Causality and Experiments

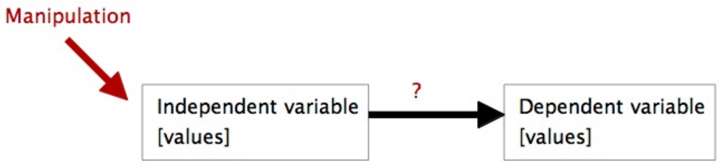
http://philosophy.ucsd.edu/faculty/wuthrich/

12 Scientific Reasoning

Acknowledgements: Bill Bechtel

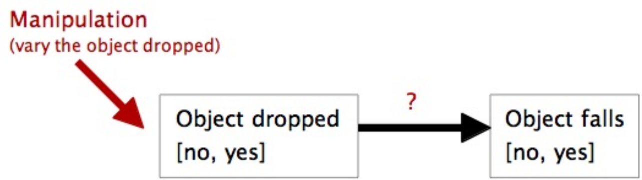
The basic idea of an experiment

- If the independent variable is the cause of the dependent variable, then a manipulation of the independent variable should produce a change in the value of the dependent variable.
- And if it were not the cause, then we would not expect such a result from manipulation.



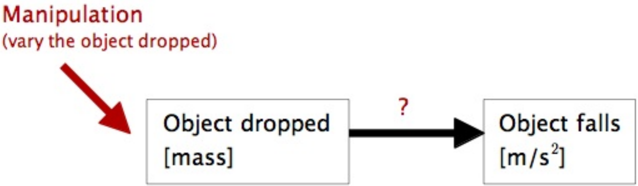
Experiments on regular deterministic systems

- When there is no variance in the population being studied, statistical analysis is not necessary.
- The main danger is affirming the consequent: The key is to test a causal hypothesis in which you would not expect the effect to occur unless you were right about the cause.



Experiments on regular deterministic systems

- When there is no variance in the population being studied, statistical analysis is not necessary.
- The main danger is affirming the consequent: The key is to test a causal hypothesis in which you would not expect the effect to occur unless you were right about the cause.



Variability in non-deterministic systems

- Different systems of the same kind will vary in their responses depending on
  - their particular composition (genetics, etc.)
  - their particular history, etc.
- The same system may respond differently on different occasions.
  - E.g. Your reaction time will differ depending on
    - how much sleep you have had
    - what you have had to drink, etc.

⇒ Challenge: how to detect causal relations in the face of background variability

Experiments on non-deterministic systems

Since complex systems (biological, cognitive, social) are not regular in their behaviour

- researchers cannot simply do an experiment on one instance and draw a conclusion about the whole population;
- must rather work with samples and draw conclusions based on statistical analysis.t
  - Are the differences in the values of the dependent variable greater than expected by chance?

Confounding variables

- The reason different individuals behave differently is that among extraneous variables, some might be related to the effect of interest.
- Such variables are called confounds and there are two kinds that are particularly important:
    1. subject variable confounds: differences between subjects in the study
    2. procedural variable confounds: differences in the way different groups are treated
- If these variables are correlated with the independent variable and are also causes of the dependent variable, the experiment is confounded.

Strategies for controlling confounding variables

1. randomization
   - most commonly used to control confounding subject variables
2. locking
   - most commonly used to control confounding procedural variables
3. matching subjects on confounding variables
4. making confounding variables into studied variables

Procedural variable confounds

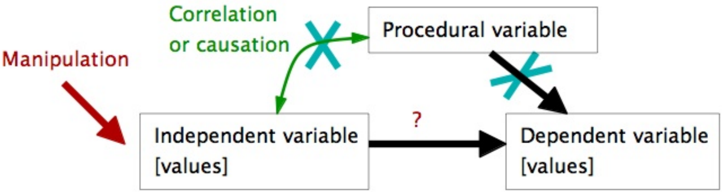
When you conduct a manipulation, generally more than one thing will be changed:

1. These variables will then be correlated with the independent variable—extraneous
2. If one of the other variables is causally related to the effect of interest, it rather than the variable you are considering may be the cause—confound

Confounding procedural variables

- The president of the AGL corporation wanted to get her workers to be more productive.
- She found that when each employee was presented with a jar of jellybeans, productivity increased.
- Was it the jellybeans that caused the increased productivity? Or was it:
  - novelty of the situation
  - attention from the president
  - desire to reciprocate

Controlling confounding procedural variables

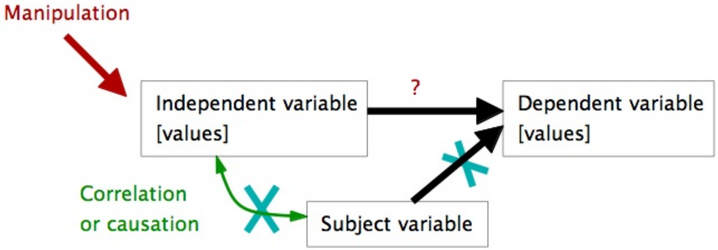


- Strategy: break the correlation—thereby breaking the effect of the confounding variable
- How to eliminate the effects of the confounding procedural variables in the jellybean case?

