Experiments

What are Experiments?

The Practice of Experimenting

- Observe

The Practice of Experimenting

- Observe
- Manipulate...

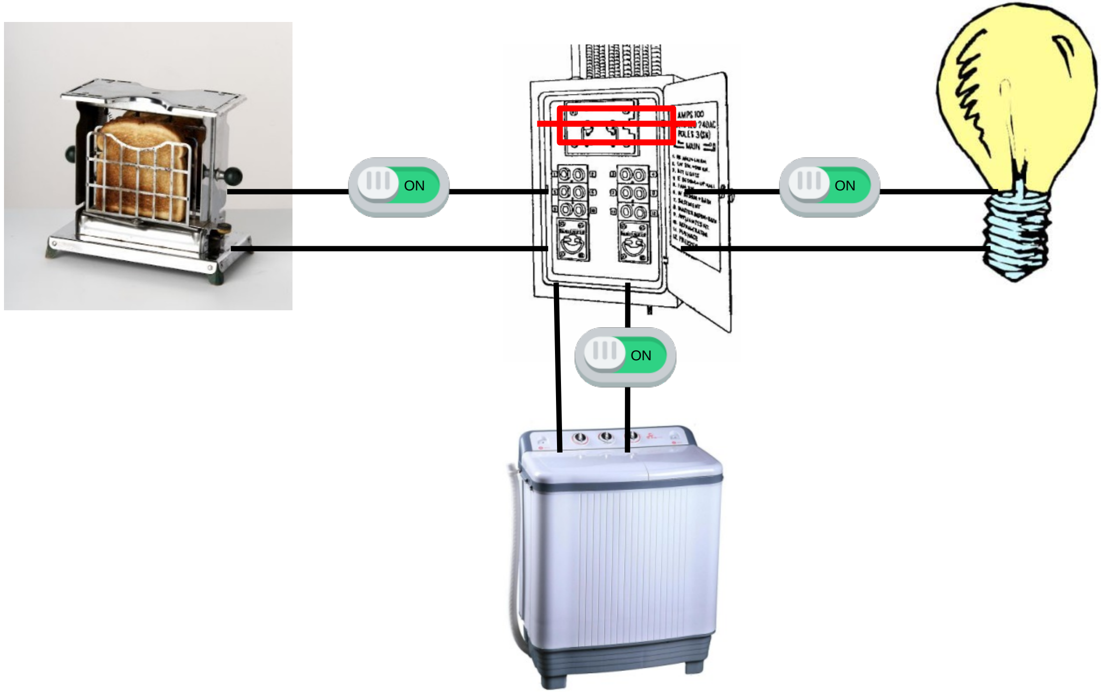
The Practice of Experimenting

- Observe
- Manipulate...

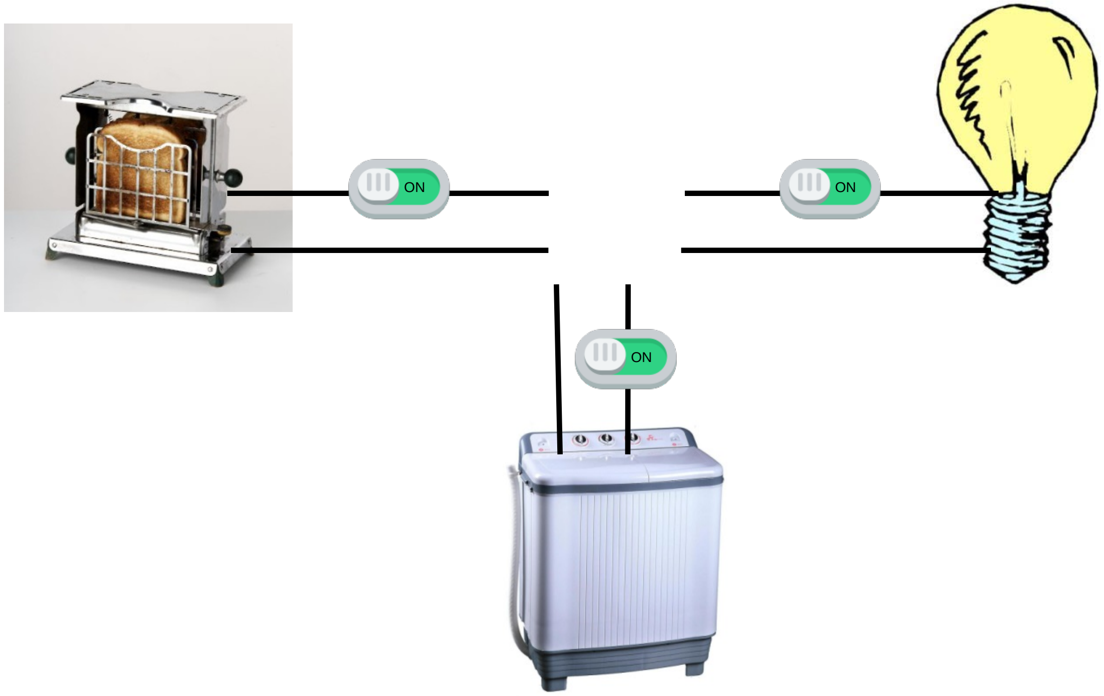
  ... in order to intervene

  ... in order to control

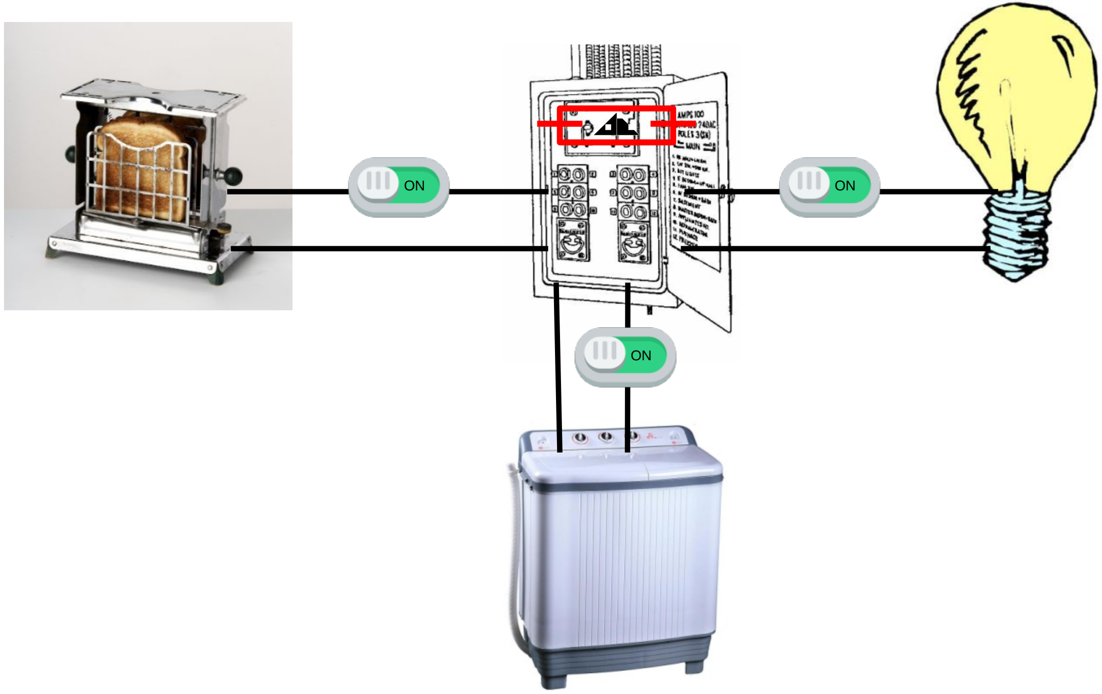
Observation



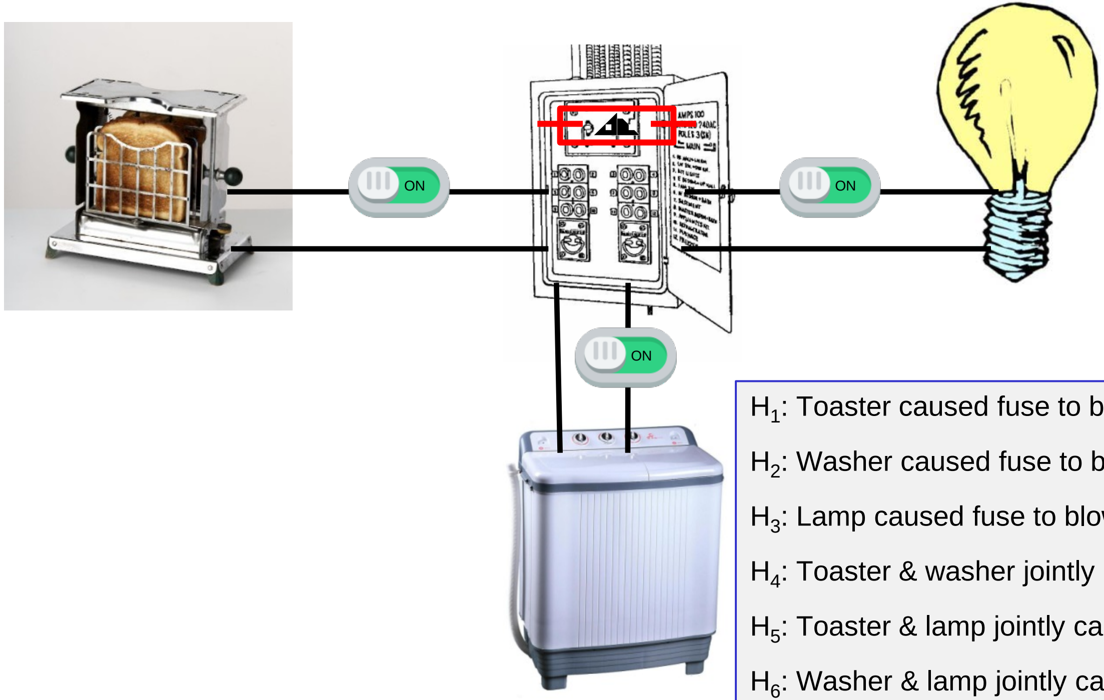
Observation



Observation



Causal Hypotheses



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

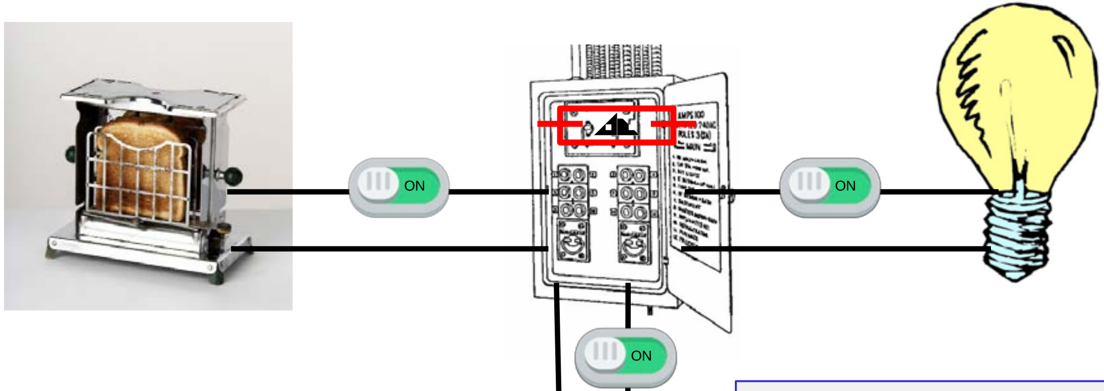
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

H_8: Other factors caused fuse to blow

Causal Hypotheses



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

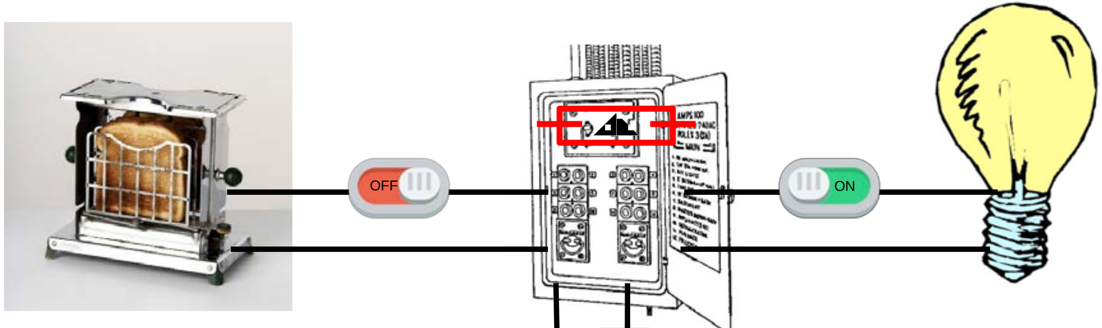
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

H_8: Other factors caused fuse to blow

Manipulation for Control



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

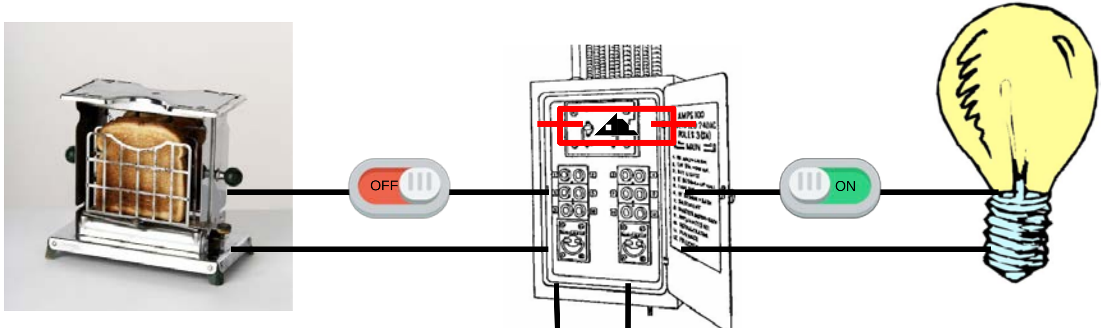
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

