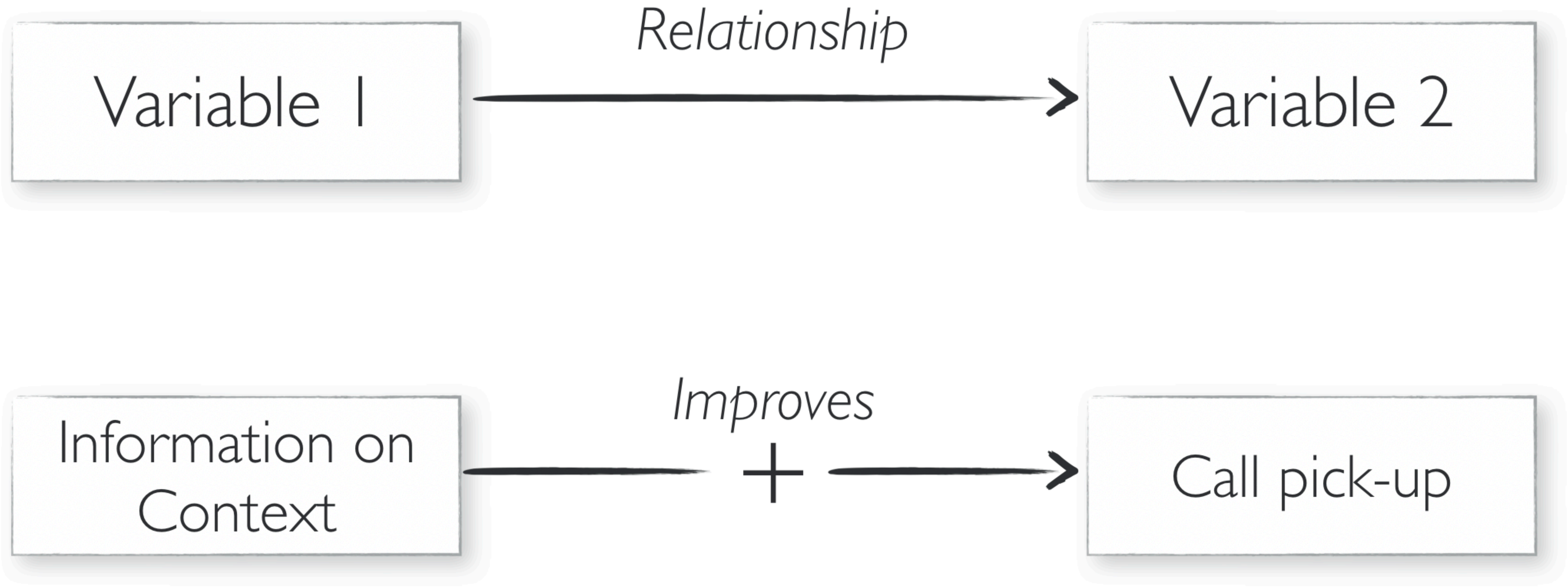
A more operational definition

A statement of the predicted or expected relationship between at least two variables (e.g., an independent and a dependent variable).

A well-formulated hypothesis should provide an provisional, testable answer to a *research question*.

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 on their callers' context than they will be when they do not.*


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

Good hypotheses should be:

1. **Testable**, such that manipulating *independent variables* and measuring *dependent variables* should be possible
2. **Falsifiable** with the data obtained from the experiment
3. **Parsimonious**, representing a prediction in the simplest form
4. **Precise** such that the researcher can operationalize it
5. **Useful** to inform existing or new theory

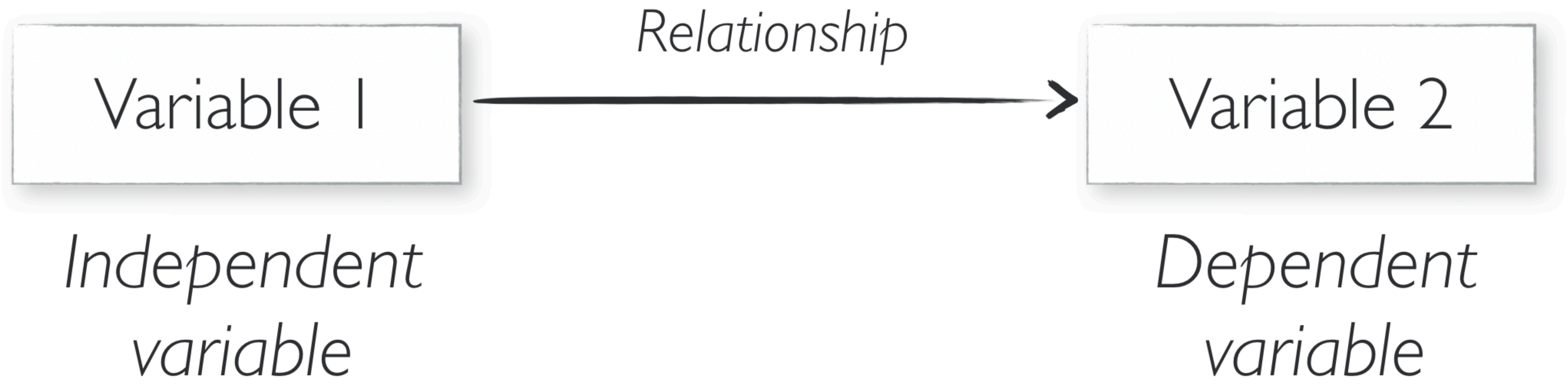
How do we think about experimental variables?

An experiment is empirical journey set to establish relationships among *variables*—ideas in the conceptual domain[2] represented as mathematical abstractions. Five key types of variables:

1. **Independent variable**: what is being manipulated
2. **Dependent variable**: what is being measured
3. **Control variable**: what is held constant
4. **Random variable**: what is allowed to vary randomly
5. **Confounding variable**: what correlates with IV/DV

