

CAUSAL GRAPHS ***WILL SAVE US ALL FROM*** ***BIG DATA***



<https://chield.excd.org>

Seán Roberts

excd.lab, University of Bristol

Learning outcomes

What are causal graphs?

Why are they useful?

- Observation vs. Intervention

- Identify confounders

- Prevent spurious correlations

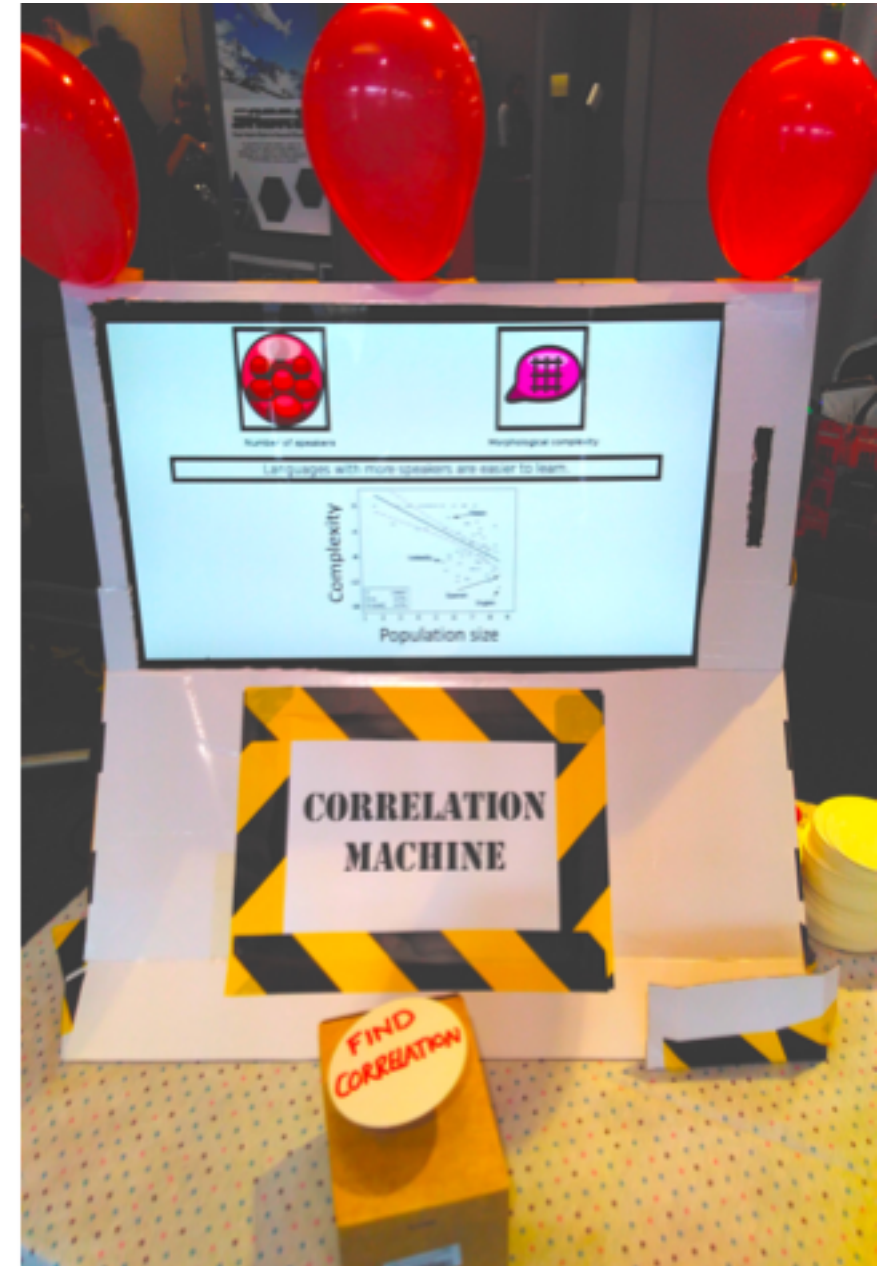
- Dealing with unobservable variables

How can causal graphs be applied in practice?

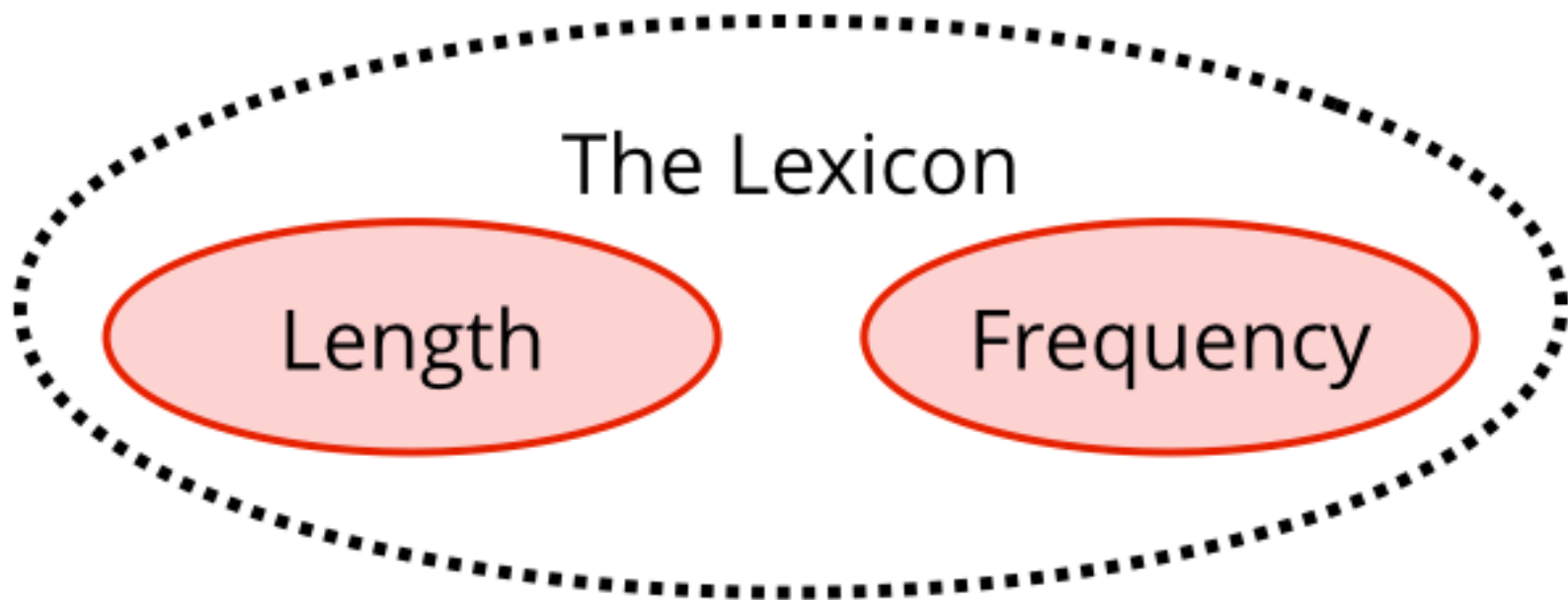
- Quick overview

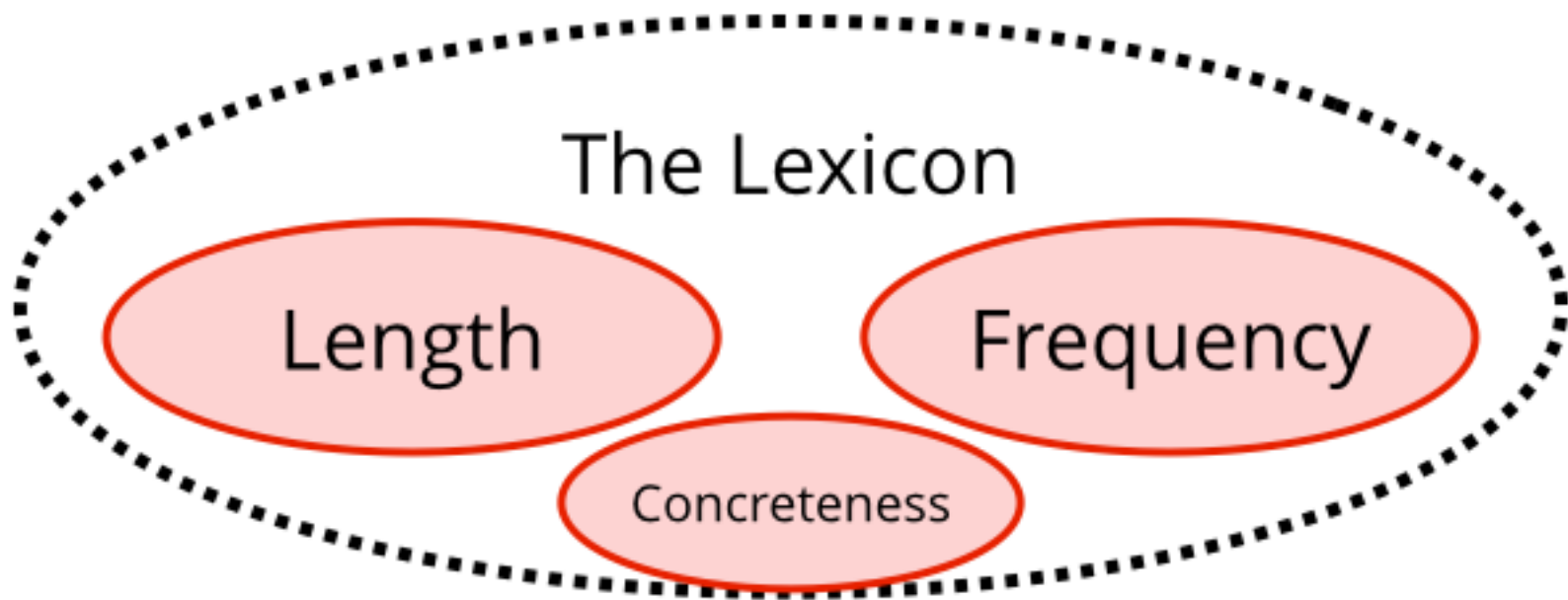
My Database of Causal Theories (CHIELD)

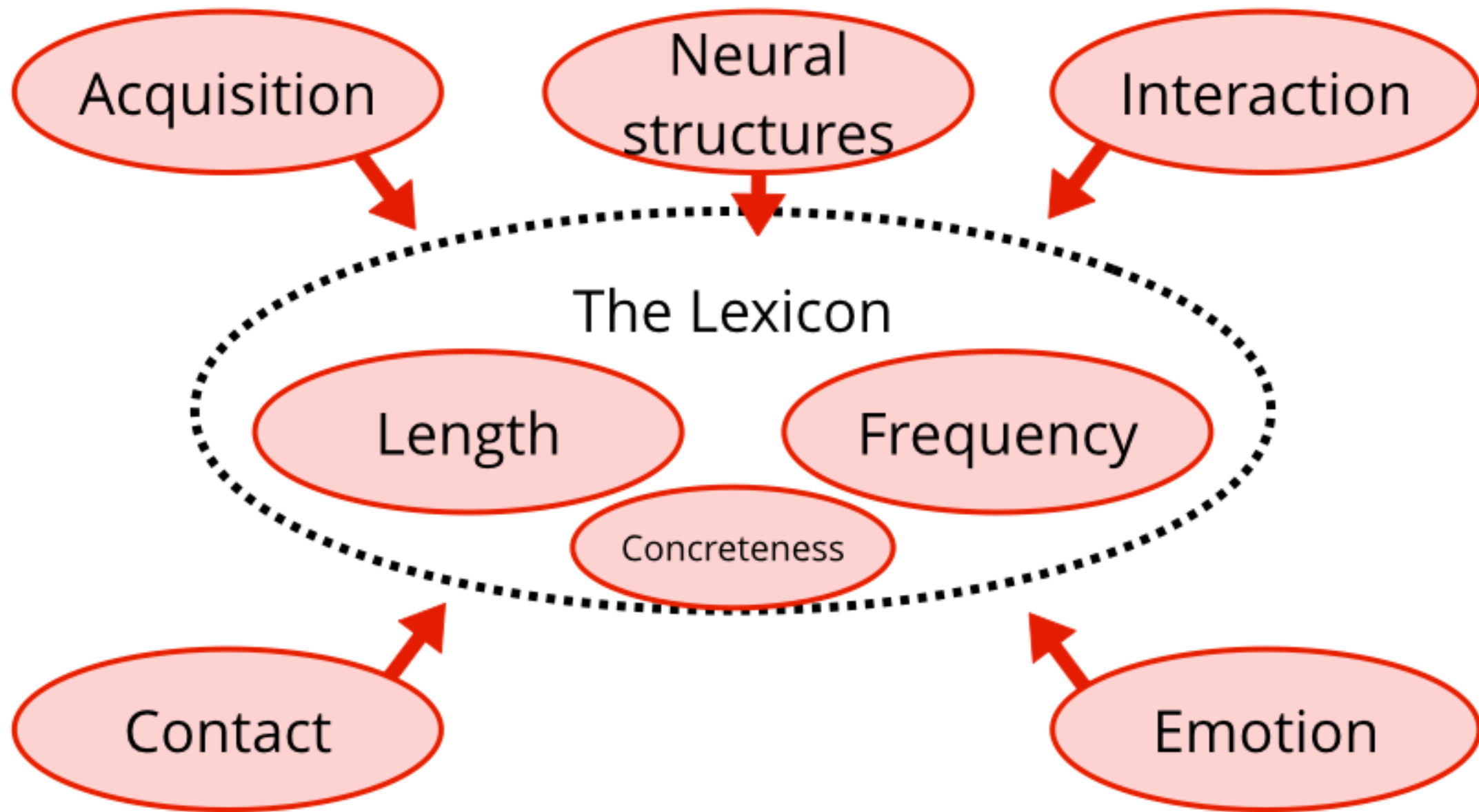
THE CORRELATION MACHINE

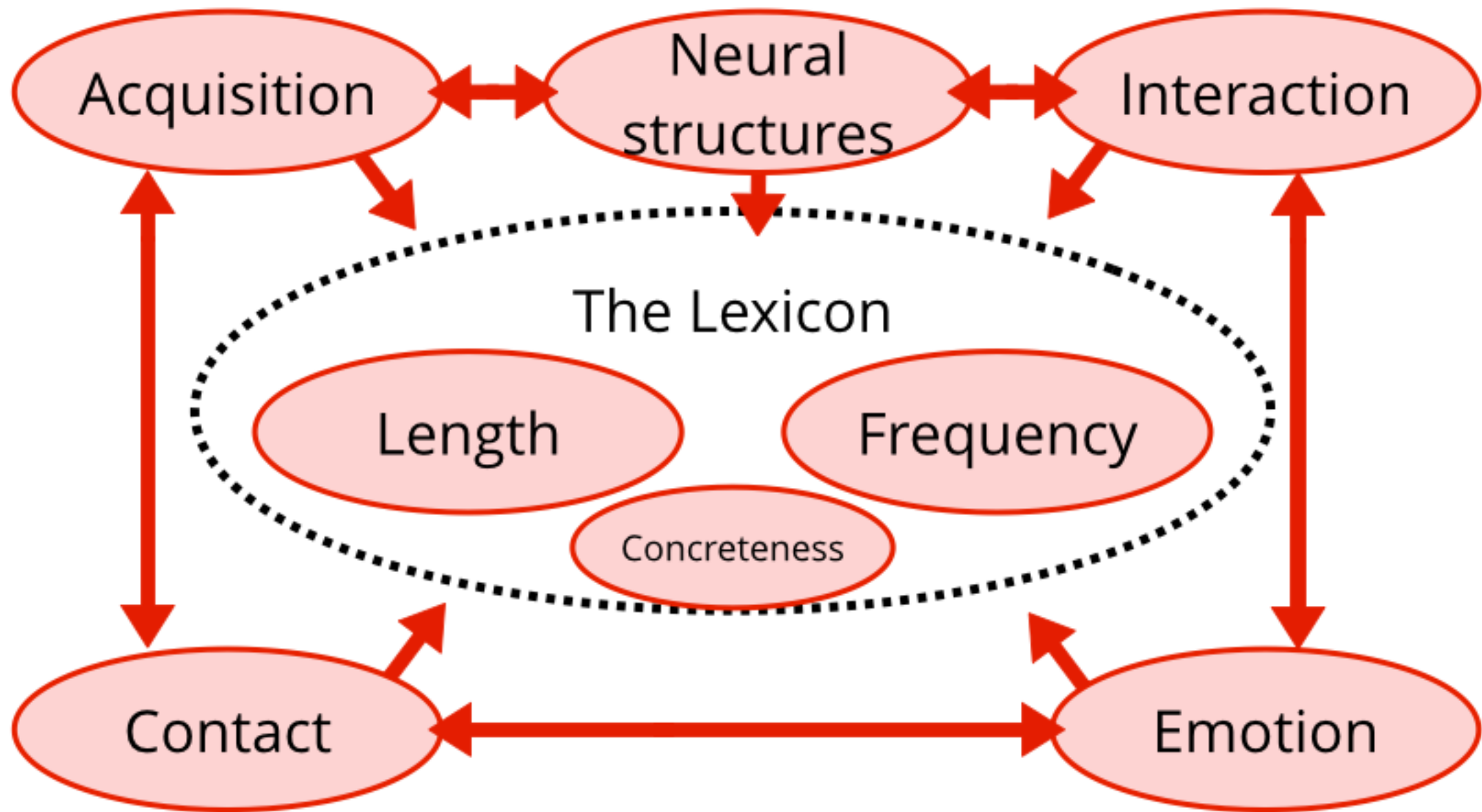




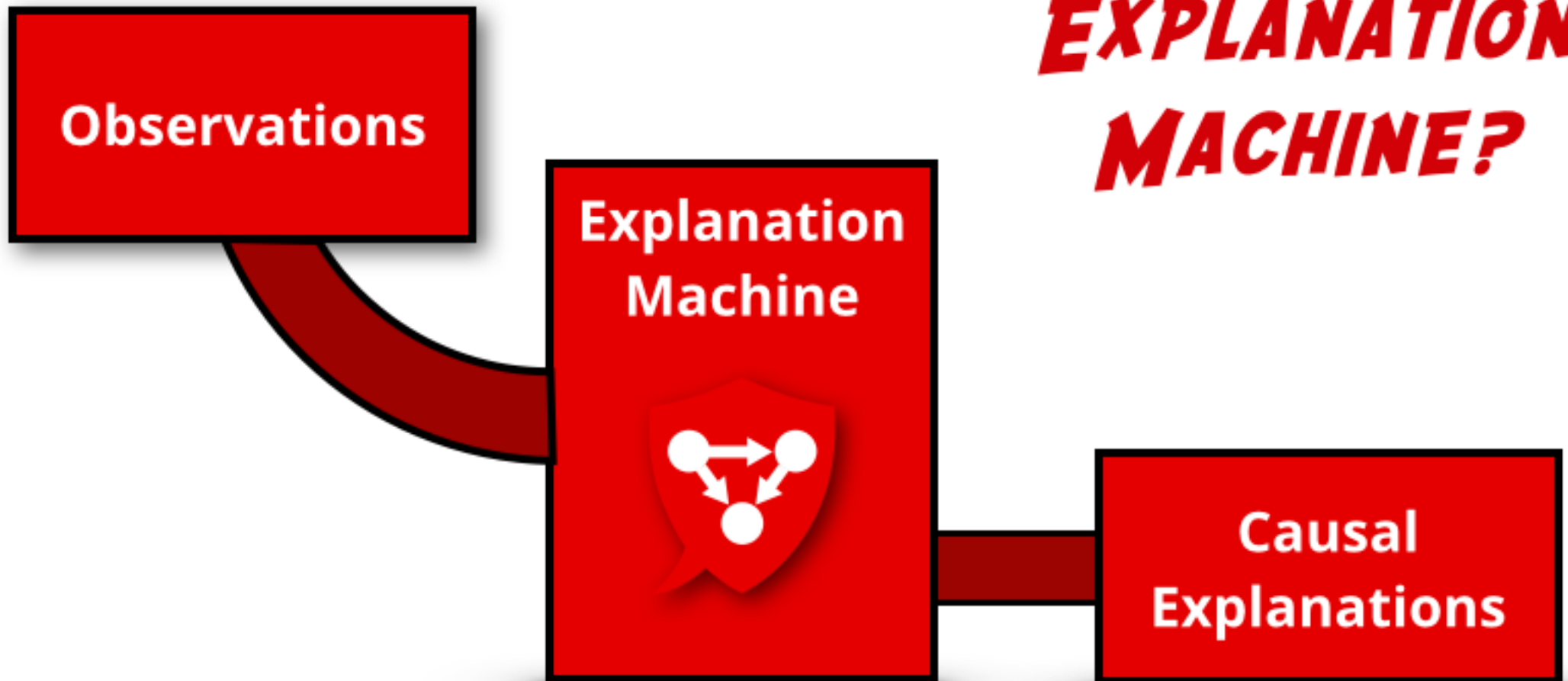








THE EXPLANATION MACHINE?