H_8: Other factors caused fuse to blow

Manipulation for Control



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

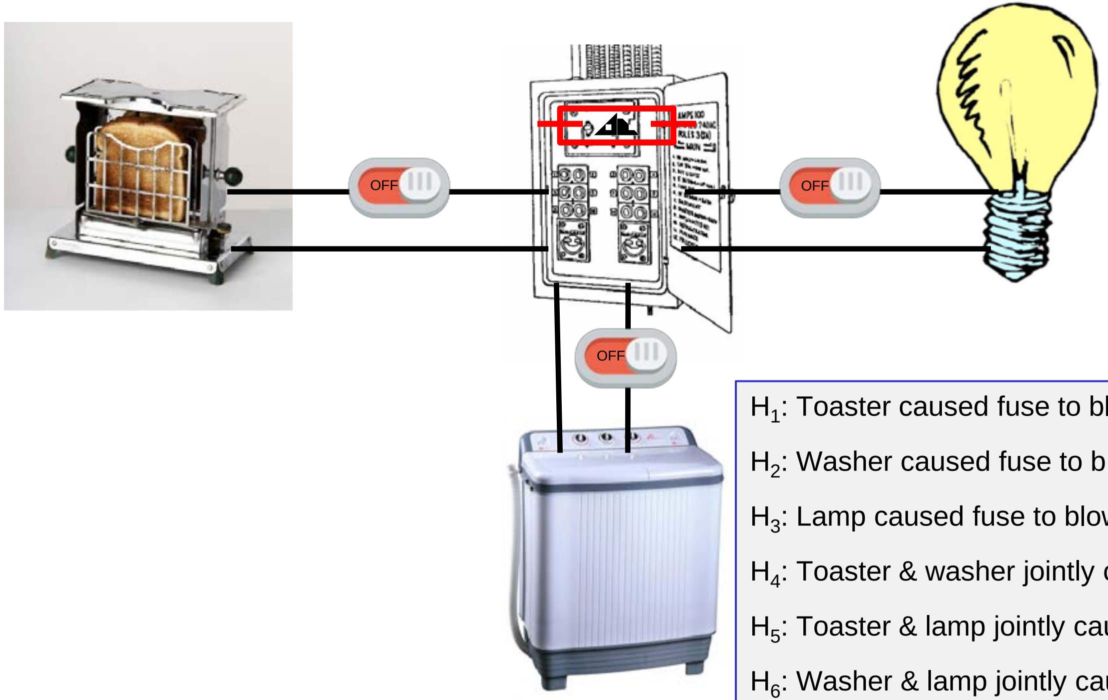
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

H_8: Other factors caused fuse to blow

Manipulation for Control



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

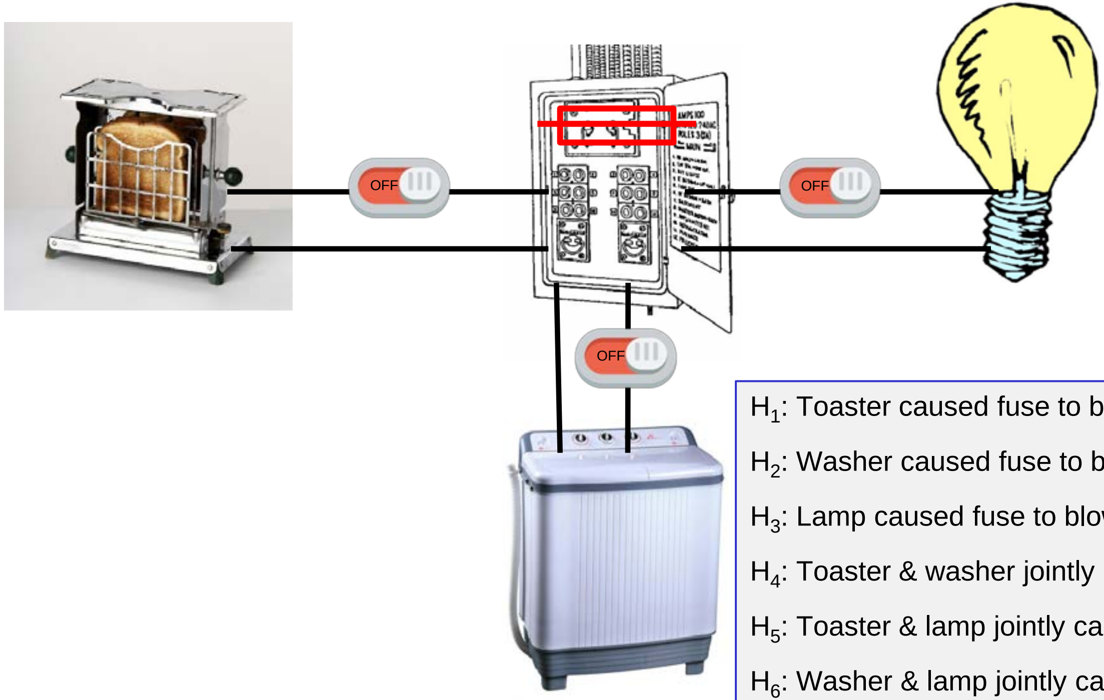
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

H_8: Other factors caused fuse to blow

Manipulation for Control



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

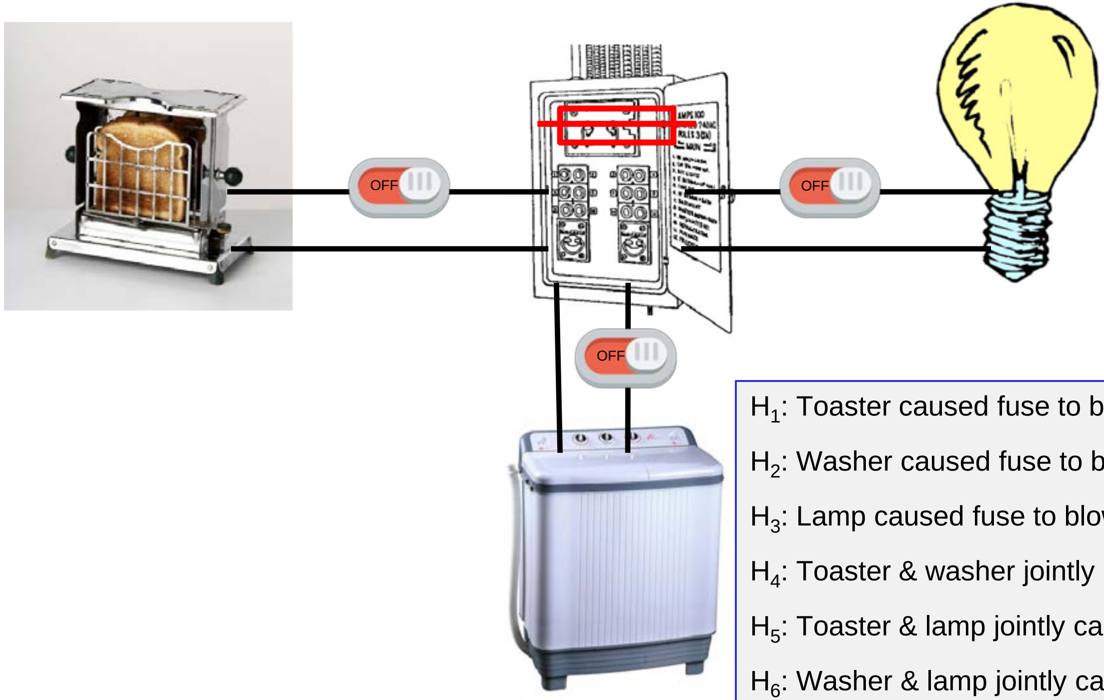
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

H_8: Other factors caused fuse to blow

Manipulation for Control



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

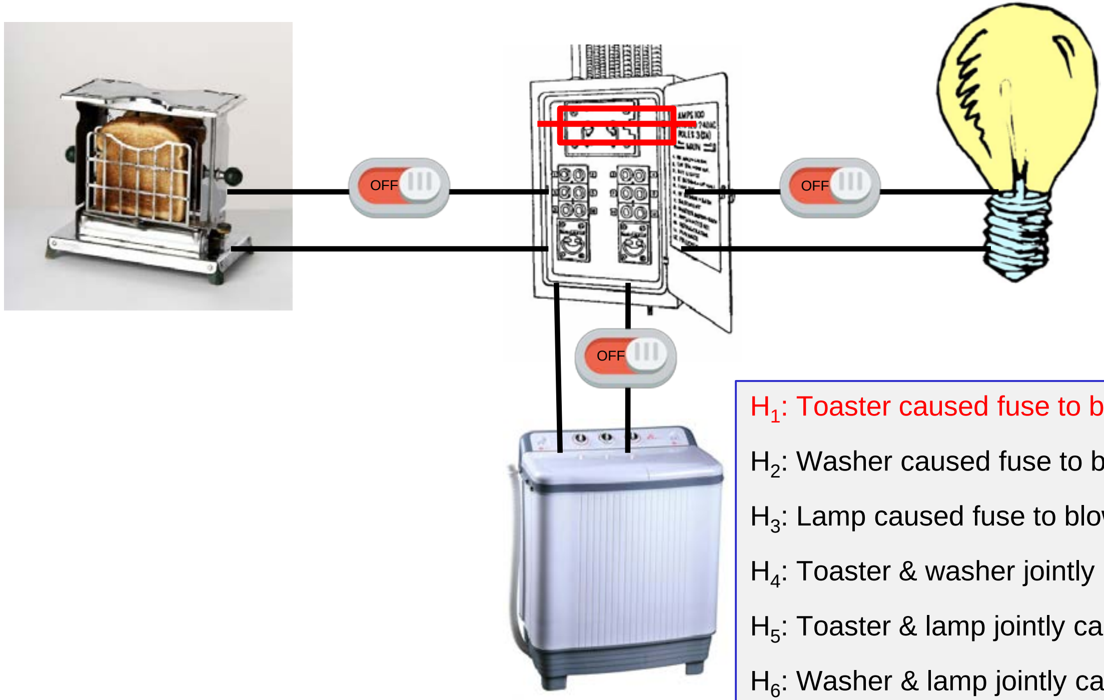
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

~~H_8: Other factors caused fuse to blow~~

Manipulation for Intervention



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

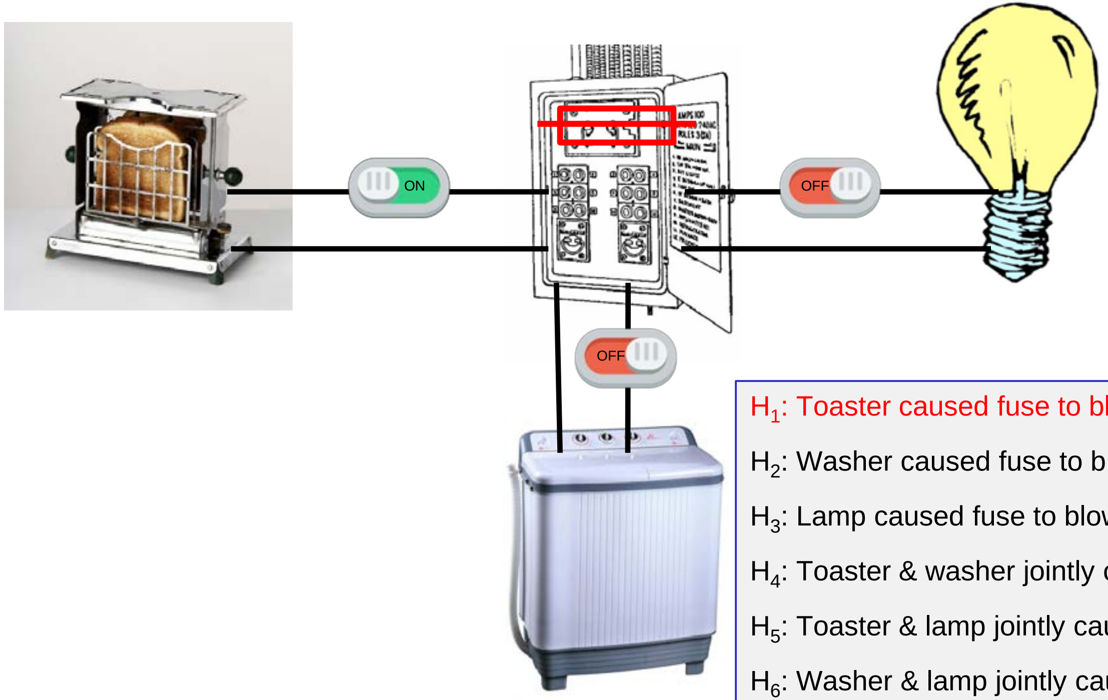
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

~~H_8: Other factors caused fuse to blow~~

Manipulation for Intervention



H_1: Toaster caused fuse to blow

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

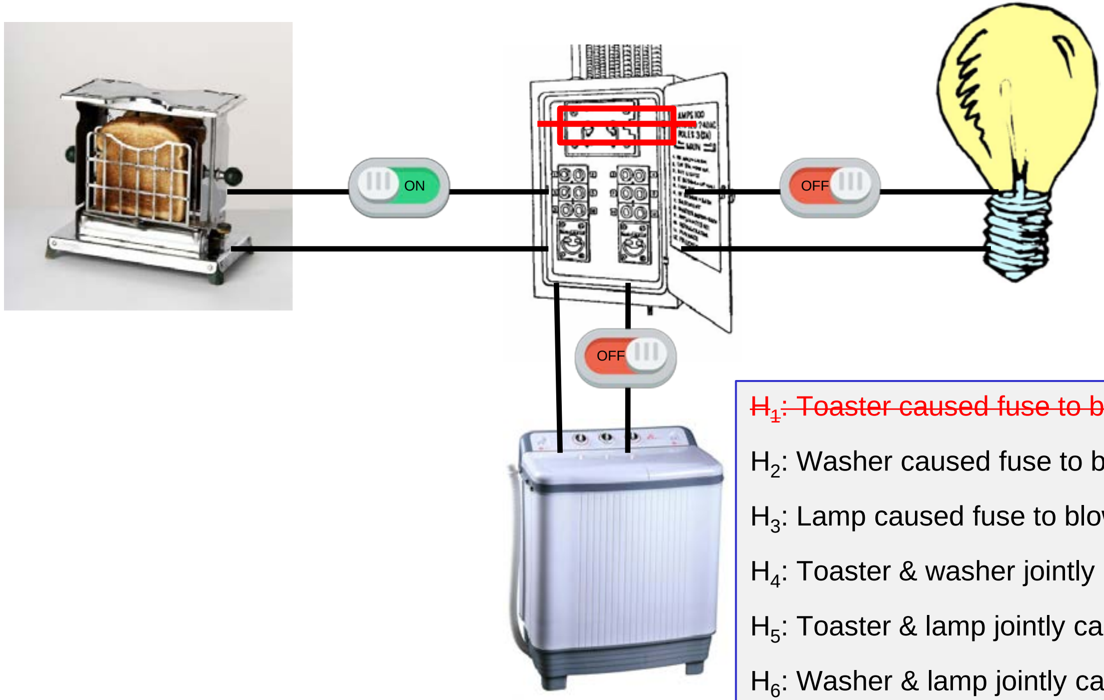
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