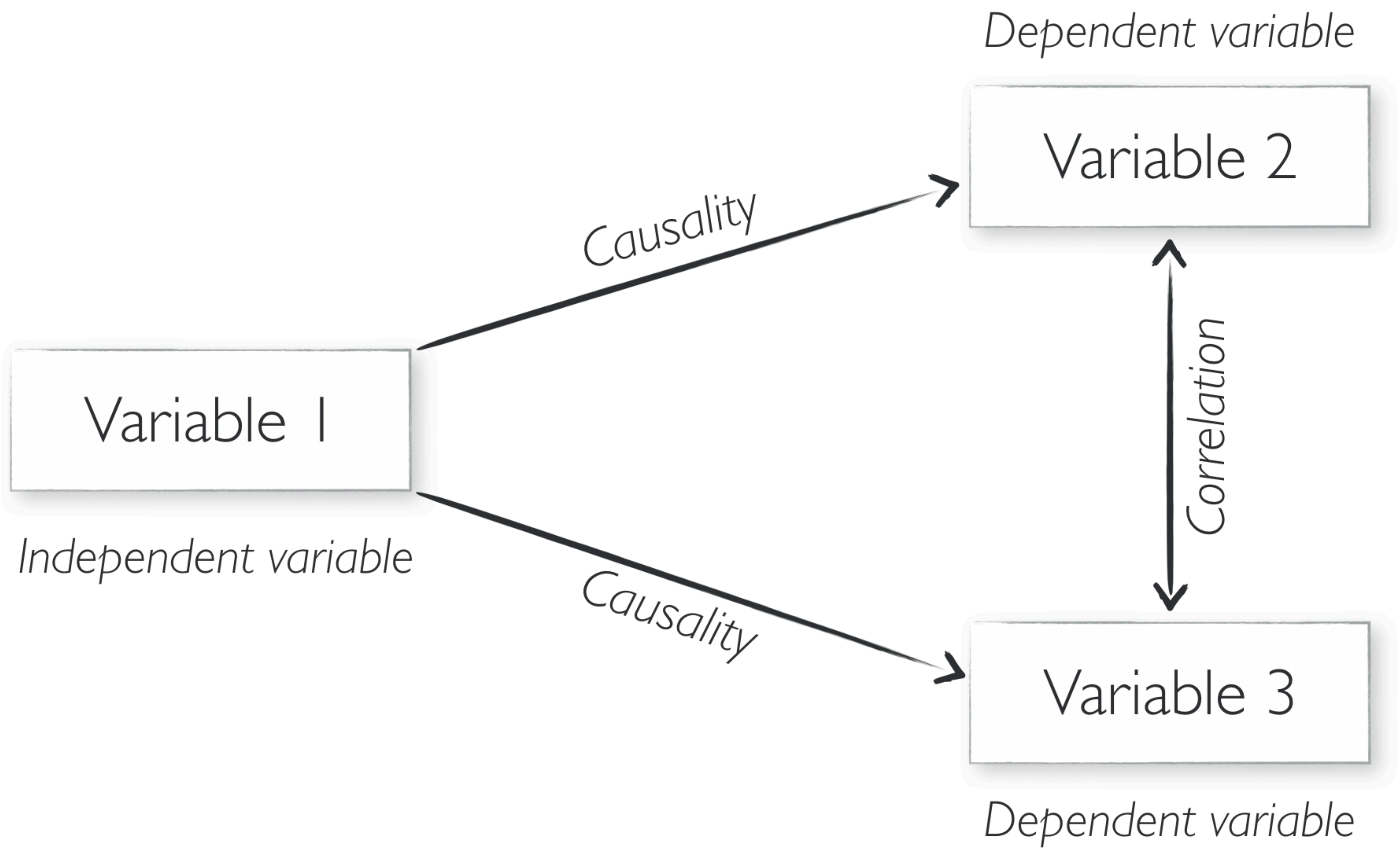
---

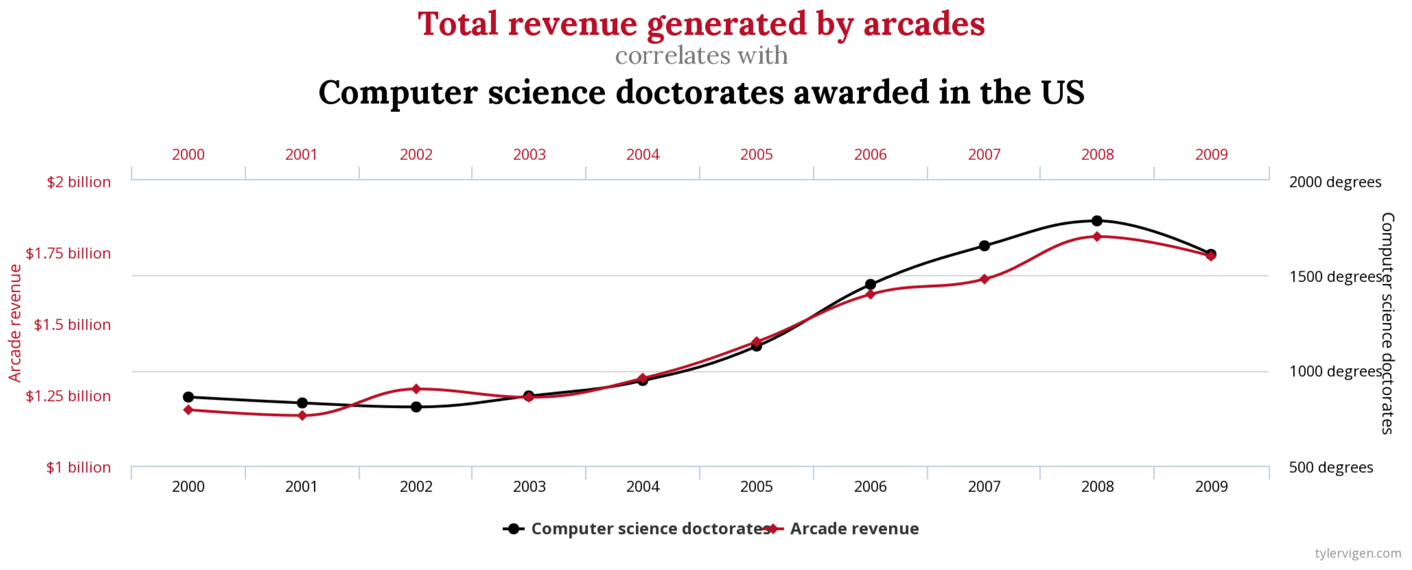
[2] McGrath, J. E. (1995). Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human–Computer Interaction* (pp. 152-169). Morgan Kaufmann.

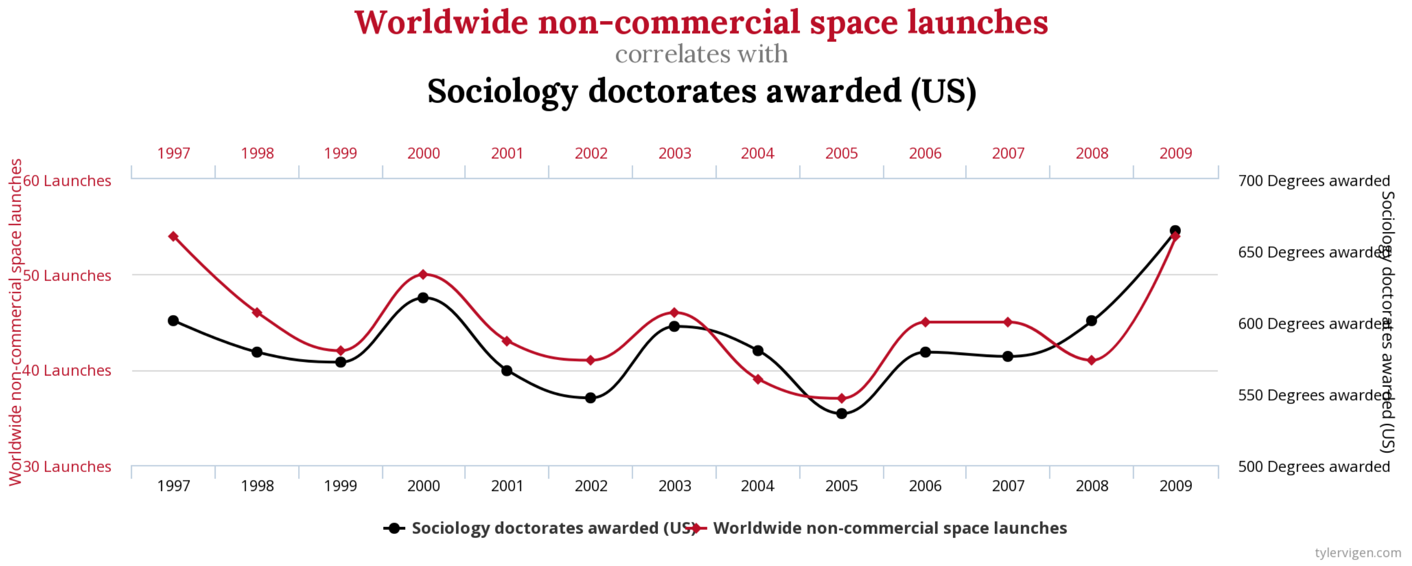

Relationship
Variable 1
Variable 2
Independent variable
Dependent variable

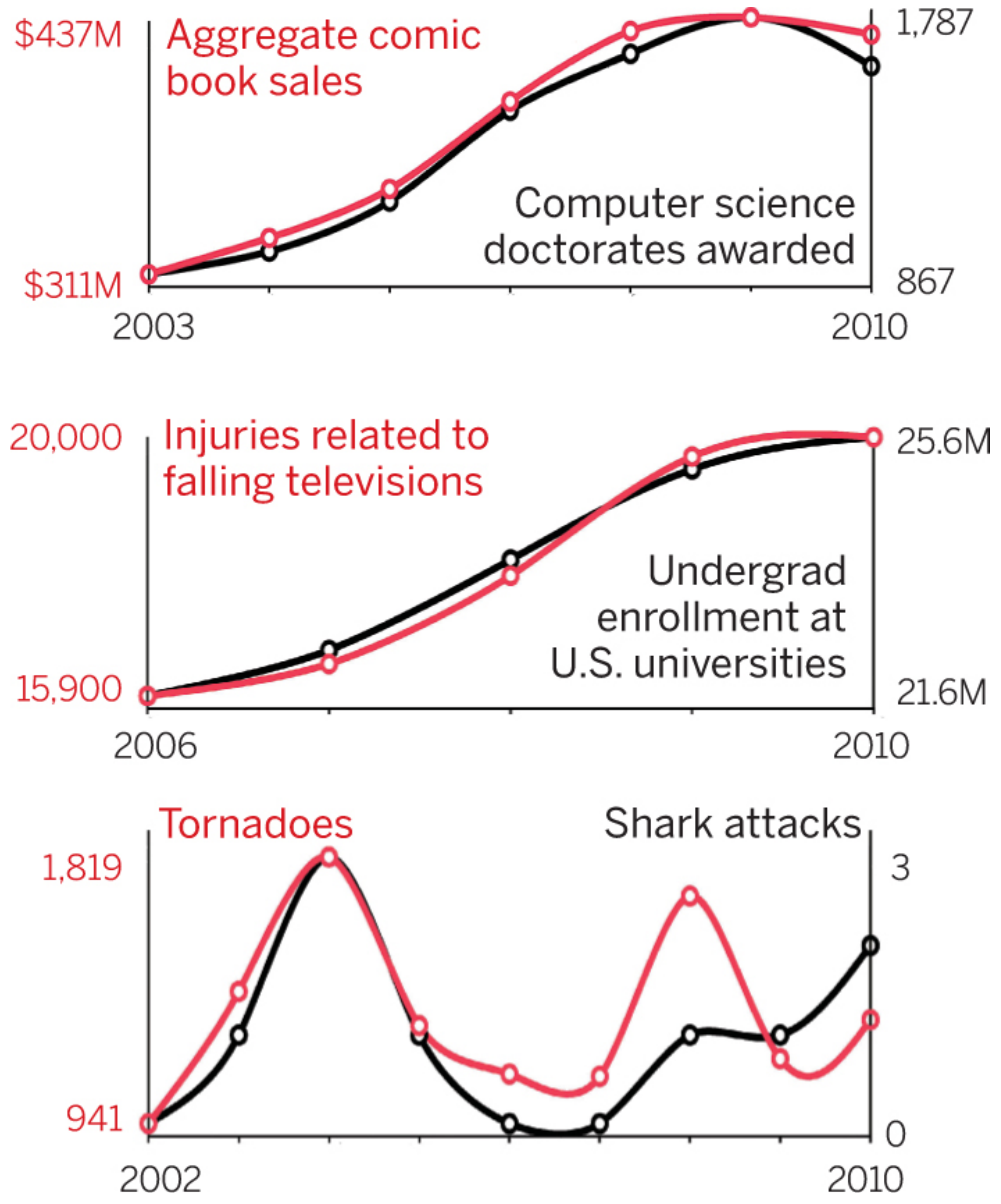
What kinds of relationships are we interested in?