Francis Galton

Correlation does
not imply
causation

Correlation,
in the absence of
alternative explanations,
does imply
causation



Judea Pearl

Observation vs. Intervention

Observation ("seeing")

Can look up in a probability table

$P(\text{stain} \mid \text{icecream})$

$$= 125/(125+25) = 83\%$$

Intervention ("doing")

What is the probability of a stain if we force someone to have an icecream?

$P(\text{stain} \mid \text{do}(\text{icecream}))$

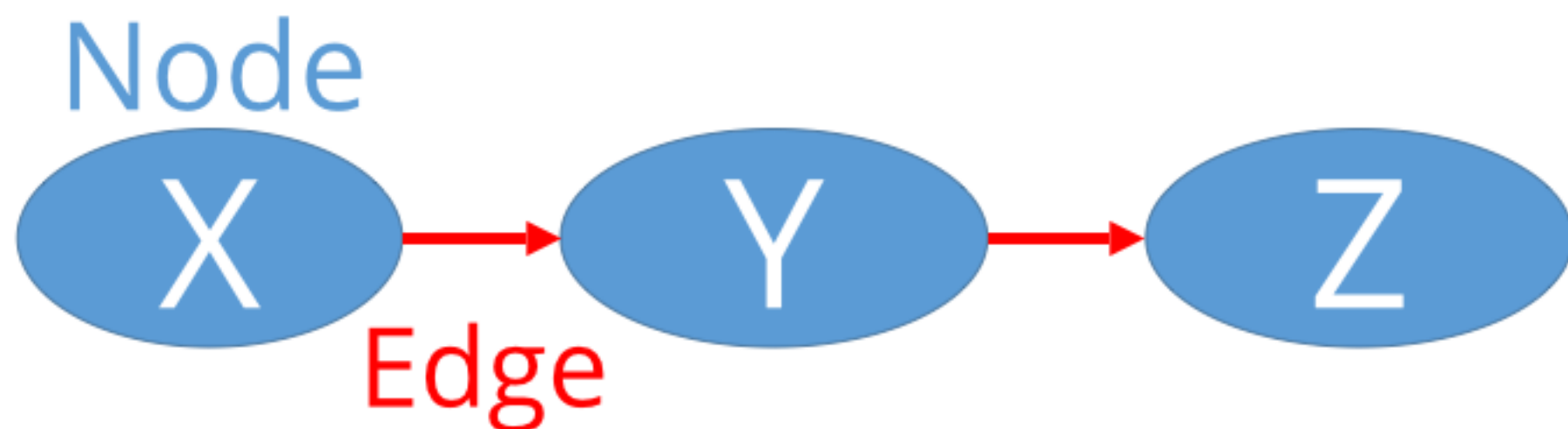
Randomised control experiment

<i>Beach survey</i>	<i>T-shirt stains</i>	<i>No T-shirt</i>
<i>Ice cream</i>	125	25
<i>No Ice cream</i>	73	77



What are causal graphs?

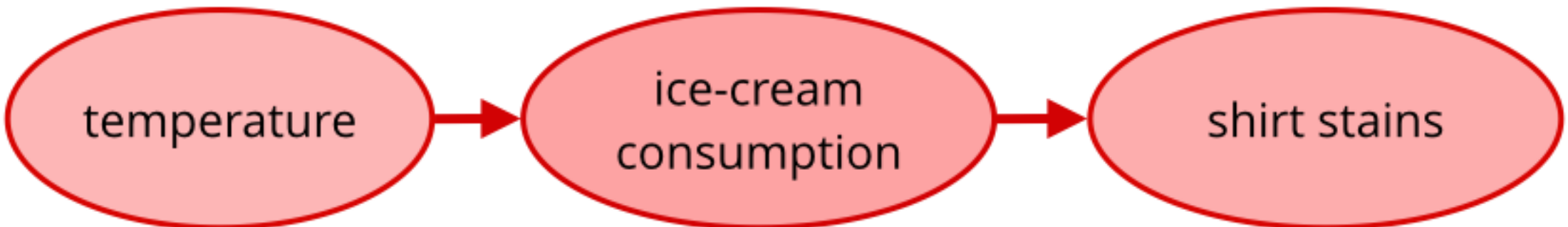
Causal graphs are a mathematical language which helps us think clearly



If we manipulated X in a suitable way, Y would change
Could be absolute, statistical, categorical ...

Causal graphs are mathematical expressions which help us think clearly about causality

"We hypothesise that high temperatures cause more ice-cream consumption, and more ice-cream consumption leads to more shirt stains."



```
graph LR; A([temperature]) --> B([ice-cream consumption]); B --> C([shirt stains]);
```

temperature

ice-cream
consumption

shirt stains

If we manipulated temperature in a suitable way,
ice-cream consumption would change.
Could be absolute, statistical, categorical ...

Causal graphs are mathematical expressions which help us think clearly about causality

temperature

ice-cream
consumption

shirt stains

temperature

ice-cream
consumption

shirt stains

Causal graphs are mathematical expressions which help us think clearly about causality

temperature

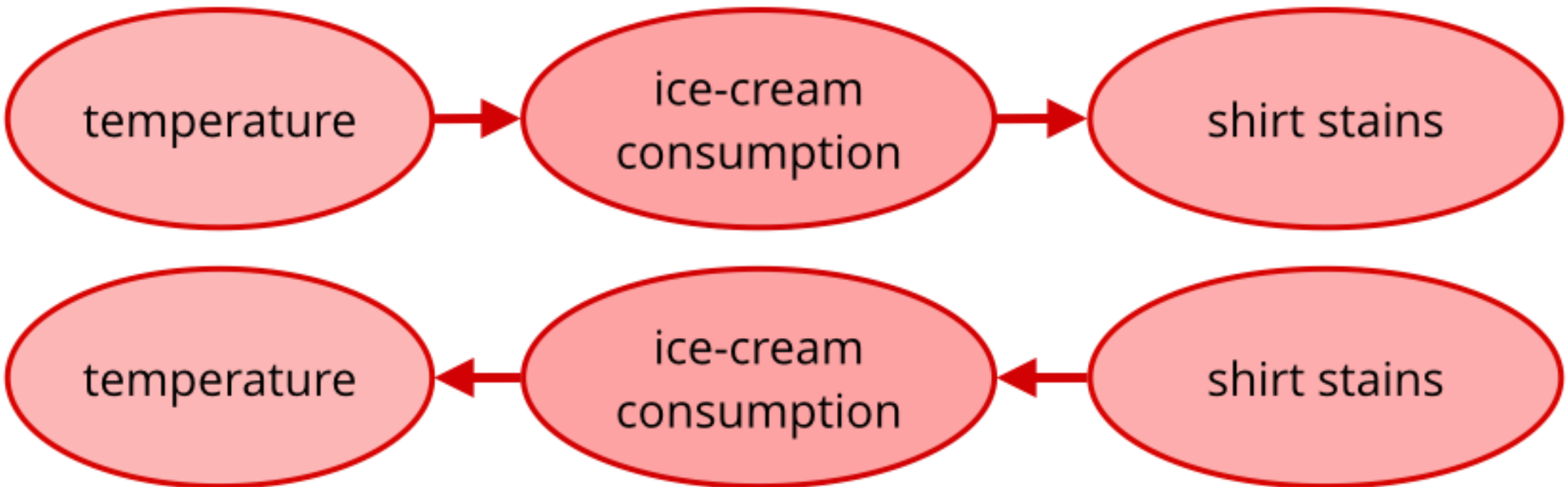
ice-cream
consumption

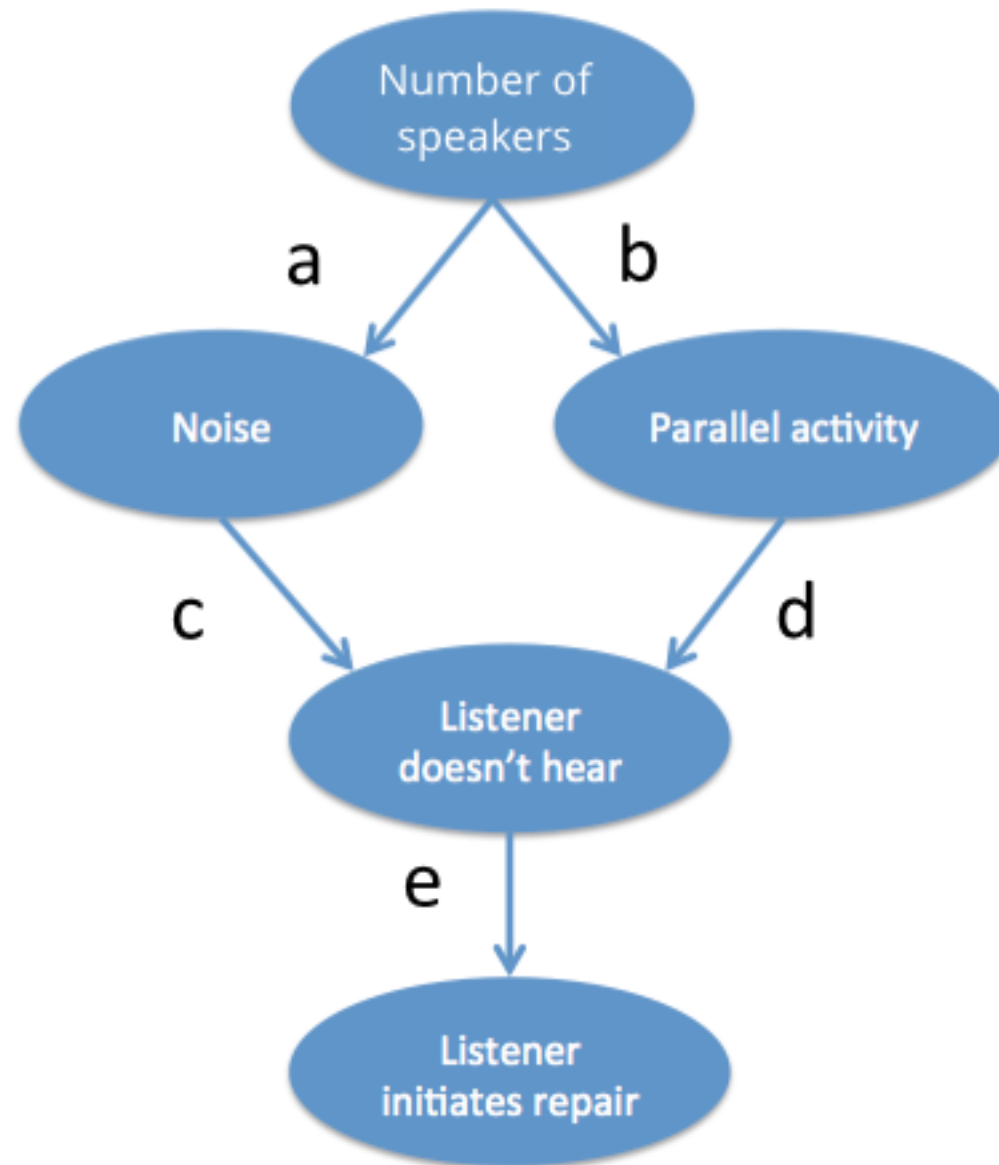
shirt stains

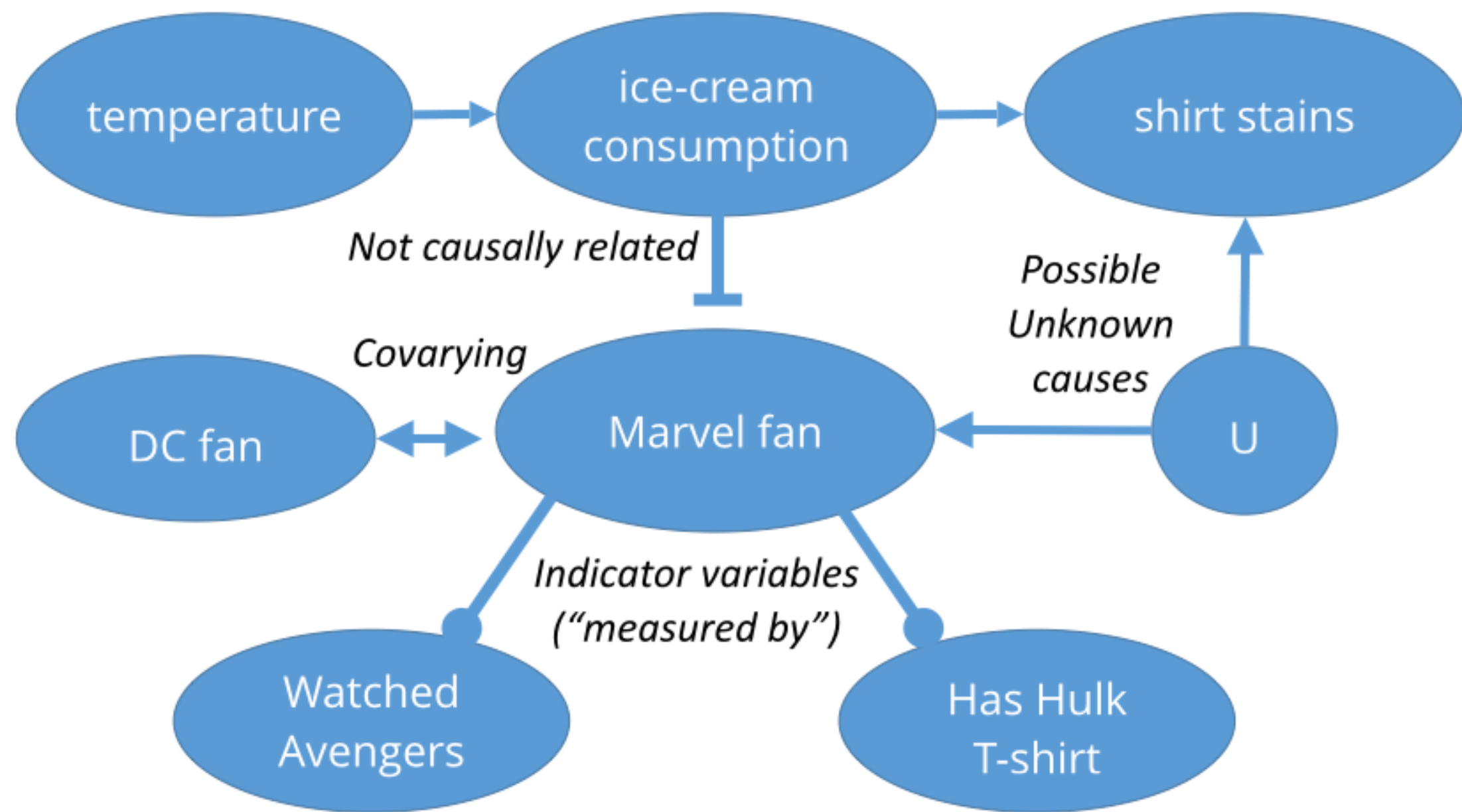
temperature

ice-cream
consumption

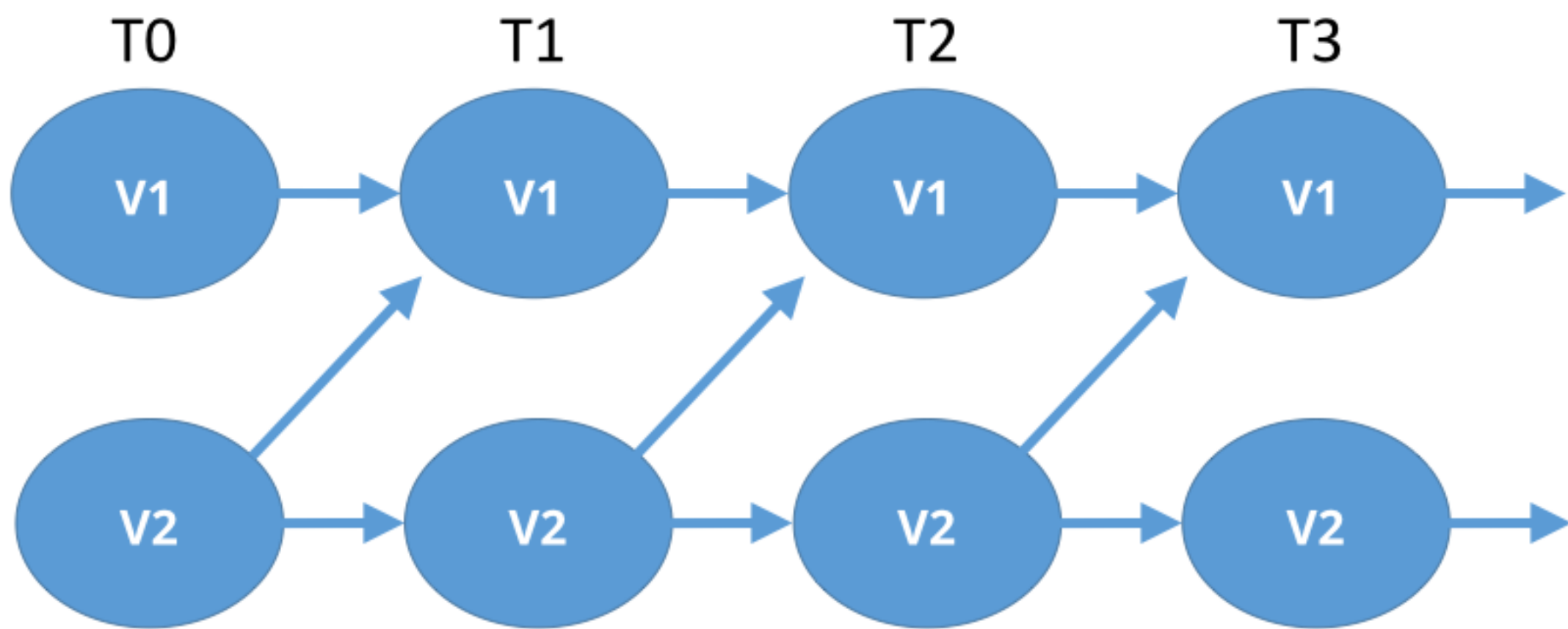
shirt stains







Time



Why use causal graphs?