Subject variable confounds

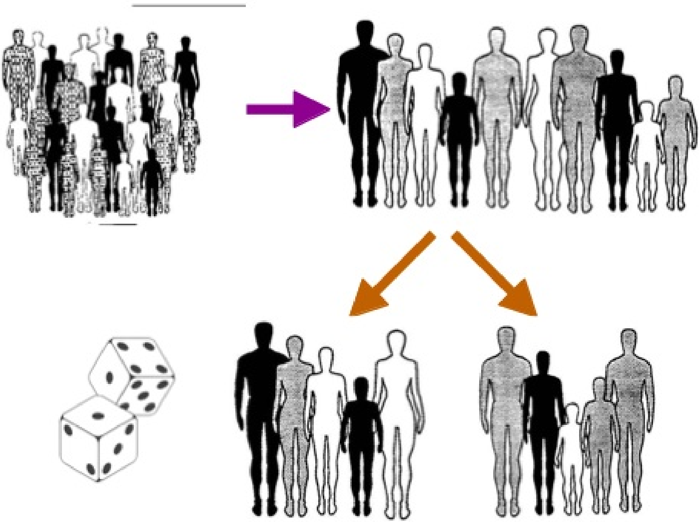
- Subjects in an experiment may be differently affected by different values of other variables.
  - people of different ages sleep different amounts
  - women might be affected differently than men
- These variables are extraneous.
- If there is a correlation between these variables and the independent variable, they, rather than the variable you are focusing on, may be what produce the change in the dependent variable.
- Such variables are confounds.

Controlling confounding subject variables

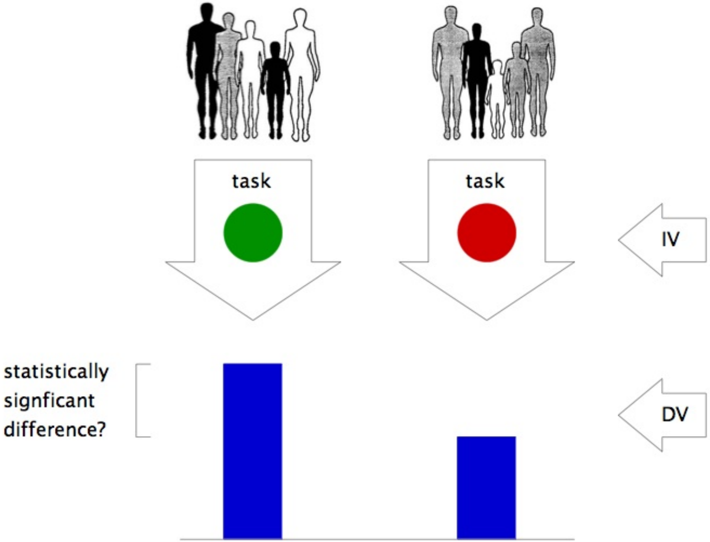


- Strategy: break the correlation—thereby breaking the effect of the confounding variable
- random assignment of subjects is a strategy for breaking the correlation

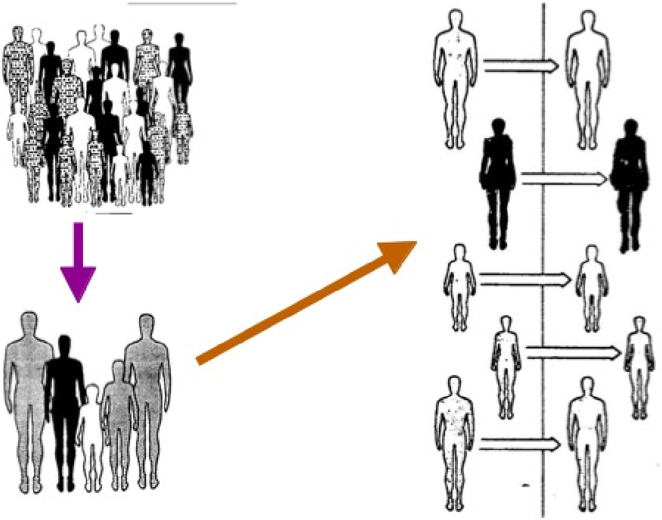
Controlling subject confounds:
Between subjects randomization