~~H_8: Other factors caused fuse to blow~~

Manipulation for Intervention



~~H_1: Toaster caused fuse to blow~~

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

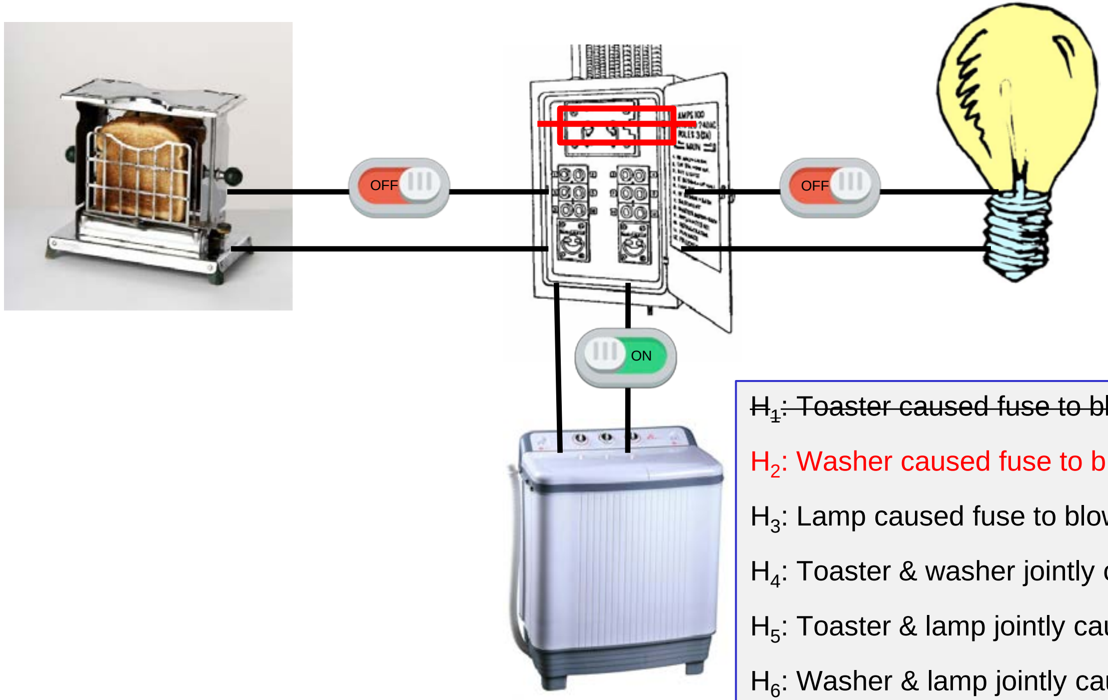
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

~~H_8: Other factors caused fuse to blow~~

Manipulation for Intervention



~~H_1: Toaster caused fuse to blow~~

H_2: Washer caused fuse to blow

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

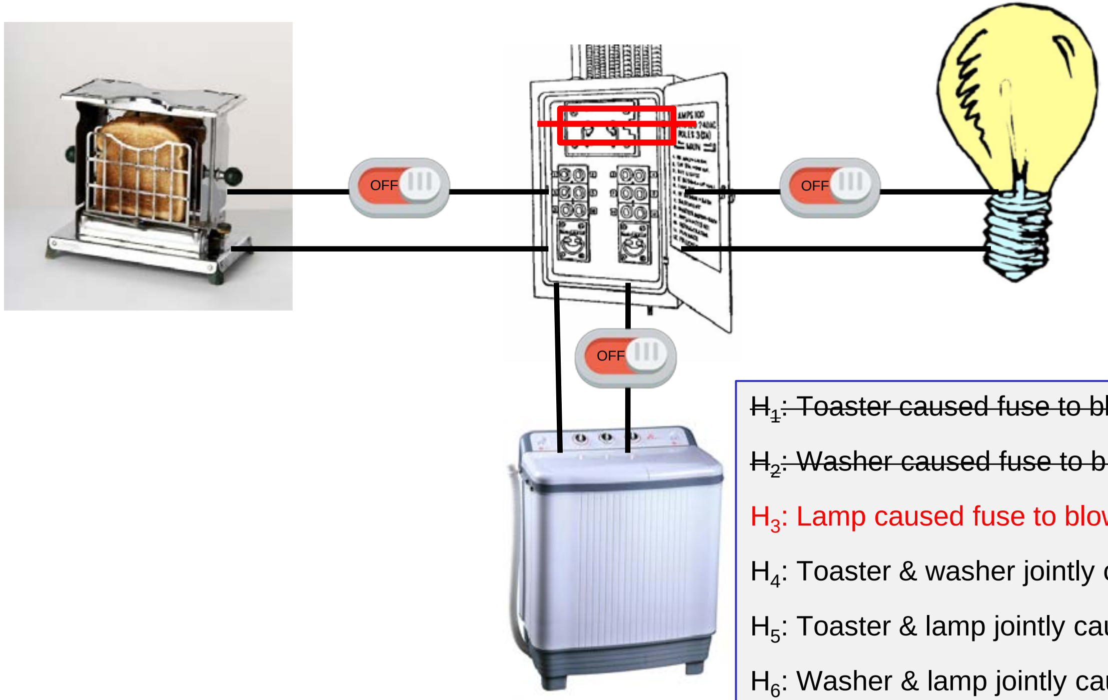
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

~~H_8: Other factors caused fuse to blow~~

Manipulation for Intervention



~~H_1: Toaster caused fuse to blow~~

~~H_2: Washer caused fuse to blow~~

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

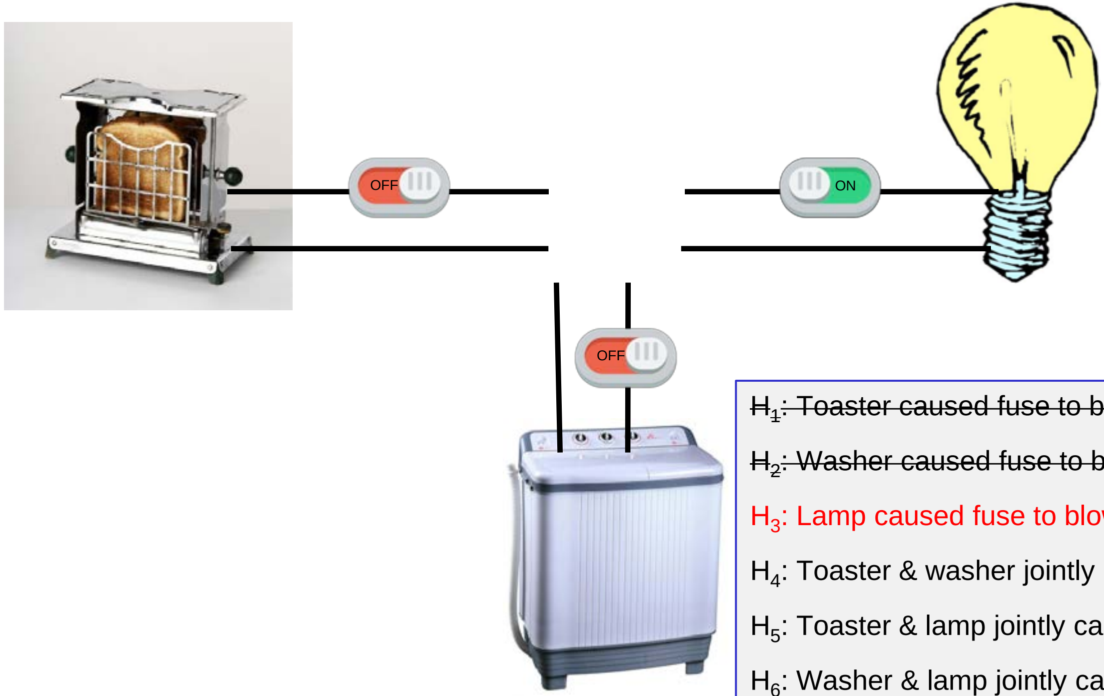
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

~~H_8: Other factors caused fuse to blow~~

Manipulation for Intervention



~~H_1: Toaster caused fuse to blow~~

~~H_2: Washer caused fuse to blow~~

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

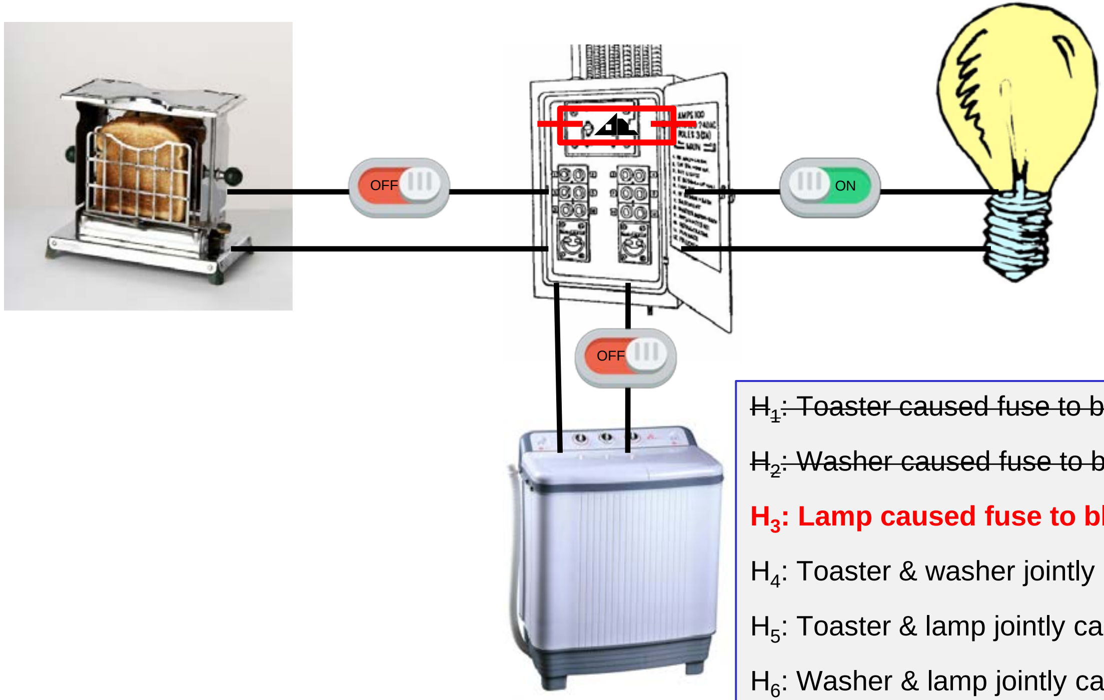
H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

~~H_8: Other factors caused fuse to blow~~

Manipulation for Intervention



~~H_1: Toaster caused fuse to blow~~

~~H_2: Washer caused fuse to blow~~

H_3: Lamp caused fuse to blow

H_4: Toaster & washer jointly caused fuse to blow

H_5: Toaster & lamp jointly caused fuse to blow

H_6: Washer & lamp jointly caused fuse to blow

H_7: All three appliances jointly caused fuse to blow

~~H_8: Other factors caused fuse to blow~~

Three Characteristic Features of Experiments

Experimenting is an observation process characterised by:

- **Control** of background variables through manipulation
- **Intervention** on target variable through manipulation
- **Observation of difference** produced by intervention

What Makes Experiments Special?

Mill’s Method of Difference

John Stuart Mill
(1806 –1873)

What Makes Experiments Special?

Mill’s Method of Difference

1. We ask: what causes phenomenon *E*?

What Makes Experiments Special?

Mill’s Method of Difference

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)

What Makes Experiments Special?

Mill’s Method of Difference

1. We ask: what causes phenomenon *E*?

2. We conjecture: *C* causes *E* (*hypothesis*)

3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)

What Makes Experiments Special?

Mill’s Method of Difference

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)
4. We activate *C* in S_1 but not in S_2. (*successful intervention)*

What Makes Experiments Special?

Mill's Method of Difference

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)
4. We activate *C* in S_1 but not in S_2. (*successful intervention)*
5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)

What Makes Experiments Special?

<u>Mill's Method of Difference</u>

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)
4. We activate *C* in S_1 but not in S_2. (*successful intervention)*
5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)
6. In S_1, something causes *E* (*from 5*)

What Makes Experiments Special?

Mill's Method of Difference

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)
4. We activate *C* in S_1 but not in S_2. (*successful intervention)*
5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)
6. In S_1, something causes *E* (*from 5*)
7. In S_2, nothing causes *E* (*from 3*)

What Makes Experiments Special?

Mill's Method of Difference

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)
4. We activate *C* in S_1 but not in S_2. (*successful intervention)*
5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)
6. In S_1, something causes *E* (*from 5*)
7. In S_2, nothing causes *E* (*from 3*)
8. The only differences between S_1 and S_2 are *C* and *E* (*from 3 and 4*)

What Makes Experiments Special?