Look at intervention, not just observation

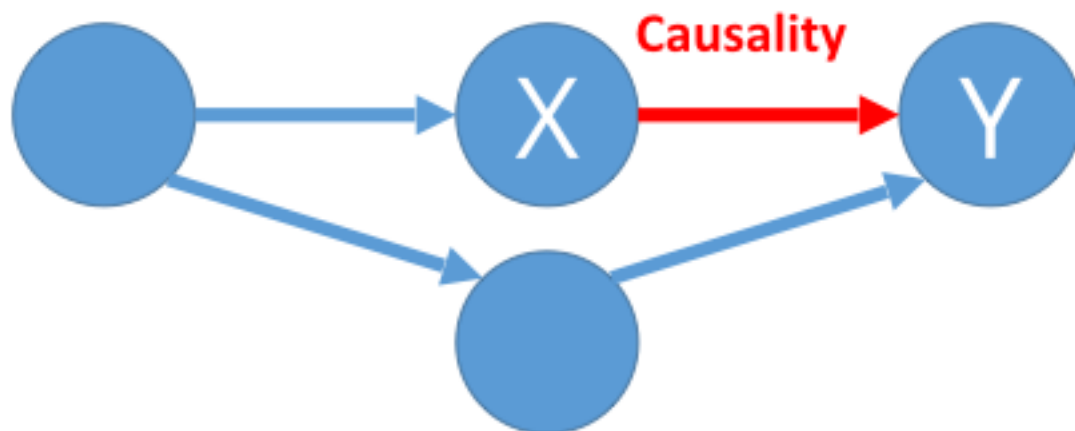
Confounding: Determine which variables to control for

Spurious correlations: Deal with colliders

Confounding

Causality follows the arrows

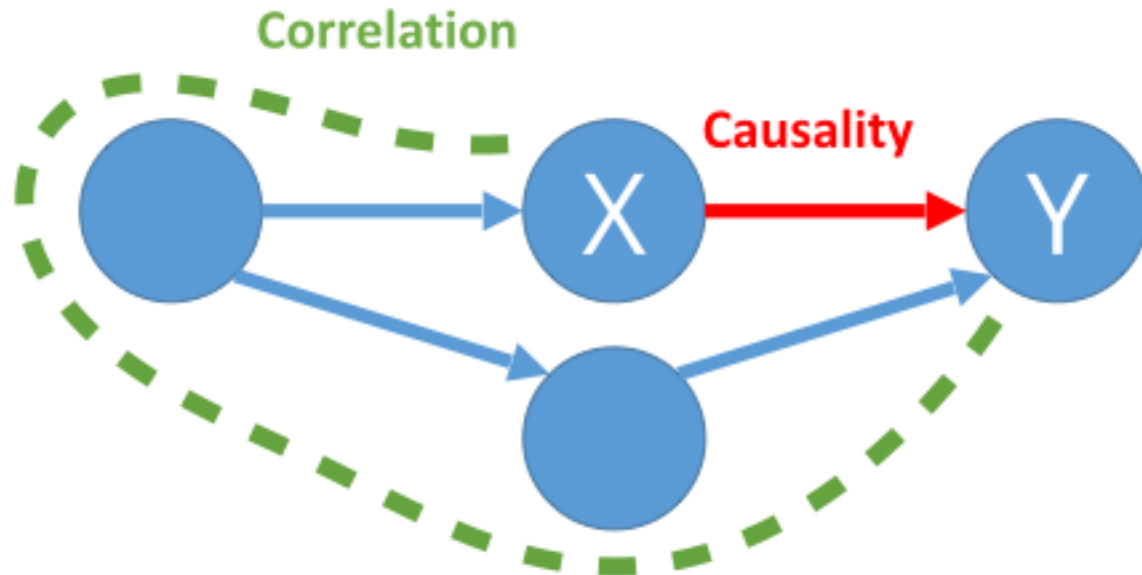
But correlation (*noncausal* information) flows both ways



Confounding

Causality follows the arrows

But correlation (*noncausal* information) flows both ways

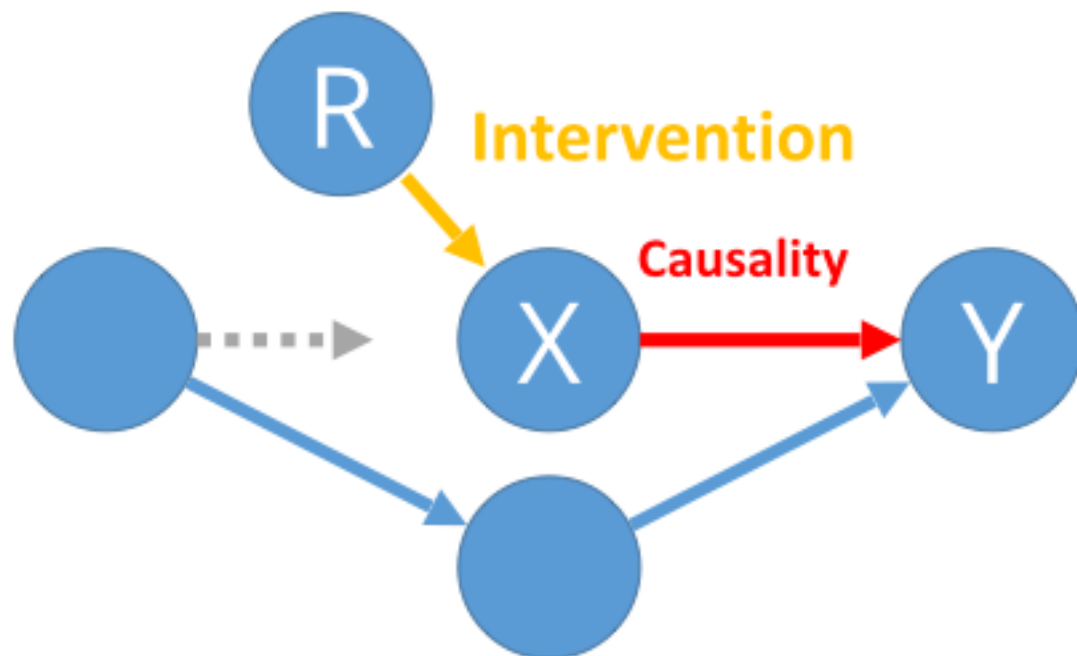


Confounding

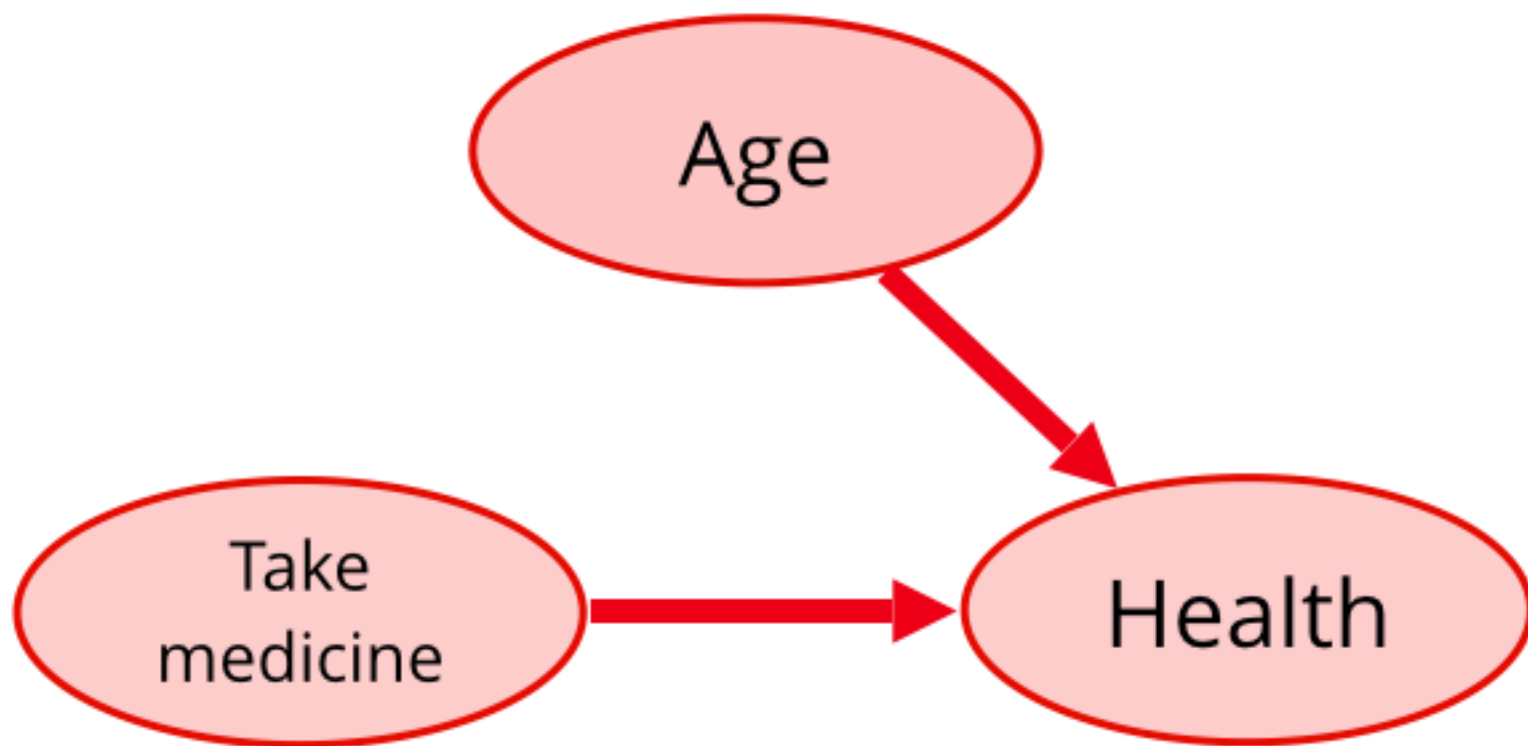
Causality follows the arrows

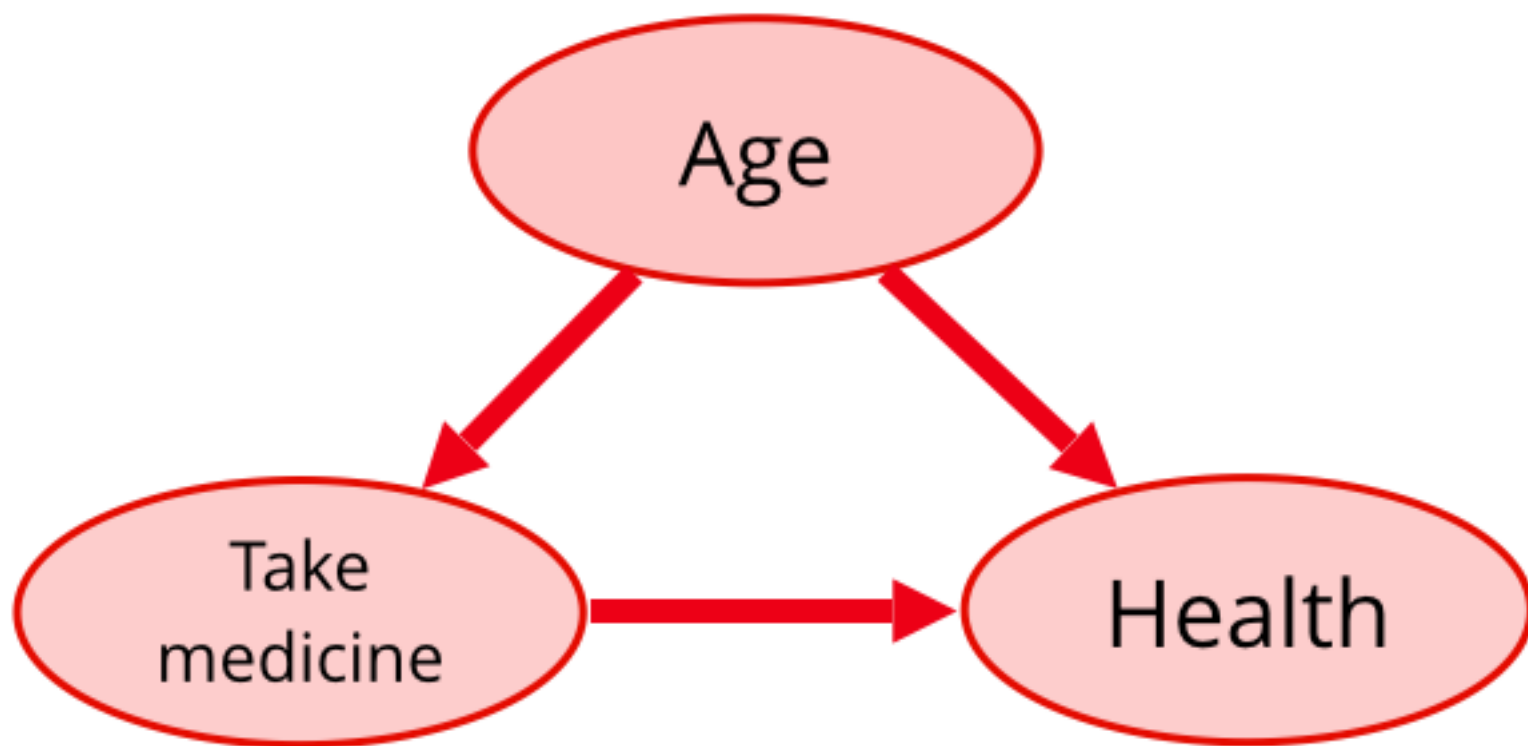
But correlation (*noncausal* information) flows both ways

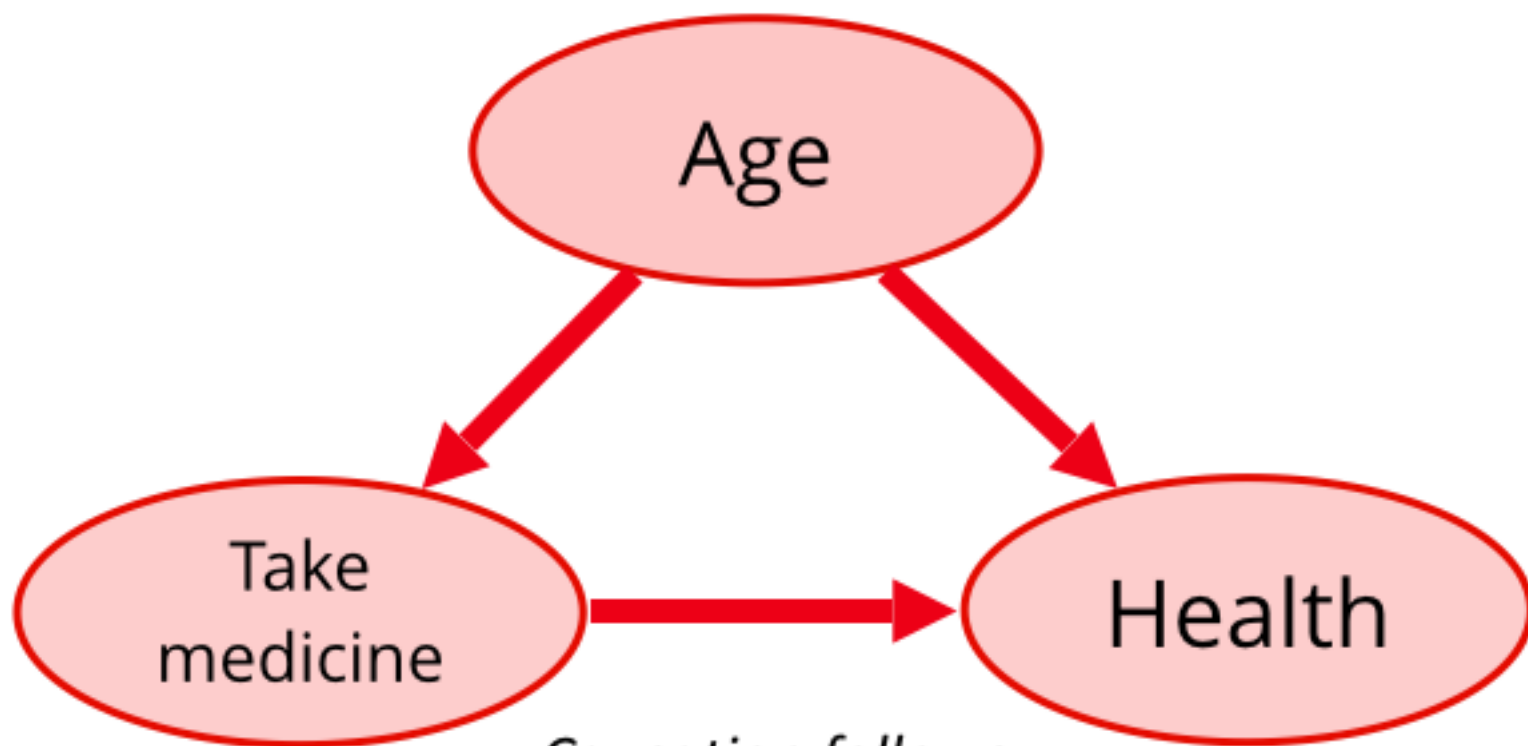
Intervention breaks the flow of correlations



