Advanced Research Methods

STATS

Paweł W. Woźniak

HI! I'M PAWEŁ

I do research in Human-Computer Interaction.
I have to run experiments.
I will talk about experiments.
I am a heavy user of stats.
I know some stats.

p.w.wozniak@uu.nl

Qualitative Methods

Problems, problems, problems

- Stats are tough
- Life is brutal
- You most likely have different skill levels
- I still want this to be fun
- Please provide constant feedback

Our big aims

- Know **when** experiments are appropriate
- Know **which** experiments are needed
- Know **how** to conduct an experiment
- Bonus round: be better at spotting good science

But why can't we do a lot of

STATS

instead?

Overview: how to conduct an experiment

1. find research question
2. formulate hypothesis
3. operationalize abstract concepts
4. design experiment
5. conduct experiment
6. analyze data
7. infer from results

I didn't come up with this course

- I learnt from the best instead.

A lot of the content in these lectures was conceived by Lewis Chuang
He happily lives in Munich now.

Rational vs. Empirical

https://pilosopiskosclcv.files.wordpress.com/2014/10/o-raphael-rooms-facebook.jpg

Rational vs. E

https://pilosopiskosclcv.files.wordpress.com/2014/10/o-raphael-rooms-facebook.jpg

What is Empiricism?

- Derive "truth" from evidence
- *Determine relationships between **variables***

What is a variable?

a) *A variable is a factor that can change*

b) *A variable is what we manipulate in controlled experiments*

c) *A variable is what we measure*

What is Empiricism?

- Derive "truth" from evidence
- *Determine relationships between **variables***

What is a variable?

a) *A variable is a factor that can change*

b) *An **independent** variable is what we manipulate in controlled experiments*

c) *A **dependent** variable is what we measure*

DO GORILLAS GET PISSED OFF AROUND HUMANS?

What is Empiricism?

- Derive “truth” from evidence
- *Determine relationships between observed **variables***

- **Observation** or Manipulation

Source: www.imdb.com

Do gorillas get pissed off around humans?

- *Non-interference*
- *Subjective accounts*
- ***What is a variable?***

Types of studies

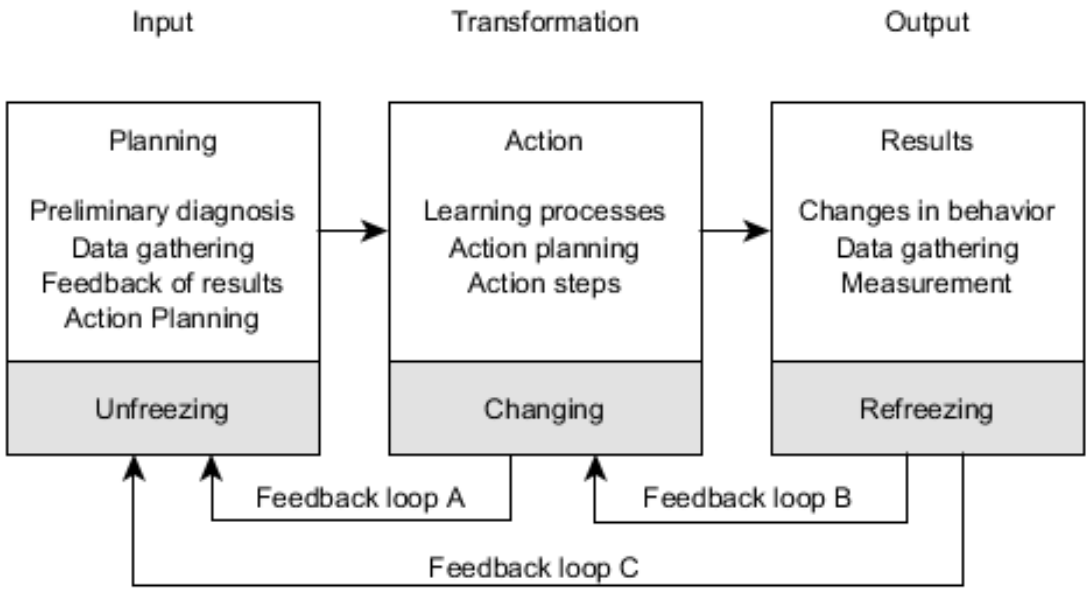
- Observational study
  - *manipulation is outside the control of the researcher*

Types of studies

- Observational study
  - *randomized experiment would be unethical*
    e.g., hypothesized relationship between drug use and academic performance
  - *researcher lacks the necessary influence*
  - *the symptoms are rare*

Types of studies

- Observational study
  - *manipulation is outside the control of the researcher*
- Action research
  - *researcher participates in the problem-solving process*



Bond M. (2014) Systems Model of Action Research Processing

Types of studies

- Observational study
  - *manipulation is outside the control of the researcher*
- Action research
  - *researcher participates in the problem-solving process*
- Contextual Inquiry
  - *one-on-one interaction with participant performing the target activity and discussions with the user*

Types of studies

- Observational study
  - *manipulation is outside the control of the researcher*
- Action research
  - *researcher participates in the problem-solving process*
- Contextual Inquiry
  - *one-on-one interaction with participant performing the target activity and discussions with the user*
- Think-Aloud
  - *participants speak their minds without interf*

Types of studies

- Think-Aloud
  - *participants speak their minds without interference*

Prof. Philip Barker, Human-Computer Interaction Laboratory,
School of Computing and Mathematics,
University of Teesside, UK

Types of studies

- Observational study
  - *manipulation is outside the control of the researcher*
- Action research
  - *researcher participates in the problem-solving process*
- Contextual Inquiry
  - *one-on-one interaction with participant performing the target activity and discussions with the user*
- Think-Aloud
  - *participants speak their minds without interference*
- Focus groups
  - interactive group settings where people are asked about their opinions, beliefs, and perception.

Examples of observational studies

- Observation
- Action research
- Contextual Inquiry
- Think-Aloud
- Interviews
- Focus group
- Diary studies

Short answers, Long answers, and Interviews

WHAT CAN WE MEASURE FROM OBSERVATIONS?

What is Empiricism?

- Derive “truth” from evidence
- *Determine relationships between observed **variables***

- **Observation** or Manipulation

Source: www.imdb.com

Do gorillas get pissed off around humans?

- *Non-interference*
- *Subjective accounts*

What is Empiricism?

- Derive "truth" from evidence
- Observation or **Manipulation**

Do gorillas get pissed off around humans?

- *Cause and effect*
- *Use of control baseline*
- *Test conditions*
- *Comparative analyses between test conditions and control*

Source: www.imdb.com

Randomized controlled trials

- randomly allocated of participants to test conditions
- control group
- test conditions for comparison to the control group
- *minimize allocation bias*
- *minimize selection bias*
- *maximize statistical power*

The Gust
Willem van de Velde

Randomized controlled trials

- randomly allocated of participants to test conditions
- control group
- test conditions for comparison to the control group

1. *cider*
2. *sulfuric acid*
3. *vinegar*
4. *seawater*
5. *oranges & lemon*
6. *spicy paste + barley*

James Lind

The Gust
Willem van de Velde

Lind, J. (1757). *A Treatise on the Scurvy: In Three Parts, Containing an Inquiry Into the Nature, Causes, and Cure, of that Disease*. A. Millar. Chicago

How to conduct an experiment

1. find research question
2. formulate hypothesis
3. operationalize abstract concepts
4. choosing an experimental design
5. define participant's task (free description, performance task etc.)
6. Analyze & interpret results
7. (Share results)

How to conduct an experiment

1. find research question
2. formulate hypothesis
3. operationalize abstract concepts
4. choosing an experimental design
5. define participant's task (free description, performance task etc.)
6. Analyze & interpret results
7. (Share results)

The research question

Two approaches in tech. usability research:

1. New technology has been developed: does it meet its design goals?
   - e.g., Does a new input device result in time-savings/likeability?

2. New technology is to be developed: which abilities and limitations do humans possess that should be taken into account in the new design?
   - e.g., How many input modalities can an (old) person remember?