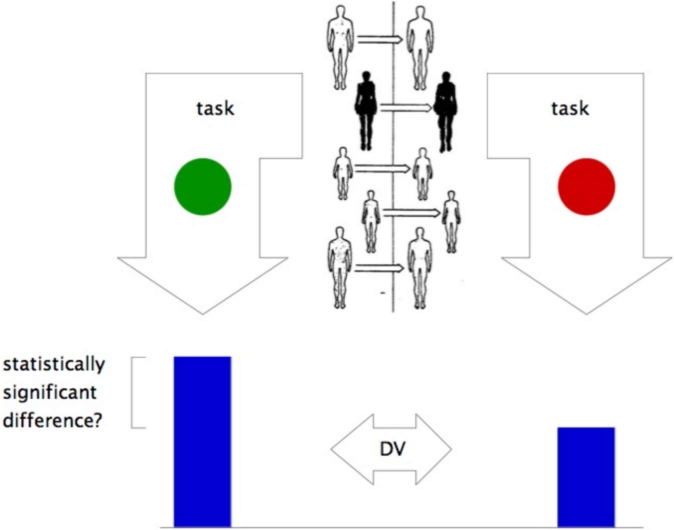
Manipulate independent variable



Controlling subject confounds:
Within subjects designs



Subjects serving as own control



Between-subject design

Good news:

- Subjects are not ‘contaminated’ in one condition as a result of having participated in the other.

Bad news:

- requires a larger number of participants
- runs the risk of non-equivalence of subject groups

Within-subject design

Good news:

- requires a smaller number of participants
- rules out any differences between subjects

Bad news:

- potential ‘contamination’ of participants’ behaviour from previous trial: carryover effect
- subjects might learn from one condition and that could alter their behaviour in the second condition
  - practice effect
  - fatigue effect

Counterbalancing

(1) within-subject counterbalancing

- reversing order: ABBA

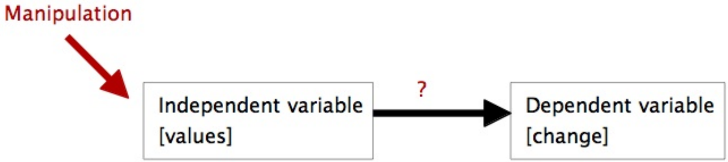
(2) across-subject counterbalancing

- complete: every possible sequence—requires *n*! runs
- partial
  - random
  - Latin square:
    - each condition appears once and only once in a given ordinal position
    - no two conditions are juxtaposed in the same order more than once

Order 1	A	B	D	C
Order 2	B	C	A	D
Order 3	C	D	B	A
Order 4	D	A	C	B

Pretest-posttest design

- There is always a danger in an experiment that the members of the two (or more) groups being studied already differ on the dependent variable.
- Best control is to focus on change, not raw value of the dependent variable.
  - Pretest: measure the dependent variable before the intervention
  - Posttest: measure the dependent variable after the intervention
  - Change = Posttest − Pretest



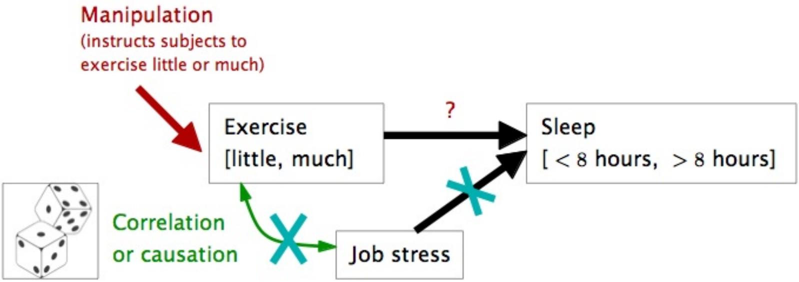
Limitations of pretest-posttest design

- Just measuring change in one group using a pretest and a posttest allows for confounds:
  - Time has elapsed and subjects have gotten older (maturation).
  - Events occurring between the pretest and posttest could affect the dependent variable (history).
  - Experience with previous test may change performance.
  - Pretest and posttest may vary in difficulty.

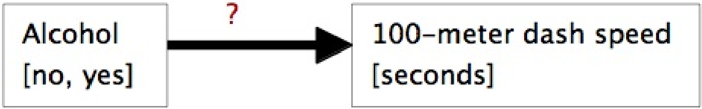
⇒ Use of pretest-posttest does not obviate the need for a control group.