- **Causal relationships:** one variable (DV) depends on and is affected by another (IV).
- **Correlational relationships:** two variables are affected by a third variable in the same direction.


Dependent variable
Variable 2
Causality
Correlation
Variable 1
Independent variable
Causality
Variable 3
Dependent variable

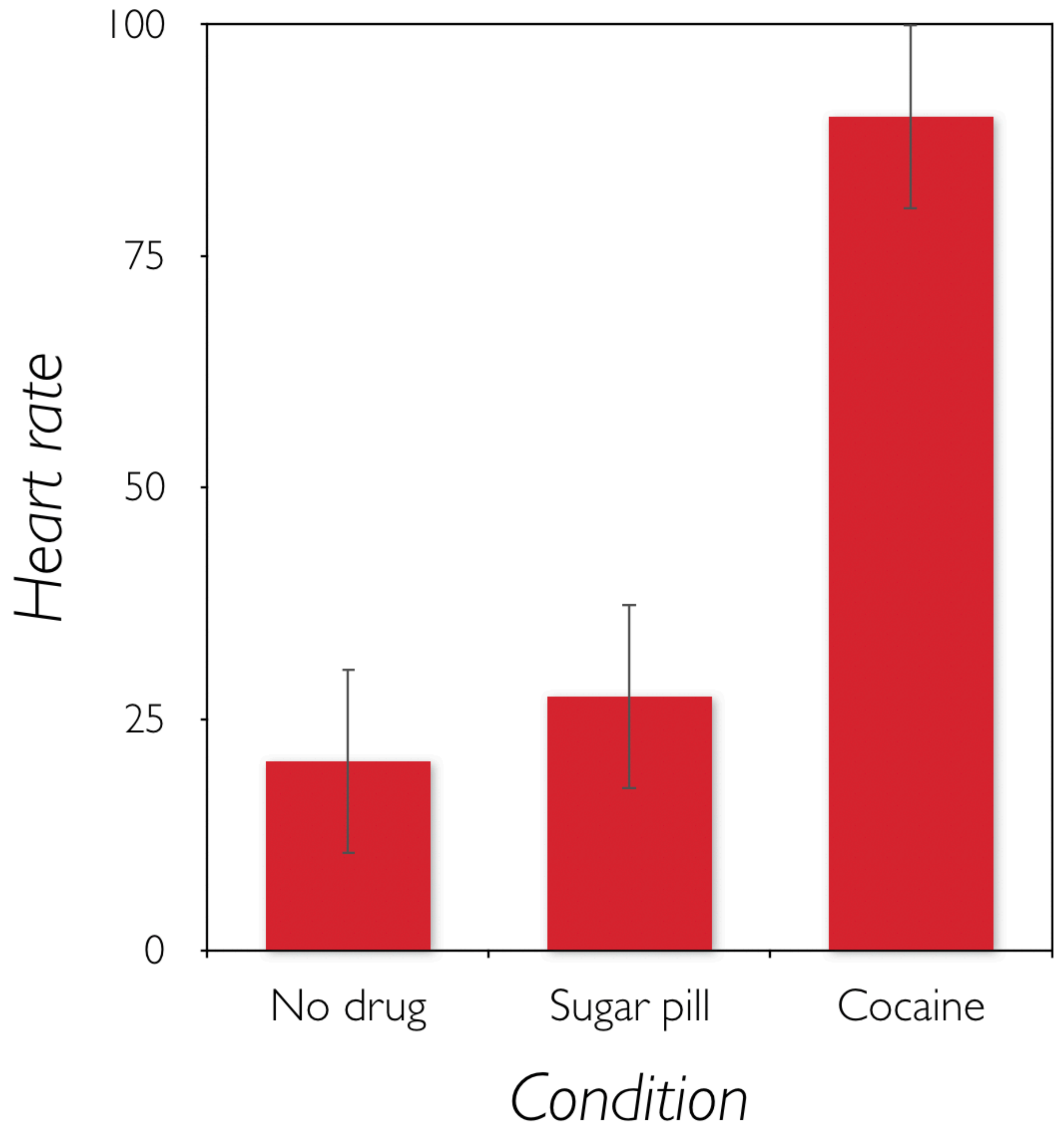


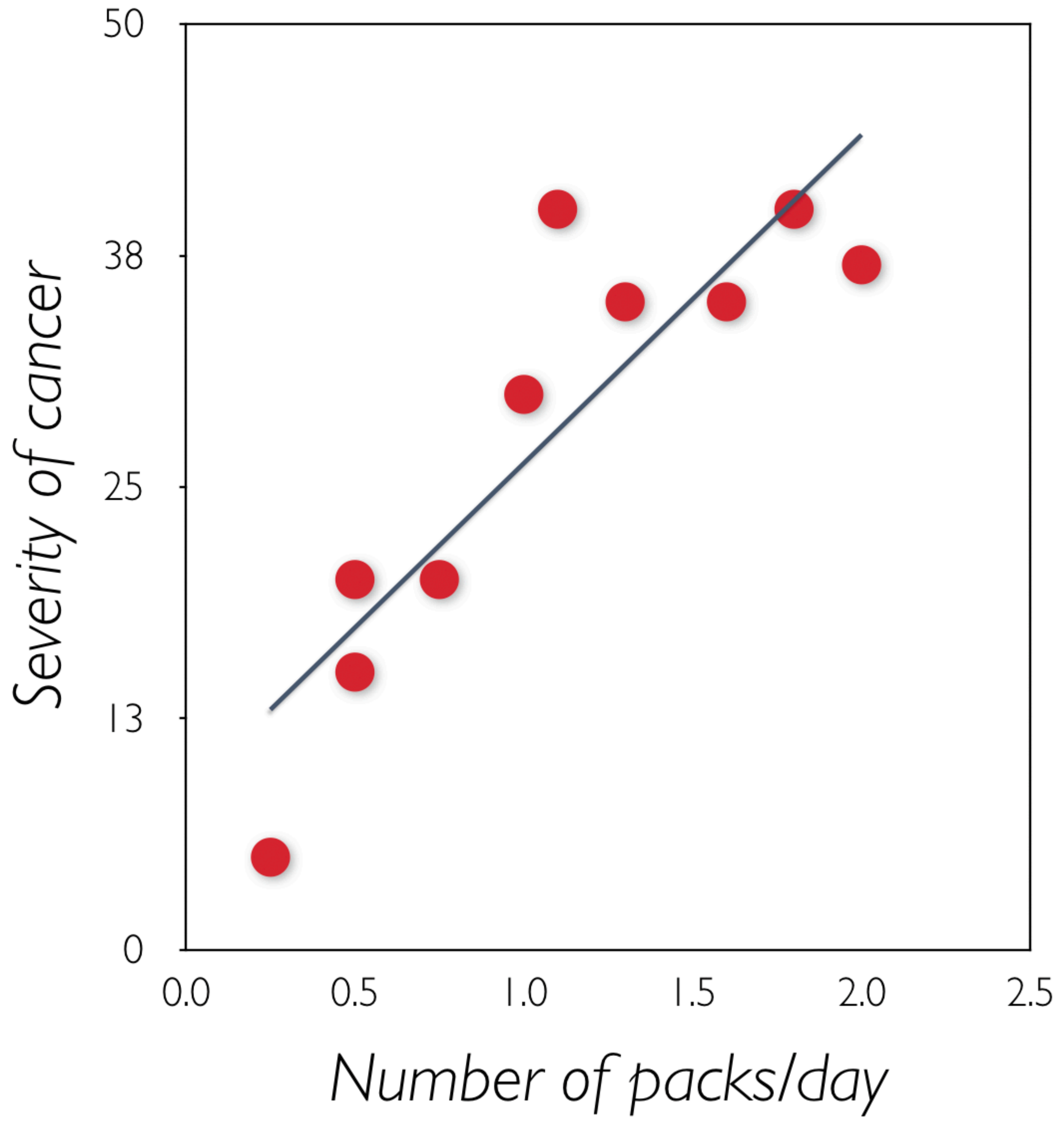




[3] Left: Tyler Vigen: Spurious correlations

[4] Right: Books et al., 2015, Spurious correlations


100
75
50
25
0
Heart rate
No drug
Sugar pill
Cocaine
Condition


50
38
25
13
0
Severity of cancer
0.0
0.5
1.0
1.5
2.0
2.5
Number of packs/day

How do we think of variables in designing experiments?

By answering the following questions:

1. Do we *know* all the variables involved?
2. Can we *control* independent variables of interest?
3. Can we *measure* all variables?