The research question

What are human capacities for tech.?

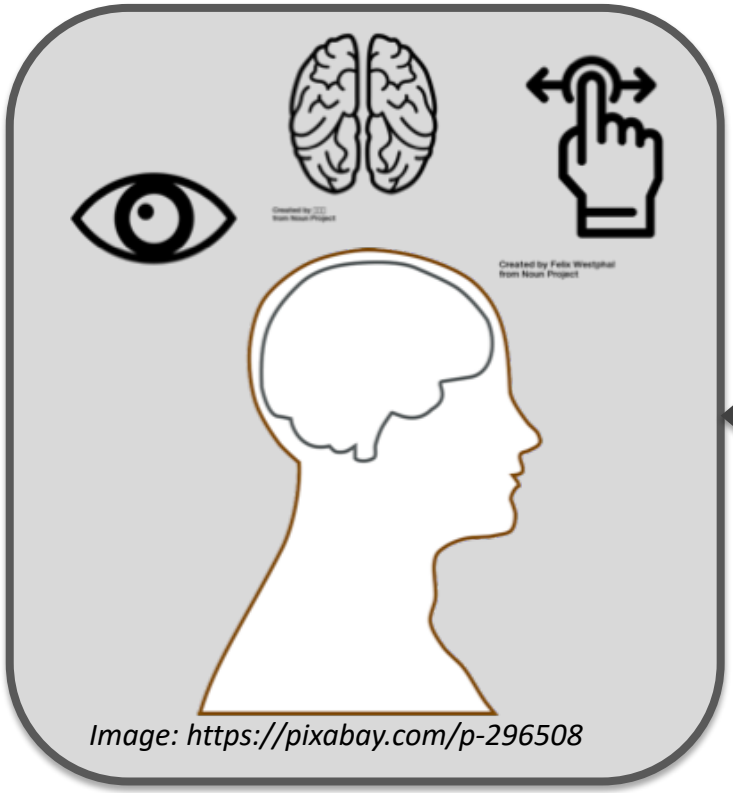
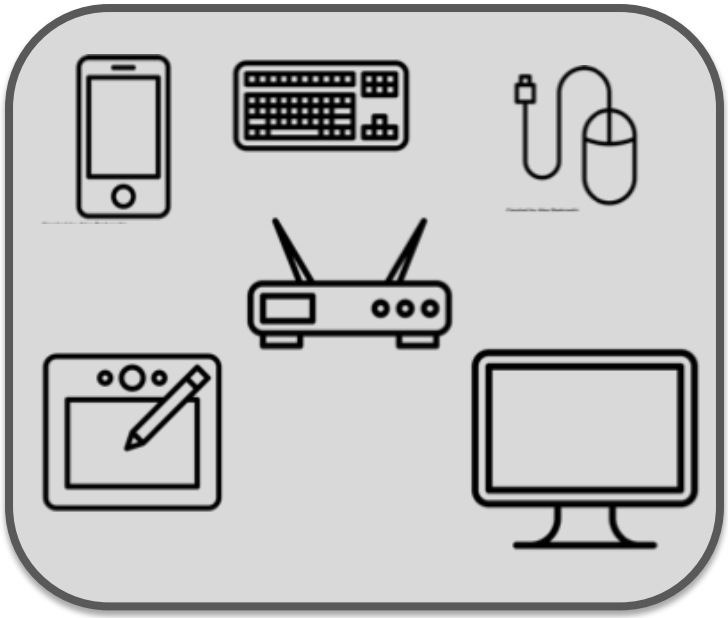


Image: https://pixabay.com/p-296508

Does tech. meet its design goals?



Images: Alex Berkowitz from the Noun Project

A hypothesis:

- is a preliminary answer to the research question

- Postulates a relationship between at least two variables
  - Independent variable(s)
  - Dependent variable(s)

Examples:

- Non-smoking mothers have healthier babies.
- Gesture inputs (as compared to a mouse) enhance the user experience of the interface.
- Coding in pairs is more productive than two people coding alone.

How do we derive a hypothesis?

- Informal observations (moving elements on a website attract attention)
- Wishful thinking (new technology should be better than existing one)
- Literature research:
  - Competing theories or;
  - Experimental findings

A scientific hypothesis:

1. can be empirically tested
2. can be formulated as a conditional sentence

   If mothers do not smoke, then they have healthier babies.

   The less mothers smoke, the healthier their babies.
3. Is generalizable: Is applicable beyond the current case.
   i. When the sun shines, it is 25 degrees. (only applicable to today)
   ii. When the sun shines, it will be warm. (applicable everyday)

https://cleanlens.files.wordpress.com/2015/06/russells-teapot.jpg

A scientific hypothesis:

…..

3. Is generalizable: Is applicability beyond the current case.
   i. Running MS Office (OS X) results in 10% reduction of productivity, compared to MS Office (Windows).
   ii. Non-native applications are less efficient than native applications.
   iii. Apps that require X will result in 10% reduction of productivity, as measured by task completion time.
   iv. Non-native applications require more computing resources than native applications because they require **X**.
4. Falsifiable

A scientific hypothesis:

…..

4. Falsifiable:
   - This swan is white. (falsified if swan is not white)
   - All swans are white. (falsified if at least one swan is not white)
   - 95% of swans are white (falsifiable, but will probably never be falsified)

Freud's hypotheses were often criticized for a lack of falsifial

A patient visits Freud with a set of symptoms.

>> Freud hypothesizes that patient suffers from potentially repressed homosexuality.

Case 1: Patient admits to homosexuality.
The hypothesis has not been falsified.

Case 2: Patient denies homosexuality.
Freud concluded that patient represses homosexuality.
The hypothesis has not been falsified.

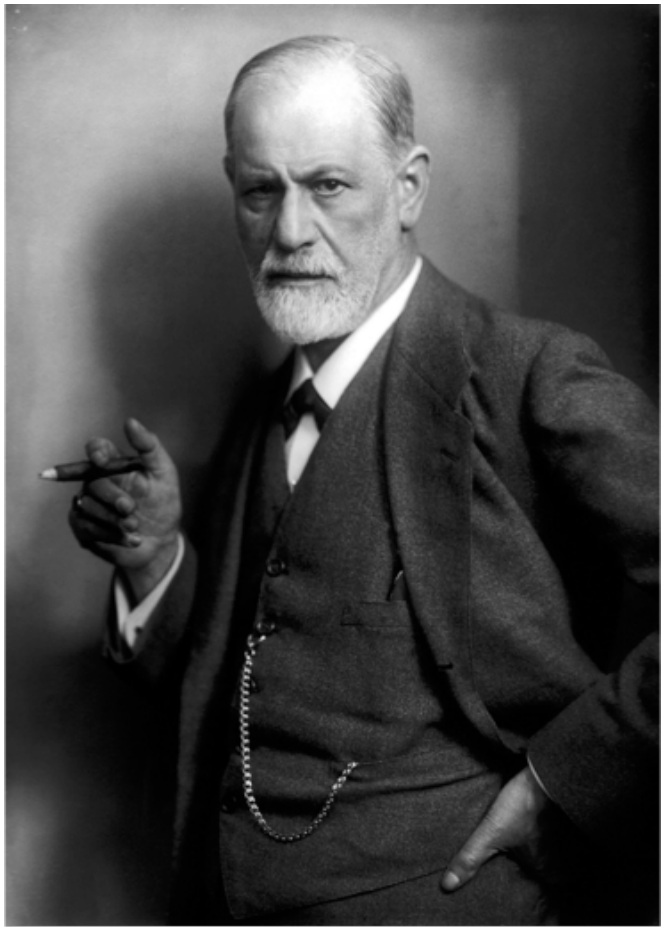


Image: wikimedia.org

Why Operationalize?