Example: exercise and sleep

Is there a causal relation between exercise and sleep?



Example: alcohol and running speed



Between subjects or within subjects

Between-subjects design:

- Different subjects would be used for the no-alcohol and alcohol condition, and each would be tested only once.

Within-subjects design:

- Each subject would be tested both under the no-alcohol and alcohol condition.

Between subjects

Jim
Angela
Megan
Tony

Roger
Shane
Sara
Jessica

Within subjects

Jim
Angela
Megan
Tony

Jim
Angela
Megan
Tony

Counterbalance

Alcohol condition ⇒	rest ⇒	No-alcohol condition
Jim		Jim
Angela		Angela

No-alcohol condition ⇒	rest ⇒	Alcohol condition
Megan		Megan
Tony		Tony

Internal validity

Definition (Internal validity)

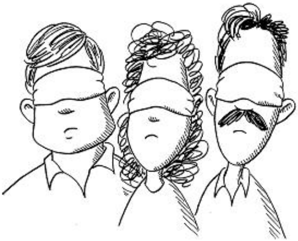
An experiment is internally valid if it was in fact the manipulation of the independent variable that produced the change in the dependent variable

- Are the effects on the dependent variable due solely to the manipulation of the independent variable?
- Was there a confounding subject variable that did not get controlled?
- Was there a confounding procedural variable that did not get controlled?

A threat to internal validity

- People may change their behavior when they are being observed:
  - People want to be liked (or not!).
  - People want to be helpful (or not!).
  - People want you to think they're a good person.
  - People want to be thought of as intelligent and normal (not crazy, stupid or obsessed).

⇒ Problem if subjects figure out the point of an experiment.



- Solution: keep subjects blind as to the point of the experiment or what is being studied (single-blind experiment)

The legend of the Western Electric Plant at Hawthorne

- test of improved working conditions on worker productivity
- legend has it that any change that was made increased worker productivity:
  - increase lighting $\Rightarrow$ work harder
  - decrease lighting $\Rightarrow$ work harder
- no published report and the data ‘has vanished’

Hawthorne mica-splitting experiment

- Experimenters monitored the output of five experienced women who split, measured, and trimmed mica chips used for insulation at their regular department workstations.
- Experimenters moved the women to a special test room and added 10-minute rest breaks at 9:30 am and 2:30 pm.
- After a brief decline in performance, the women's output increased by an average of 0.15 percent and remained at that level.
- When the women returned to their department and lost the rest periods, their output dropped back to the original rate.
- Since they thought no other conditions had changed, the researchers attributed the increase in output to the beneficial effects of rest periods—not an example of the Hawthorne effect.
- And there were confounds aplenty in this study.

Experimenter bias

- Danger that experimenters will see what they want to see.
  - Mendel's data is too perfect—there should be more variability
  - most likely explanation is that he did not deliberately cheat (remember, he was a monk!)
  - but he reported the best cases and subjectively biased his counting of plants

$\Rightarrow$ keep the data-tabulator blind as to which group different subjects are in

$\Rightarrow$ double-blind study

Planning an experiment

Say the colour the following words are written in:

Blue	Pink
Brown	Yellow
White	Orange
Red	Green

- Does it seem harder to name the colour when the words name a different colour?

How might we test the claim that it is the meaning of the word that makes it harder to say the colour it is written in?

1. operationalize the notion of being hard to read
    - slower reaction time when incongruent words
2. identify a sample population
    - college undergraduates in psychology courses
3. pick study design
    - between subject
    - within subject

Controlling subject variable confounds

Blue	Pink
Brown	Yellow
White	Orange
Red	Green

- What subject variables might you have to worry about as confounds?
- How to control for these confounds?
  - if between subject ⇒ randomize
  - if within subject ⇒ counterbalance

Controlling for procedural variables

Blue Pink
Brown Yellow
White Orange
Red Green

- What procedural variables should be controlled to avoid confounds?
  - context of presentation
  - illumination of the stimuli
  - length of words
  - familiarity and frequency of words

$\Rightarrow$ need to lock these variables so that they do not vary across conditions