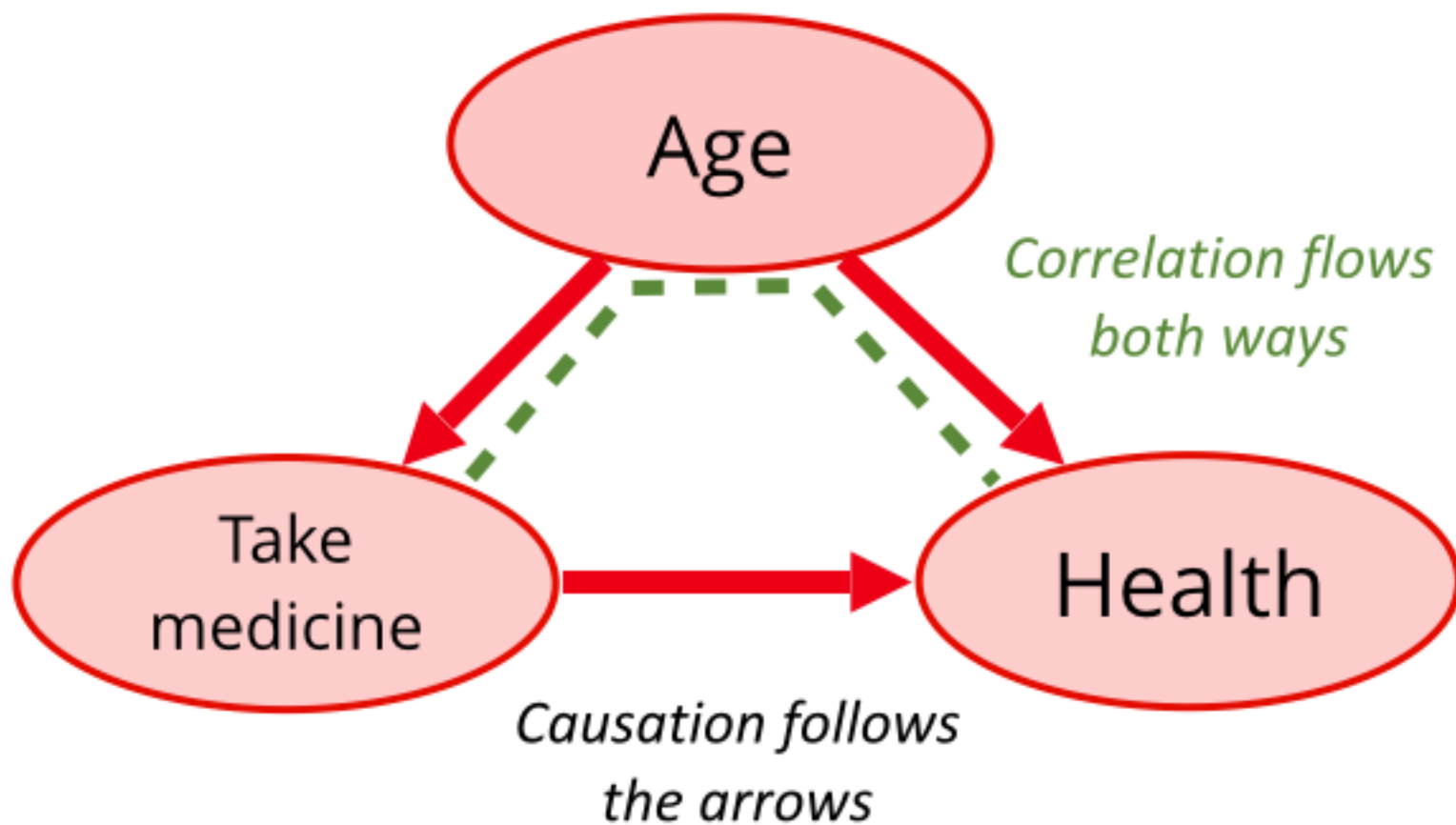




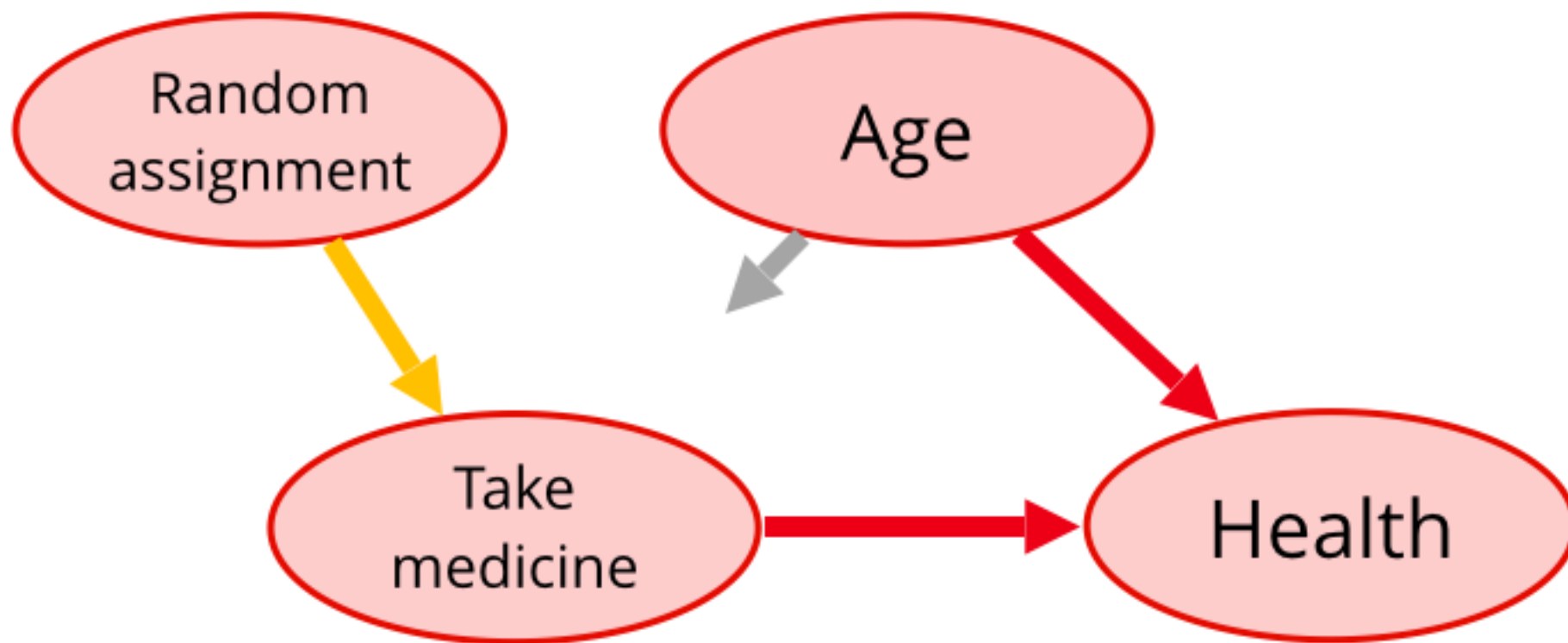




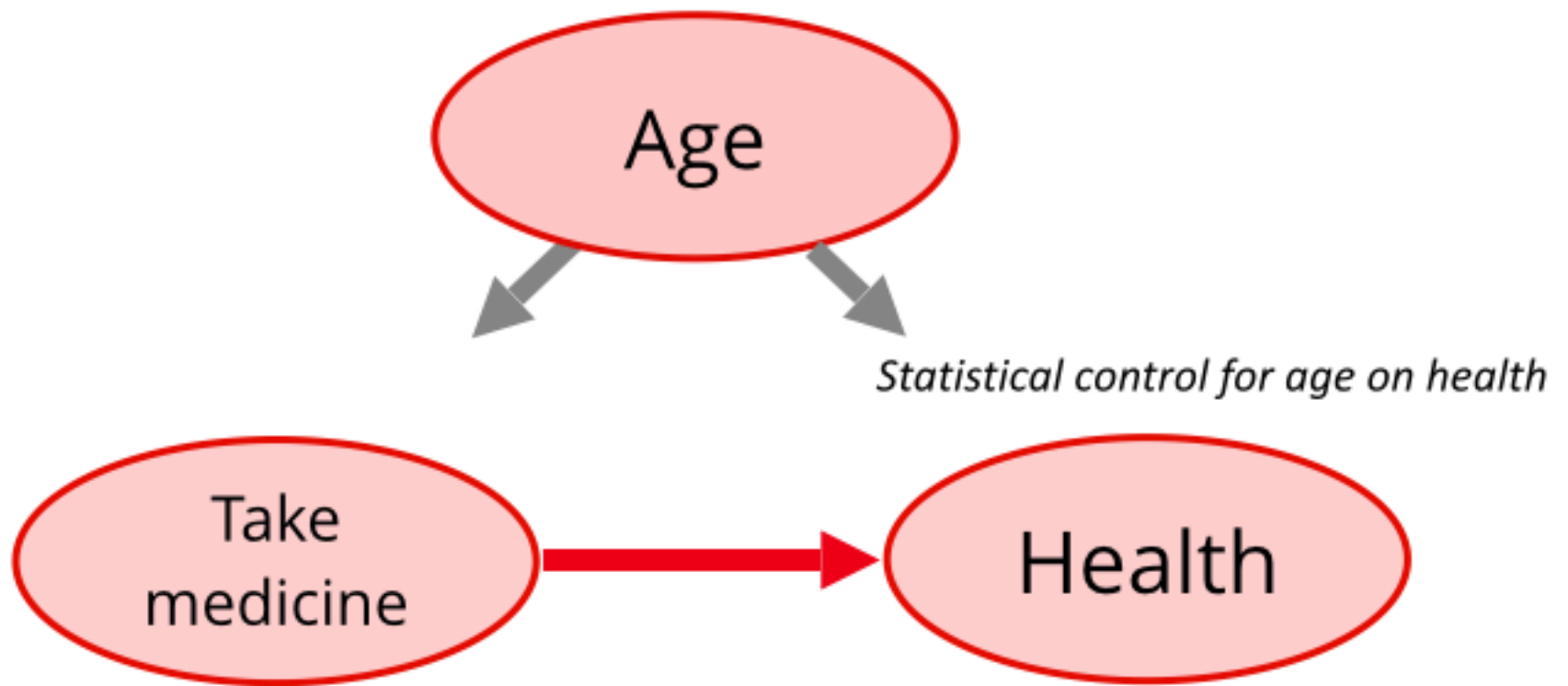
*Causation follows
the arrows*



Randomised control experiment



Statistical control



X is correlated with Y?

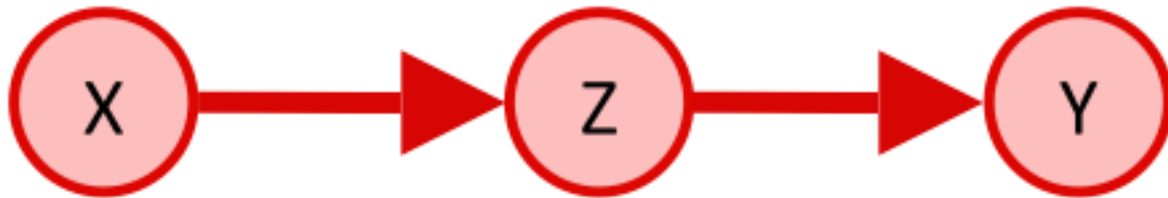


Yes

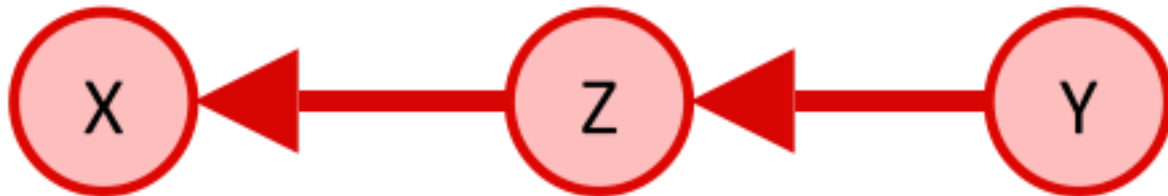


Yes

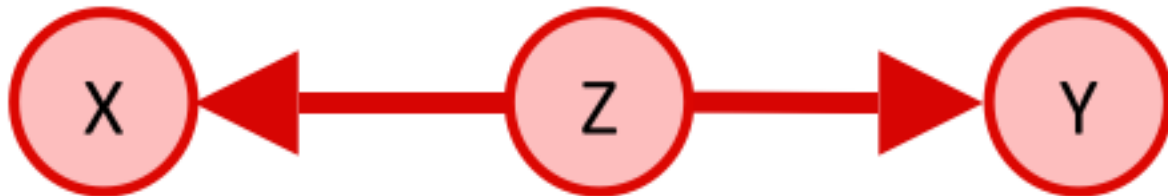
X is correlated with Y?



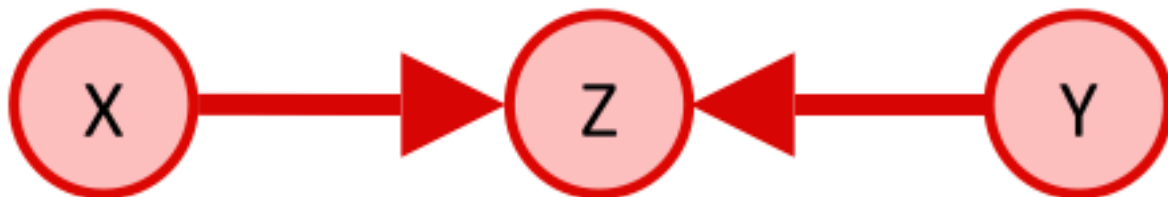
Yes



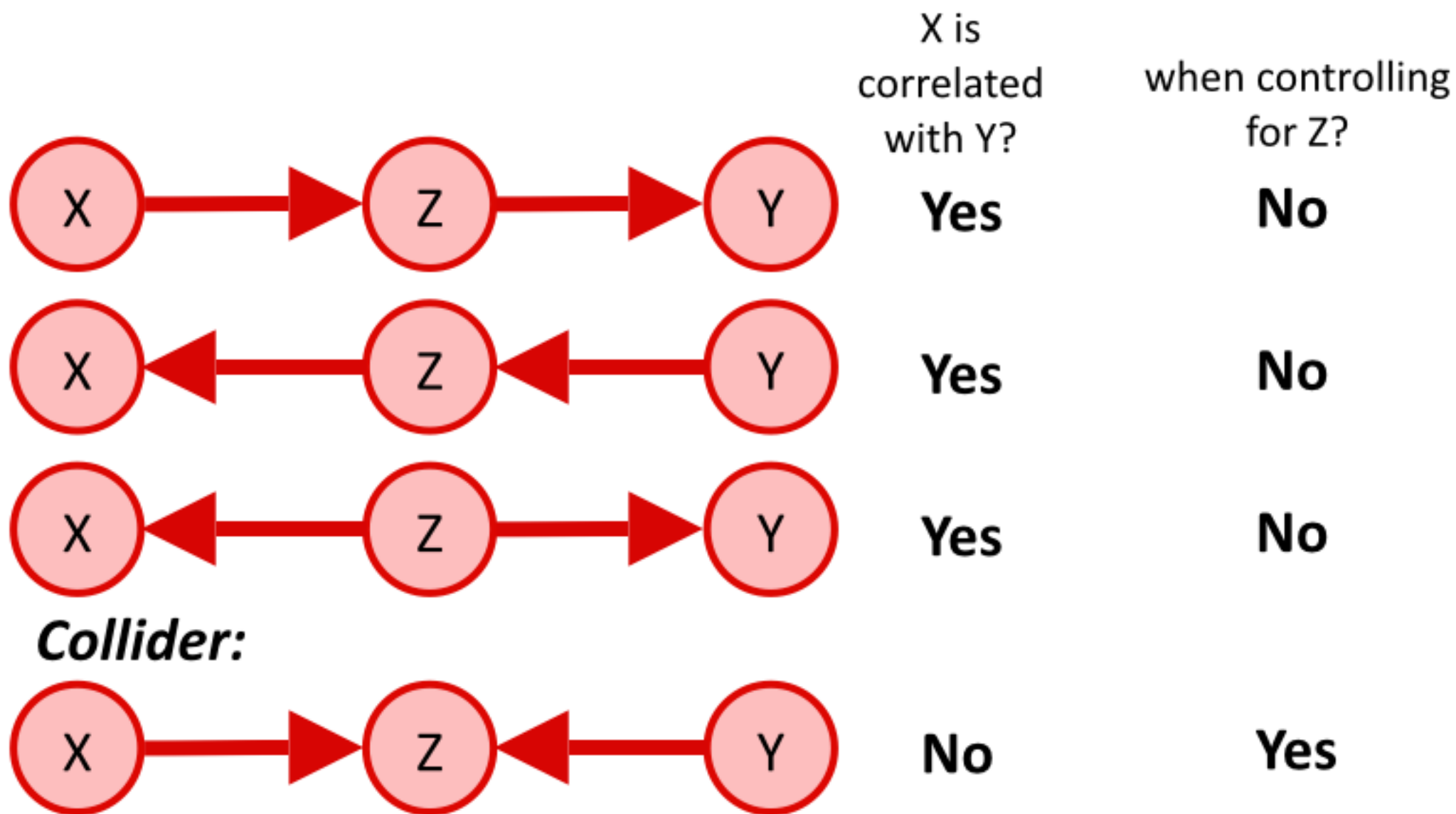
Yes



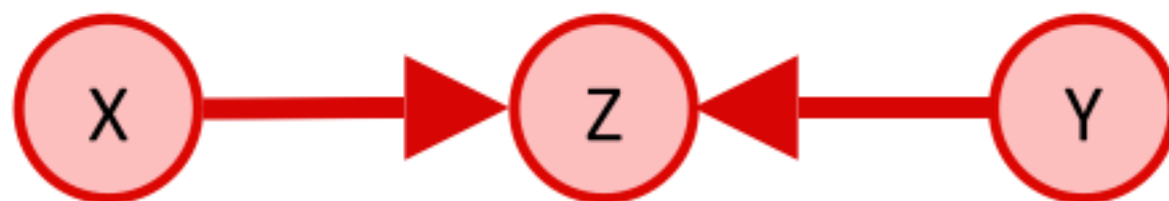
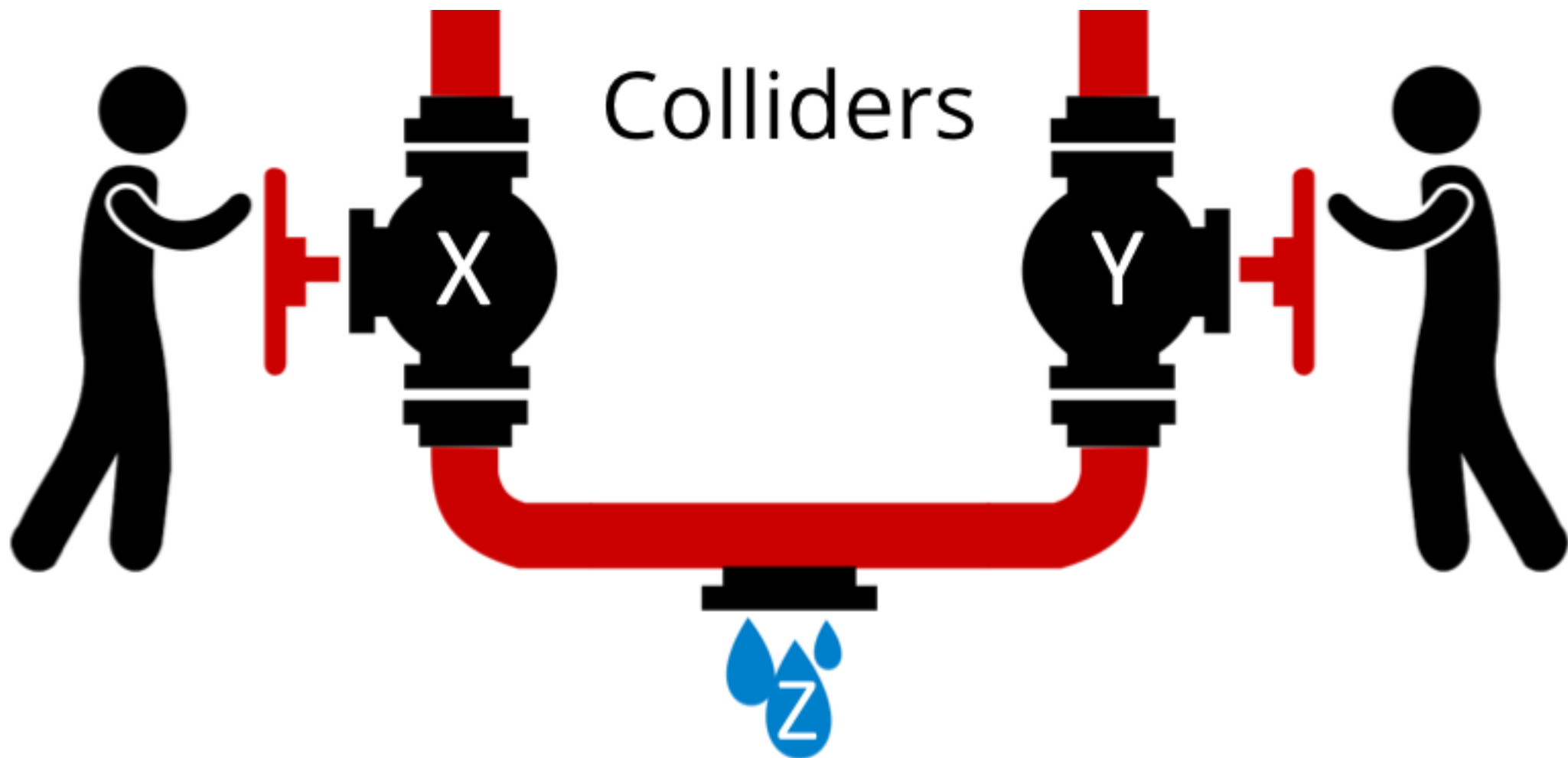
Yes

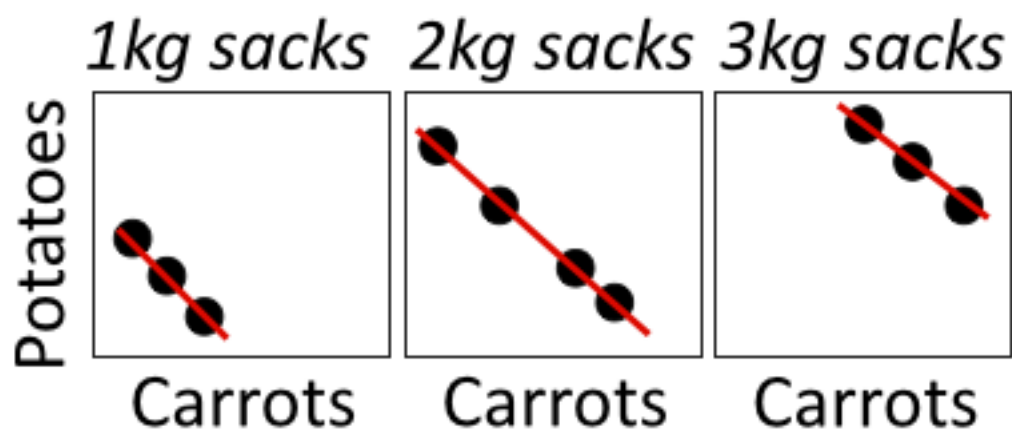
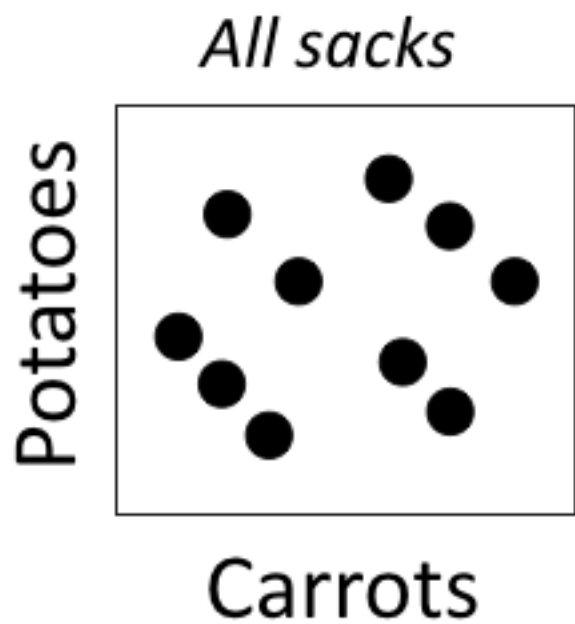
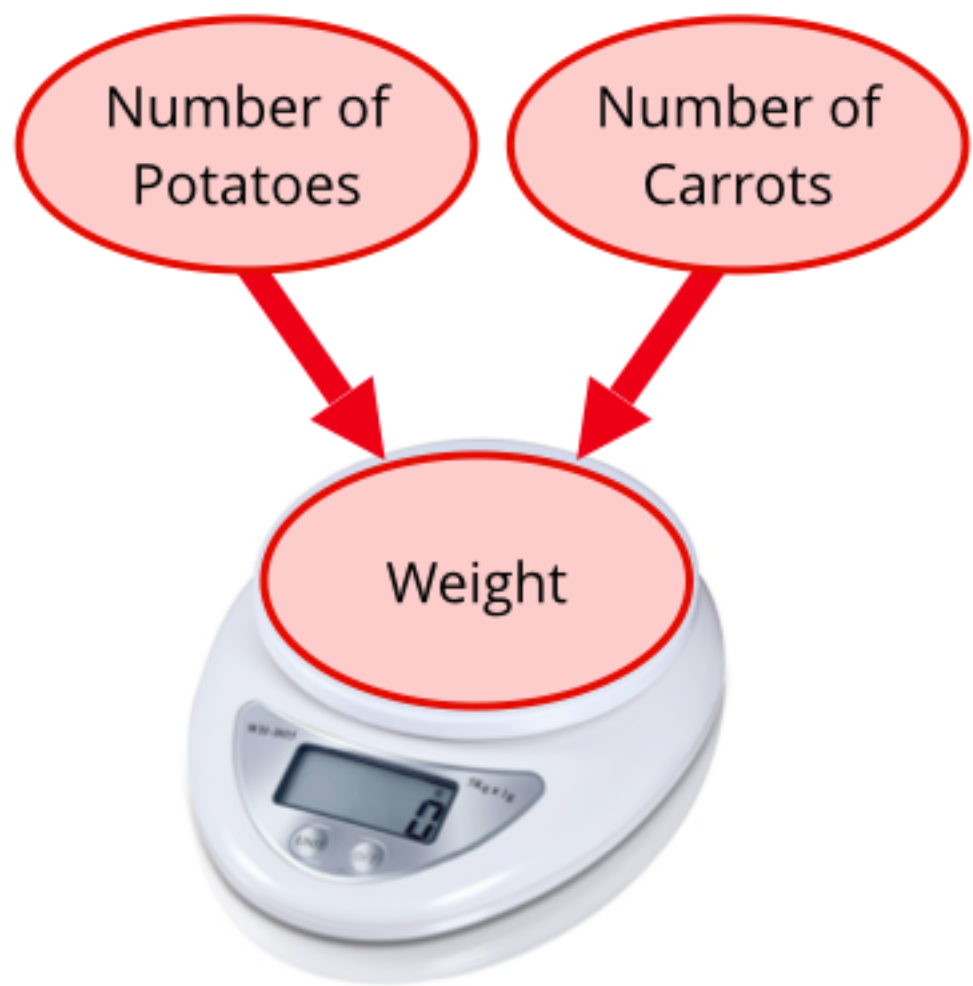