- Allows measurement of a fuzzy question or theoretical construct (see V01).
- Examples:
  - “healthy baby”
  - “good driver”
  - “improves user-friendliness of interface”
  - “improves situational awareness of pilots”

Hypothesis: Operationalizing

- Operationalize = make a fuzzy question/theoretical construct measurable
- Examples:
  - *"healthy baby"* = body weight/size at birth
  - *"good driver"* = # years driven without accident
  - *"improves user-friendliness of interface"* = high satisfaction rating on questionnaire, shorter task completion time than on the old interface
  - *"improves situational awareness of pilots"* = fewer missed targets (braking car in front, passenger crossing street) in a driving simulator task

How to operationalize an abstract concept?

1. Check literature/previous research for commonly accepted operationalizations
   - Task-completion times, Accuracy, …

2. Operational definition
   - "in the present study, we define driver aptitude as XY”
   - clear approach: (i) helps to explain contradictory findings in the field or, (ii) allow generalization of concept to new field.
   - potential disadvantage: your work cannot be easily compared to previous work and other researchers might disagree with your definition.

3. Inter-rater consensus

How to operationalize an abstract concept?

…..

3. Inter-rater consensus
   - For difficult concepts (e.g. aggression, user-acceptance)

Hypothesis:

"MS Office assistant increases aggressive user-behavior, compared to regular MS Office."

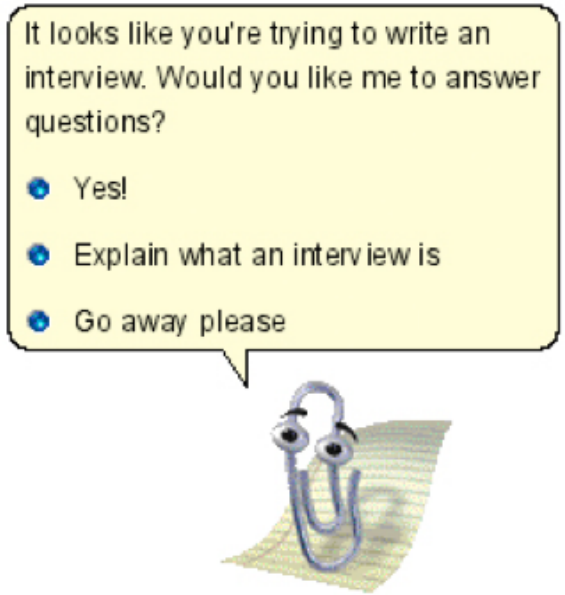
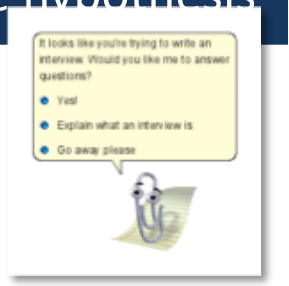


Image: http://blogdailyherald.com/wp-content/uploads/2015/03/CLIPPY3.jpg

Inter-rater consensus



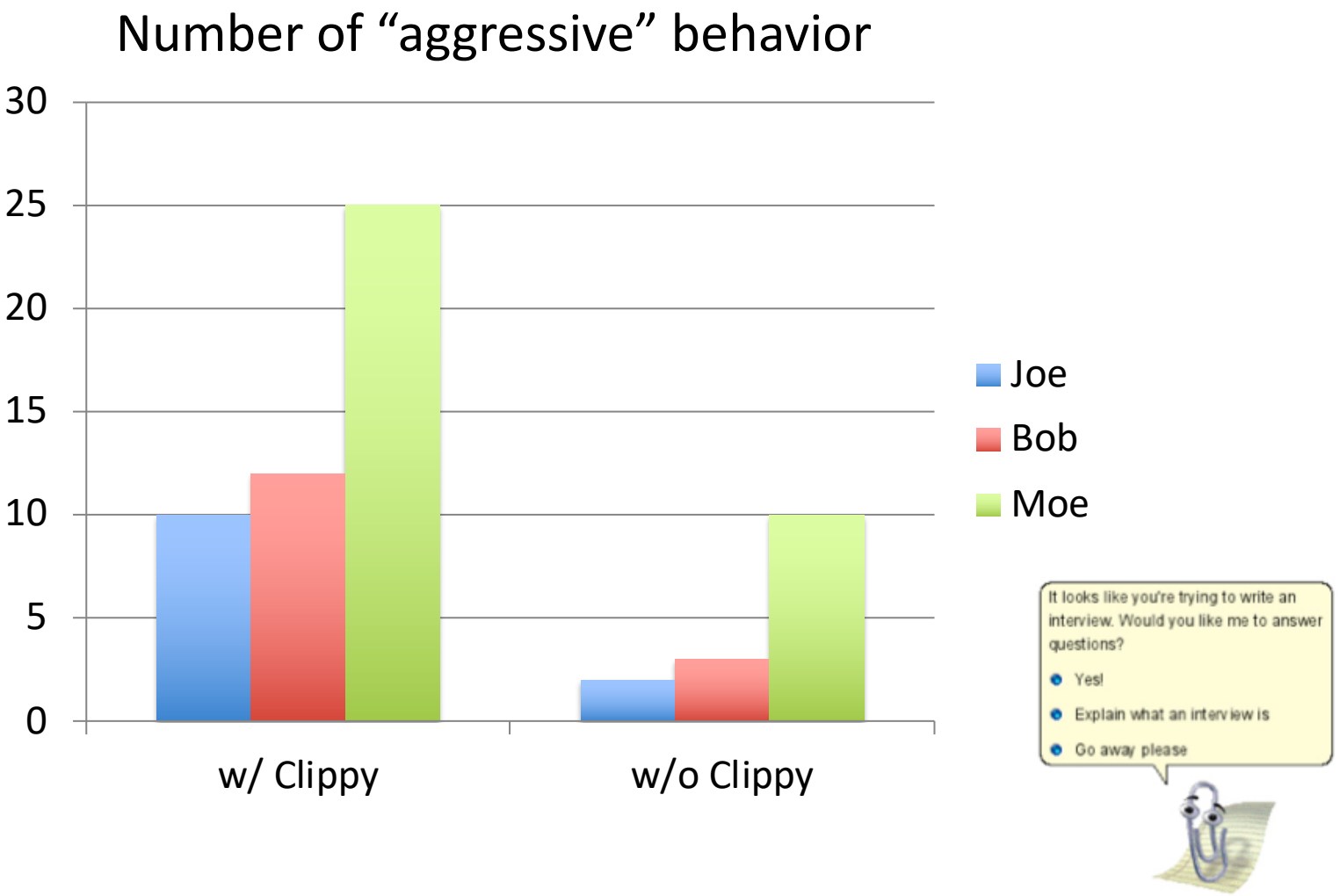
…..

- film an participant's reactions to the different experimental manipulations
  "Incidents of visible emotional responses were isolated by two researchers and used as video-clip stimuli for analysis.".

- (At least two) independent raters judge each reaction for its aggressiveness/user-acceptance/…
  "Three independent raters viewed each clip and labeled them as being either "aggressive", "satisfaction", "confused", "undefined"."

- Count the number of times that "aggressive" behavior is demonstrated and compare across levels of IV.
  "An incident was given one of the four possible labels if 2/3 raters provided the same label. The number of each type of occurrence was summed for each condition.

Inter-rater consensus



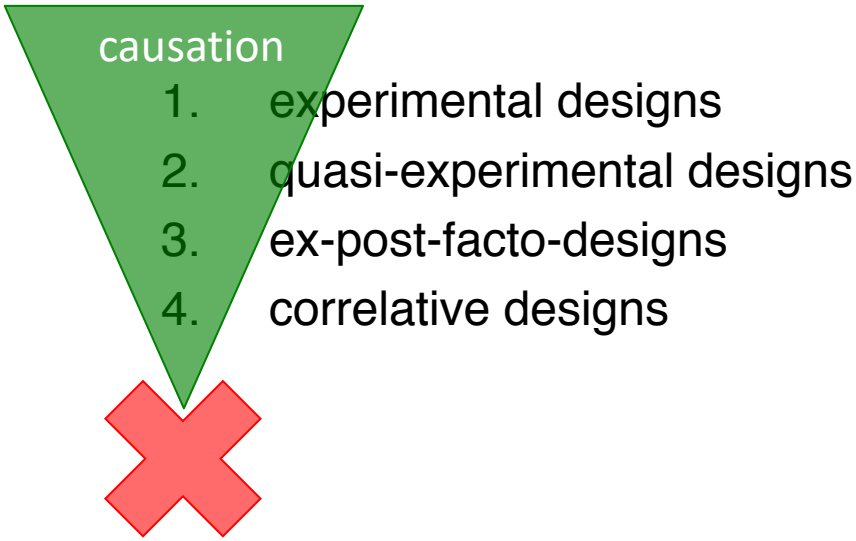
What is an experimental design?