If the answers are:

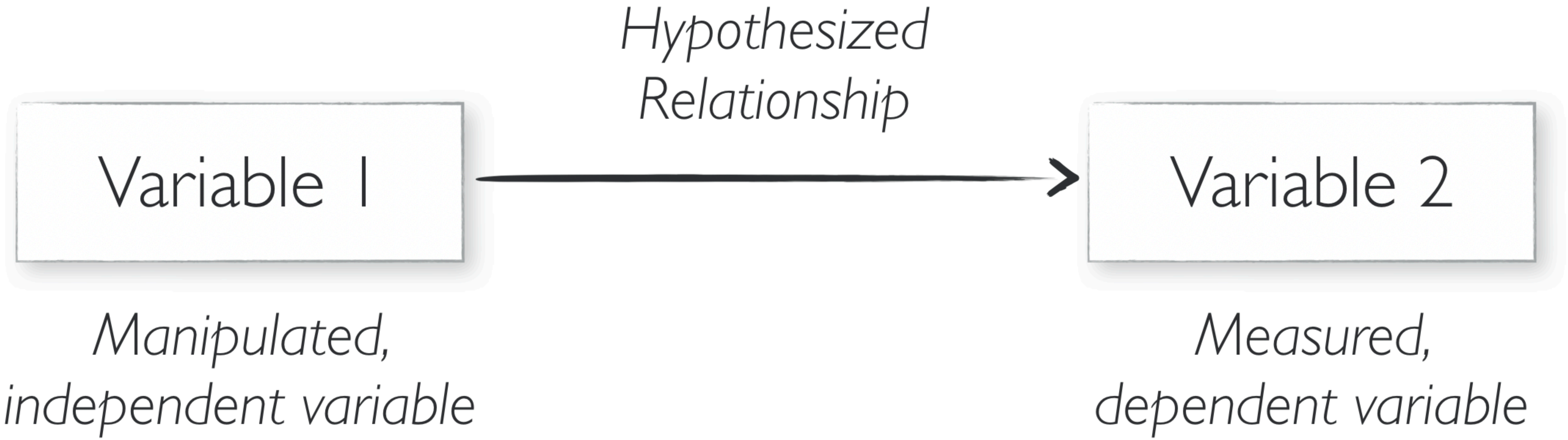
Yes. Yes. Yes.

We are designing a **true experiment** where independent variables are manipulated, dependent variables are measured, and the hypothesis is a cause-effect statement.

Example hypotheses:

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

Reading speed (words/minute) will change with font size, increasing as font size is increased from 4 point to 20 point, but decreasing as font size is increased above 20 point.


Hypothesized
Relationship
Variable 1
Variable 2
Manipulated,
independent variable
Measured,
dependent variable

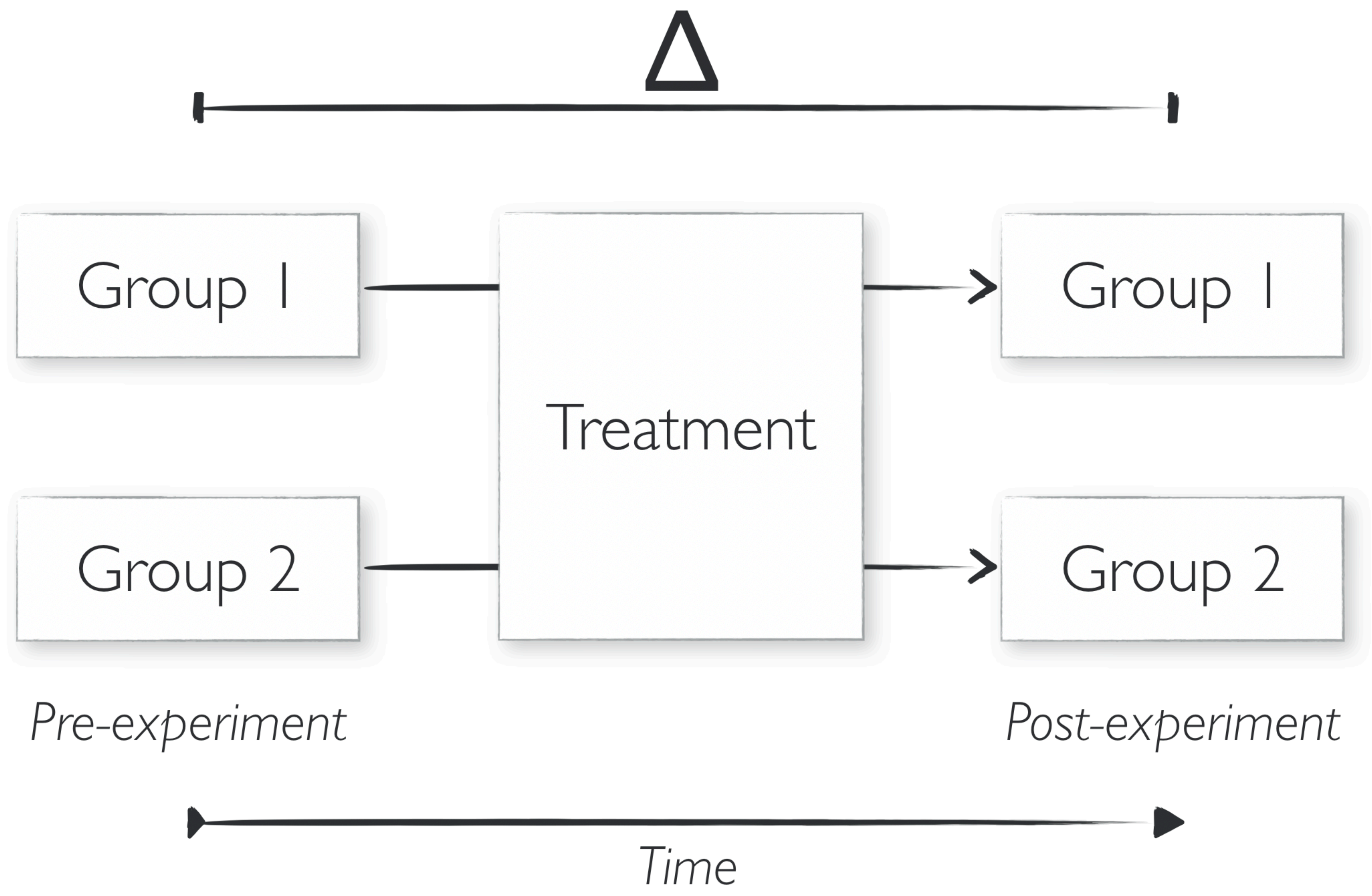
If the answers are:

Yes. No. Yes.

We are designing a **quasi experiment** where controlling independent variables or random assignment is not practical or possible and where we separate participants based on some characteristics (e.g., expertise, verbal ability) and measure them before and after an intervention.

Example hypothesis:

Students with high verbal ability will show higher vocabulary retention after the second-language-learning intervention than students will low verbal ability.


Δ
Group 1
Group 2
Treatment
Group 1
Group 2
Pre-experiment
Post-experiment
Time

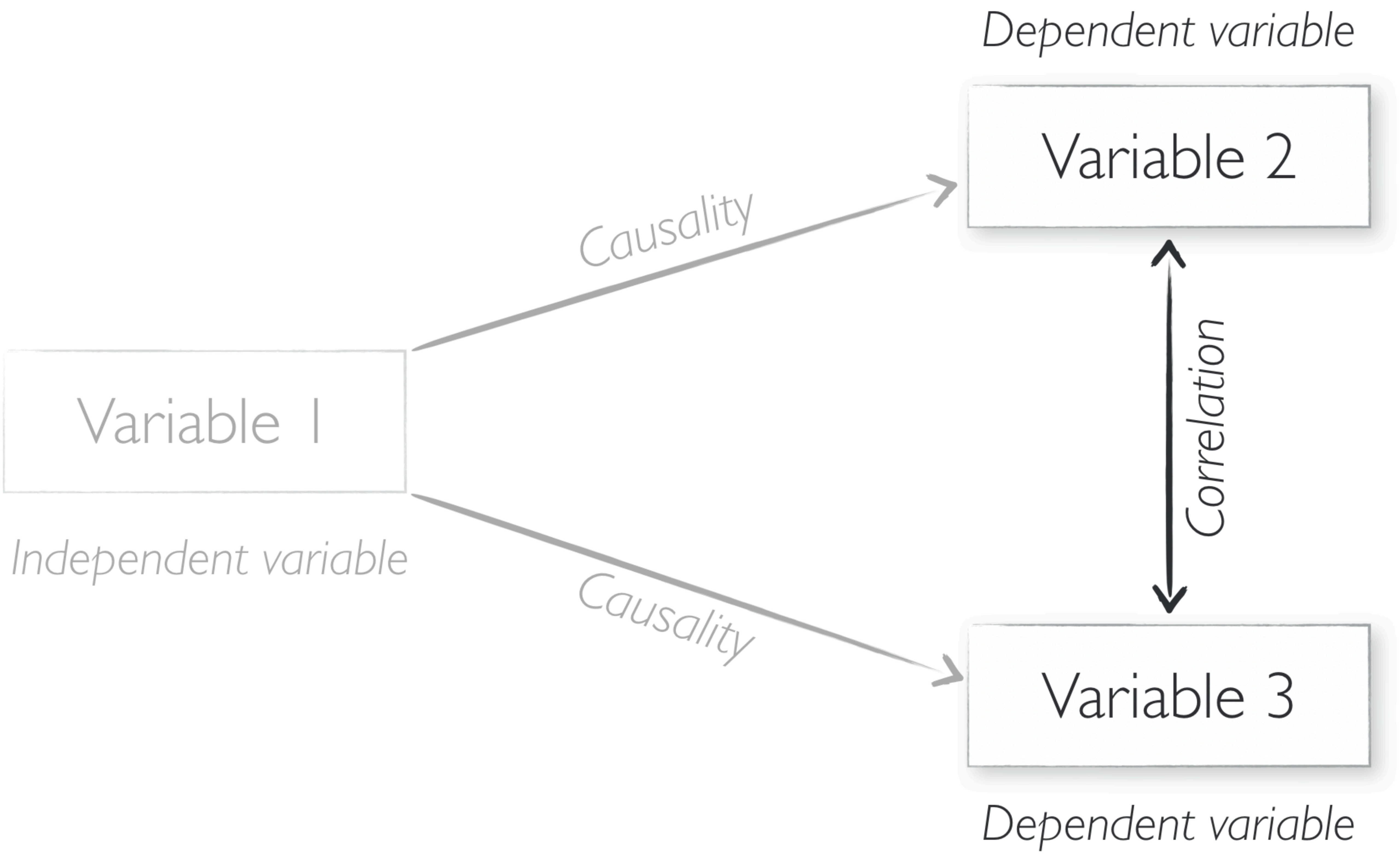
If the answers are:

No. Yes/no. Yes.

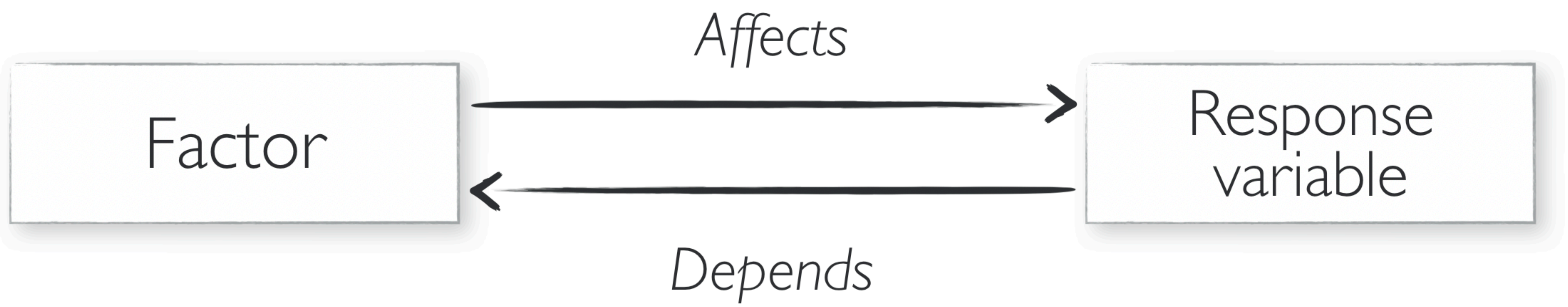
We are designing a correlational study that examines the relationships among dependent variables such as individual traits, behavior, outcomes and the hypotheses are specific statements about these relationships.

Example hypotheses:

We will observe a strong positive correlation between test-taking anxiety and test-taking performance.

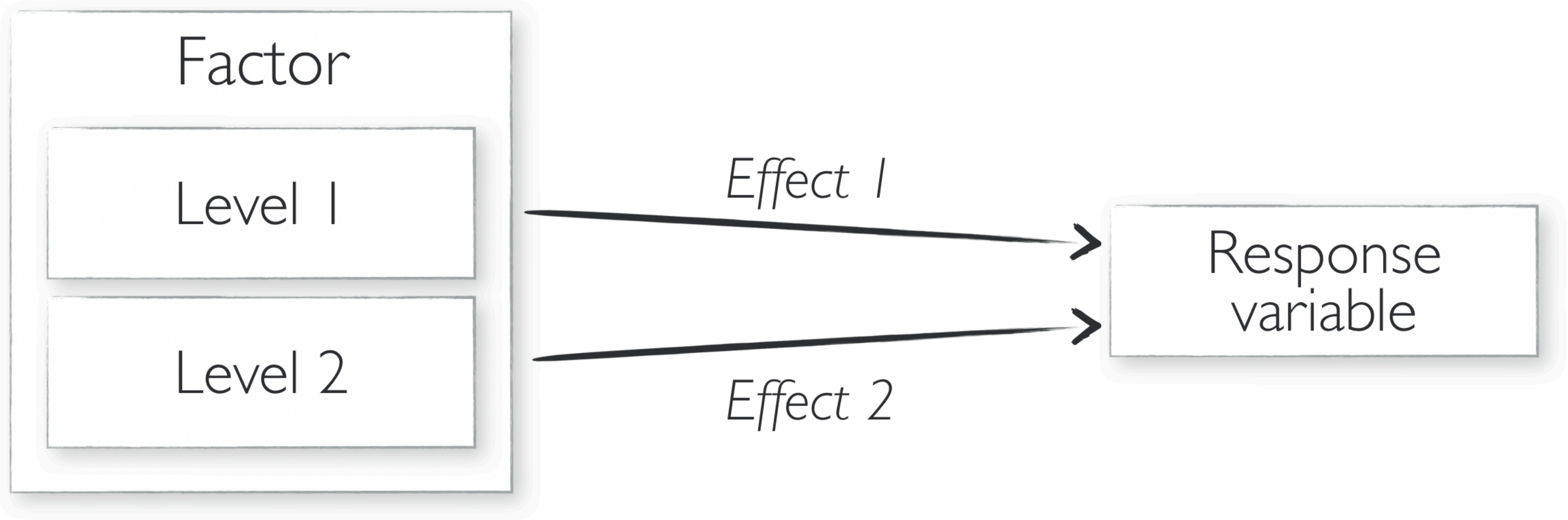

Dependent variable
Variable 2
Causality
Correlation
Variable 1
Independent variable
Causality
Variable 3
Dependent variable

Independent and dependent variables are also called **factors** and **response** or **outcome**.



How do we perform random assignment of the factor to a population?

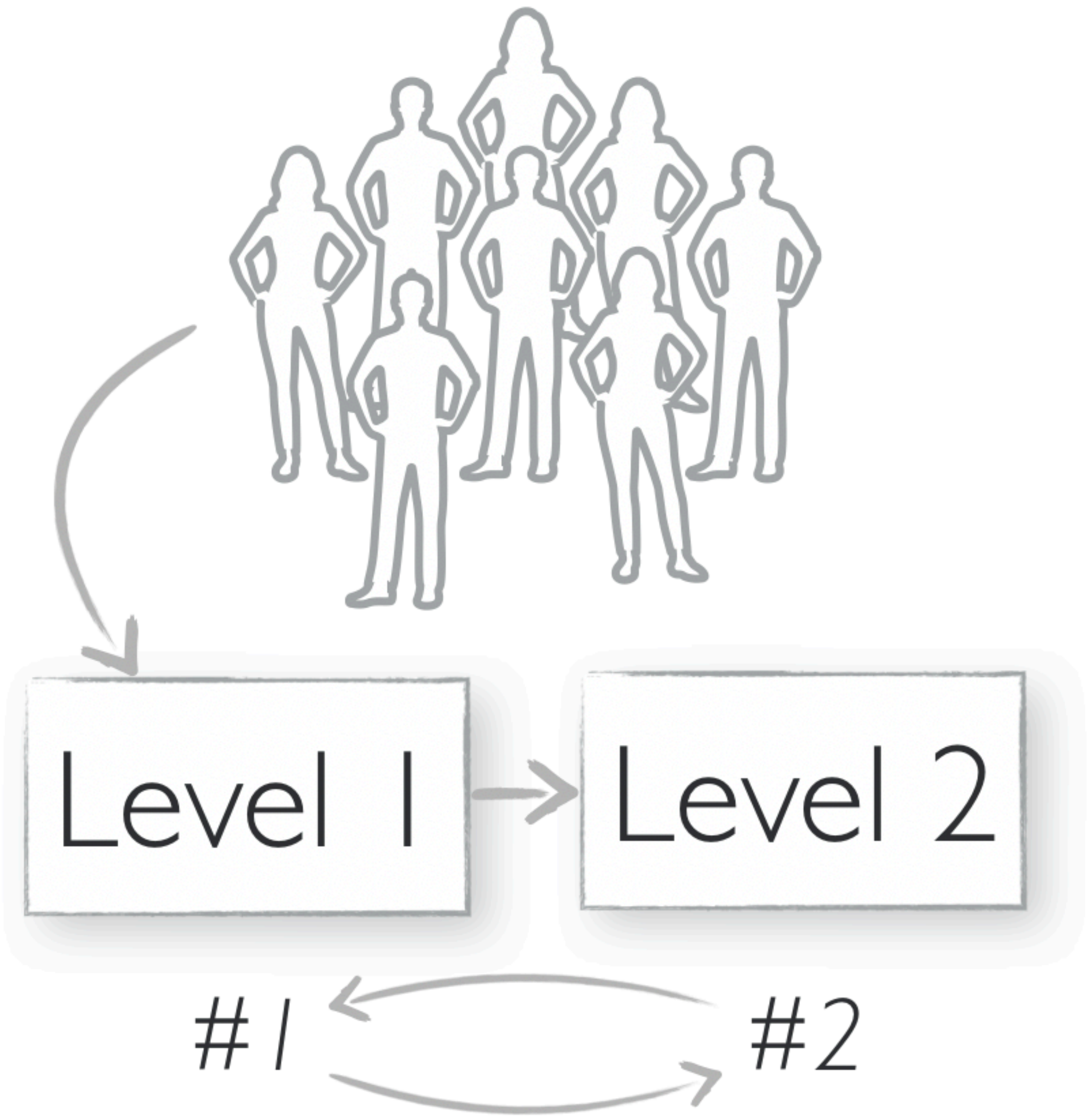
We create multiple **levels** of the factor (also called **experimental conditions**), e.g., interface 1 vs. interface 2; low verbal ability vs. high verbal ability; novice vs. expert. The independent *variable* takes different values. We measure and compare the differential effects of the different levels on the response variable.