Mill's Method of Difference

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)
4. We activate *C* in S_1 but not in S_2. (*successful intervention)*
5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)
6. In S_1, something causes *E* (*from 5*)
7. In S_2, nothing causes *E* (*from 3*)
8. The only differences between S_1 and S_2 are *C* and *E* (*from 3 and 4*)

Conclusion: *C* causes *E*

Potential Sources of Error

Mill's Method of Difference

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)
4. We activate *C* in S_1 but not in S_2. (*successful intervention)*
5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)
6. In S_1, something causes *E* (*from 5*)
7. In S_2, nothing causes *E* (*from 3*)
8. The only differences between S_1 and S_2 are *C* and *E* (*from 3 and 4*)

Conclusion: *C* causes *E*

Potential Sources of Error

<u>Mill's Method of Difference</u>

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)
4. We activate *C* in S_1 but not in S_2. (*successful intervention*)
5. We observe that *E* occurs i ... *ation of differences*)
6. In S_1, something causes *E* (
7. In S_2, nothing causes *E* (*from 3*)
8. The only differences between S_1 and S_2 are *C* and *E* (*from 3 and 4*)

Conclusion: *C* causes *E*

We might fail to control all causally relevant factors

Potential Sources of Error

<u>Mill's Method of Difference</u>

1. We ask: what causes phenomenon *E*?

2. … *E* (*hypothesis*)

We might fail to activate C in S_1, or inadvertently activate it in S_2

3. … and S_2, in which neither *C* nor *E* occur and in which all causally relevant factors are the same. (*successful control*)

4. We activate *C* in S_1 but not in S_2. (*successful intervention*)

5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)

6. In S_1, something causes *E* (*from 5*)

7. In S_2, nothing causes *E* (*from 3*)

8. The only differences between S_1 and S_2 are *C* and *E* (*from 3 and 4*)

Conclusion: *C* causes *E*

Potential Sources of Error

<u>Mill's Method of Difference</u>

1. We ask: what causes phenomenon *E*?

2. We conjecture: *C* causes *E* (*hypothesis*)

3. We p $_2$, in which neither *C* nor *E* occur and in whic re the same. (*successful control*)

We might be wrong in observing a difference between S_1 and S_2

4. We activate *C* in S_1 but not in S_2. (*successful intervention*)

5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)

6. In S_1, something causes *E* (*from 5*)

7. In S_2, nothing causes *E* (*from 3*)

8. The only differences between S_1 and S_2 are *C* and *E* (*from 3 and 4*)

Conclusion: *C* causes *E*

Potential Sources of Error

<u>Mill's Method of Difference</u>

1. We ask: what causes phenomenon *E*?
2. We conjecture: *C* causes *E* (*hypothesis*)
3. We produce two situations S_1 and S_2, in which neither *C* nor *E* occur and in which all causally relevant
4. We activate *C* in S_1 but not
5. We observe that *E* occurs in S_1 but not in S_2. (*observation of differences*)
6. In S_1, something causes *E* (*from 5*)
7. In S_2, nothing causes *E* (*from 3*)
8. The only differences between S_1 and S_2 are *C* and *E* (*from 3 and 4*)

Conclusion: *C* causes *E*

We assume that *E* is caused, or that it is part of a deterministic system, or that it doesn't cause itself

Summary

- Experiments are observational processes
- Experiments are characterised by manipulation, intervention, control and observation
- Experimental observations often offer a justification for accepting or rejecting a hypothesis
- The argument from experimental observation to causal claims is fallible: it might contain errors
- In order to minimize these potential errors we need to design experiments carefully

Distinguishing Experiments from other Empirical Practices

Non-Experimental Empirical Practices

Observational studies


Edwin P. Hubble (1889–1953)

Non-Experimental Empirical Practices

Observational studies

worker	w wage	h_1 age	h_2 education	x_1 job type	p risk of injury	q risk of death
1	$3500	46	K-12	skilled	0.2	0.007
2	$2775	28	K-8	office	0.05	0.0003
3	$3000	32	B.S.	office	0.09	0.0002
…	…	…	…	…	…	…
n	$2100	31	K-12	unskilled	0.4	0.01

Observation of *n* workers – more specifically of 6 properties of these workers

Non-Experimental Empirical Practices

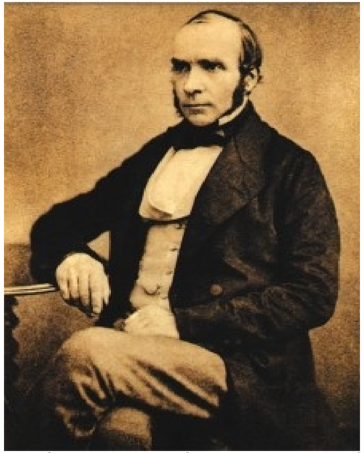
Observational studies

- no manipulation
- no intervention on target variable
- no control of background variable

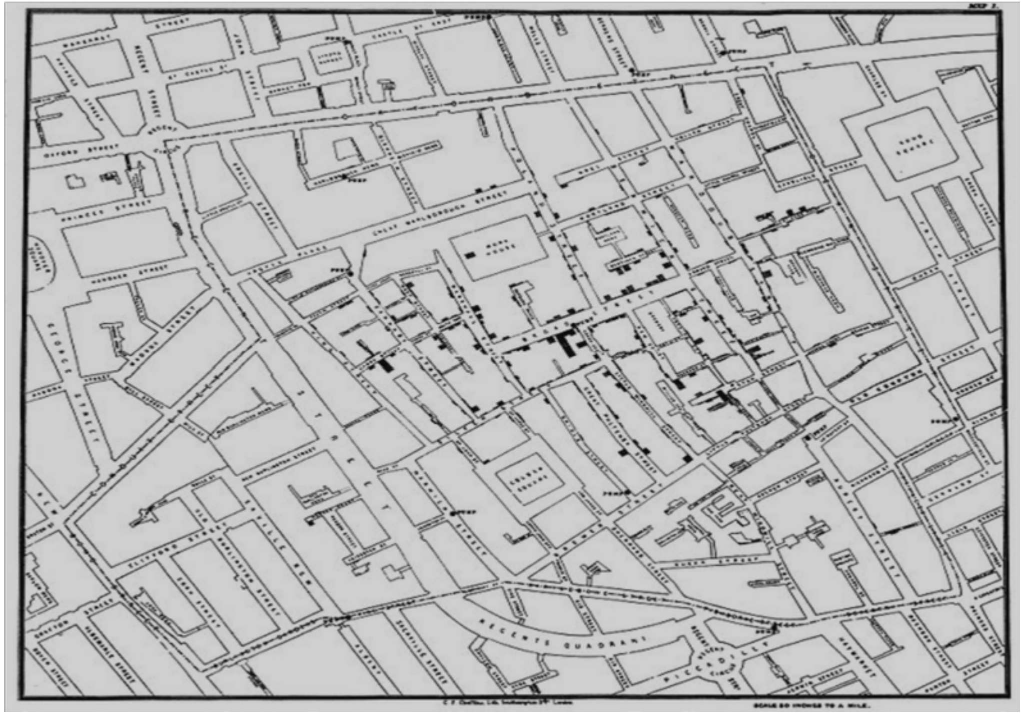
Note: this relates to the example in the lecture. There can be control in observational studies, see course literature.

Non-Experimental Empirical Practices

Natural experiments



John Snow (1813 –1858)



Non-Experimental Empirical Practices

Natural experiments

- no manipulation
- no intervention on target variable
- control of background variable (not achieved through manipulation)

Field Experiments

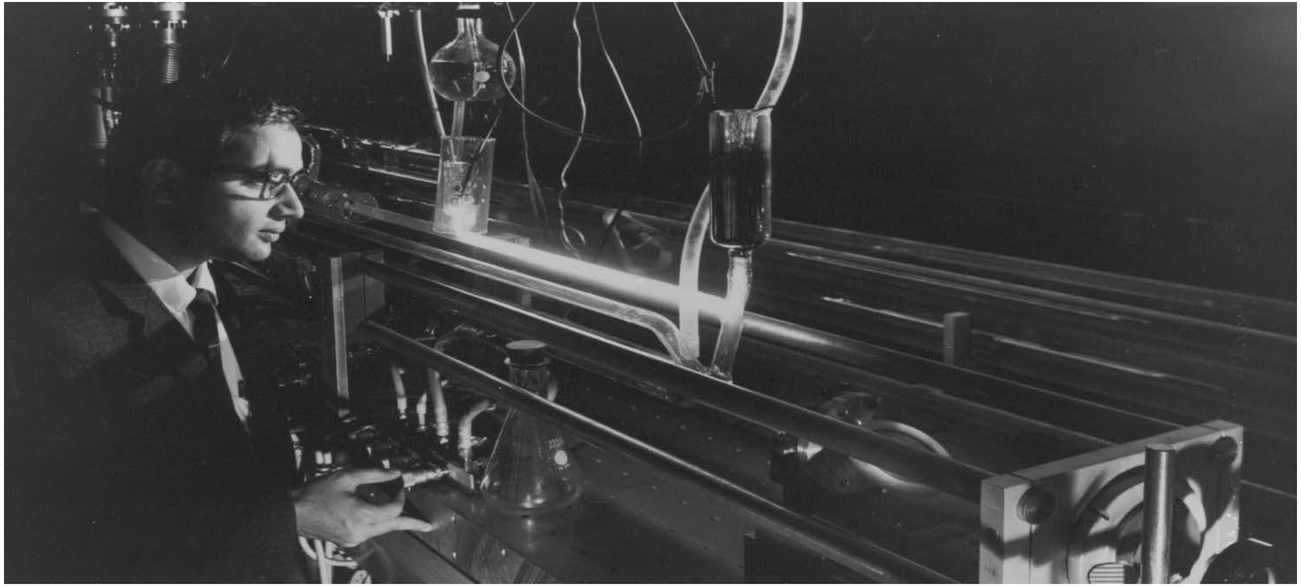
Maintain the same **found** conditions in both treatment and control group, varying only the intervention on the factor of interest.

Experimental Empirical Practices

Field experiments

- manipulation
- intervention on target variable
- control of background variable (not achieved through manipulation)

Laboratory Experiments



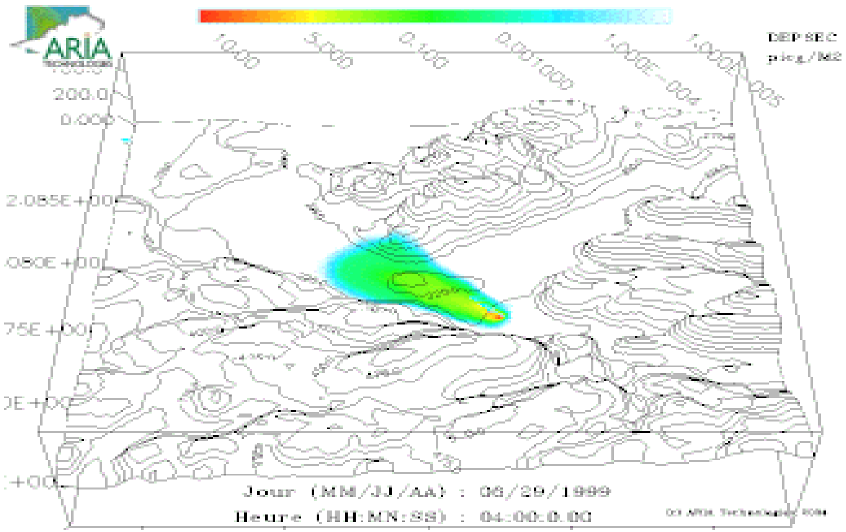
Construct the same background conditions in both treatment and control group (through manipulating the environment), varying only the intervention on the factor of interest.

Experimental Empirical Practices

Laboratory experiments

- manipulation
- intervention on target variable
- control of background variable, achieved through manipulation

Simulation Experiments



Construct a **representation** of a real system on a computer, and then perform various interventions on that representation

Non-Experimental Empirical Practices

Simulation studies

- Manipulation of representations, not real variables
- intervention on representation
- control of background variable representations

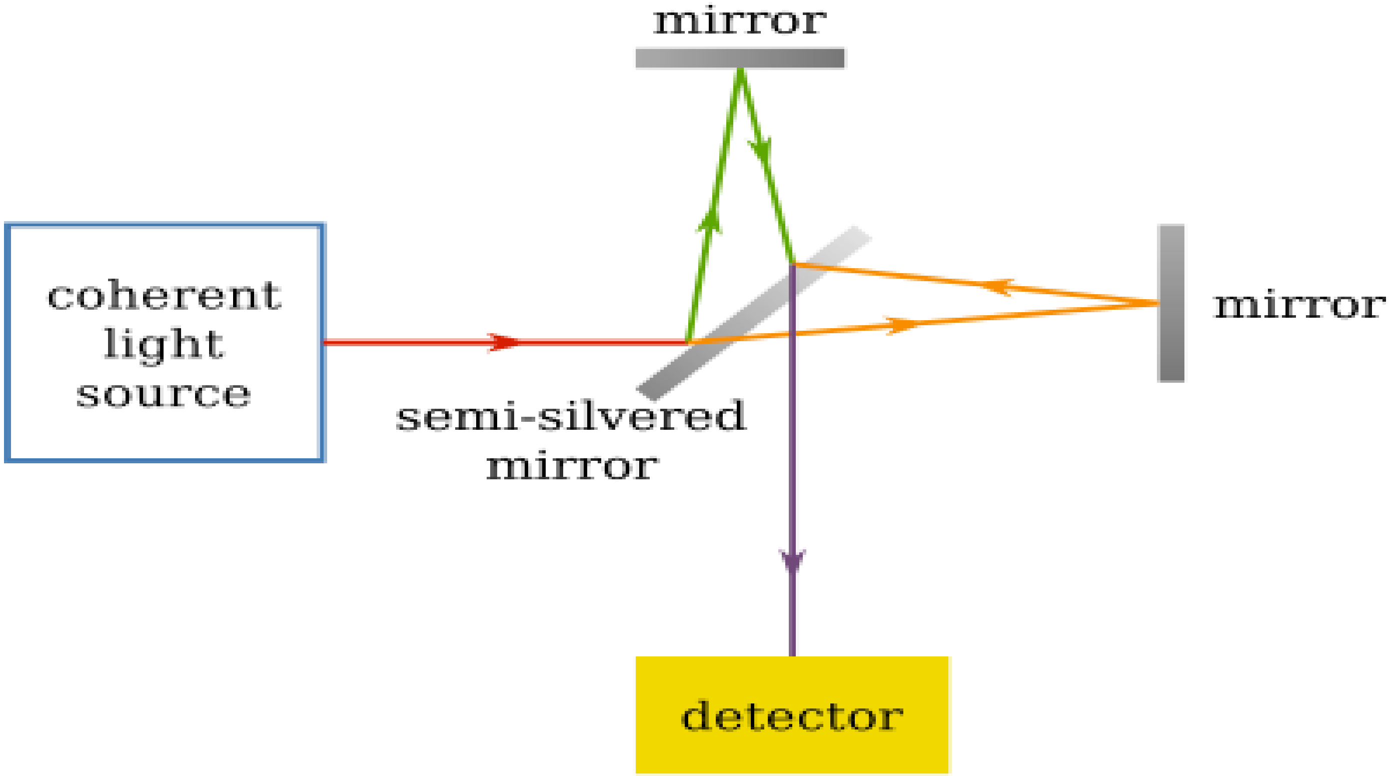
Experiments: A Definition

An experiment is a *controlled observation* in which the observer *manipulates* the *real variables* that are believed to influence the outcome, both for the purpose of intervention and control.

Errors in Experiment

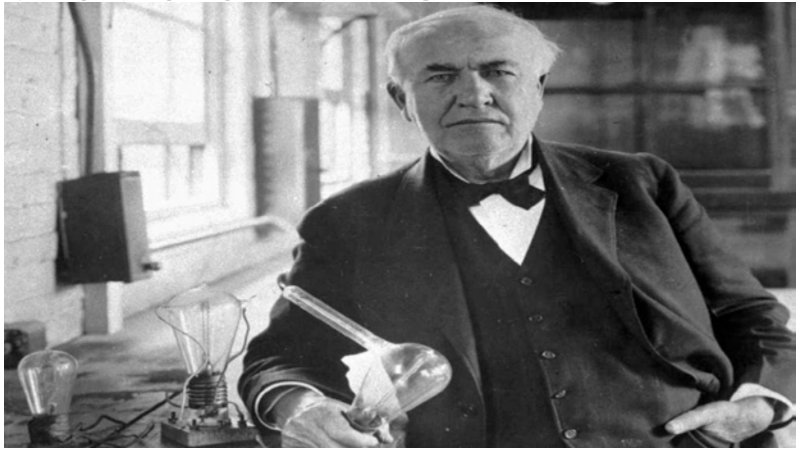
Failed Experiments?