- it is a subgroup of research designs that fulfills certain criteria
- there are at least 4 groups of research designs that differ in the degree in which they allow conclusions about cause-effect relationships:

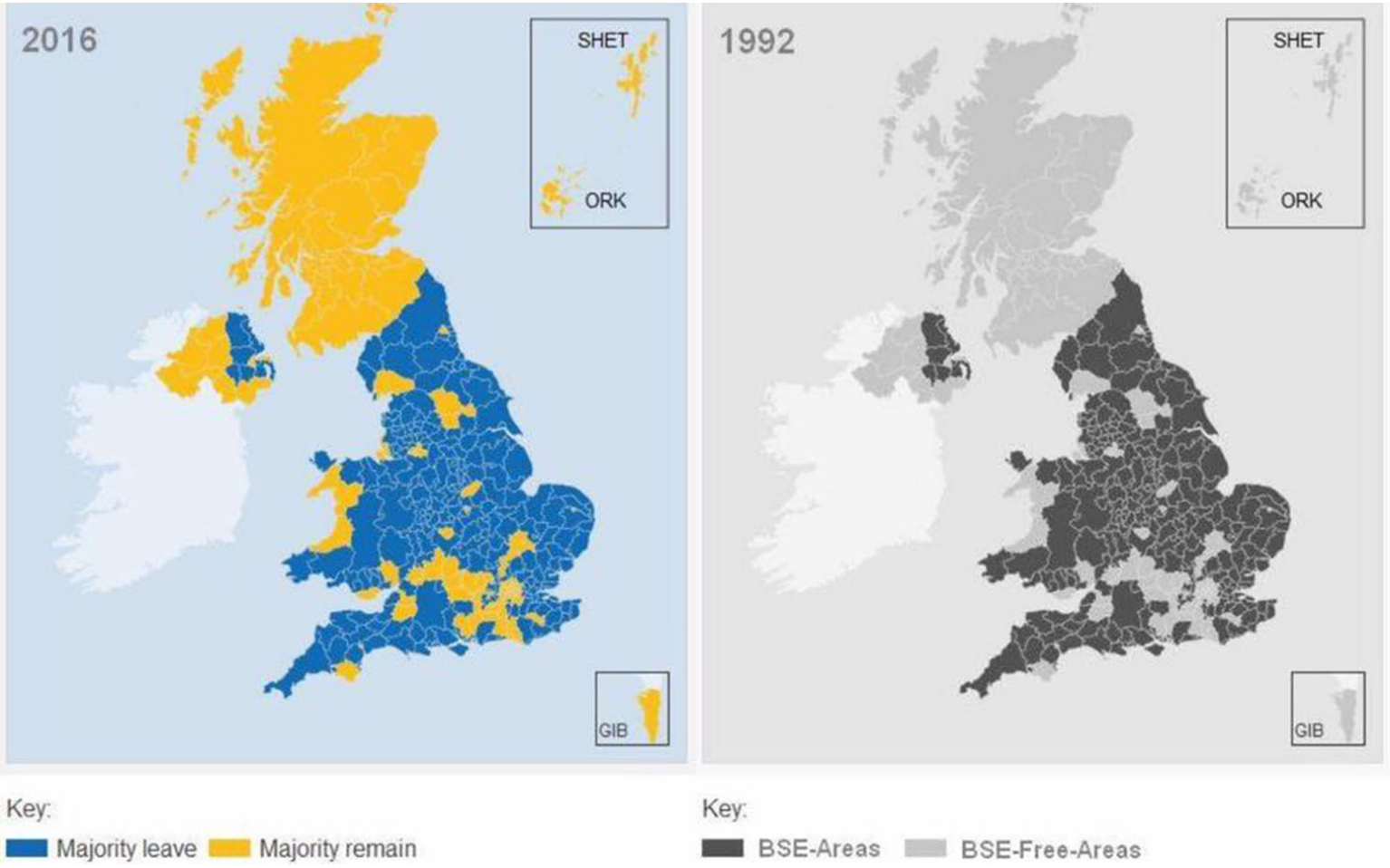
1. experimental designs
2. quasi-experimental designs
3. ex-post-facto-designs
4. correlative designs

What is an experimental design?

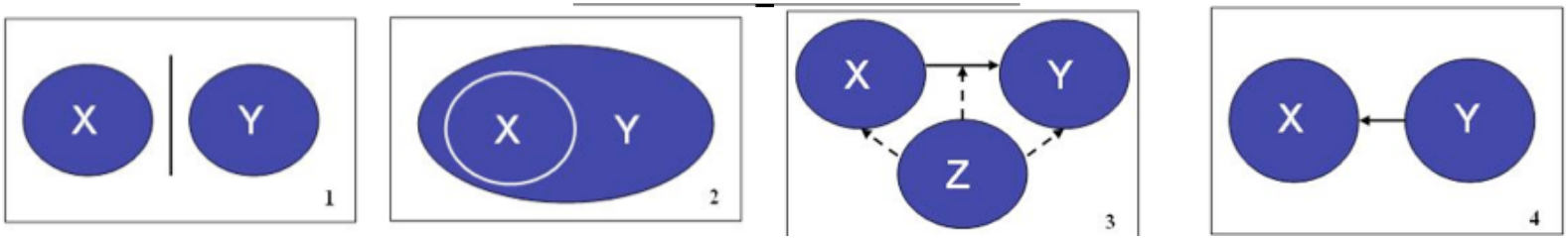
- it is a subgroup of research designs that fulfills certain criteria
- there are at least 4 groups of research designs that differ in the degree in which they allow conclusions about cause-effect relationships:



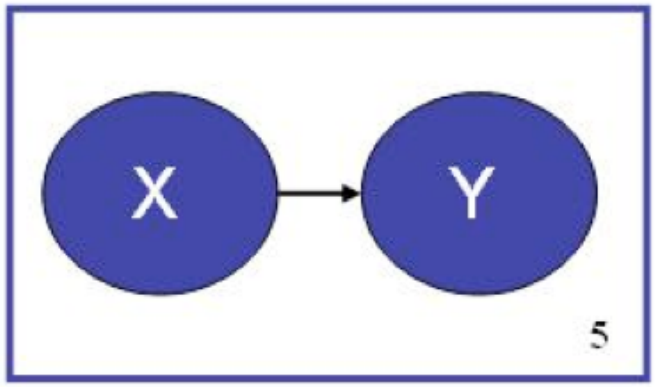
Brexit correlates with Mad Cow Disease



Why is correlation not (necessarily) causation?



1. Coincidence
2. X is an subset of Y
3. Moderator- or Mediator-effect?
   a hidden third variable underlies the relationship.
4. Opposite causation
5. **Causal relationship**



A Story:
How simply measuring stuff is not enough to answer questions

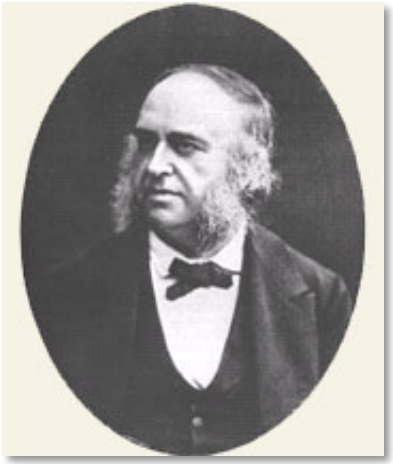
" In the most intelligent races, as among the Parisians, there are a large number of women whose brains are closer in size to those of gorillas than to the most developed male brains... (all psychologists) recognise today that they represent the most inferior forms of human evolution and that they are closer to children & savages than to an adult, civilised "

This is an example of the Observational Method (as opposed to the Experimental Method).