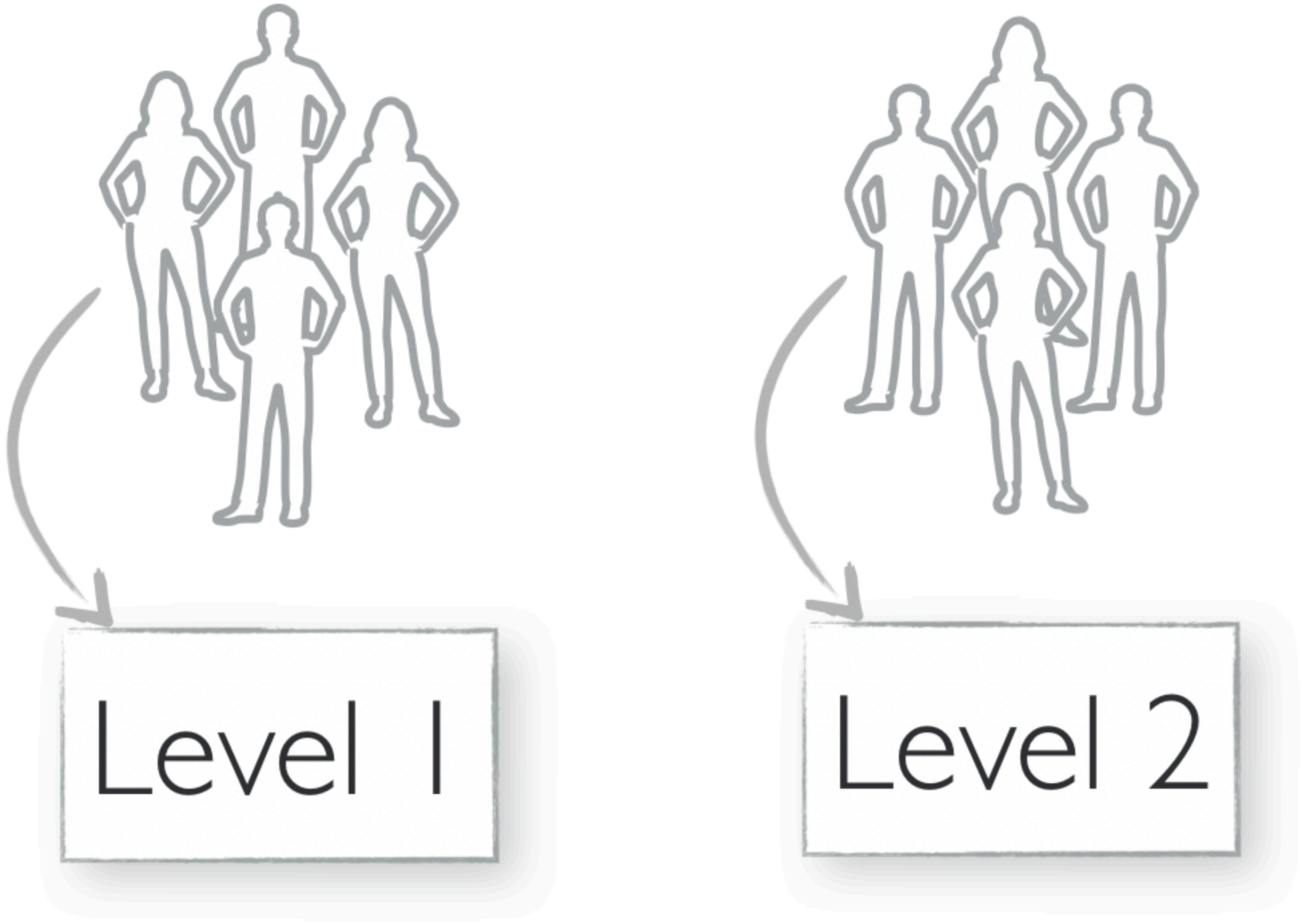
Each level of the factor needs to be assigned to a sub-population. This assignment can be done in two ways:

1. **Within-participants** assignment, where all participants will observe each level in a specified order.
2. **Between-participants** assignment, where mutually exclusive subgroups of the population will each observe one of the levels.

Within-participants



Between-participants

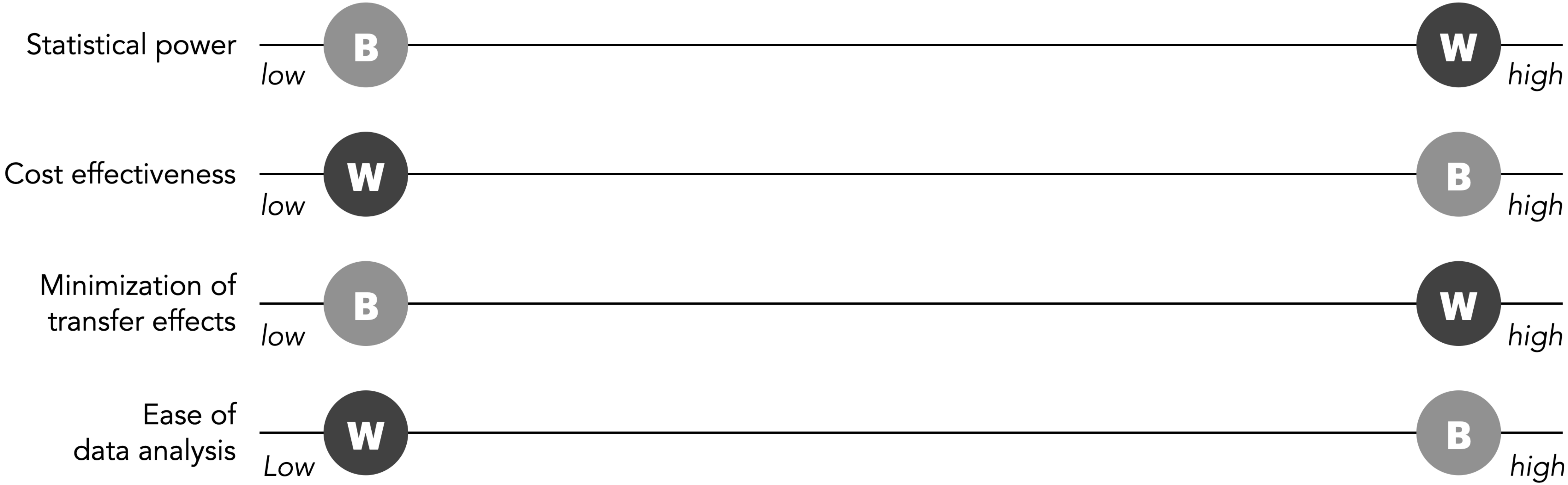


Within-participants designs:

1. Offers higher statistical power
2. Takes more time
3. Suffers from transfer effects
4. Data analysis can be more complex

Between-participants designs:

1. Minimizes transfer effects
2. Requires larger samples
3. Offers less power
4. Easier to analyze data


Statistical power
low
B
W
high
Cost effectiveness
low
W
B
high
Minimization of
transfer effects
low
B
W
high
Ease of
data analysis
Low
W
B
high

What if I have multiple factors?

Experiments can involve manipulating **single** or **multiple** independent variables (factorial designs).

Factorial designs:

1. Examine multiple variables at once
2. Analyze interactions
3. Will require larger samples

In general, factorial designs are *more efficient* but also *more demanding.*

How does gaming and technology savviness affect perceptions of robots?