Find two major confounds

Test for verbal (lecture) versus visual (reading) learners

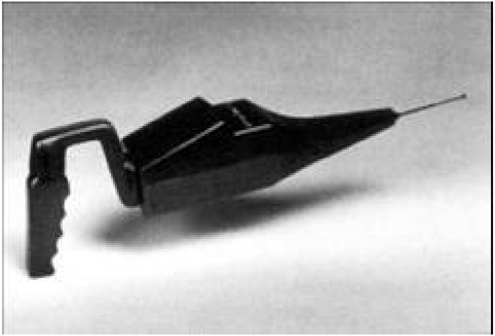
- in a different context, have students self-identify as verbal versus visual learners
- students then tested for memory of 20 different words presented orally (tape recorder) or visually (paper)
- lists were presented first in the subject's preferred format, then other
- oral and visual lists the same for all subjects
- all other variables (length of presentation, delay before memory test held constant)

Detecting pseudo-causal relations

- Can dowsers reliably locate water or metal objects?
- Ray Hyman tested an experienced dowser on the PBS program *Frontiers of Science* (19 Nov 1997):
  - used random numbers to pick which of the buckets that were placed upside down in a field would have a metal object under it
  - no one going with the dowser knew which buckets had been selected

- If there are 100 buckets and 10 of them have a metal object, then getting 10% correct would be expected by chance.
- The dowser walked up and down the lines of buckets with his rod but complained that he couldn't get any strong readings.
- Occasionally he selected a bucket but qualified his selection by expressing doubt that he'd be right.
- He was right about never being right—he didn't find a single metal object despite several attempts...
- The dowser was genuinely surprised by his failure.

A miracle device?



Detect human beings hidden in building or behind objects from almost two football fields away...

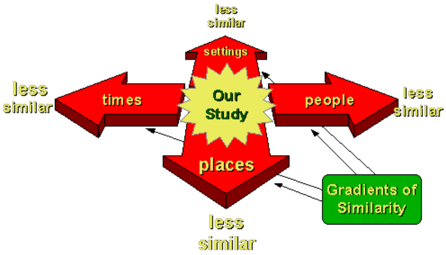
The DKL LifeGuard

- according to the company can detect human heartbeat
- range of up to 500 meters (540 yards)
- no effective electronic or other countermeasures
- no natural and synthetic material it cannot penetrate
- no battery or any power sources required!
- repeatability of 99%
- can penetrate:
  - 10-meter wide earthen barrier
  - 10 feet of water
  - concrete walls, steel bulkheads
- can distinguish a man, woman or child from any other animal, even a gorilla or an orangutan
- requires only one day of operator training
- work as well in darkness as in daylight
- like no other technology on the market today... literally in a league of its own

Double-blind test

- five large plastic packing crates were set up in a line a 30-foot intervals
- a DKL representative, using the DKL LifeGuard Model 2, tried to detect which of the five crates contained a human being
- On preliminary tests in which the operator knew which crate contained the person, the DKL LifeGuard found the person 10 out of 10 times.
- In the real, double-blind test, the operator found the person 6 out of 25 times (and took much longer to find the person).

External validity



- To what extent can you generalize the results of your study?
- Are they specific to a particular sample?
  - college sophomores or the general population
- Do they only apply in a particular (laboratory) setting?
- Do they generalize beyond the details of the manipulation?
  - ecological validity

Population generalization

Question

Will a study using one population generalize to another population?

- Will a study of college sophomores generalize to middle-aged adults?
- Will a study of chronically depressed patients generalize to patients who are acutely depressed?
- Will a study of captive raised dolphins generalize to wild dolphins?
- Will a study on mice generalize to humans?

Setting generalization

Question

Will a study conducted in one laboratory or clinical setting generalize to the setting of interest?

- Will results obtained in a flight simulator generalize to an actual cockpit?
- Will results obtained in an outpatient setting generalize to a psychiatric hospital?
- Will results obtained in a laboratory generalize to customers in a store?

Manipulation generalization

Question

Will a result obtained with one task generalize to other tasks or stimuli?

- Will studies of perceiving visual illusions generalize to perception of ordinary objects?
- Will a survey of consumer attitudes generalize to consumer behaviour?

Assessing external validity

Must make a plausibility judgment in assessing external validity:

- Is the target population different from the studied population in ways that are likely to matter for the causal claim?
- Is the target setting different from the studied setting in ways that are likely to matter for the causal claim?
- Is the manipulation used in the experiment different from the target process in nature in ways that are likely to matter for the causal claim?

Example: rats and saccharine

- 1977 Canadian study which fed pregnant rats up to 20% of their body weight per day in saccharine showed an increase in bladder tumors
- Saccharine was banned in Canada and the FDA was about to ban its use in the US when Congress intervened
- Assessing external validity:
  - Are rats relevantly like humans?
  - Is eating in the laboratory like eating at home, etc.?
  - Is feeding up to 20% of body weight like eating as part of diet?