The Michelson–Morley experiment (1887)



Productive “Failures”

“I haven’t failed. I’ve found 10,000 ways that didn’t work.”



Thomas Edison (1847 –1931)

Non-productive Failures

- Experimental observations are powerful evidence *only if* the process is designed correctly

Non-productive Failures

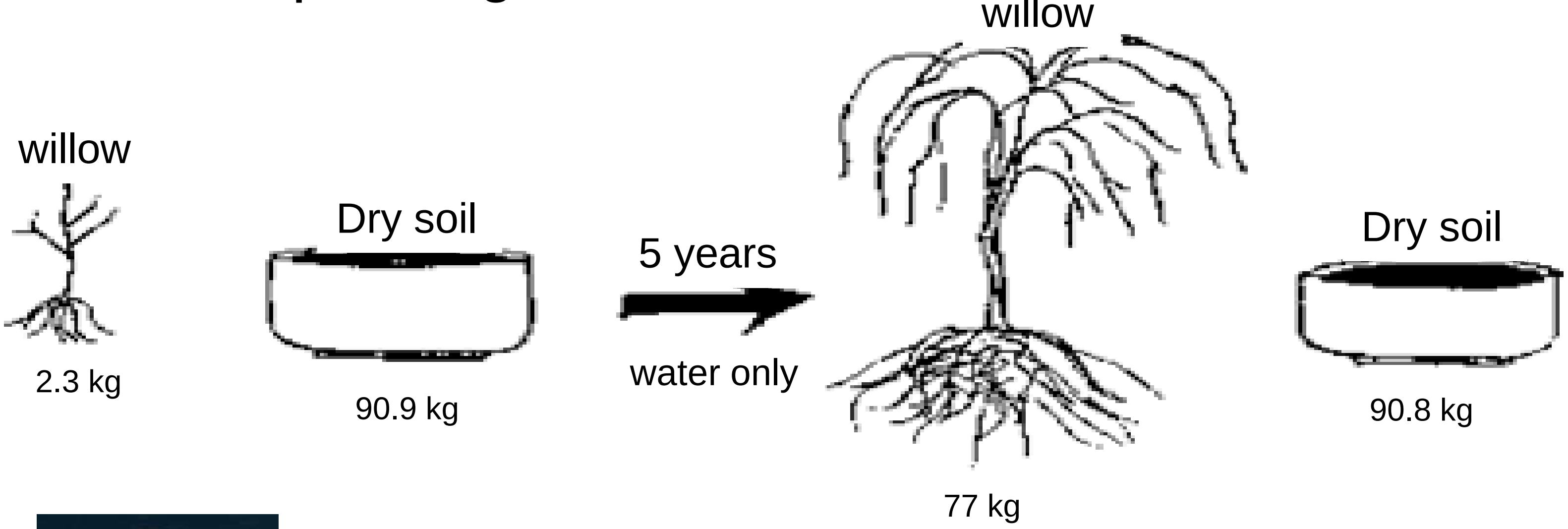
- Experimental observations are powerful evidence *only if* the process is designed correctly
- Mill’s *Method of Difference*: justifies conclusion *only if* assumptions about process are correct.

Non-productive Failures

- Experimental observations are powerful evidence *only if* the process is designed correctly
- Mill's *Method of Difference*: justifies conclusion *only if* assumptions about process are correct.
- Most important: *all* relevant factors are *controlled*

Failing to Control for a Relevant Factor

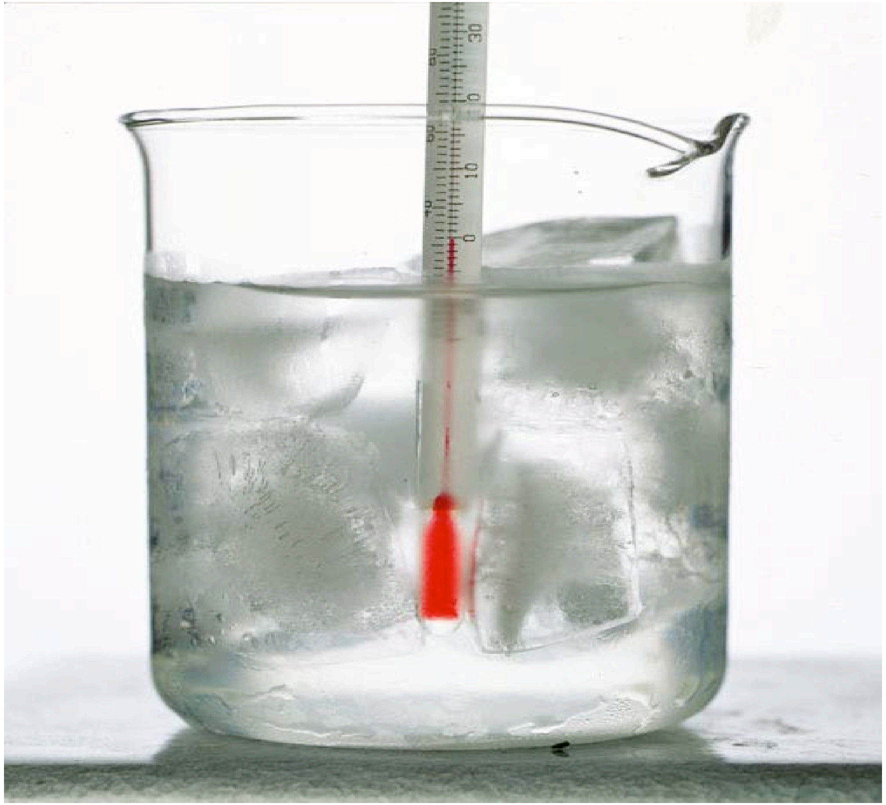
Where do plants get their mass?



Jan Baptist van Helmont
(1580 – 1644)

Failing to Control for a Relevant Factor

Observer Effect



Failing to Control for a Relevant Factor

<u>Confirmation Bias</u>

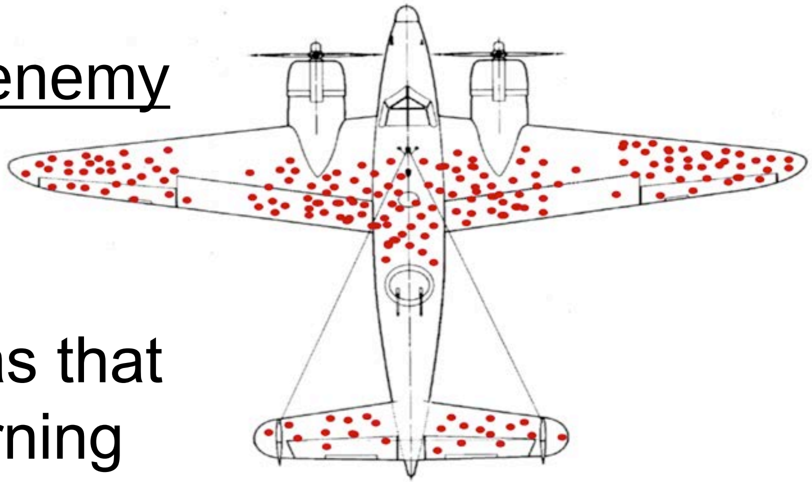
Failing to Control for a Relevant Factor

Placebo Effect

Selection Bias

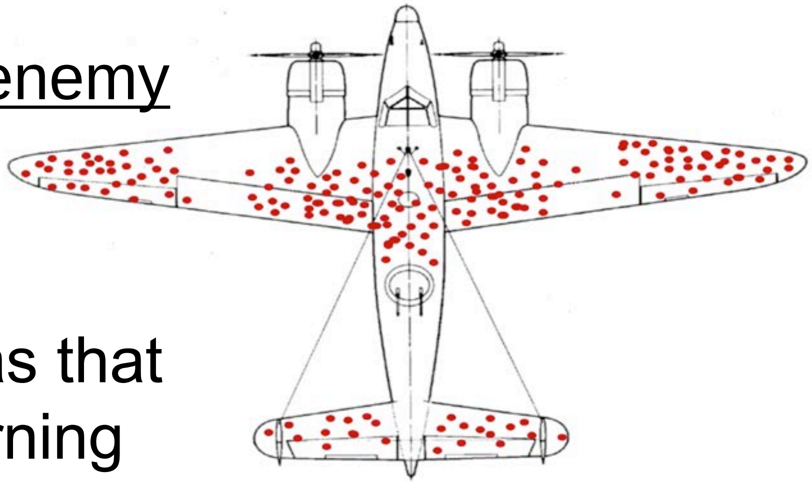
How to minimize bomber loss to enemy fire?

Answer 1: add armour to the areas that showed the *most* damage in returning aircraft



Damage density on returning aircraft

Selection Bias



How to minimize bomber loss to enemy fire?

Answer 1: add armour to the areas that showed the *most* damage in returning aircraft

Answer 2: add amour to areas that showed the *least* damage in returning aircraft – because airplanes hit there do *not* return!

Abraham Wald (1902 – 1950)

Internal Validity

The above errors compromise the process of experimentation:

- Control of background variable
- Intervention *only* on target variable
- Correct observation

Internal Validity

The above errors compromise the process of experimentation:

- Control of background variable
- Intervention *only* on target variable
- Correct observation

Only if such errors are absent can an experimental observation justify accepting or rejecting a hypothesis

In that case: conclusion from experimental observation are ***internally valid***

Detecting Experimental Errors

Apply prior knowledge

- Theoretical knowledge
- Tacit experimenter’s knowledge

Detecting Experimental Errors

Investigate previously performed experiments

- Repetition

  An experiment is **repeatable** if there is enough information available about an experiment so that a competent person other than the original experimenter can repeat the experimental procedures

- Reproduction

  An experimental result is **reproduced** if a competent repetition of the original experiment yields the same result

- Replication

  An experimental result is **replicated** if a competent independent experiment, in the spirit of the original experiment, with independent data, analytical methods, laboratories, and instruments, yields the same result

How to implement experimental control?

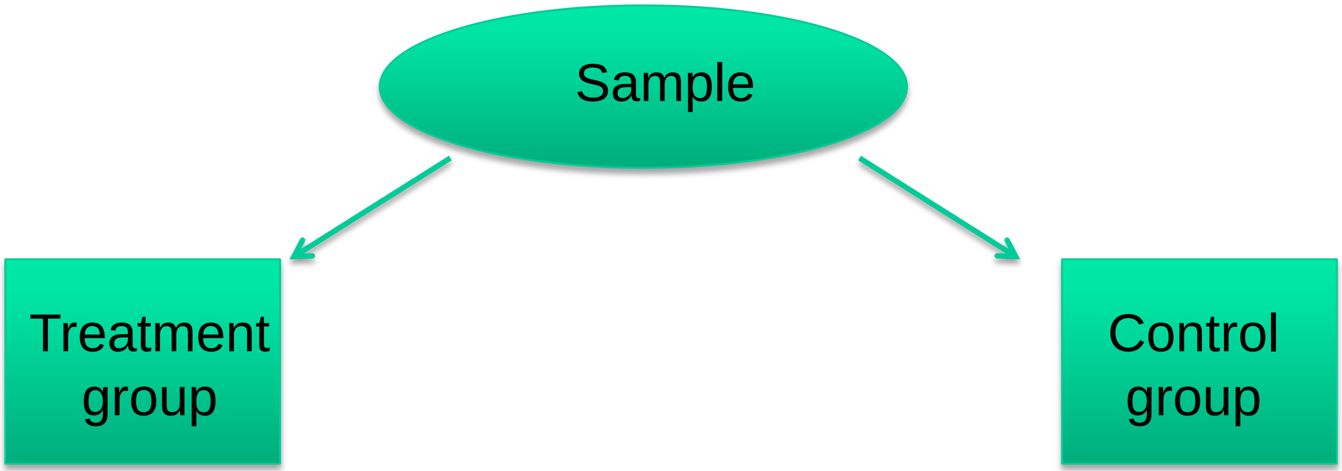
Internal Validity

Experimental Control consists in

- accurately *identifying* the features that are relevant for an experimental result, and also
- in being able to *influence* these features in such a way that alternative explanations of the experimental result can be ruled out.

How to Influence/Control Relevant Features?

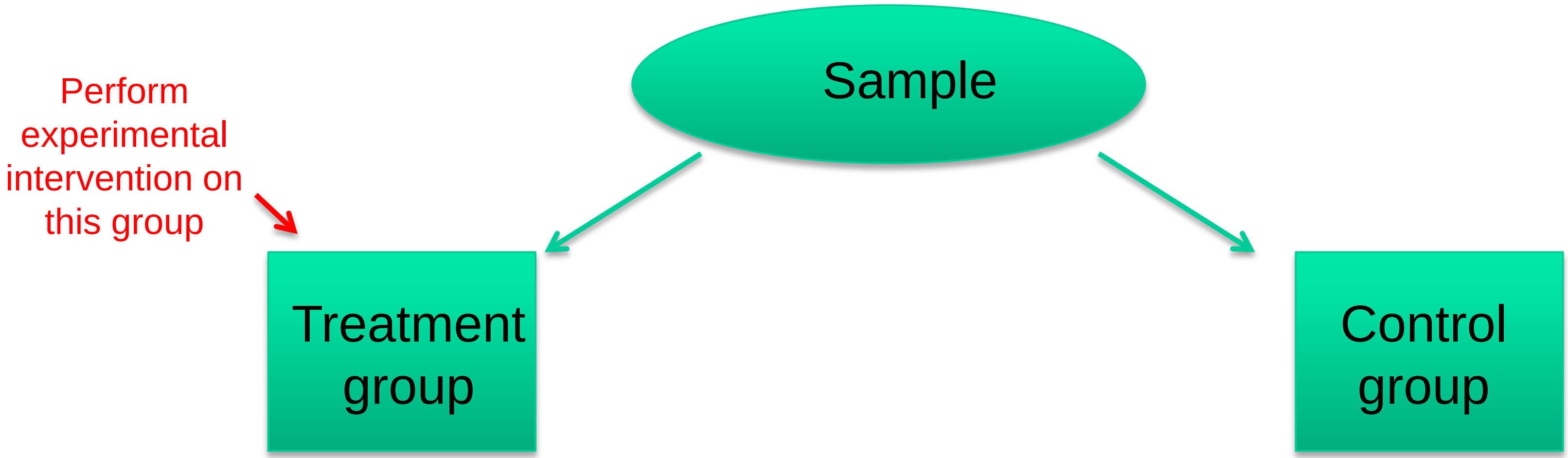
1. Divide experimental subjects/objects into ***treatment*** *and* ***control group***



[Beginning at 0:42]

How to Influence/Control Relevant Features?