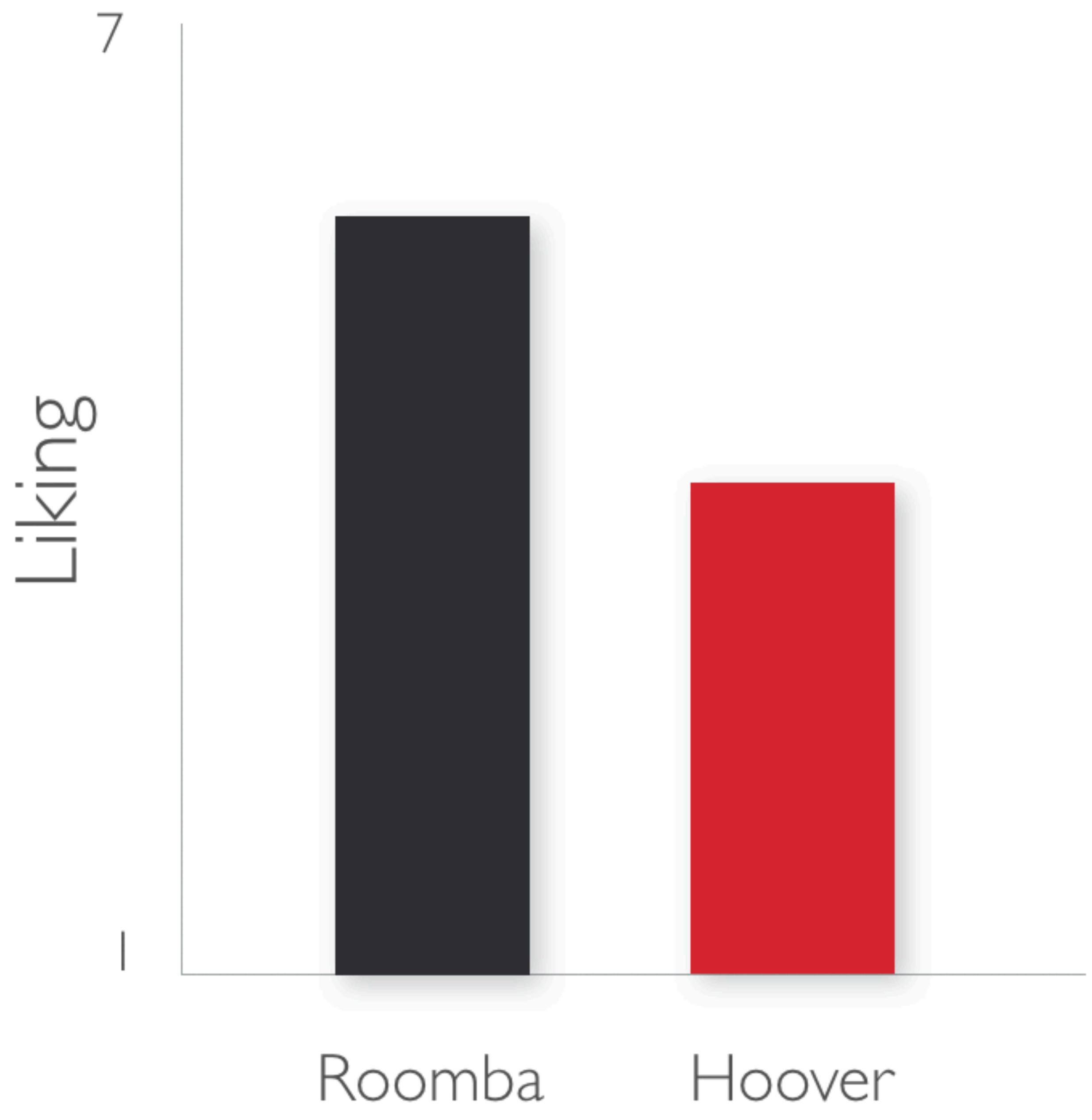
Single-factor experiment (2 two-population experiments):

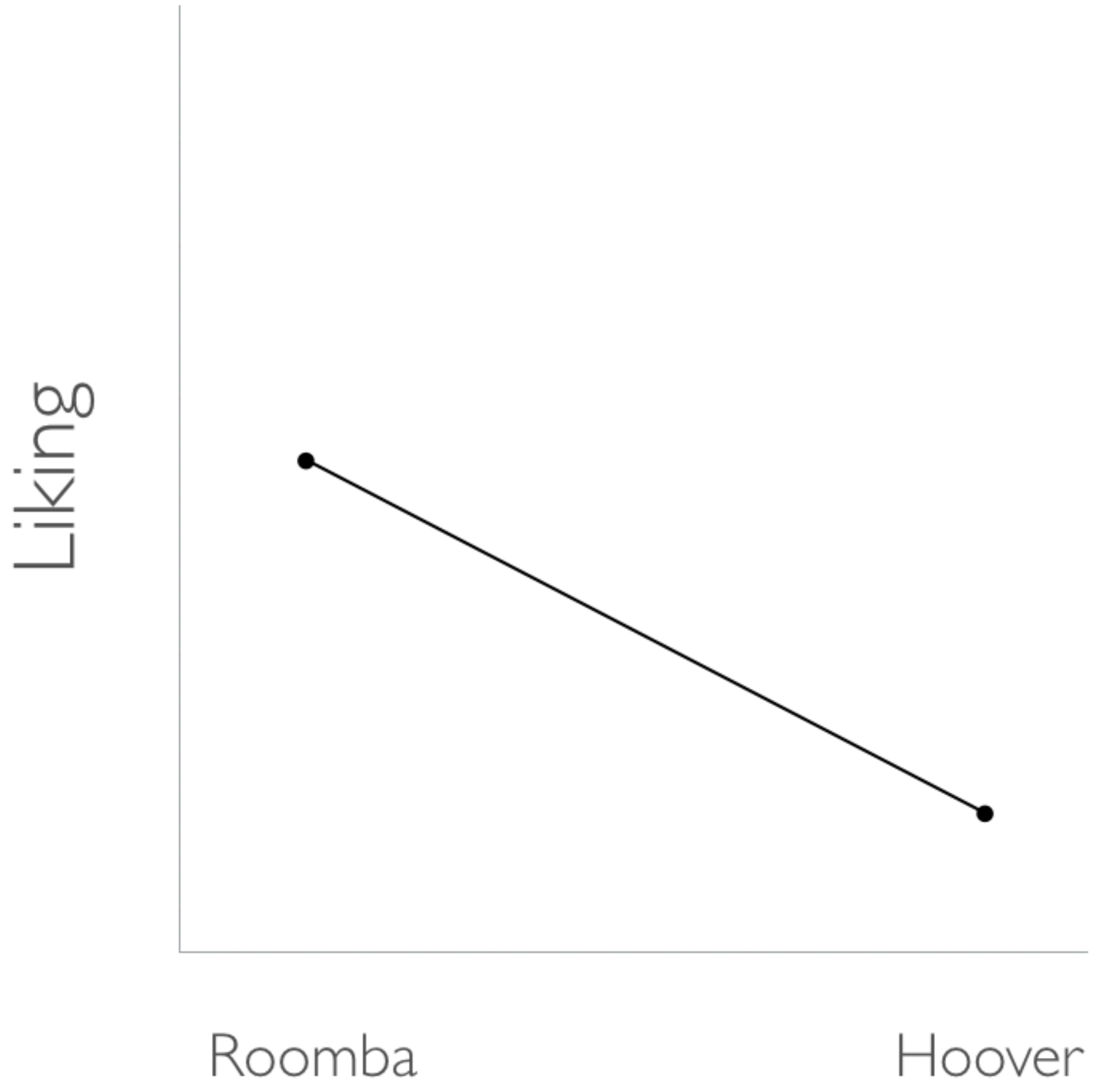
- *Experiment 1:* low/high gaming ↠ perceptions of robots
- *Experiment 2:* low/high tech savviness ↠ perceptions of robots

Factorial experiment (1 four-population experiment):

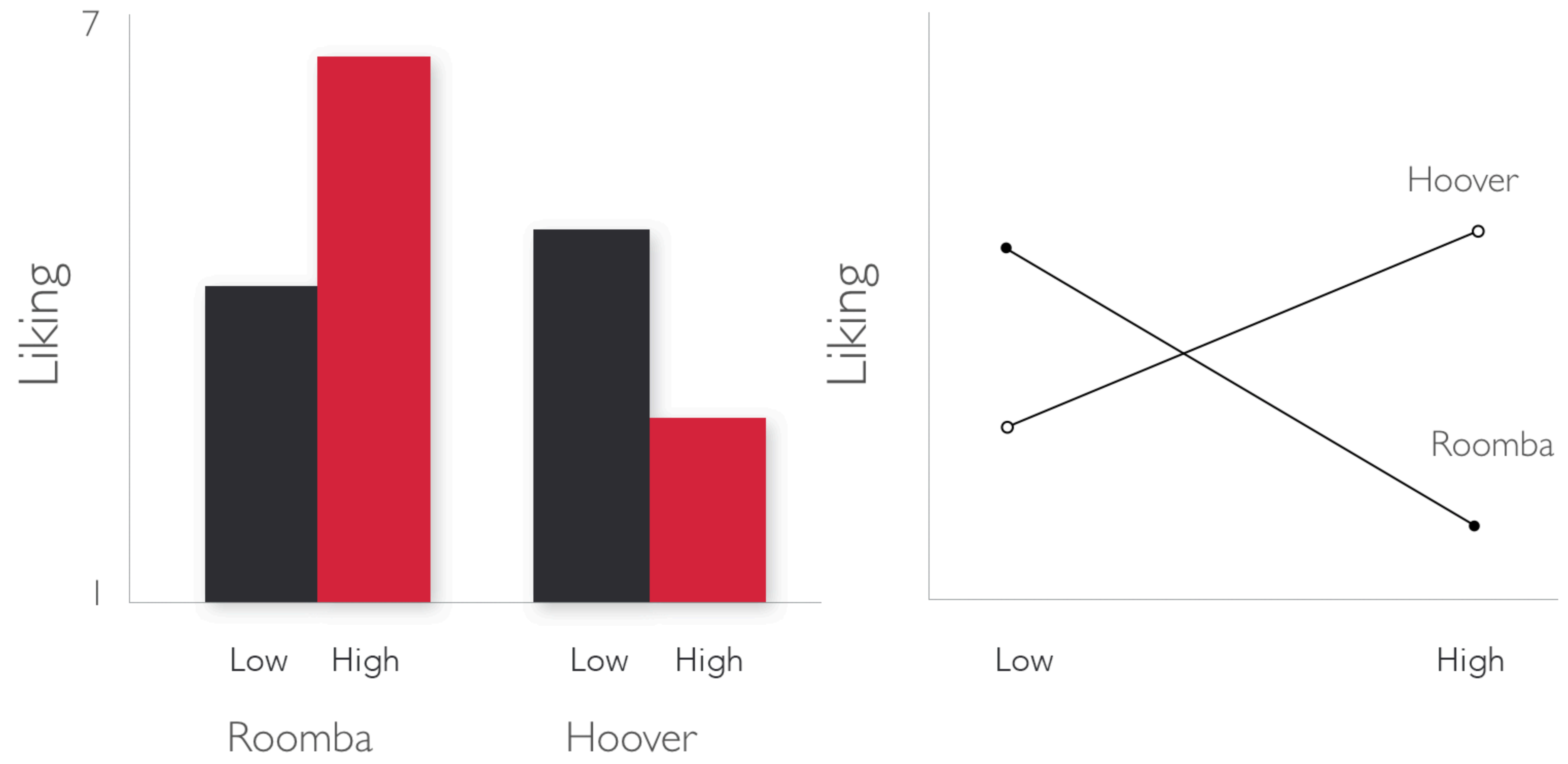
- low/high gaming × low/high tech savviness ↠ perceptions of robots