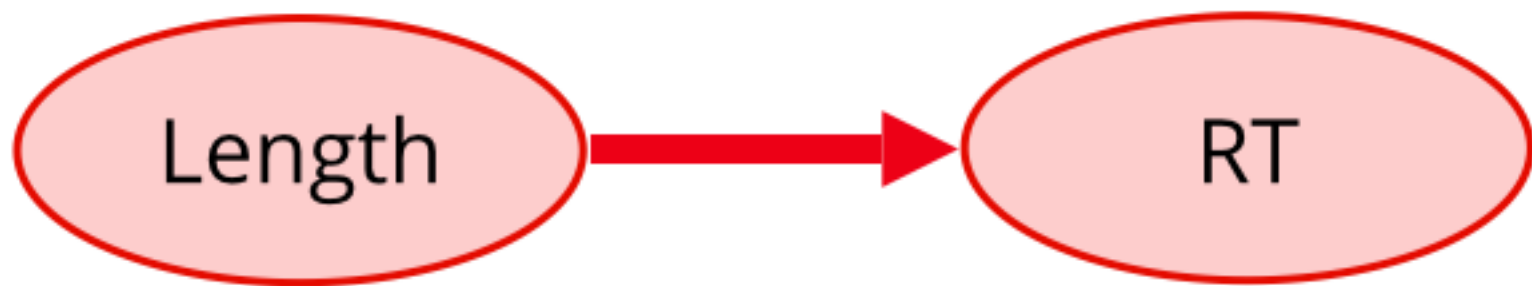
No

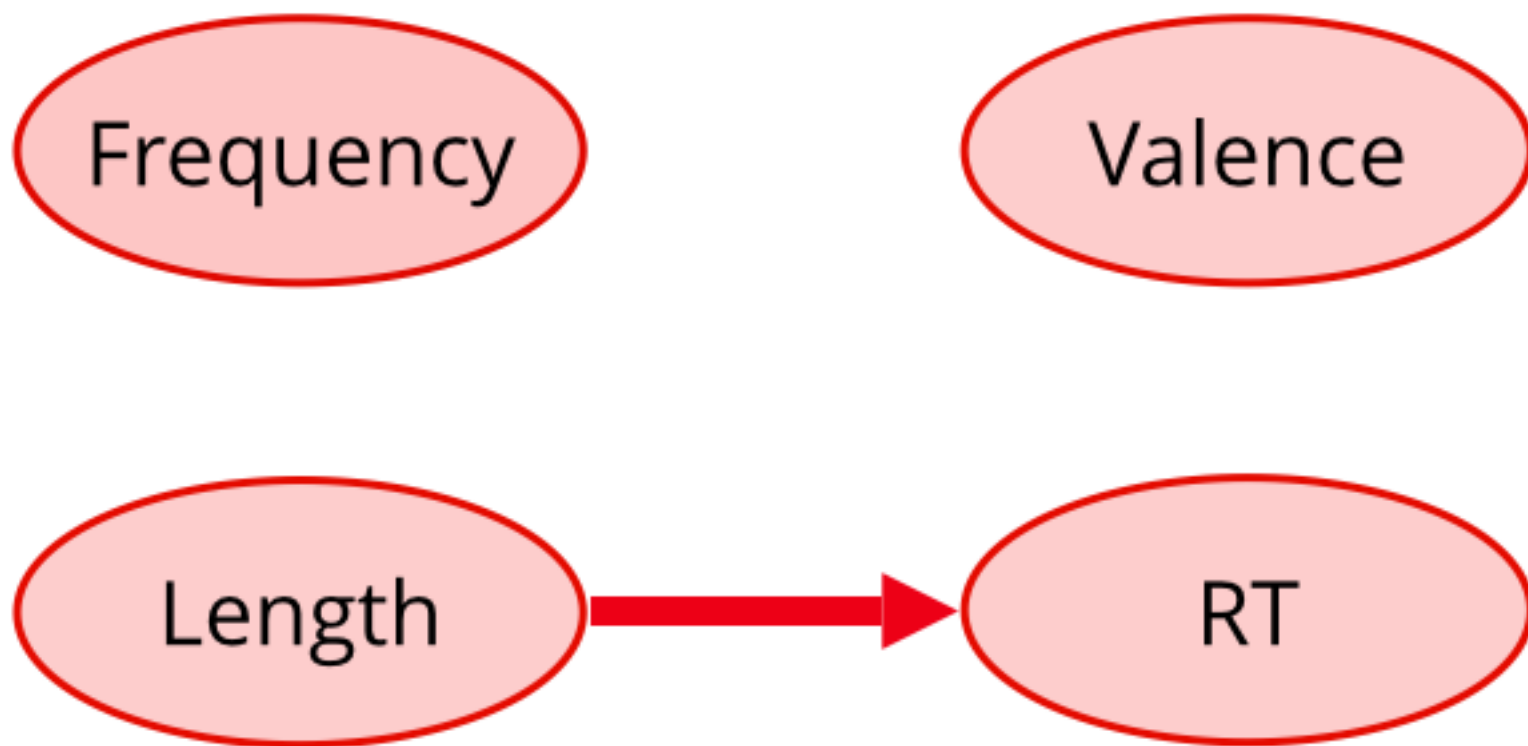


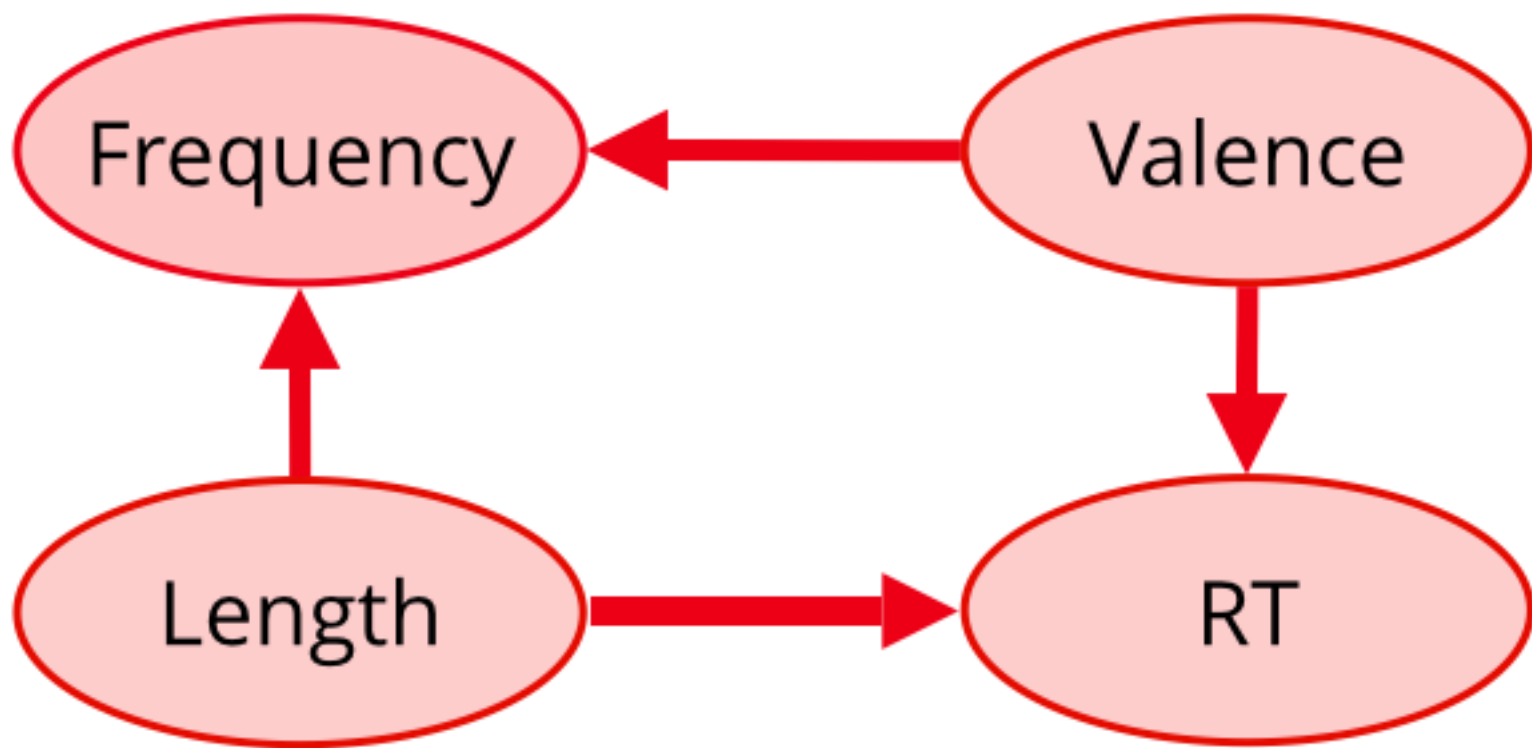
Colliders



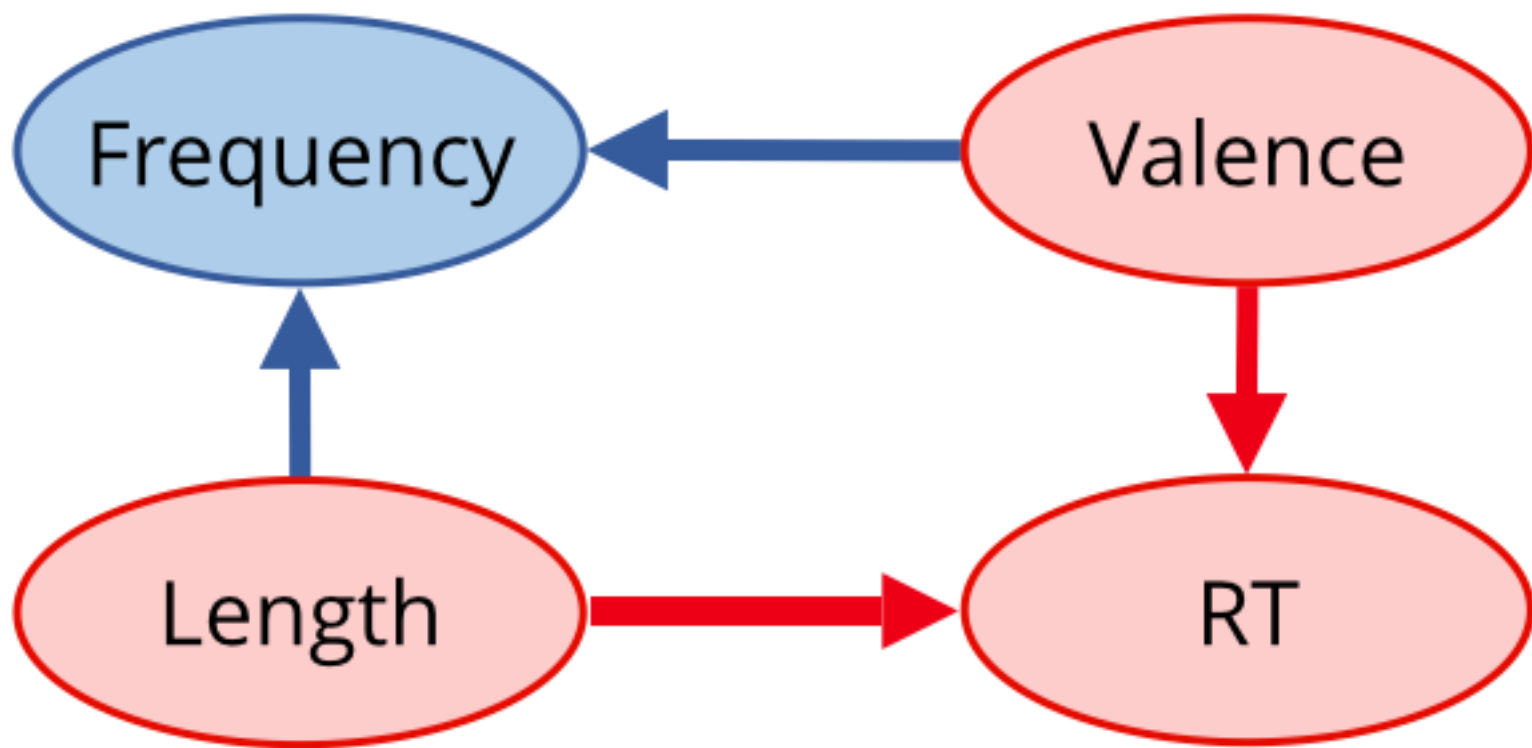




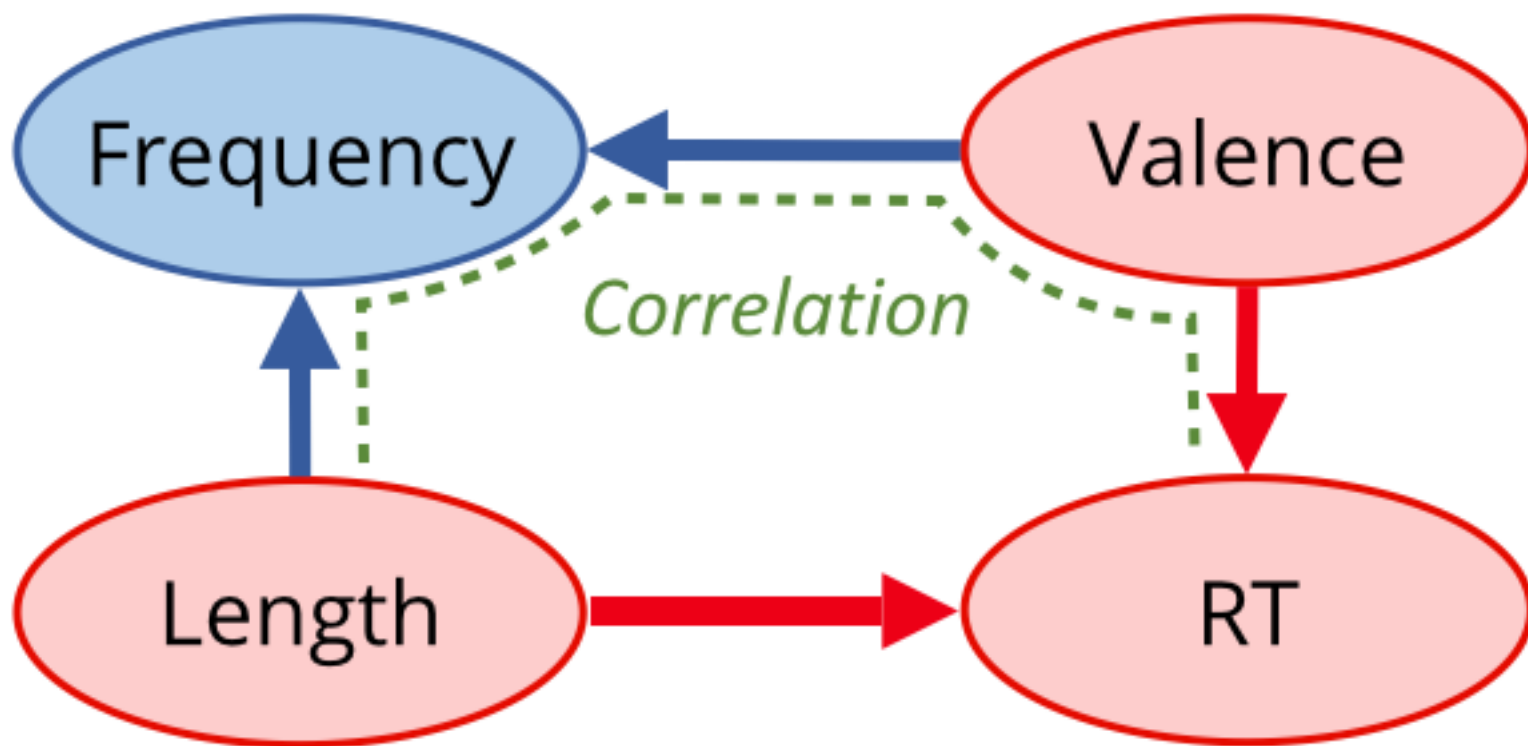


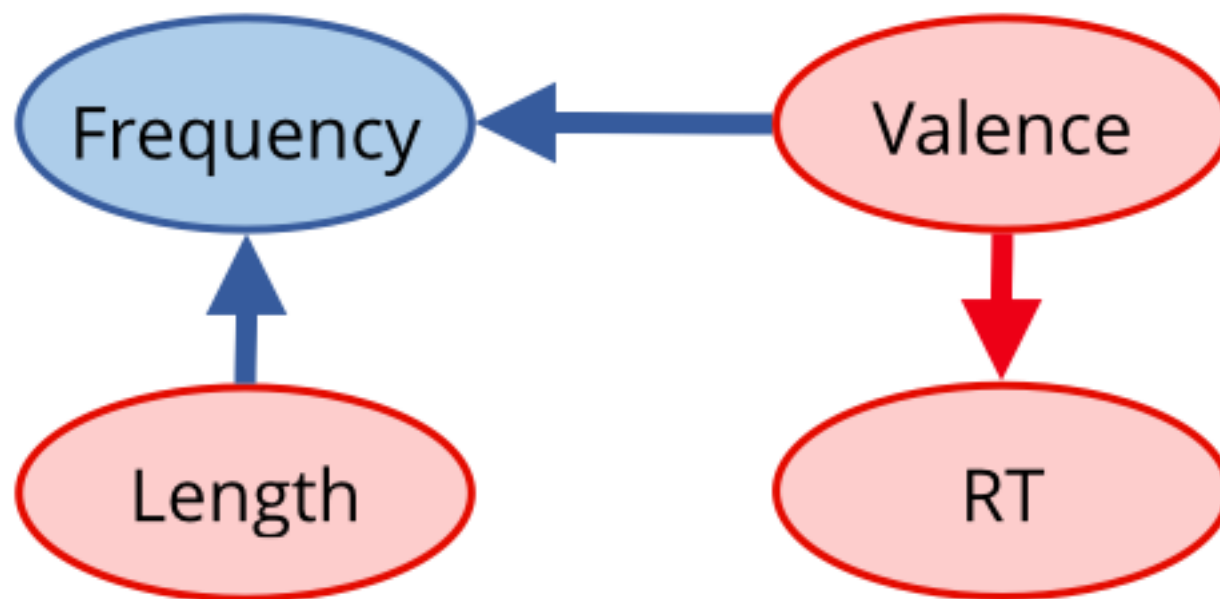


Collider



Collider





Create some hypothetical variables

```

n = 200
length = sample(1:7, n, replace = T)
valence = sample(1:7, n, replace = T)
freq = length + valence + rnorm(n)
RT = valence + rnorm(n)
  
```

Model without controls:

```

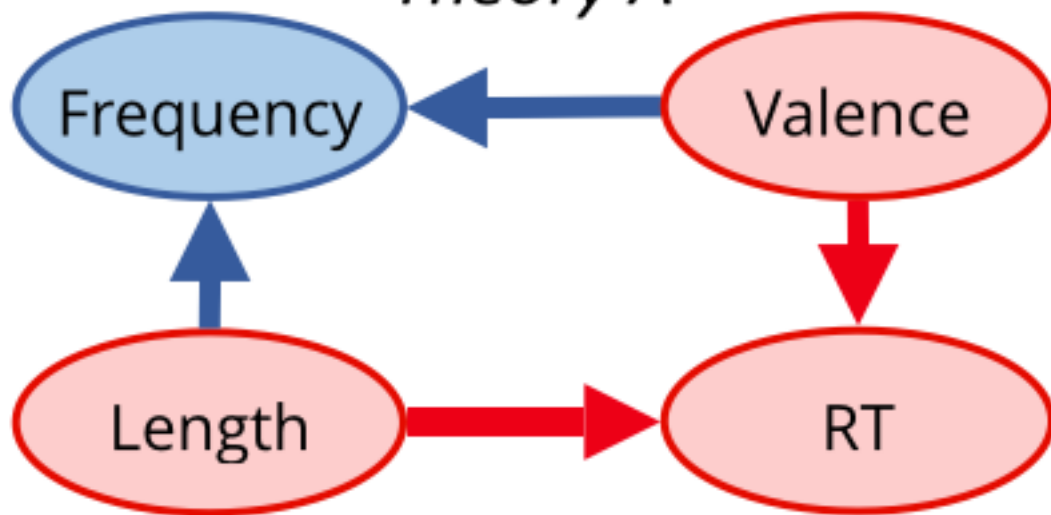
summary(lm(RT ~ length))
      Estimate Std. Error t value Pr(>|t|)
length  -0.03436    0.07971  -0.431    0.667
  