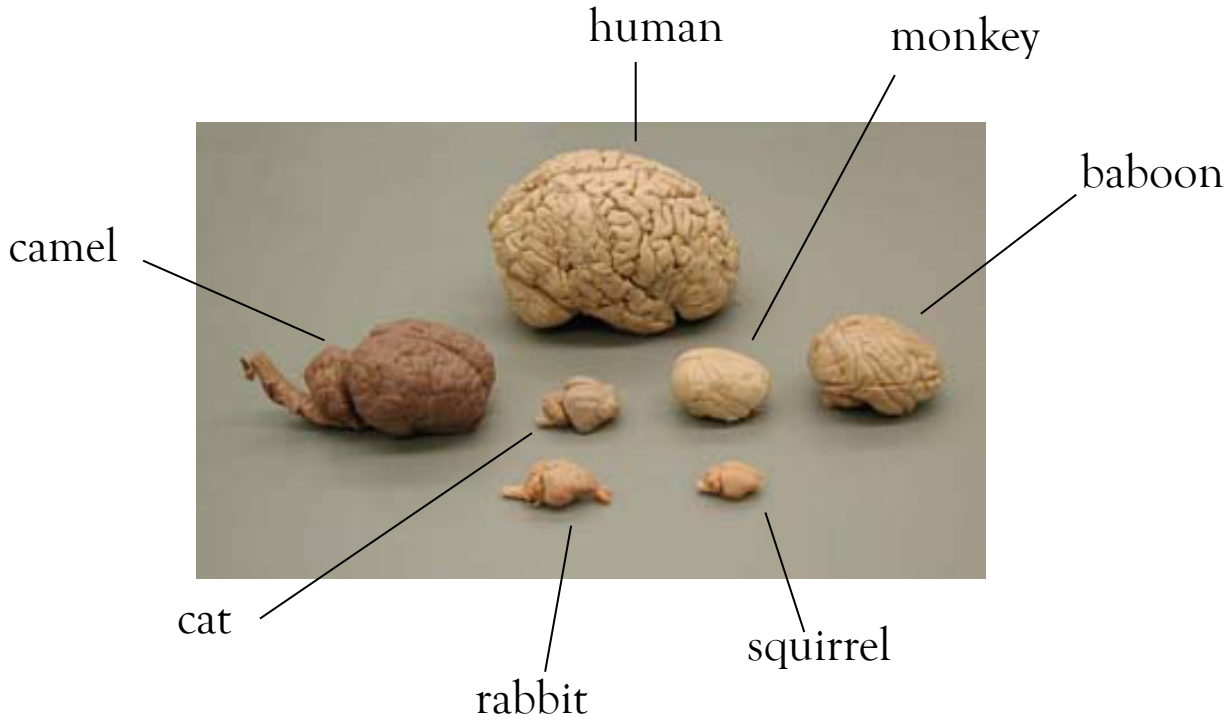
Women, like gorillas, have smaller brains than men

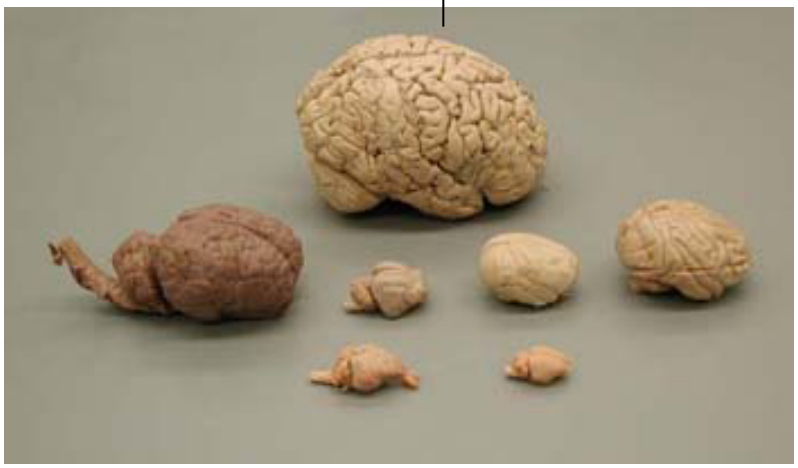
Therefore, women are more like gorillas than men in terms of intelligence

Gustave Le Bon (1841-1931)



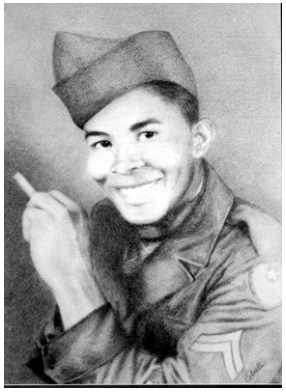
P. Paul Broca
(1824-1925)





Stephen J. Gould
(1941-2002)

VS

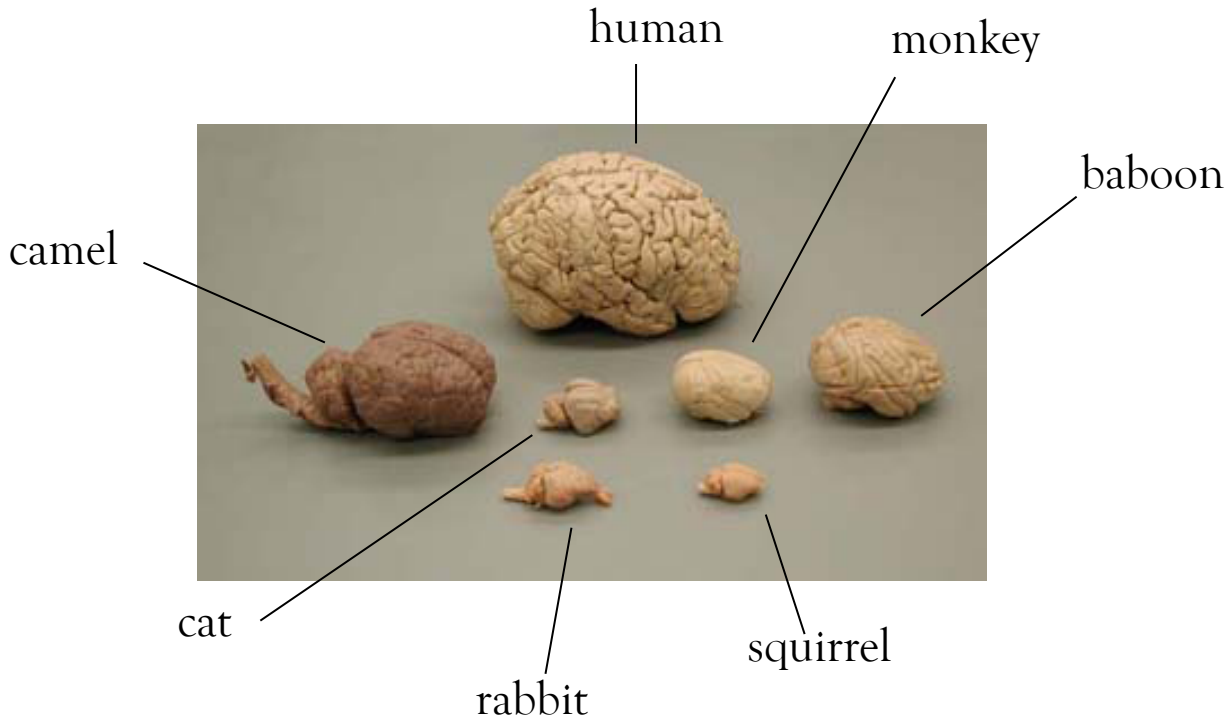


confounding factors (eg, sex, age) refer to factors whose effects on a DV (eg, brain size) cannot be distinguished from each other.