1. Divide experimental subjects/objects into ***treatment*** *and* ***control group***



[Beginning at 0:52]

A before/after quasi-experiment

last year

Change
lighting design

this year

time

Sample 1

Sample 2

TEST: do students in sample 2 perform
better than in sample 1?

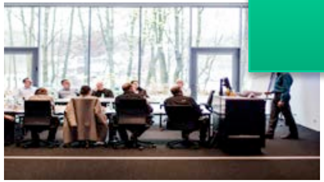
[Beginning at 1:50]

Using treatment & control groups

Sample

Treatment group

Control group



Change
lighting design

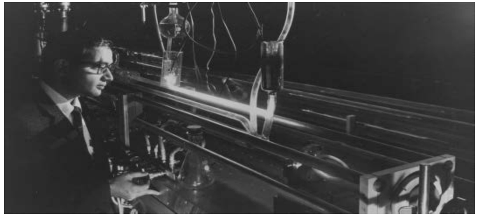
Don’t change
lighting design

[Beginning at 2:20]

How to Influence/Control Relevant Features?

2. Holding things constant

- *Finding* situations with same background variables
- *Constructing* situations with same background variables



How to Influence/Control Relevant Features?

3. Eliminate disturbing factors

[Beginning at 3:18]

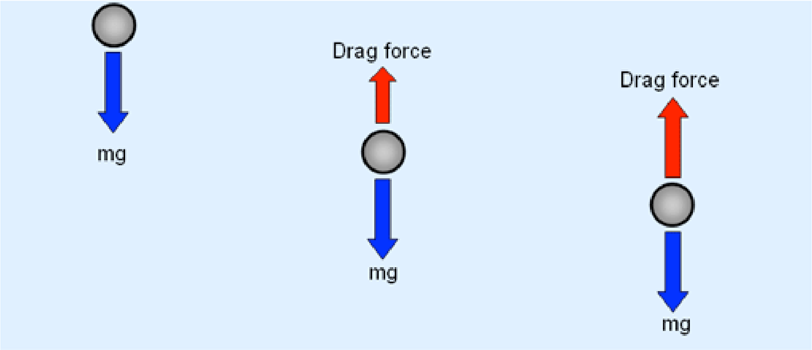
How to Influence/Control Relevant Features?

3. Eliminate disturbing factors

[Beginning at 3:24]

How to Influence/Control Relevant Features?

3. Eliminate disturbing factors



Body released from rest

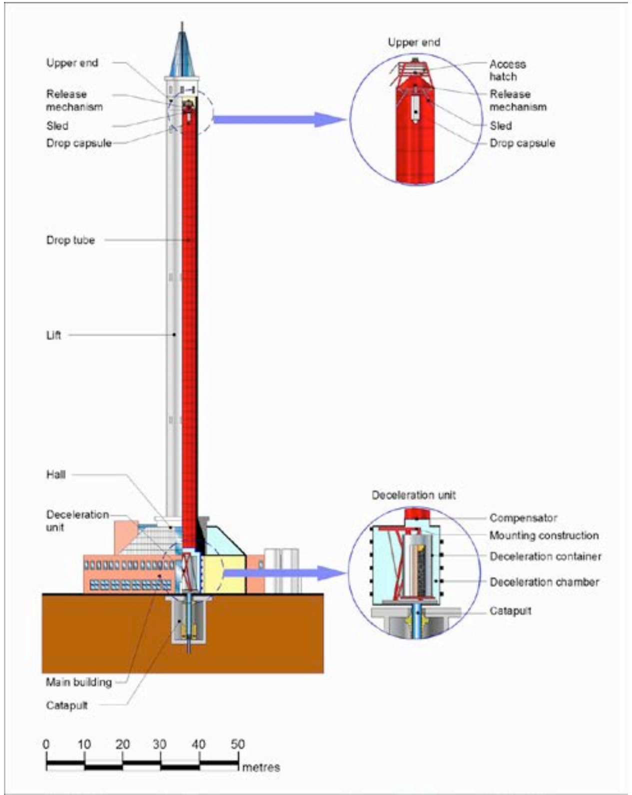
Forces on body during acceleration

Forces on body at terminal velocity

[Beginning at 3:24]

How to Influence/Control Relevant Features?

3. Eliminate disturbing factors



[Beginning at 3:39]

How to Influence/Control Relevant Features?

3. Eliminate disturbing factors

- Vacuum
- Faraday cage
- Outer space

[Beginning at 3:56]

How to Influence/Control Relevant Features?

Special case of elimination: ***Blinding***

[Beginning at 4:16]

How to Influence/Control Relevant Features?

Special case of elimination: ***Blinding***

- Single Blinding: Subjects do not know which treatment they are getting

[Beginning at 4:55]

How to Influence/Control Relevant Features?

Special case of elimination: ***Blinding***

- Single Blinding: Subjects do not know which treatment they are getting
- Double Blinding: Experimenters do not know which treatments are administered to what subjects

[Beginning at 5:58]

How to Influence/Control Relevant Features?

4. Separating factors

[Beginning at 6:35]

How to Influence/Control Relevant Features?

4. Separating factors

- Gravity Probe A experiment

Clock 1 stays on earth

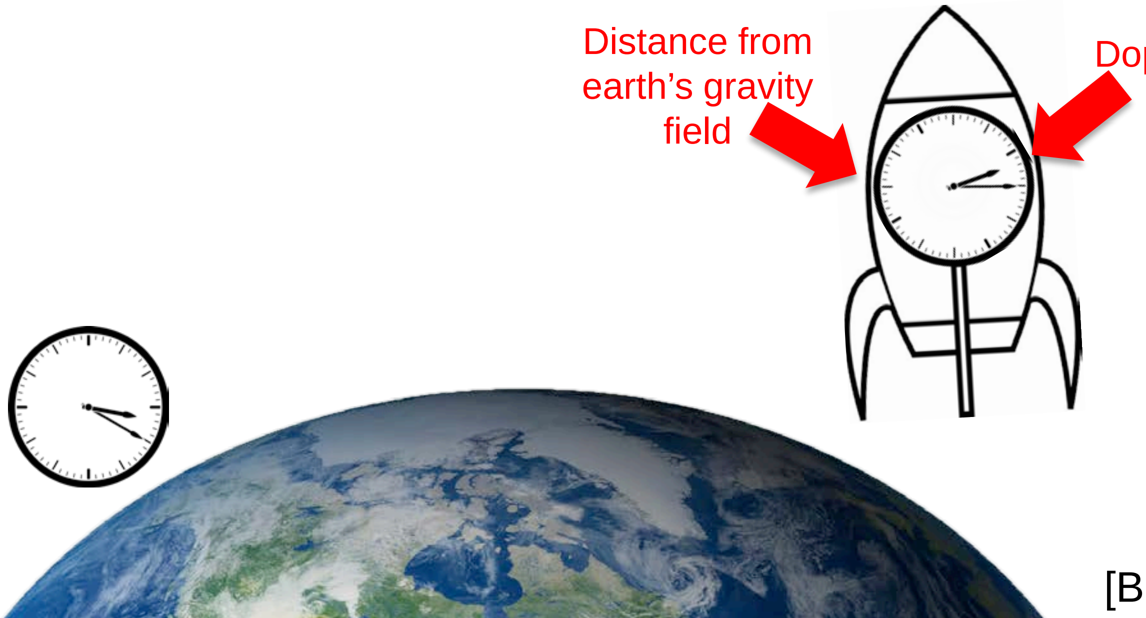
Clock 2 is launched into space

[Beginning at 6:50]

How to Influence/Control Relevant Features?

4. Separating factors

- Gravity Probe A experiment



[Beginning at 7:00]

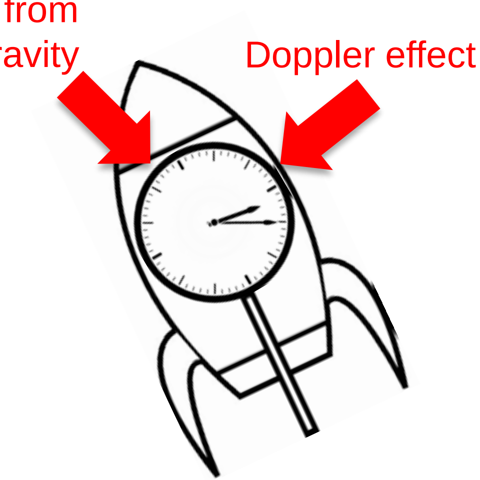
How to Influence/Control Relevant Features?

4. Separating factors

- Gravity Probe A experiment

Distance from earth's gravity field

Doppler effect



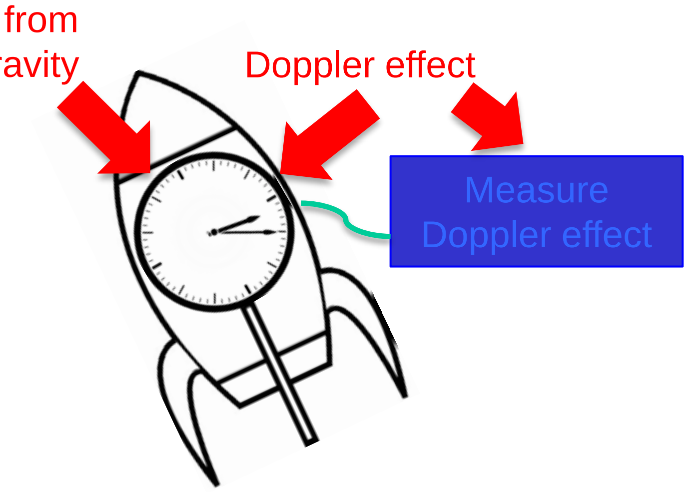
Clock 1 stays on earth

[Beginning at 7:03]

How to Influence/Control Relevant Features?

4. Separating factors

- Gravity Probe A experiment



[Beginning at 7:28]

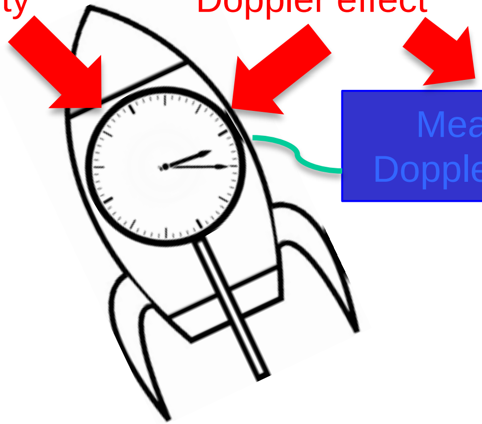
How to Influence/Control Relevant Features?

4. Separating factors

- Gravity Probe A experiment

Distance from earth's gravity field

Doppler effect



Measure Doppler effect

Time dilation = difference clock 1 – (clock 1 – Doppler effect)

Clock 1 stays on earth

[Beginning at 7:42]

How to Influence/Control Relevant Features?

1. *Treatment/control group*
2. *Holding things constant*
3. *Elimination (& blinding)*
4. *Separation*

Randomisation in Experiments

Randomised Controlled Trial (RCT)

Sample

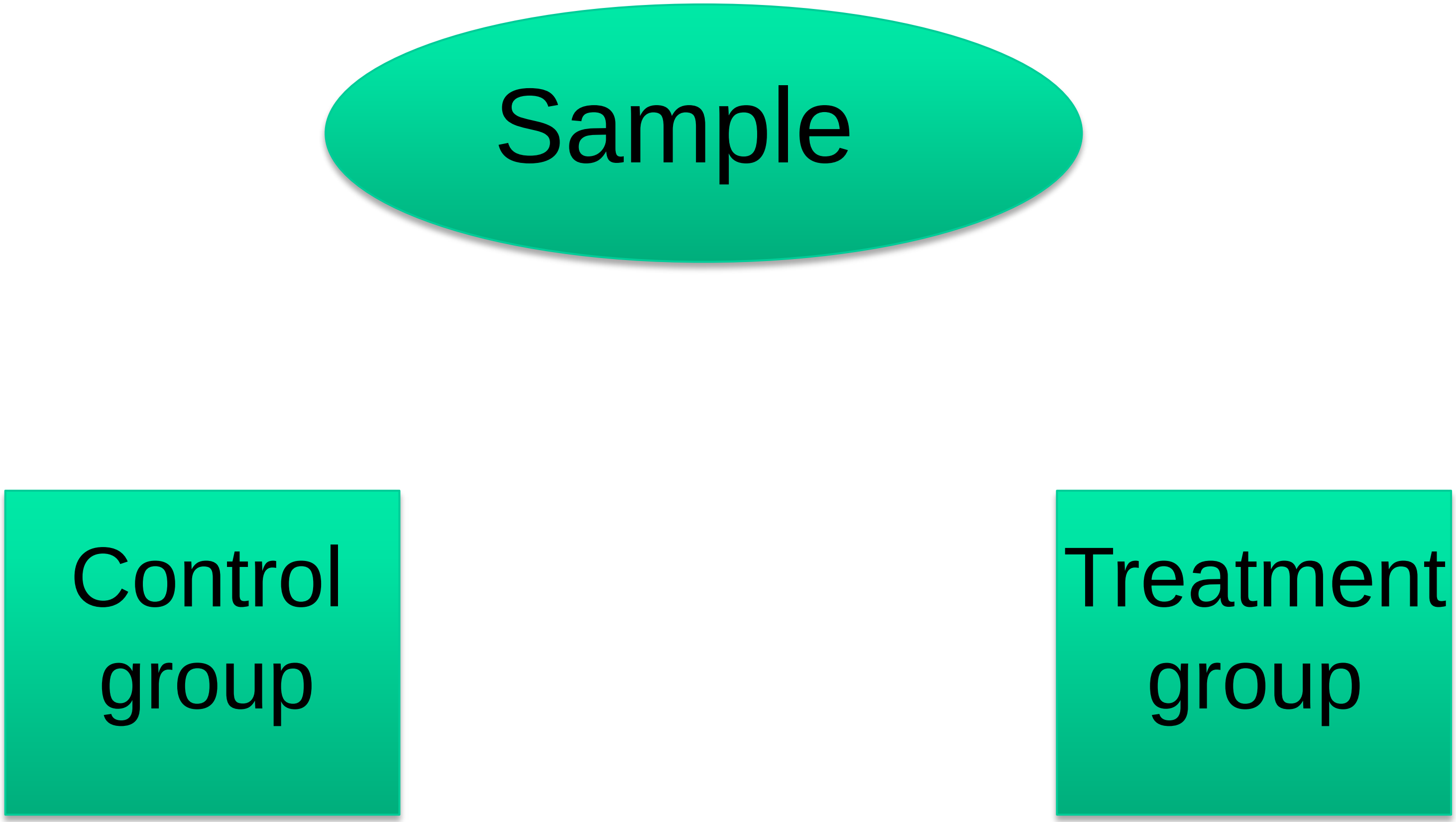
Randomised Controlled Trial (RCT)