```

Model controlling for frequency:

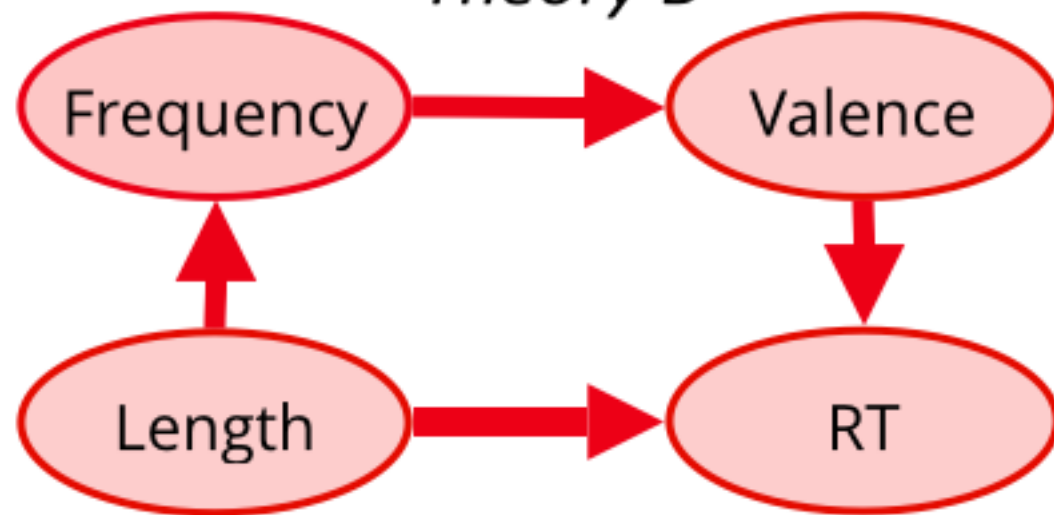
```

summary(lm(RT ~ length + freq))
      Estimate Std. Error t value Pr(>|t|)
length  -0.83004    0.06520 -12.730 <0.001 ***
freq      0.85081    0.04647  18.310 <0.001 ***
  
```

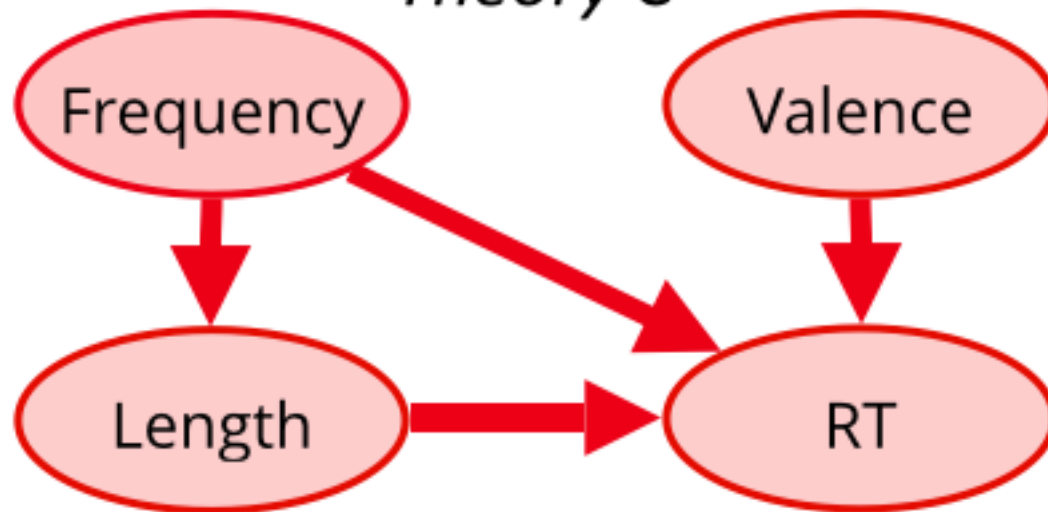
Theory A



Theory B



Theory C



Identifying confounding variables

Causal (but vague)

Cartwright: Control for anything that is “causally relevant”

Declarative

Any variable that is correlated with both X and Y.

Procedural

Noncollapsibility: Try controlling for Z. If it makes a difference, then Z is a confounder.

Identifying confounding variables



We can observe $P(Y | X)$

We want to find $P(Y | \text{do}(X))$

Confounds are anything that leads to a difference between these.

Block every noncausal path between X and Y

Without blocking any causal paths

Block all back-door paths:

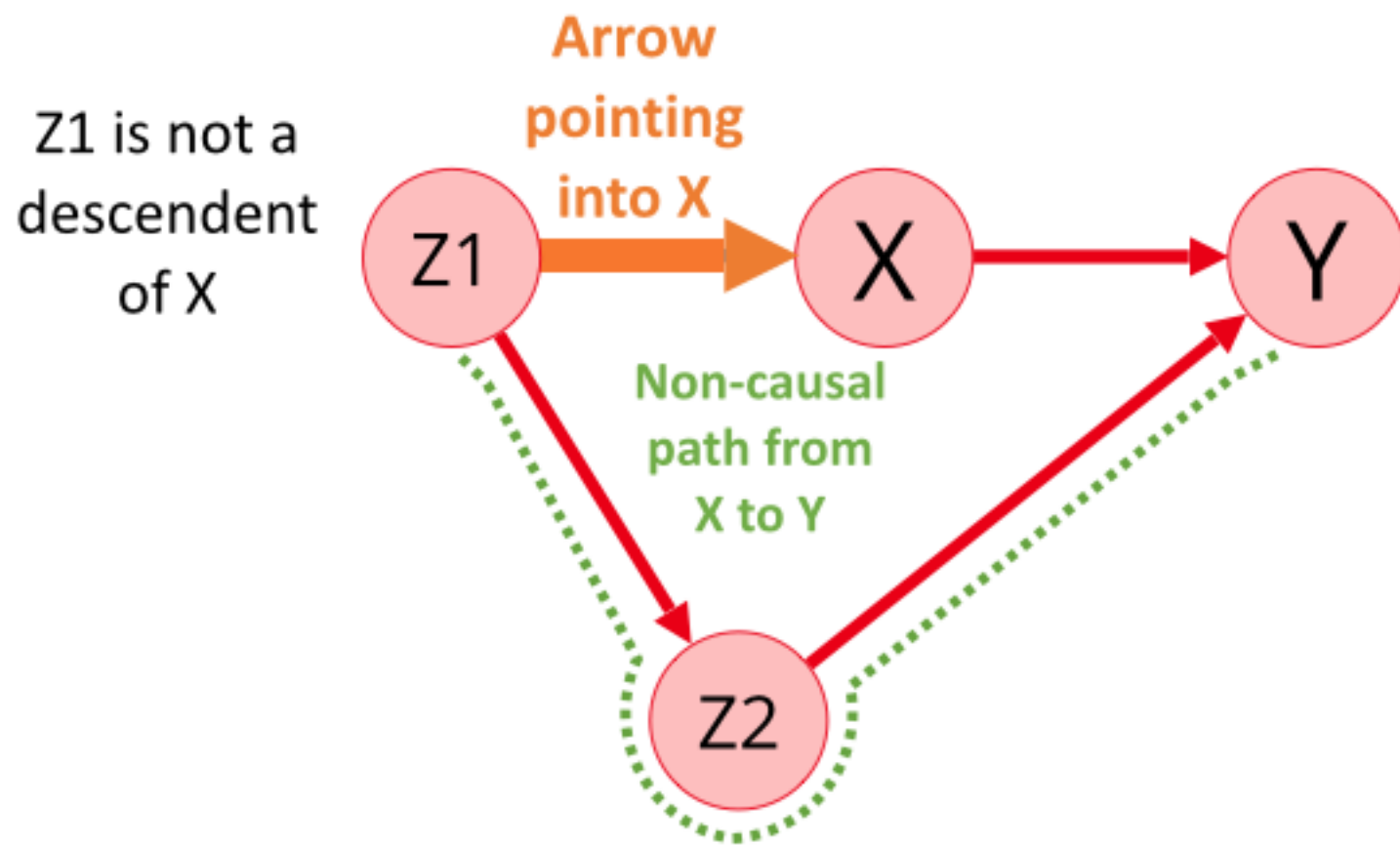
Control for all variables on a path from X to Y that starts with an arrow pointing to X.

And where the variable is not a descendant of X

See Shrier & Platt (2008)

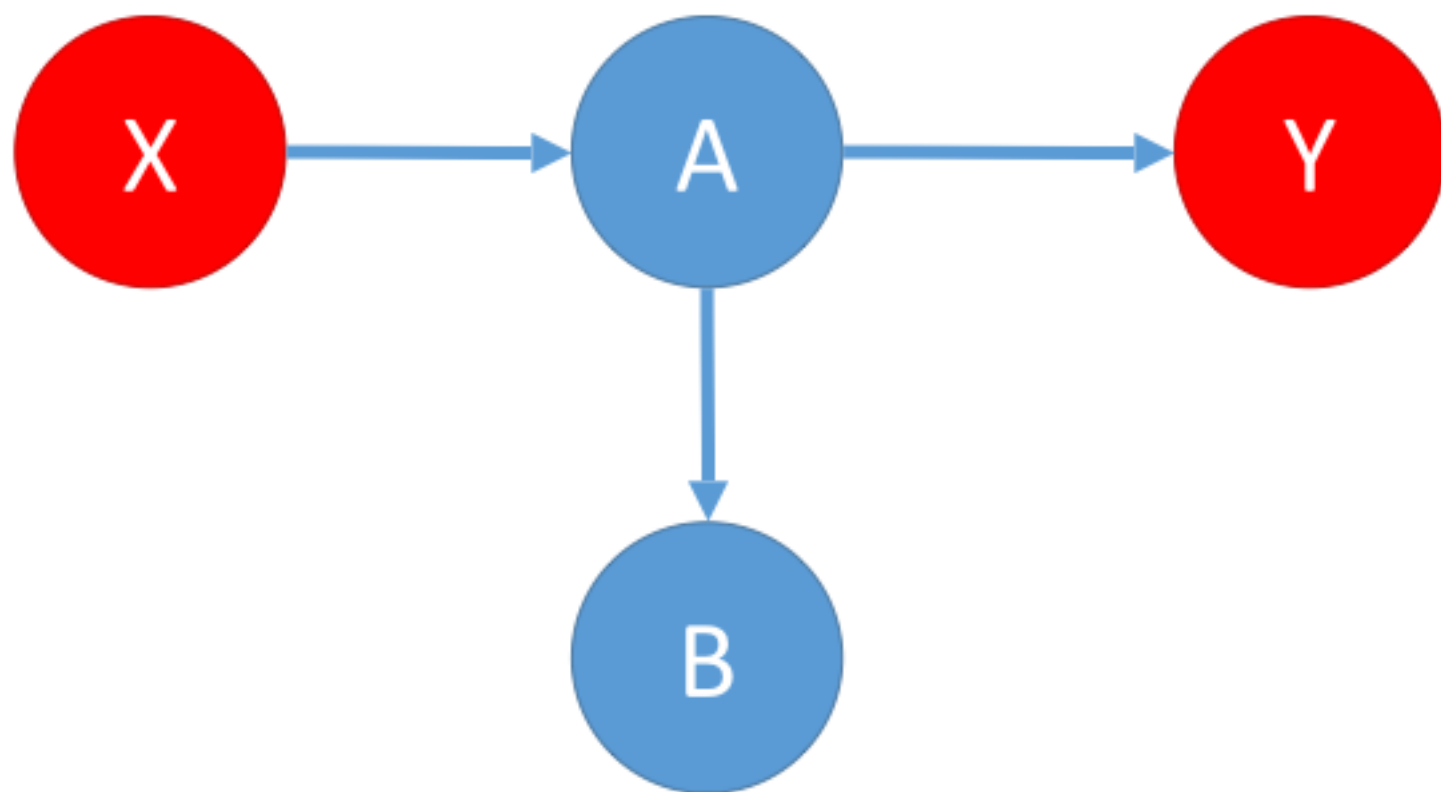
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



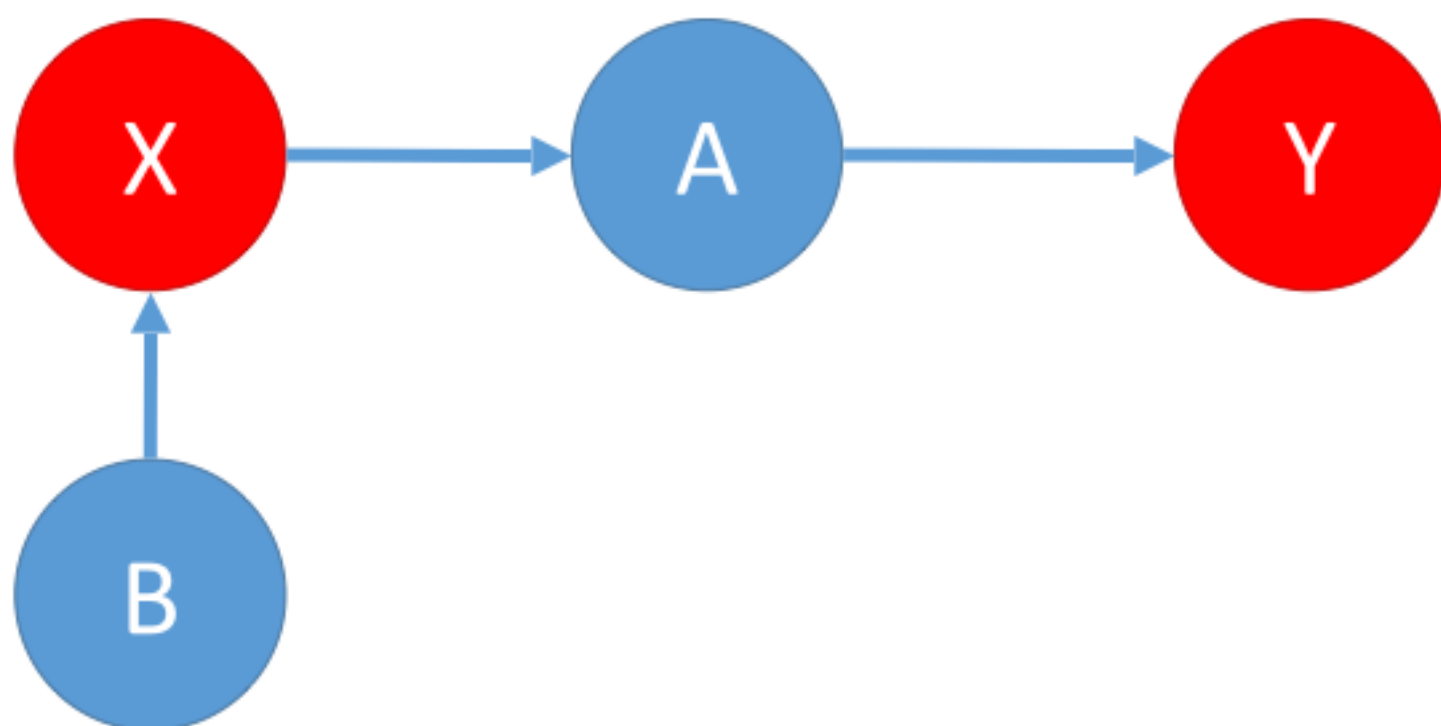
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



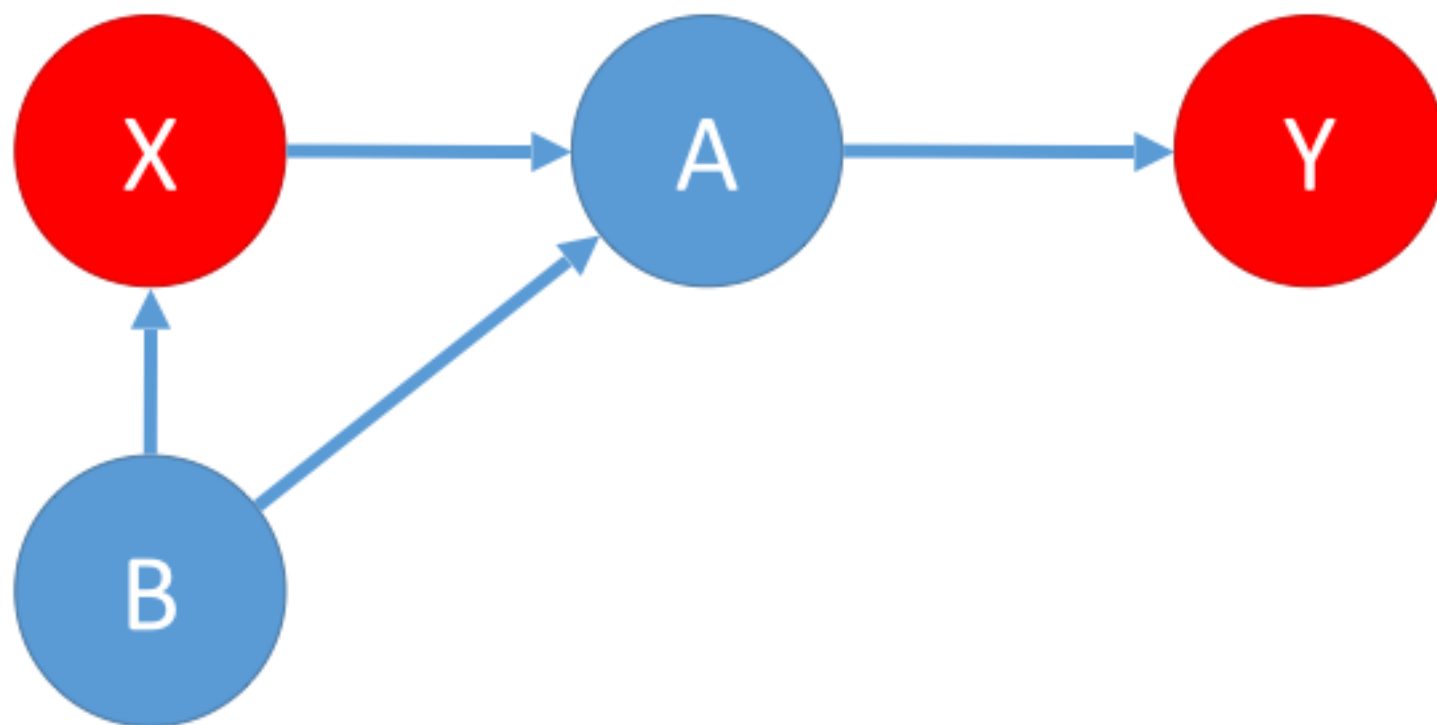
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



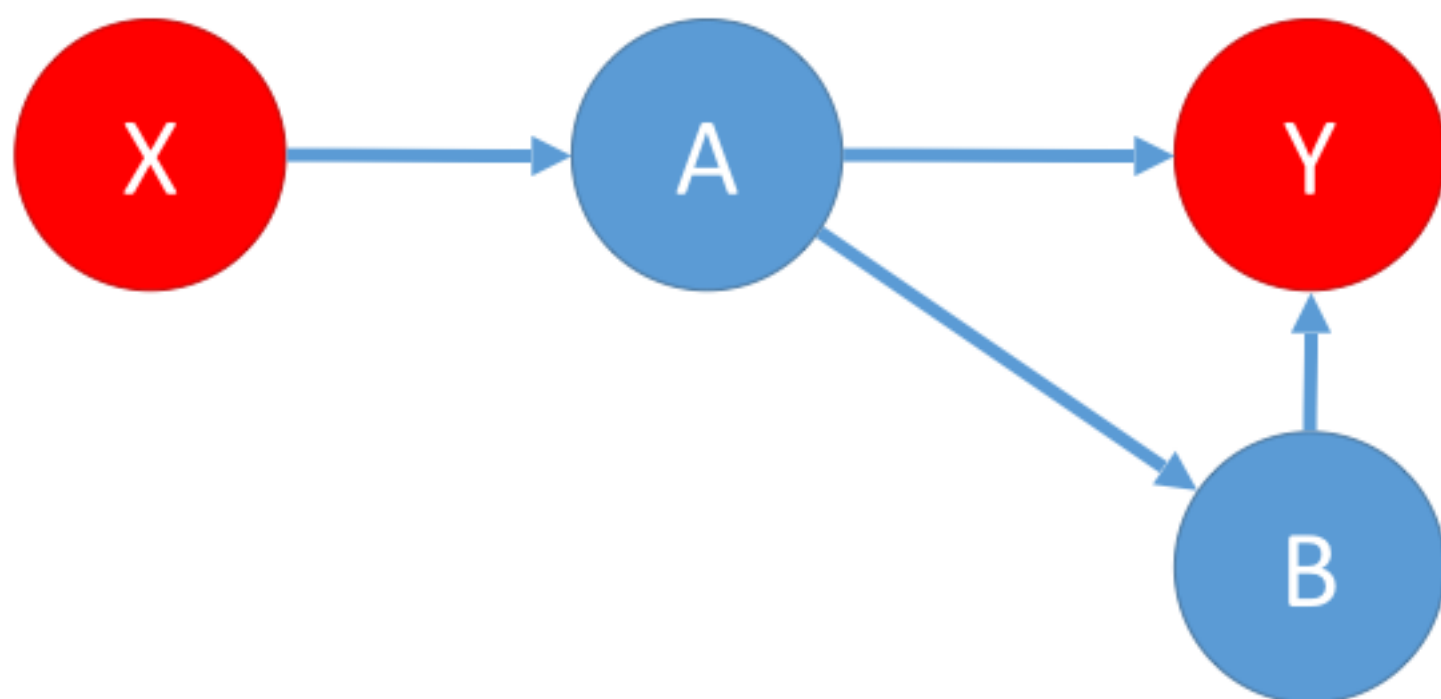
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