Resistance to appropriate operationalizations

Aristotle
(384 – 322 BC)

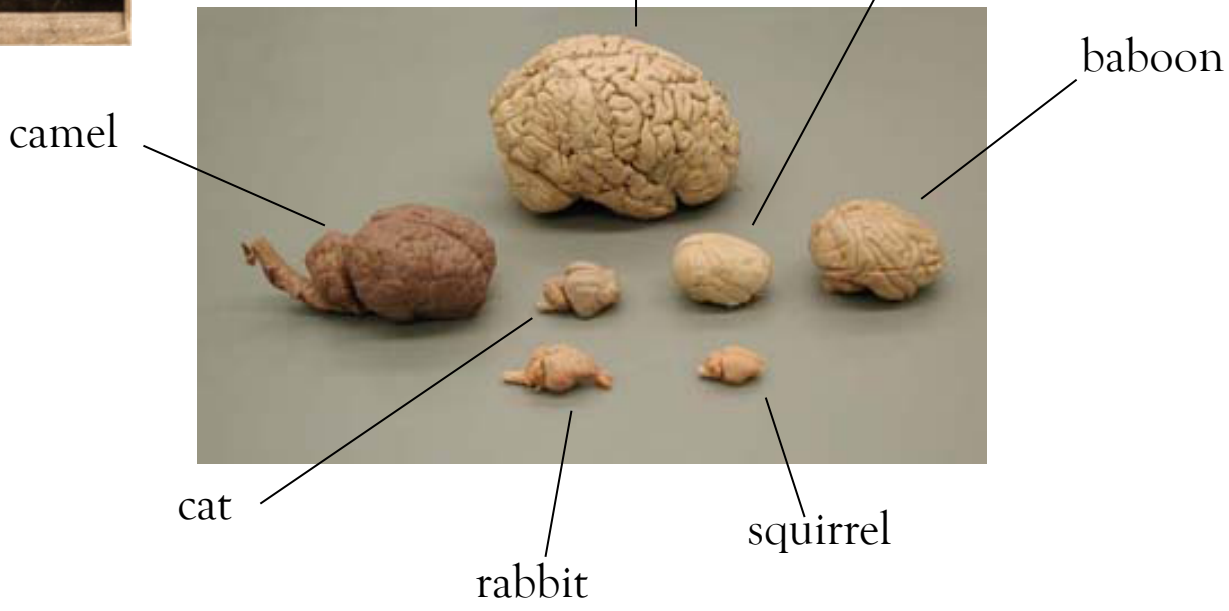
the heart is the seat of intelligence,
the brain is a cooling mechanism…



Effective operationalization for X

Erasistratus (260 BC)

Maybe it is cortical folding of the brain



Resistance to appropriate operationalizations

Maybe it is cortical folding of the brain

Erasistratus (260 BC)

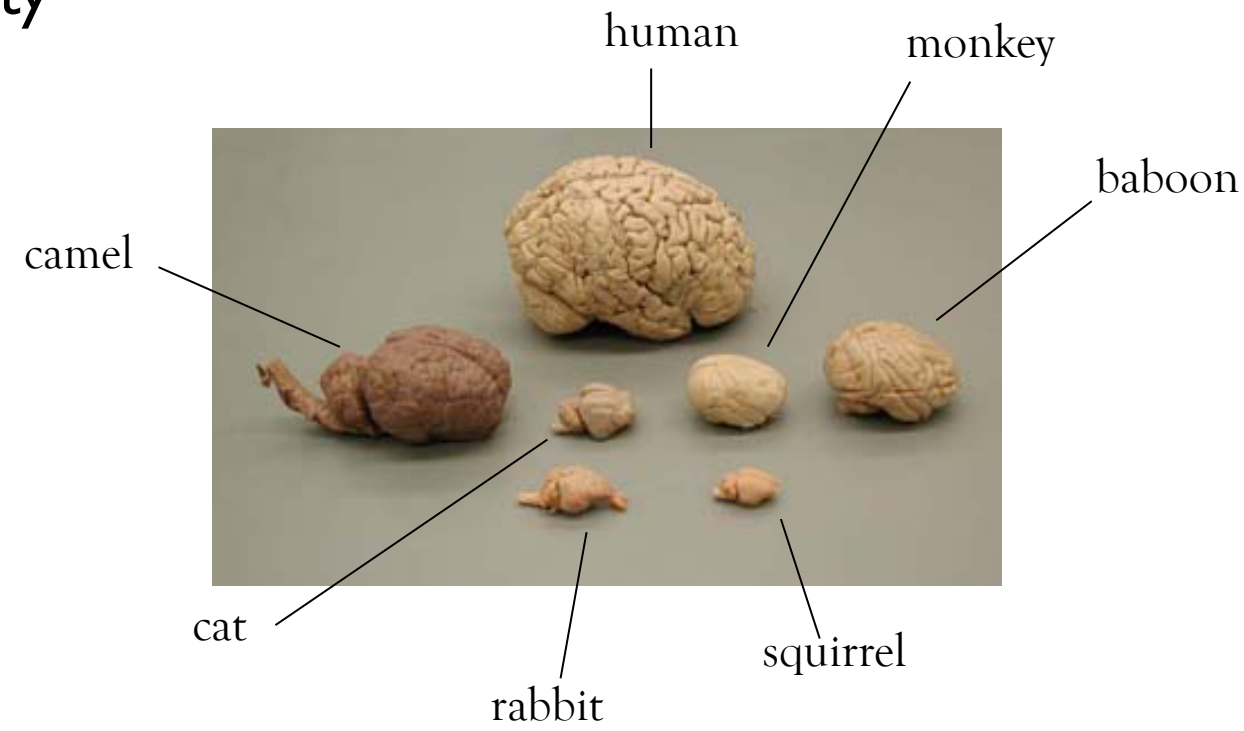
"a donkey has more brain convolutions than humans "

Galen
(AD 129—200)

Effective operationalization for X

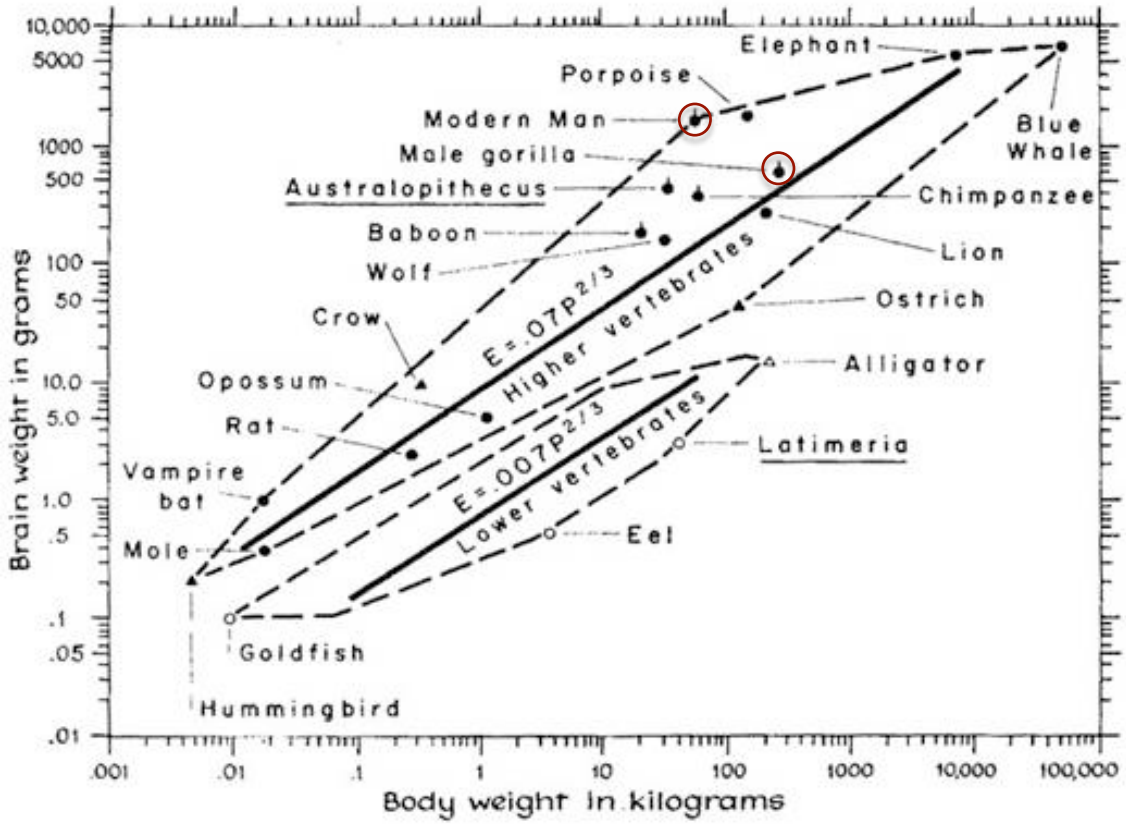
Thomas Willis (1621–1675) and others

- cortical folding:
- organization
- vol. of comparable regions
- functional activity