Sample

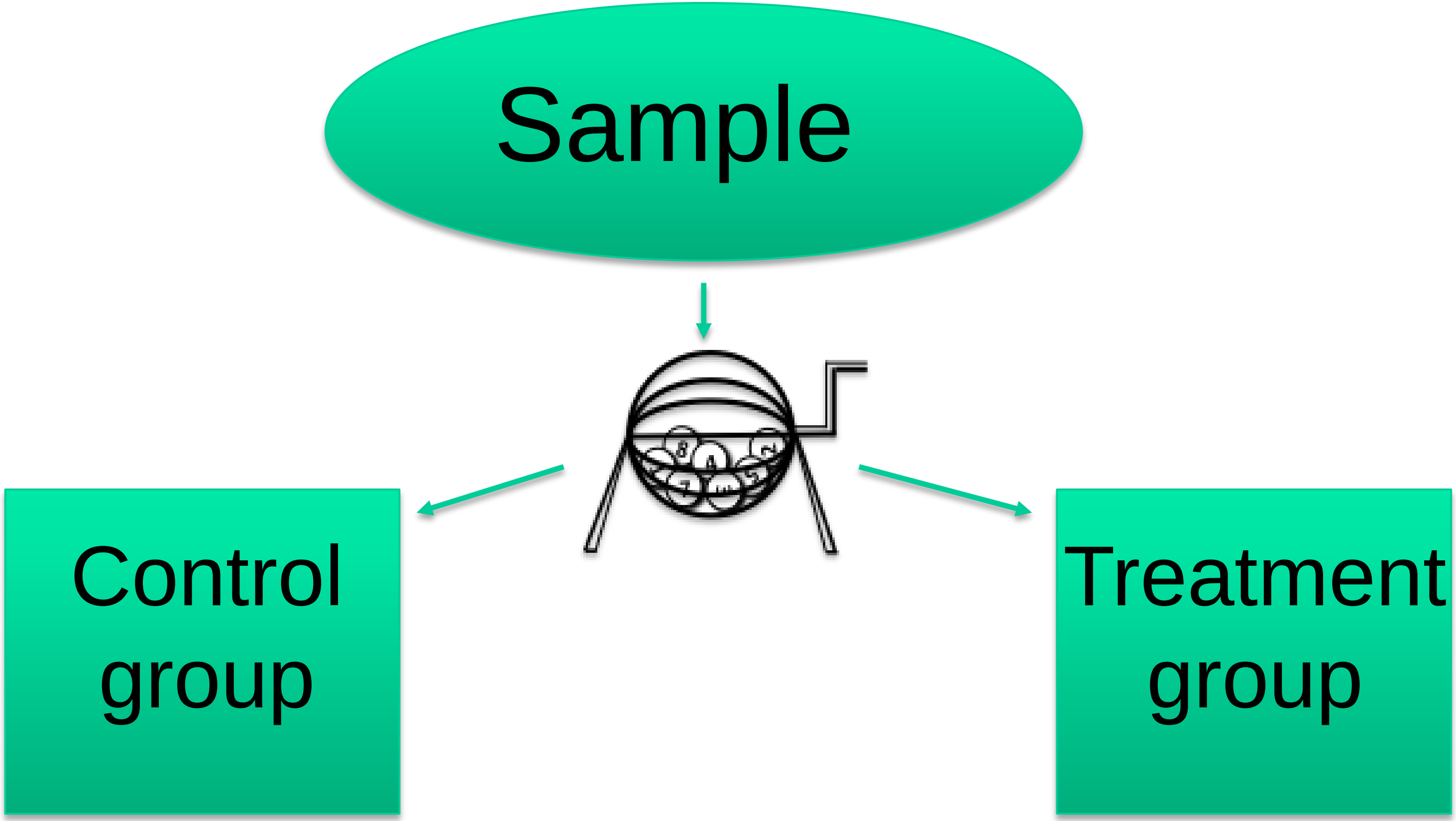




Randomised Controlled Trial (RCT)

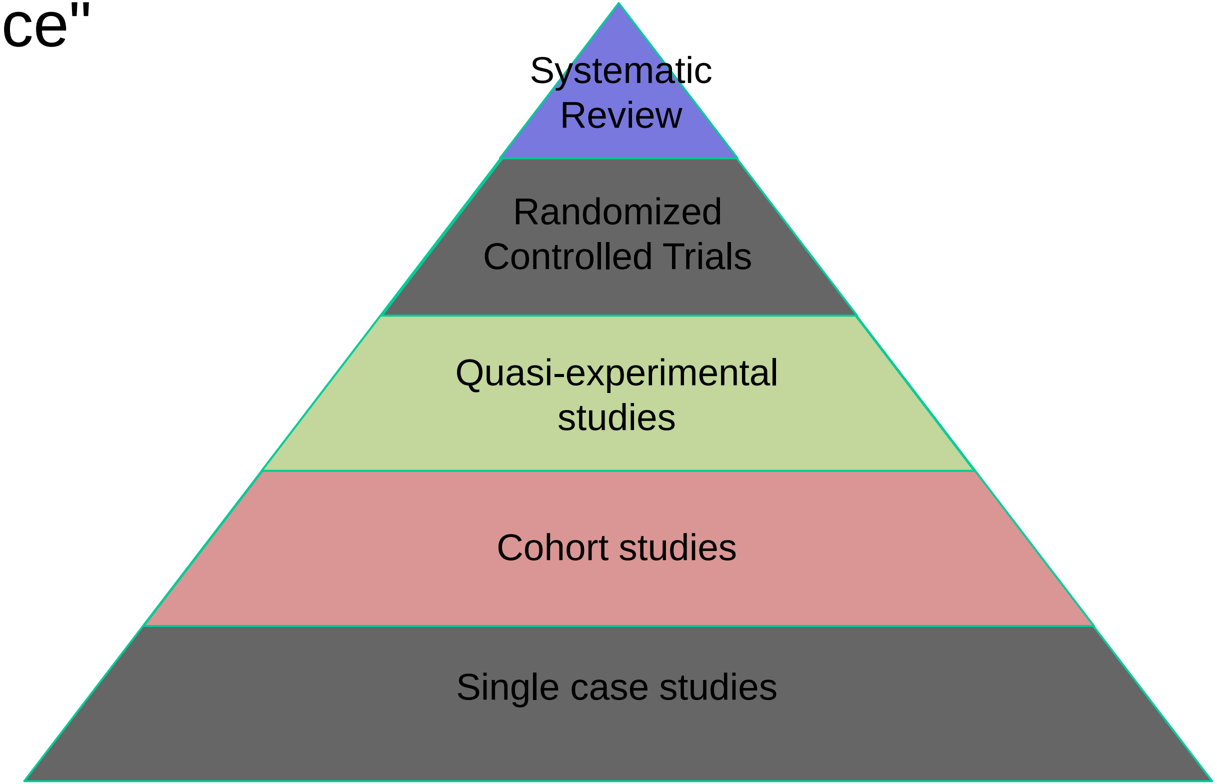


Randomised Controlled Trial (RCT)



Importance of RCTs

"Hierarchy of Evidence"



Why are RCTs considered so important?

Why are RCTs considered so important?

- Randomization eliminate selection bias

Why are RCTs considered so important?

- Randomization eliminate selection bias

- Randomization helps convince others that you have not rigged the treatment/control division in favor of the outcome you want

Why are RCTs considered so important?

- Randomization eliminate selection bias
- Randomization helps convince others that you have not rigged the treatment/control division in favor of the outcome you want
- Randomization facilitates blinding of the identity of treatments from investigators, participants, and assessors

Randomization helpful, but not necessary

- Avoid selection bias by other controlled assignment procedures
- Other controlled assignment procedures are also arguments against rigging selection
- Many other available blinding procedures that do not require randomization

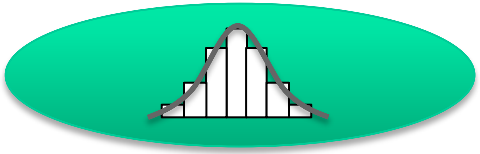
Debunking stronger claims for RCTs

<u>Claim:</u> Randomization ensures that background factors are equally distributed in treatment and control group

Debunking stronger claims for RCTs

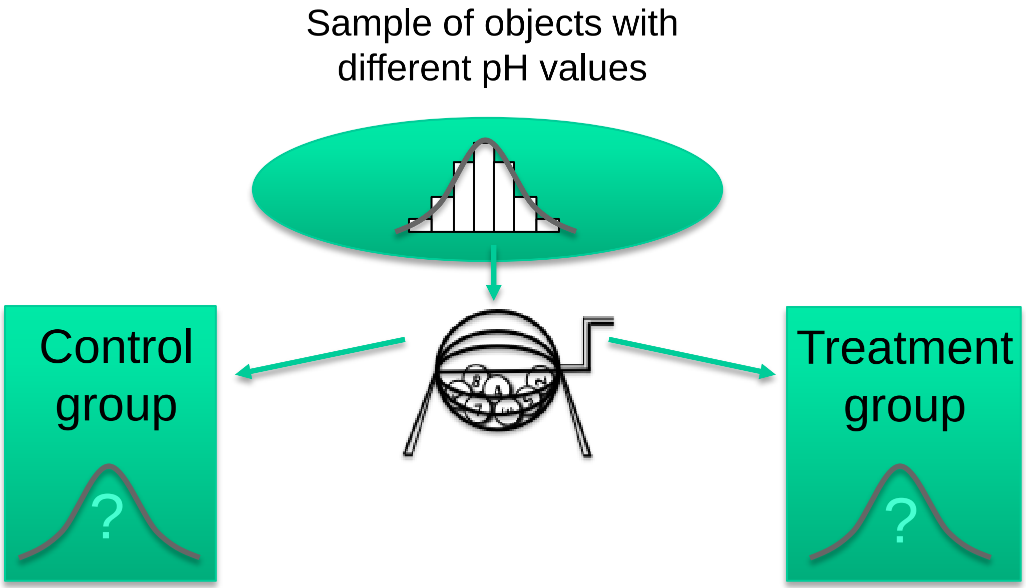
Claim: Randomization ensures that background factors are equally distributed in treatment and control group

Sample of objects with different pH values



Debunking stronger claims for RCTs

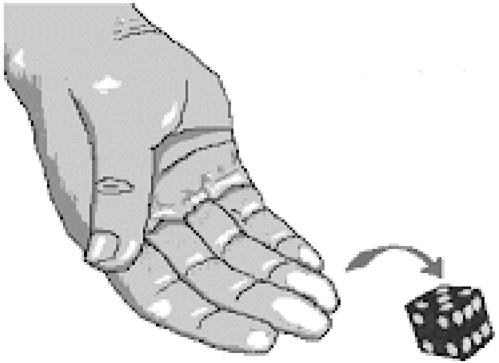
Claim: Randomization ensures that background factors are equally distributed in treatment and control group

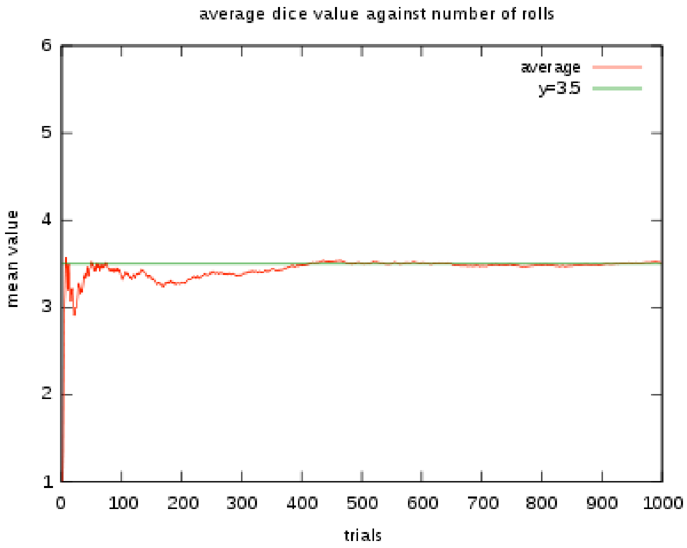


Debunking stronger claims for RCTs

Claim: Randomization ensures that background factors are equally distributed in treatment and control group

Analogy: rolling a die



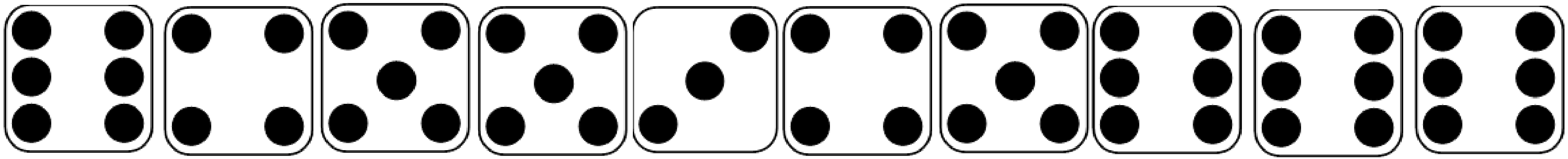


Debunking stronger claims for RCTs

Claim: Randomization ensures that background factors are equally distributed in treatment and control group

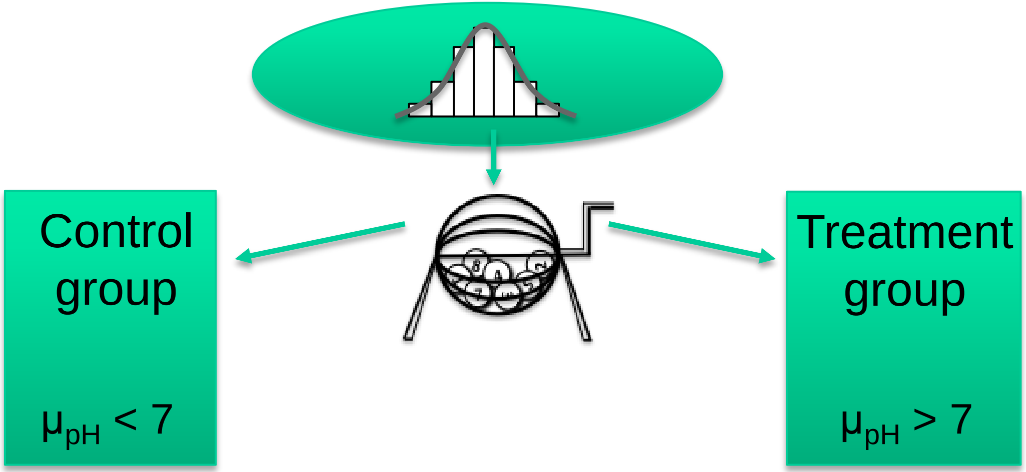
Analogy: rolling a die

Rolling it 10 times, a result like this should not be too surprising:



Debunking stronger claims for RCTs

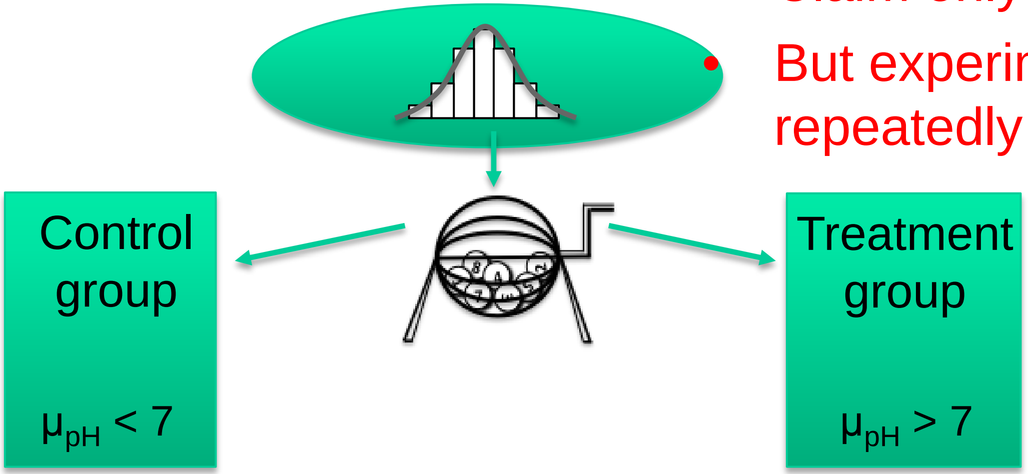
Claim: Randomization ensures that background factors are equally distributed in treatment and control group



Debunking stronger claims for RCTs

Claim: Randomization ensures that background factors are equally distributed in treatment and control group

- Any particular randomization might yield differing distributions in treatment and control
- Claim only true in infinite sequence
- But experimenters only randomize once, not repeatedly



Debunking stronger claims for RCTs

<u>Claim:</u> Randomization ensures that background factors are equally distributed in treatment and control group

- Any particular randomization might yield differing distributions in treatment and control
- Claim only true in infinite sequence
- But experimenters only randomize once, not repeatedly

Ronald Fisher
1890-1962

"Most experimenters on carrying out a random assignment will be shocked to find how far from equally the plots distribute themselves"

Fisher (1926)

Debunking stronger claims for RCTs

~~Claim: Randomization ensures that background factors are equally distributed in treatment and~~ control group

- Any particular randomization might yield differing distributions in treatment and control
- Claim only true in infinite sequence
- But experimenters only randomize once, not repeatedly

Ronald Fisher
1890-1962

"Most experimenters on carrying out a random assignment will be shocked to find how far from equally the plots distribute themselves"

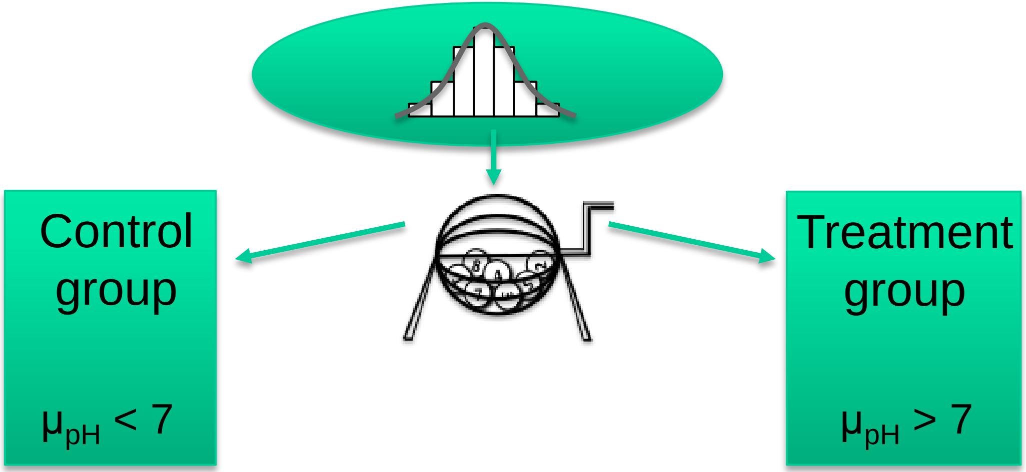
Fisher (1926)

How to avoid such imbalances?

1. Check for imbalances *in known factors* post-randomization

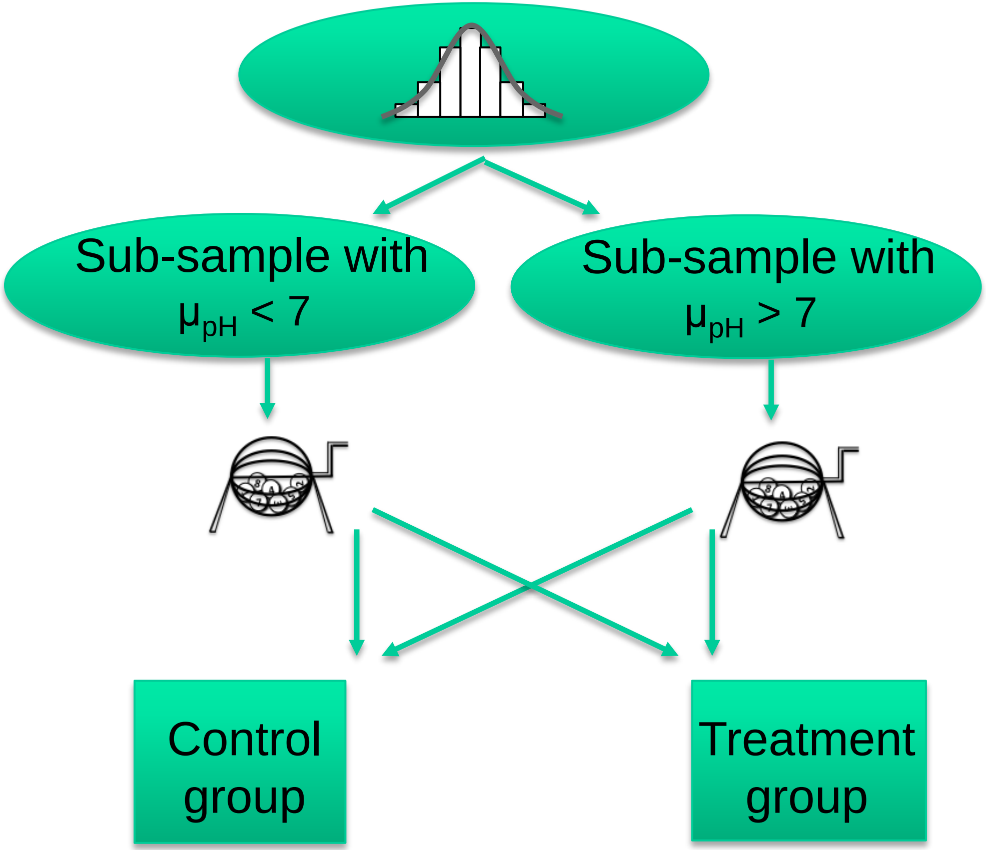
How to avoid such imbalances?

1. Check for imbalances *in known factors* post-randomization



How to avoid such imbalances?

2. Stratified randomization

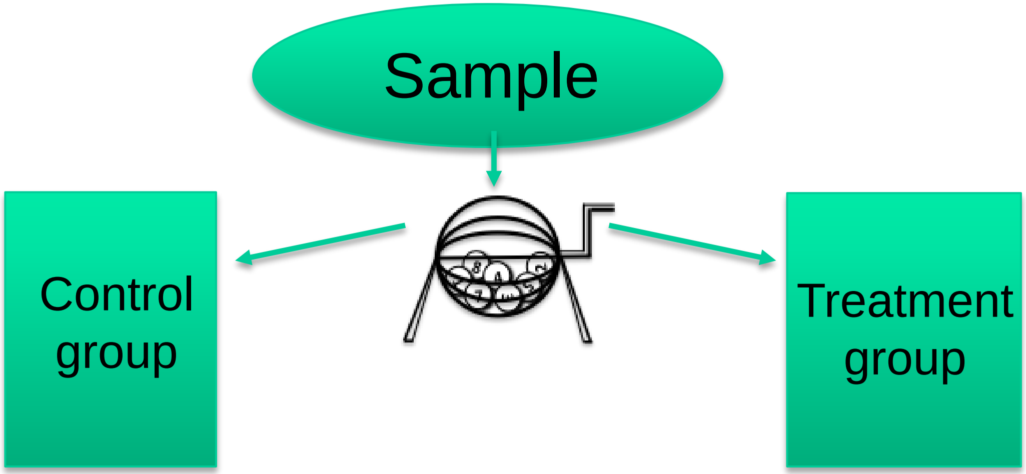


How to avoid such imbalances?

- For those two balancing strategies, factors to be balanced *need to be known*

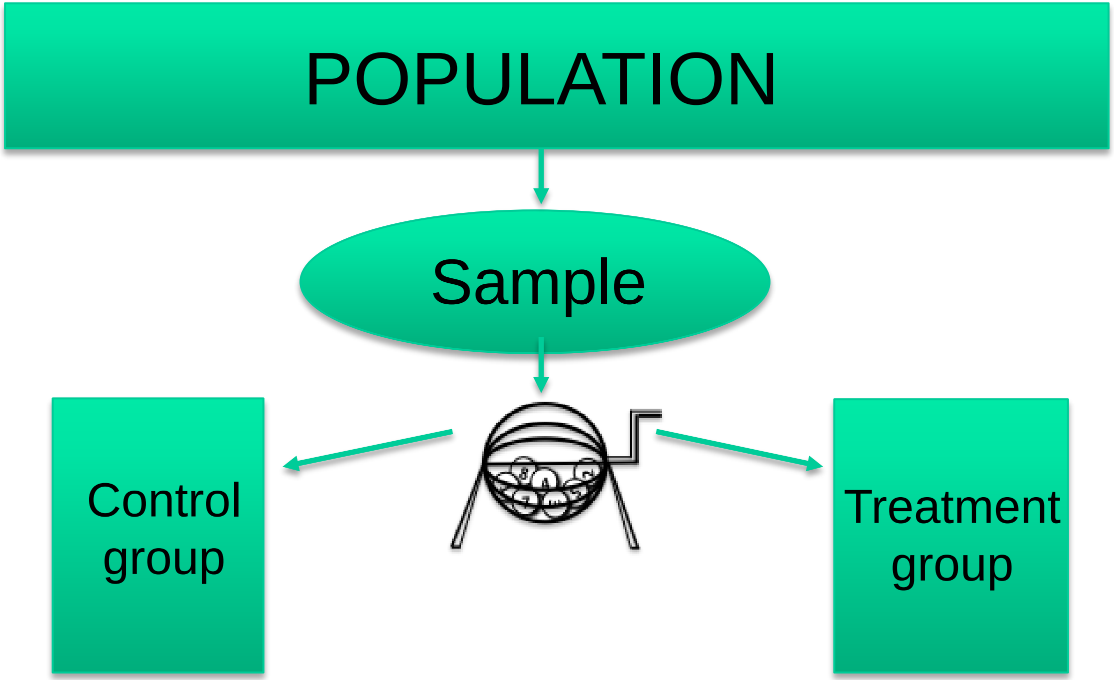
- Consequence: randomization does *not* guarantee control of unknown factors

Transferring RCT Results



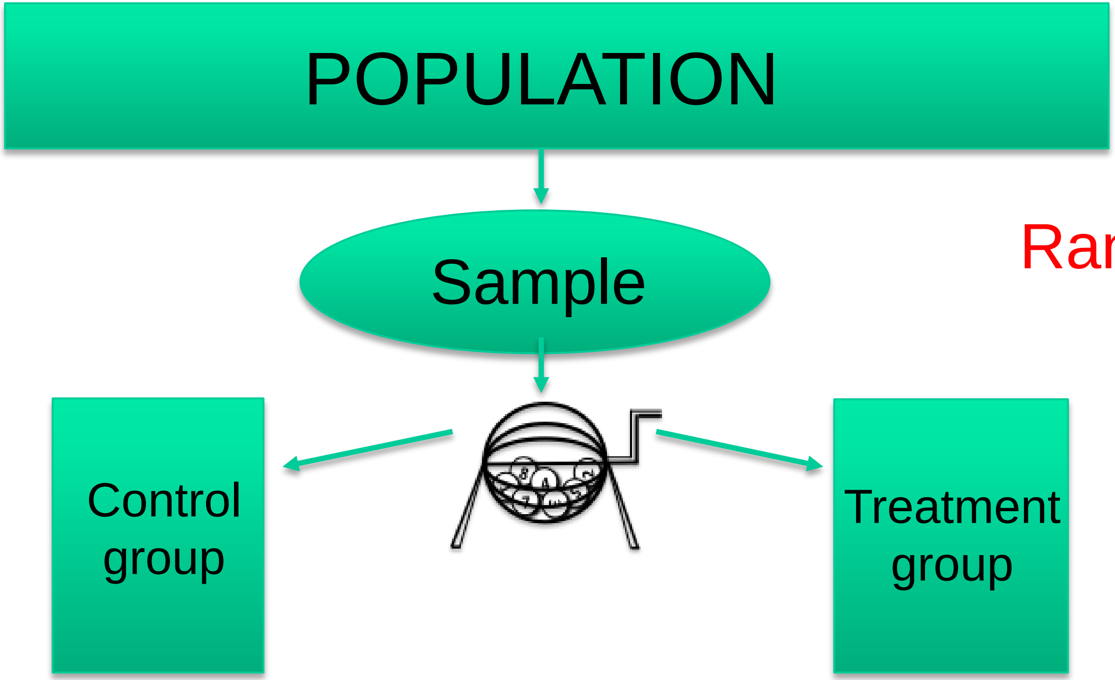
Result: e.g. average treatment effect (ATE): 50% increase in mean outcome in treatment vs. control group

Transferring RCT Results



Result: e.g. average treatment effect (ATE): 50% increase in mean outcome in treatment vs. control group

Transferring RCT Results



Randomization is *not* random sampling from population!

Result: e.g. average treatment effect (ATE): 50% increase in mean outcome in treatment vs. control group

Summary

- Good reasons for randomization
- Randomizing not *necessary* for these objectives
- Randomization *not* a guarantee for control of known or unknown background factors
- RCT results cannot be generalized without further argument
- Good RCTs require background knowledge
- RCTs not better evidence in principle than other experiments