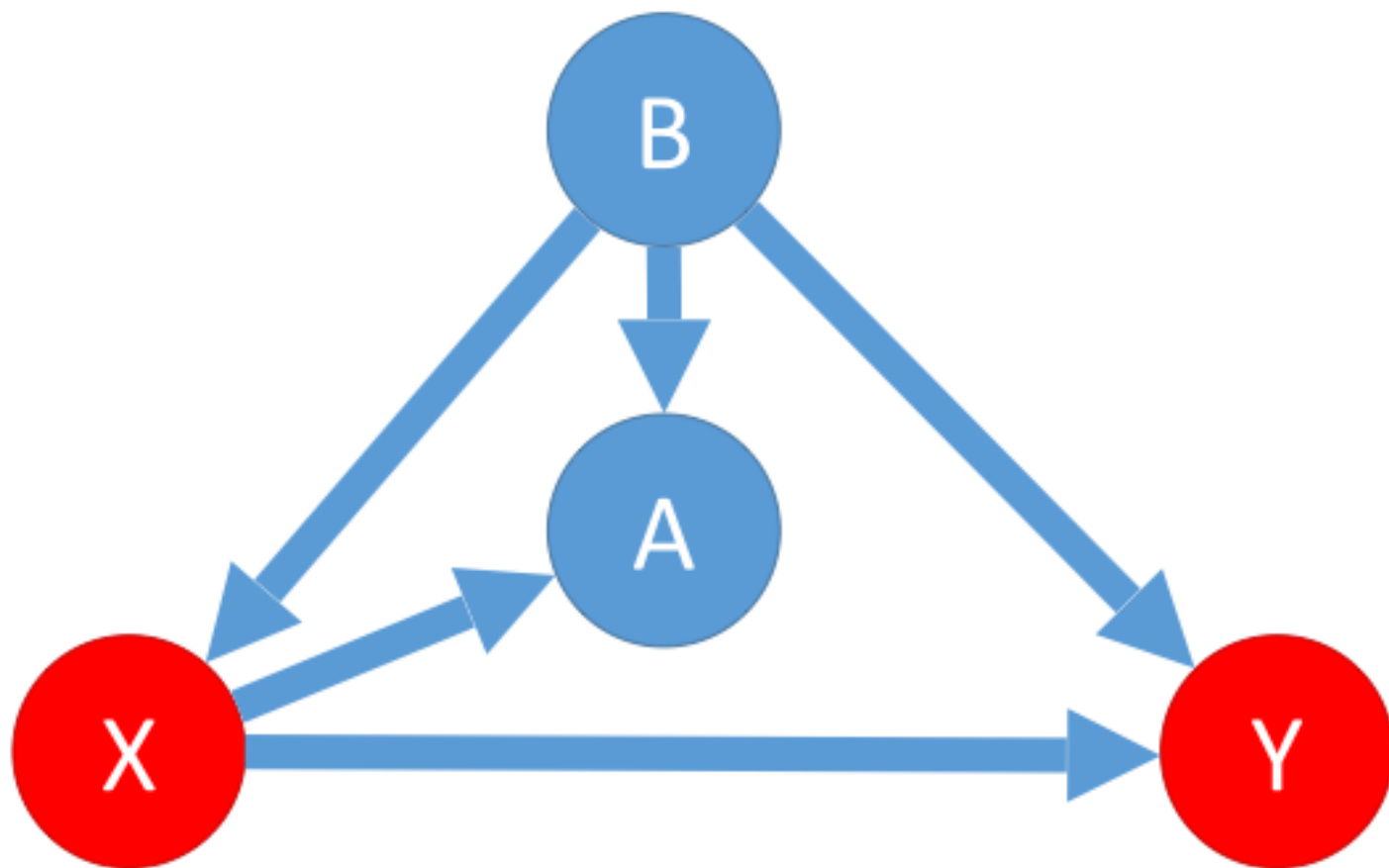
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



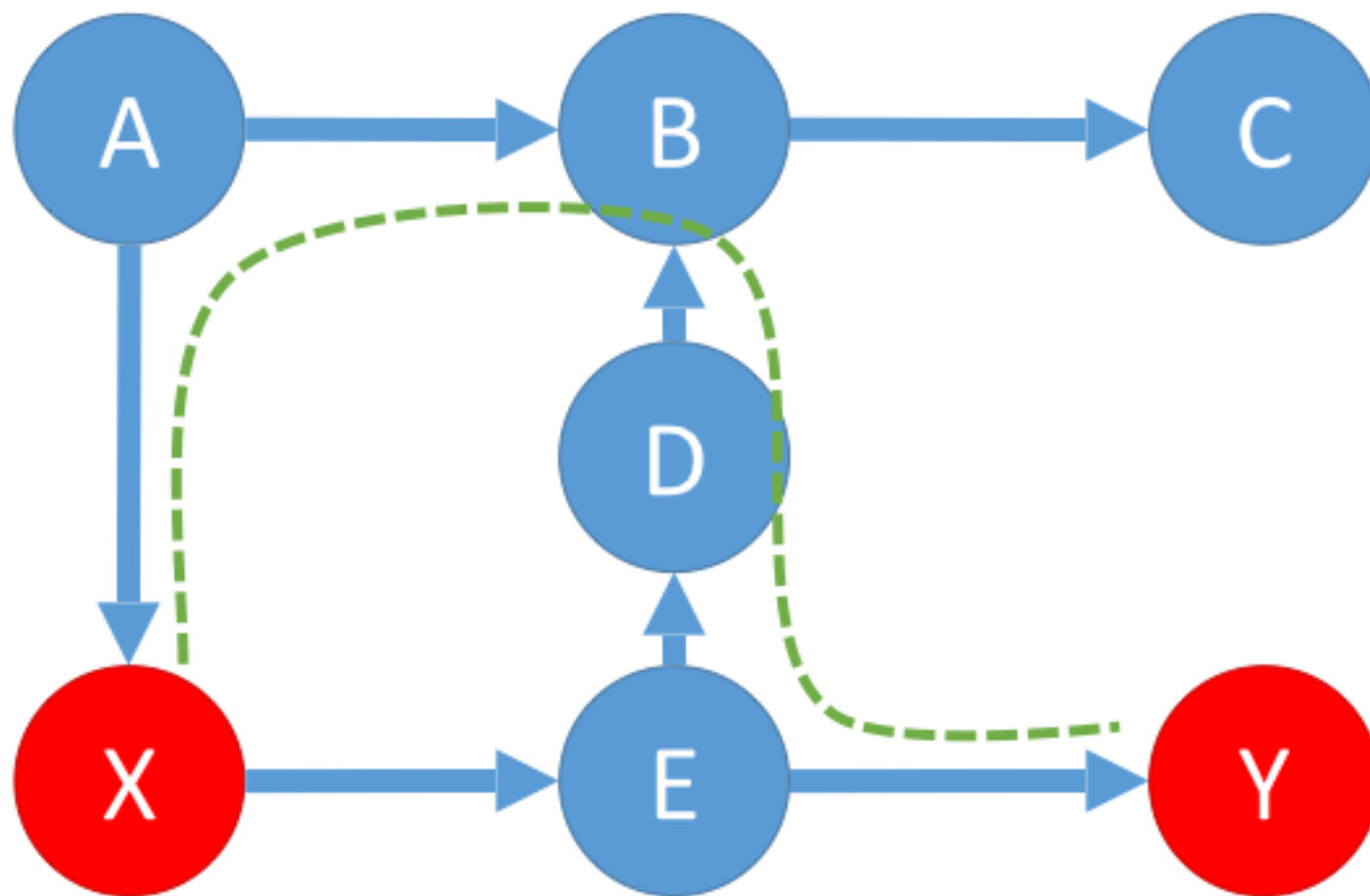
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



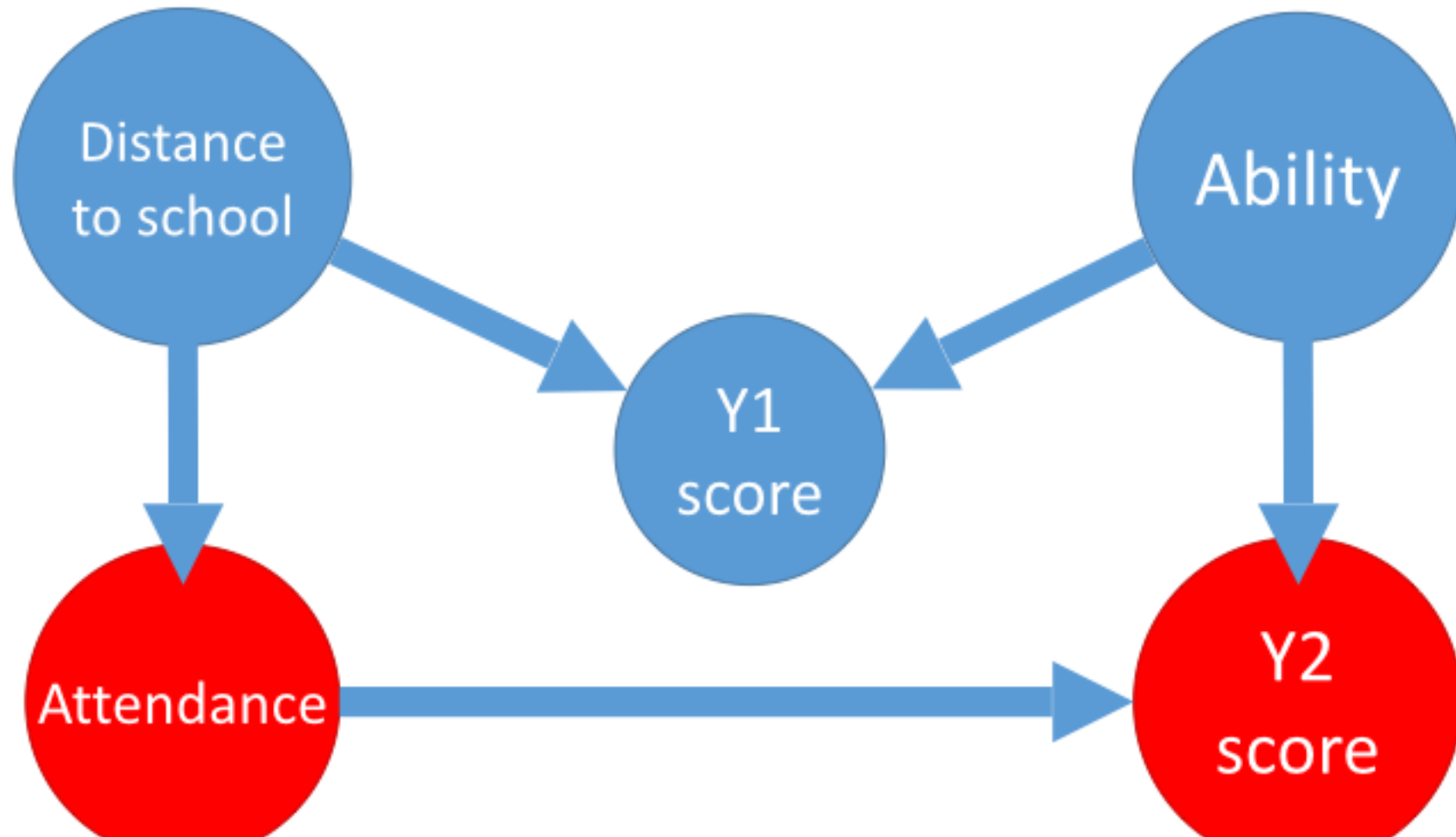
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



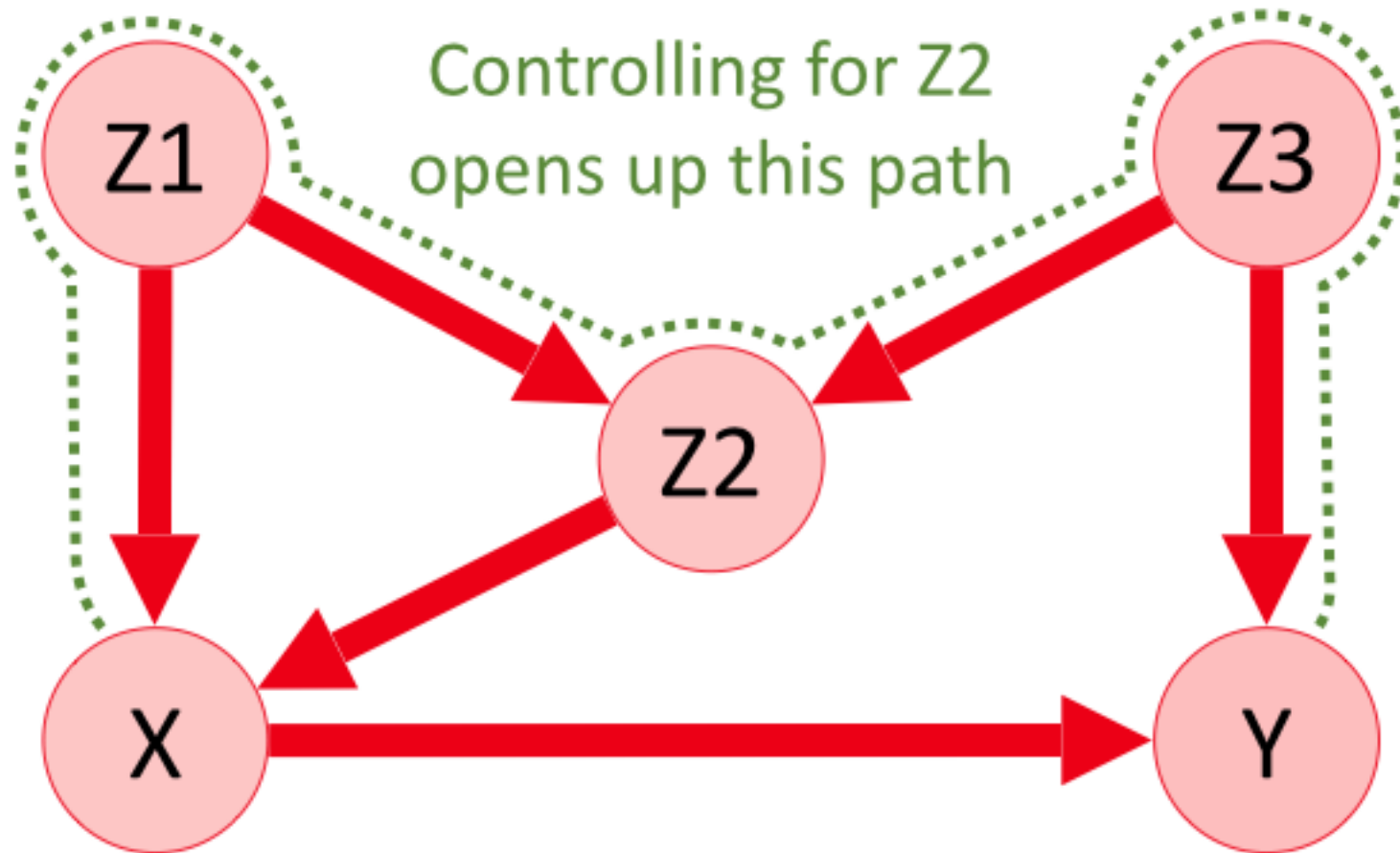
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



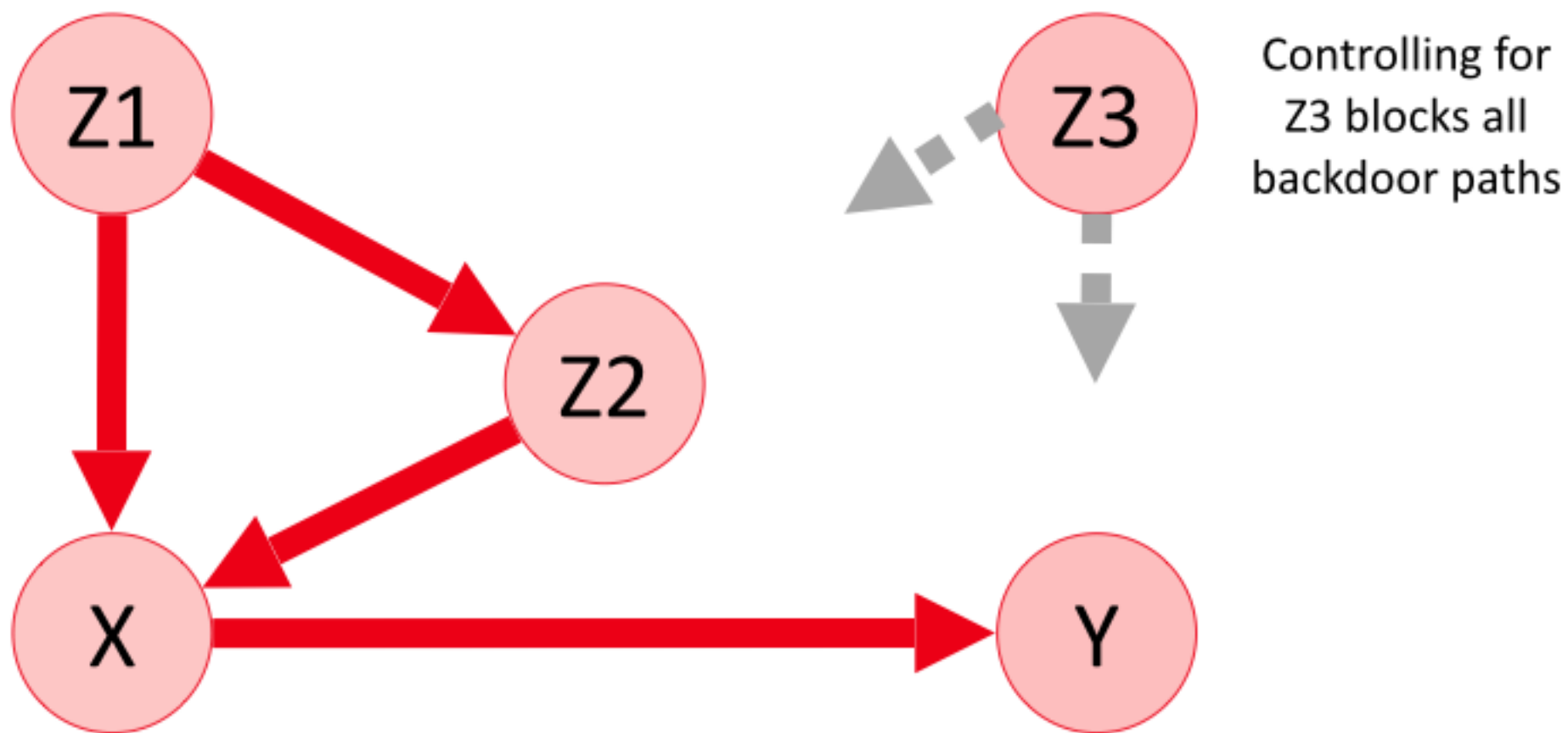
Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



Block Back Door Paths:

Variables on a path from X to Y that starts with an arrow pointing to X.
And where the variable is not a descendant of X