On the validity of behavioral measures

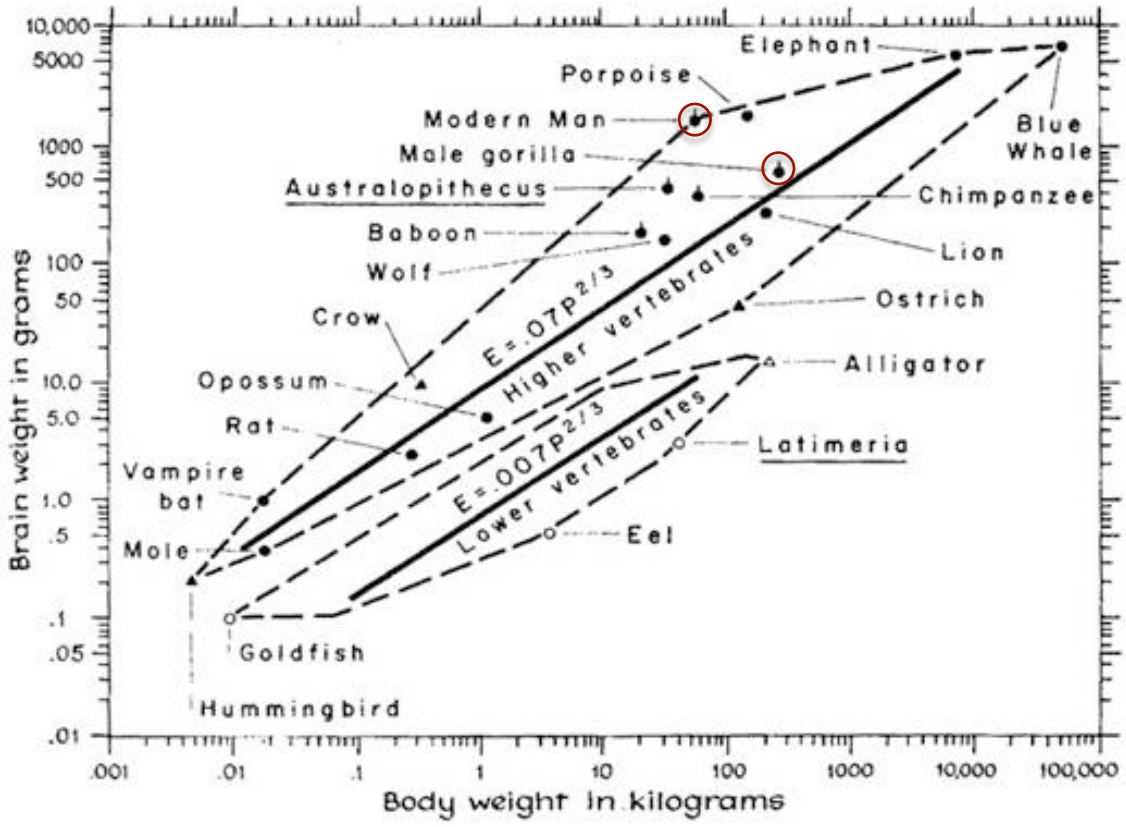
Remove the uninteresting covariate of body weight by measuring brain:body ratio for intelligence, instead of raw brain size



Jerison, H. F. (1983). The evolution of the mammalian brain as an information-processing system. pp. 113-146 IN Eisenberg, J. F. & Kleiman, D. G. (Ed.), Advances in the Study of Mammalian Behavior (Spec. Publ. Amer. Soc. Mamm. 7). Pittsburgh: American Society of Mammalogists.

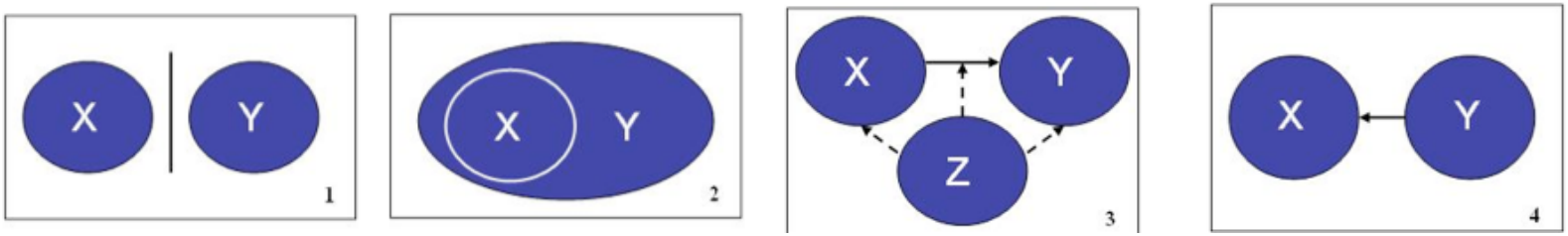
Correlations between body weight and brain size

Does larger body weight cause larger brain weight or does larger brain weight cause larger body weight?

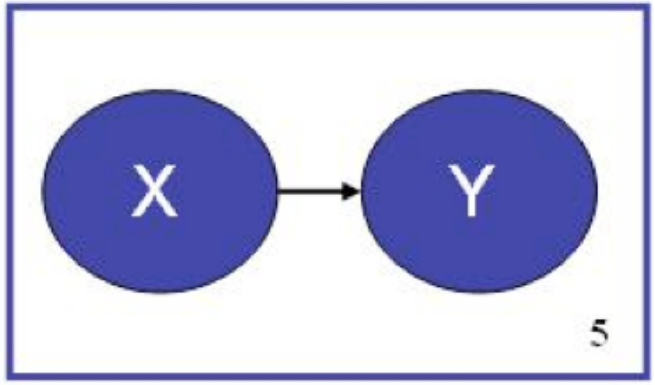


Jerison, H. F. (1983). The evolution of the mammalian brain as an information-processing system. pp. 113-146 IN Eisenberg, J. F. & Kleiman, D. G. (Ed.), Advances in the Study of Mammalian Behavior (Spec. Publ. Amer. Soc. Mamm. 7). Pittsburgh: American Society of Mammalogists.

Why is correlation not (necessarily) causation?