A **main effect** is the effect of a single factor on the response variable.





An **interaction effect** is the interaction among multiple factors over the response variable.



How do we achieve random assignment?

Random sampling and **assignment** are critical for experimental validity and generalizability. To minimize *selection bias* or *experimental bias*, we need to:

1. *Choose* participants randomly from the population

2. *Assign* participants randomly to experimental conditions

Methods for random assignment: *random assignment at arrival*, *counterbalancing*, *matching*

How do we assign randomly at arrival?

1 2 3 4 5 6 7 8 9 10, ten participants, *between-participants* design

7 4 8 6 9 1 2 10 3 5 ↠ **Level 1** [7 4 8 2 9], **Level 2** [1 6 10 3 5]

L2 L1 L2 L1 L2 L2 L1 L1 L1 L2

How do we balance order in within-participants designs?

Conterbalancing involves randomizing the order in which participants observe different levels across individuals to minimize *transfer effects*.

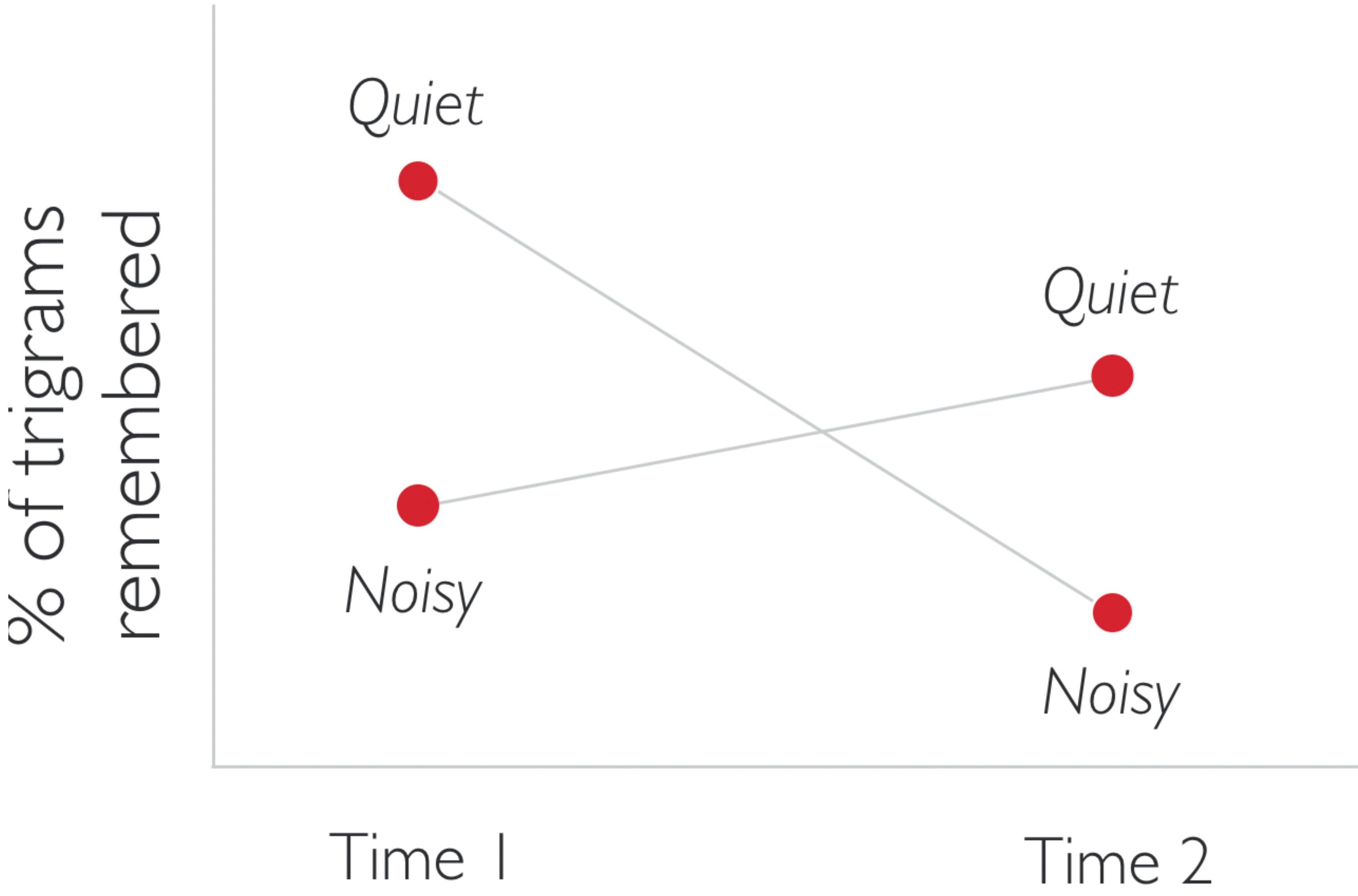
Transfer effects result from observing earlier levels affecting participant performance/behavior in later levels due to learning, fatigue, etc.

Transfer effects can be **linear**, where observing each level changes user behavior/performance to the same extent, or **non-linear**, where different levels have different effects on user behavior/performance (e.g., most of the learning taking place in the first trial).

Counterbalancing can address linear and non-linear transfer effects.

Counterbalanging cannot address **asymmetrical** transfer effects, where total magnitude of transfer effects are different across different orderings of levels (e.g., people sticking with a strategy they choose).

A *between-participants* design is a better fit to situations with asymetrical transfer effects.



How do we do counterbalancing?

1 2 3 4 5 6 7 8 9 10, ten participants, *within-participants* design

7 4 8 6 9 1 2 10 3 5 ↠ **L1 ↠ L2** [7 4 8 2 9], **L2 ↠ L1** [1 6 10 3 5]

L2 ↠ L1, L1 ↠ L2, L2 ↠ L1, L1 ↠ L2, L2 ↠ L1, L2 ↠ L1, L1 ↠ L2, L1 ↠ L2, L1 ↠ L2, L2 ↠ L1

This is an example of a process called **blocking** or **randomized block design** where the experimenter defines blocks and ensures random assignment of conditions for each block.

The **latin square design** is special type of blocking where unique ordering of levels appear once within each cell of an $n \times n$ block.[5]

2 × 2

A	B
B	A

3 × 3

A	B	C
B	C	A
C	A	B

4 × 4

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

(1)

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

(2)

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

(3)

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

(4)

5 × 5

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

(1)

5 × 5

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A

(2)

6 × 6

A	B	C	D	E	F
B	C	F	A	D	E
C	F	B	E	A	D
D	E	A	B	F	C
E	A	D	F	C	B
F	D	E	C	B	A

[5] Kirk, 2013, Experimental Design: Procedures for the Behavioral Sciences

How do we minimize experimental bias?

Blinding, where stakeholders are intentially kept naive to experimental conditions, ensures that knowledge of experimental conditions do not affect the behavior/performance of participants or experimenters.

Single-blind designs: when participants are not told what condition is being administered.

Double-blind designs: when participants or researchers are not told what condition is being administered.

Hands-on Activity: Reverse Engineer an Experiment

Choose an **experimental** paper:

- CHI 2023 (all types)
  - Search: CHI 2023,CHI 2022
- CSCW 2023 (more empirical)
- DIS 2023 (more design-based)
- UIST 2023 (more systems)

- 30-min activity to **reverse engineer an experiment**
  - Work in pairs
  - Submit completed paper handout to TA