1. Coincidence
2. X is an subset of Y
3. Moderator- or Mediator-effect?
   a hidden third variable underlies the relationship.
4. Opposite causation
5. **Causal relationship**



Brexit correlates with Mad Cow Disease

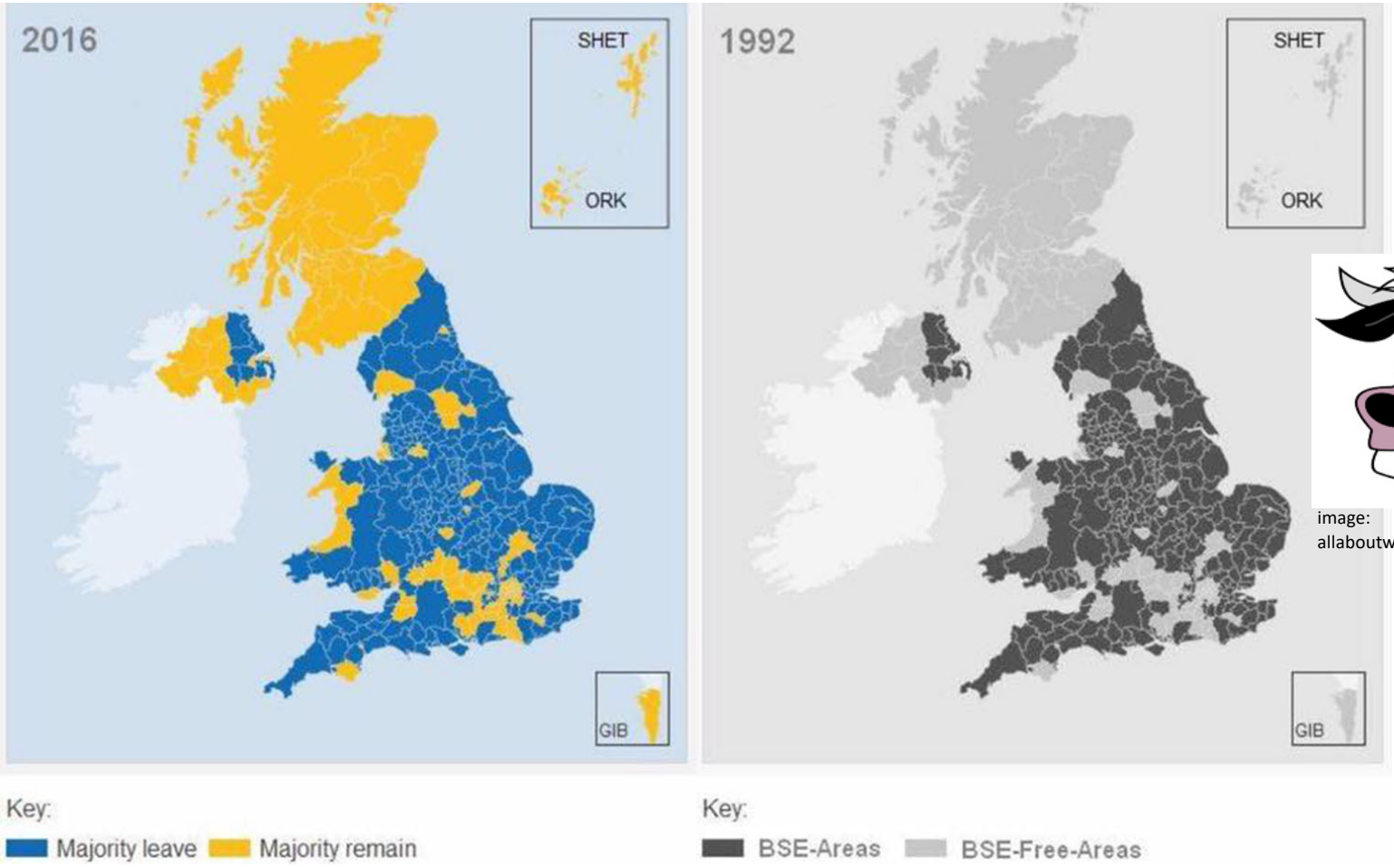


image:
allaboutwindowsphone.com

Correlation is not necessarily causation

Between- & within- subjects designs

- Between-subjects design
  - One partcipant only completes the task in one condition
- Within-subjects design
  - Participants complete the task in multiple conditions

Between- & within- subjects designs

- Between-subjects design
  - One partcipant only completes the task in one condition
- Within-subjects design
  - Participants complete the task in multiple conditions
- But it could also be a mixed design
- And don’t forghet repetiton

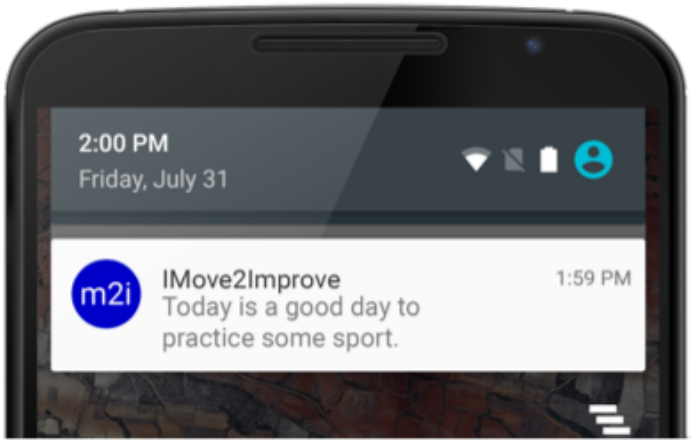
Between- & within- subjects designs

- Between-subjects design
  - One partcipant only completes the task in one condition
- Within-subjects design
  - Participants complete the task in multiple conditions
- But it could also be a mixed design
- And don't forget repetitons

Between-subjects design

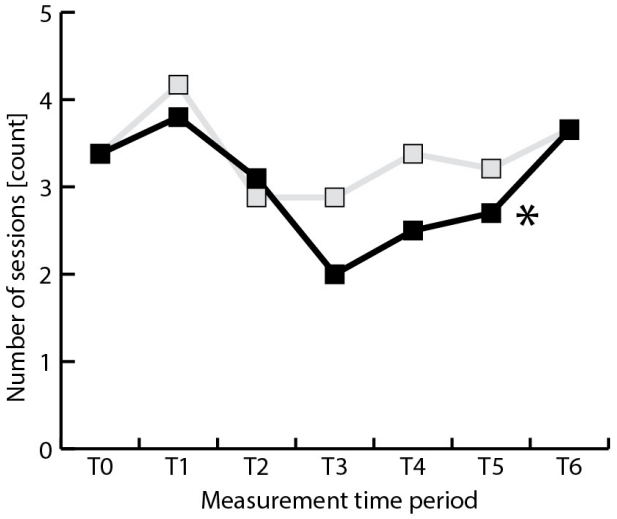
Within-subjects design

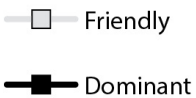
Mixed design

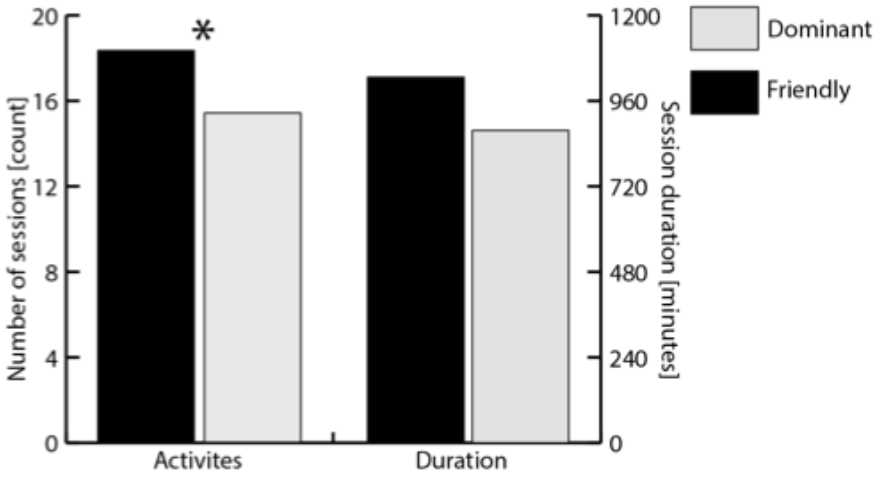


The picture can't be displayed.

Mixed design







Order effects

- People get better at stuff
- People get tired
- This can really mess with our experiment!

- How do we deal with that?

Order effects

- No obvious answer
- Classic way: Latin squares

$$[1] \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \\ 3 & 1 & 4 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 5 & 1 & 4 \\ 3 & 5 & 4 & 2 & 1 \\ 4 & 1 & 2 & 5 & 3 \\ 5 & 4 & 1 & 3 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 5 & 3 \\ 3 & 5 & 4 & 2 & 1 \\ 4 & 1 & 5 & 3 & 2 \\ 5 & 3 & 2 & 1 & 4 \end{bmatrix}$$

Order effects

The picture can't be displayed.

Order effects