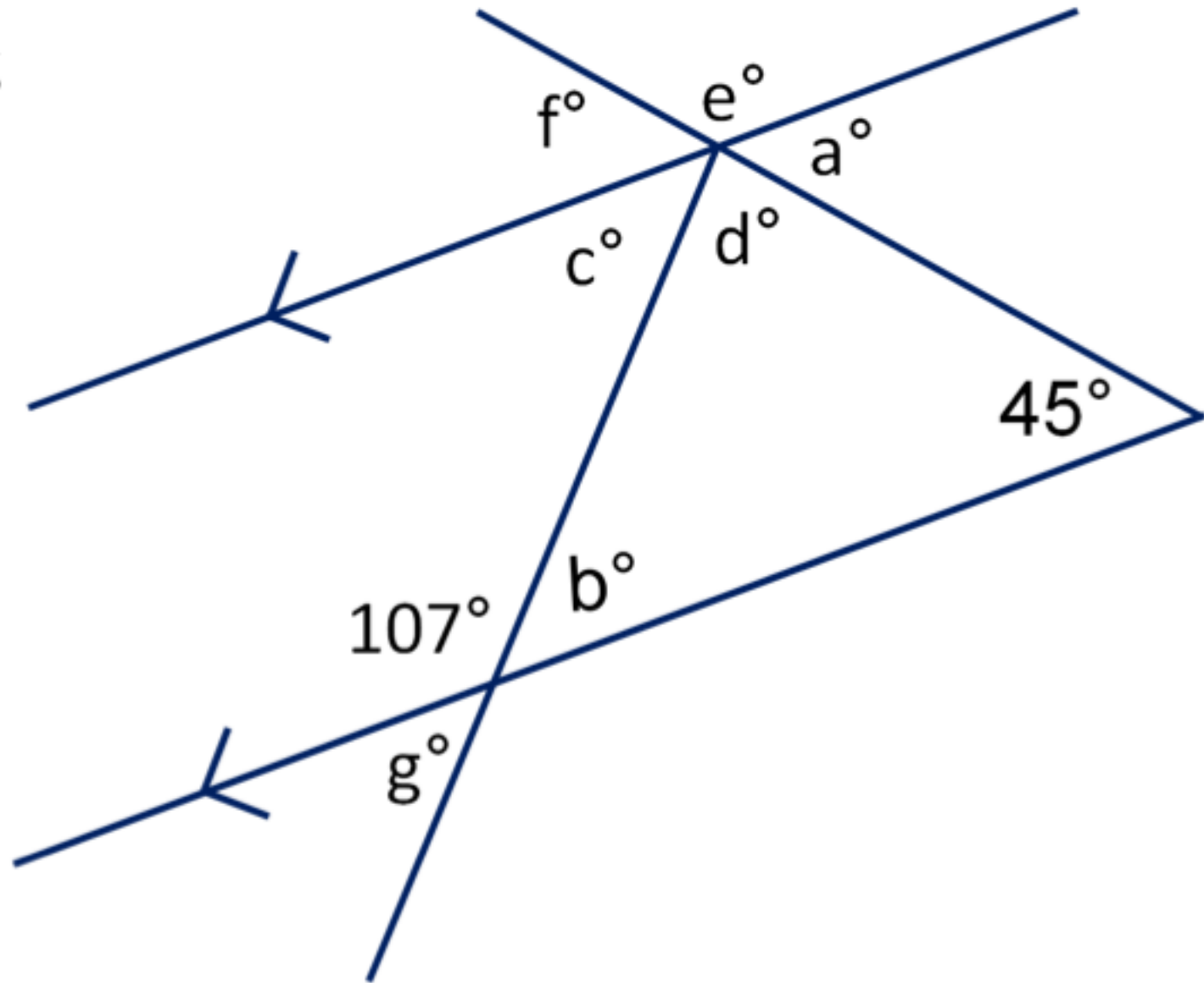
Dagitty <http://www.dagitty.net/>

Diagram style	Model	Examples	How to ...	Layout	Help	Causal effect identification
<div>Diagram style</div> <div> <input checked="" type="radio"/> classic <input type="radio"/> SEM-like </div> <div>View mode</div> <div> <input checked="" type="radio"/> normal <input type="radio"/> moral graph <input type="radio"/> correlation graph </div> <div>Coloring</div> <div> <input checked="" type="checkbox"/> causal paths <input checked="" type="checkbox"/> biasing paths <input checked="" type="checkbox"/> ancestral structure </div> <div>Effect analysis</div> <div> <input type="checkbox"/> atomic direct effects </div> <div>Legend</div> <div> <div></div> exposure <div></div> outcome <div></div> ancestor of exposure <div></div> ancestor of outcome <div></div> ancestor of exposure and outcome <div></div> adjusted variable <div></div> unobserved (latent) <div></div> other variable <div></div> causal path <div></div> biasing path </div> <div>Summary</div> <div> exposure(s) WarmUpExerc outcome(s) Injury </div>	<pre> graph TD Coach((Coach)) --> FitnessLevel((FitnessLevel)) Genetics((Genetics)) --> FitnessLevel Genetics --> NeuromuscularFatigue((NeuromuscularFatigue)) Genetics --> ConnectiveTissueDisorder((ConnectiveTissueDisorder)) FitnessLevel --> TeamMotivation((TeamMotivation)) FitnessLevel --> NeuromuscularFatigue TeamMotivation --> PreGameProprioception((PreGameProprioception)) TeamMotivation --> WarmUpExercises((WarmUpExercises)) PreGameProprioception --> PreviousInjury((PreviousInjury)) ContactSport((ContactSport)) --> PreviousInjury ContactSport --> IntraGameProprioception((IntraGameProprioception)) PreviousInjury --> IntraGameProprioception NeuromuscularFatigue --> IntraGameProprioception NeuromuscularFatigue --> Injury((Injury)) ConnectiveTissueDisorder --> TissueWeakness((TissueWeakness)) TissueWeakness --> Injury IntraGameProprioception --> Injury WarmUpExercises --> IntraGameProprioception WarmUpExercises --> Injury </pre>					<div>Causal effect identification</div> <div>Adjustment (total effect)</div> <div>Minimal sufficient adjustment sets for estimating the total effect of WarmUpExercises on Injury:</div> <ul style="list-style-type: none"> Coach, FitnessLevel Coach, PreGameProprioception ConnectiveTissueDisorder, NeuromuscularFatigue FitnessLevel, Genetics FitnessLevel, TeamMotivation NeuromuscularFatigue, TissueWeakness PreGameProprioception, TeamMotivation <div>Testable implications</div> <div>The model implies the following conditional independences:</div> <ul style="list-style-type: none"> WarmUpExercises \perp Injury IntraGameProprioception, NeuromuscularFatigue, TissueWeakness WarmUpExercises \perp Injury ConnectiveTissueDisorder, IntraGameProprioception, NeuromuscularFatigue WarmUpExercises \perp Injury FitnessLevel, Genetics, IntraGameProprioception, NeuromuscularFatigue WarmUpExercises \perp Injury Coach, FitnessLevel, IntraGameProprioception, NeuromuscularFatigue WarmUpExercises \perp Injury Coach

<https://www.youtube.com/watch?v=pJhU4fimHBQ>

Causal Axioms

Geometric axioms +
geometric calculus
= ability to solve
problems that looked
impossible before.



Causal Axioms

Causal calculus aims to do the same:

Exchange probabilities based on intervention ($\text{do}(X)$)
for probabilities based on observation ($\text{see}(X)$)

Rule1: $P(Y \mid \text{do}(X), Z, W) = P(Y \mid \text{do}(X), Z)$

if W is irrelevant to Y

Rule 2: $P(Y \mid \text{do}(X), Z) = P(Y \mid X, Z)$

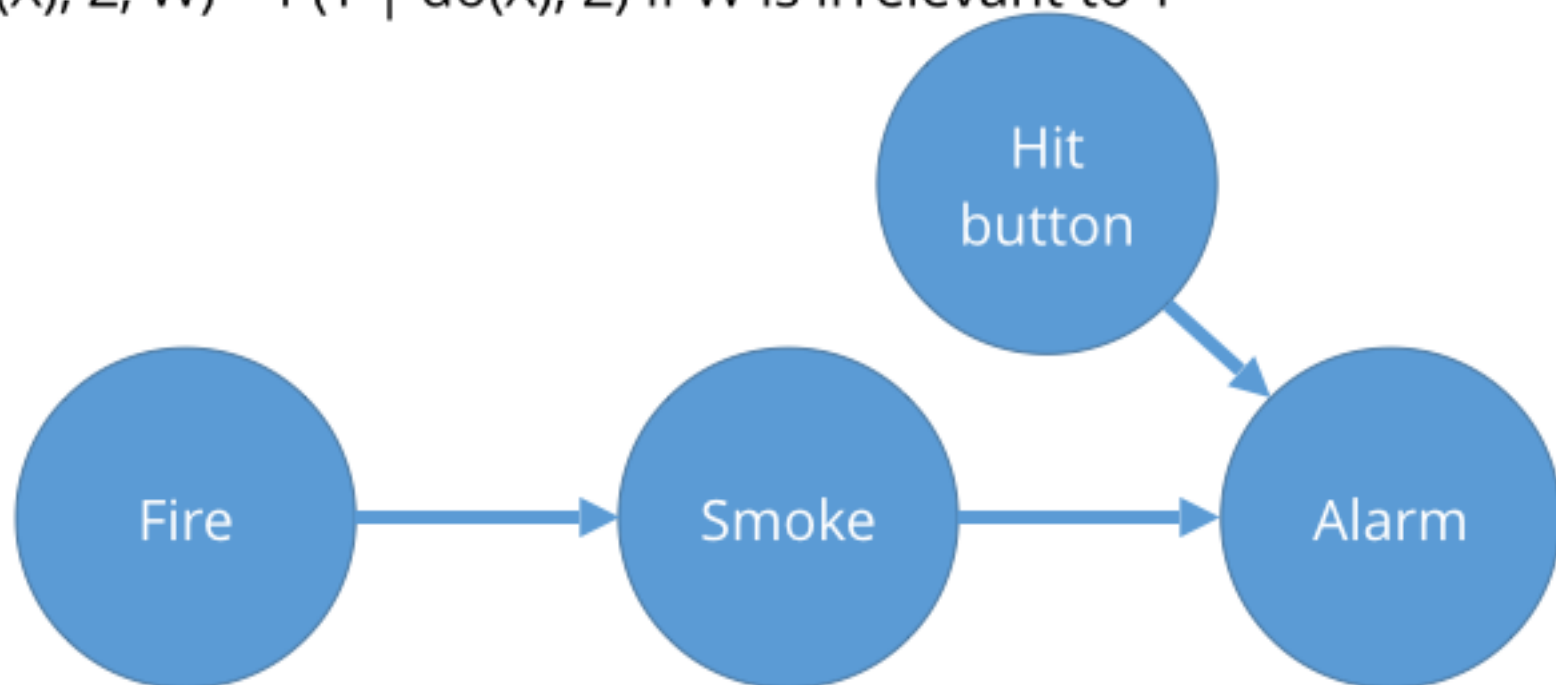
if Y is independent of Z given X and W if all connections are severed from X

Rule 3: $P(Y \mid \text{do}(X)) = P(Y)$

if there is no path from X to Y with only forward-directed arrows

Rule 1: Delete observations

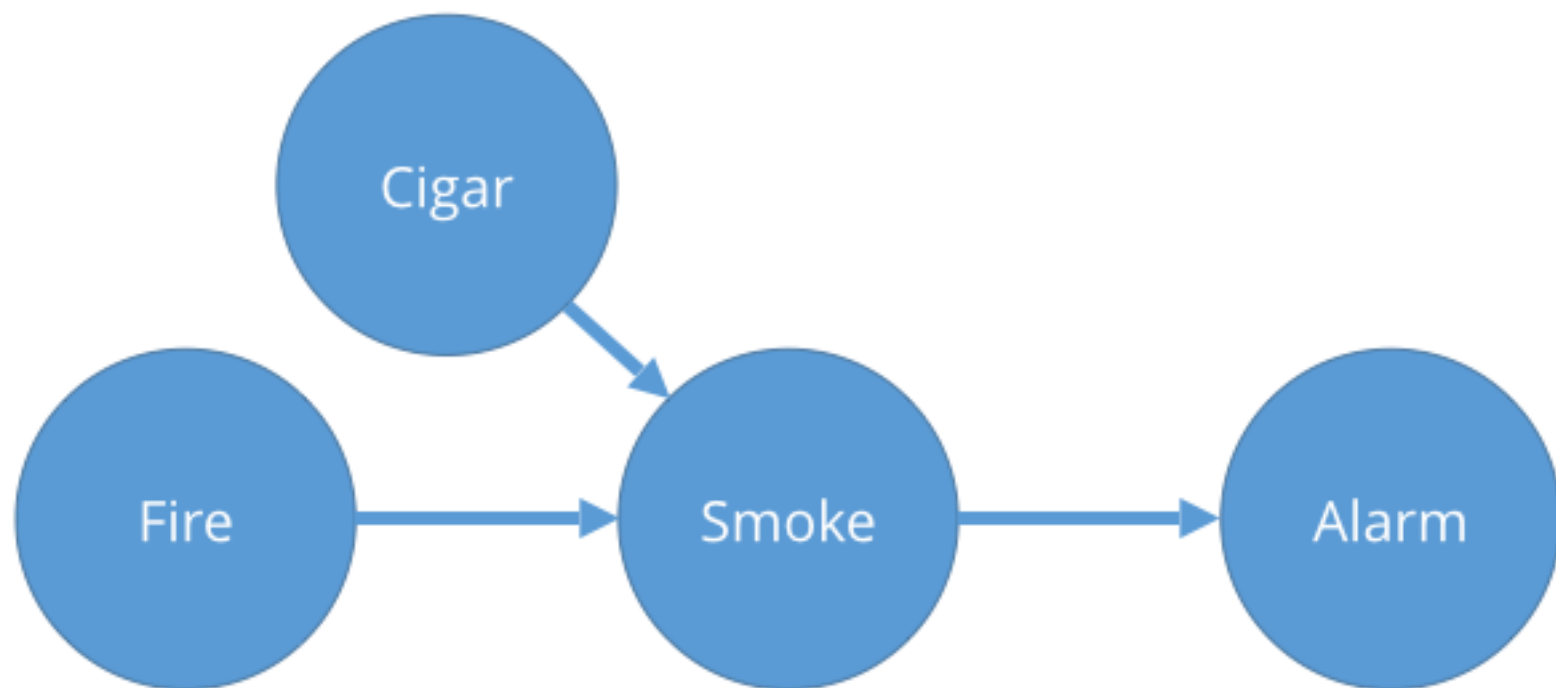
$P(Y \mid \text{do}(X), Z, W) = P(Y \mid \text{do}(X), Z)$ if W is irrelevant to Y



$P(\text{Alarm} \mid \text{do}(\text{hit button}), \text{Smoke}, \text{Fire}) =$
 $P(\text{Alarm} \mid \text{do}(\text{hit button}), \text{Smoke})$

Rule 2: Exchange interventions with observations

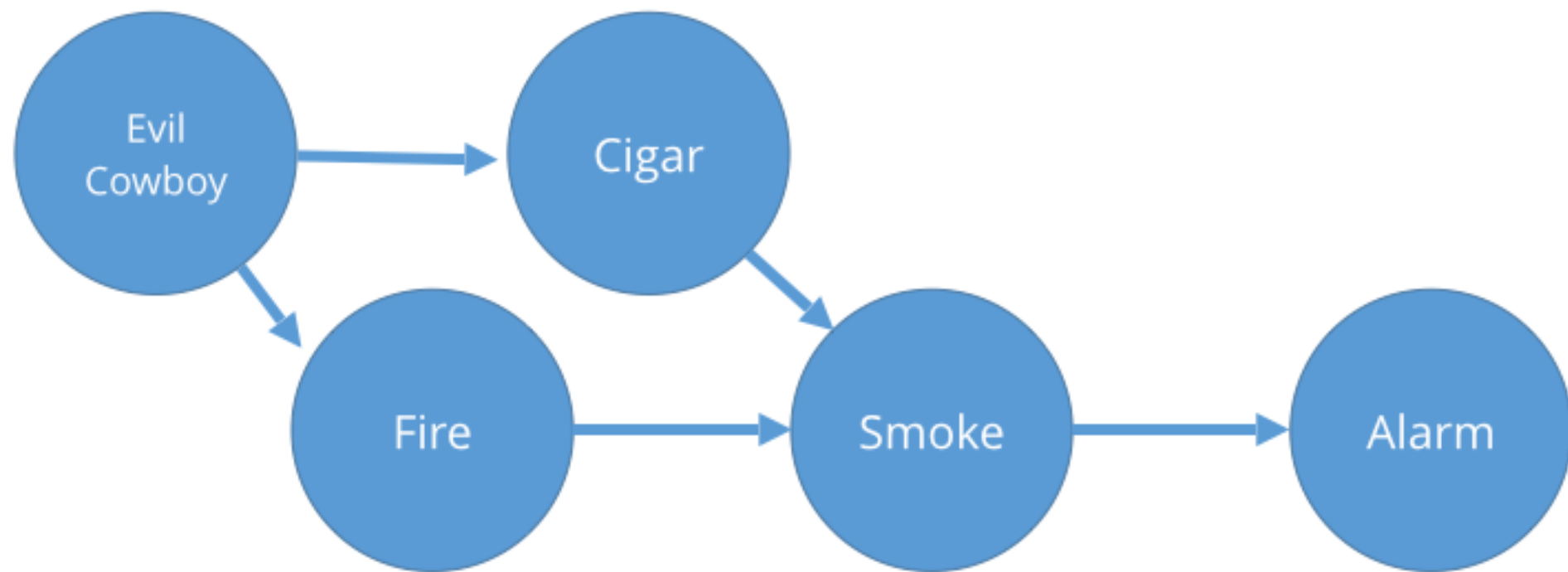
$P(Y \mid \text{do}(X), Z) = P(Y \mid X, Z)$ if Y is independent of Z given X and W if all connections are severed from X



$$P(\text{Alarm} \mid \text{do}(\text{Cigar}), \text{Smoke}) = \\ P(\text{Alarm} \mid \text{Cigar}, \text{Smoke})$$

Rule 2: Exchange interventions with observations

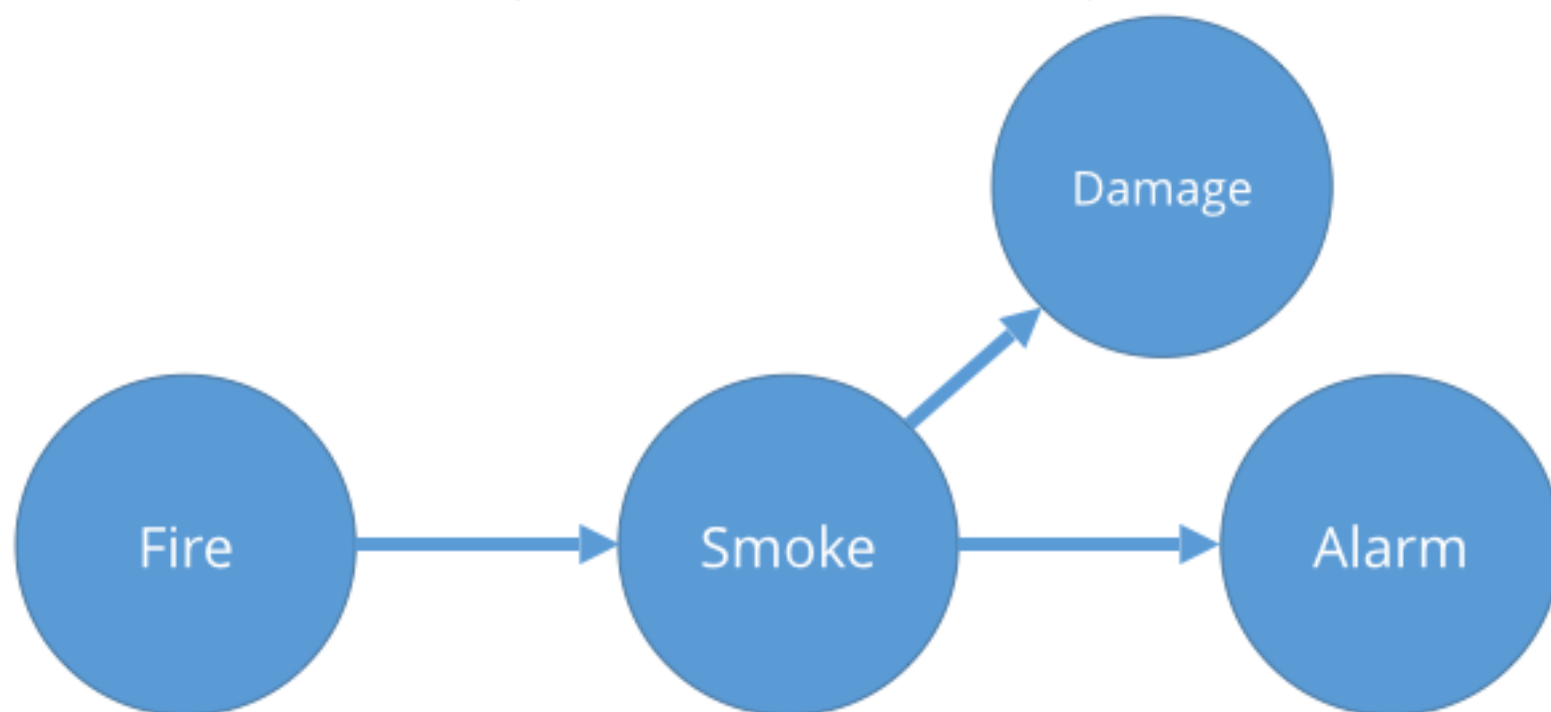
$P(Y \mid \text{do}(X), Z) = P(Y \mid X, Z)$ if Y is independent of Z given X and W if all connections are severed from X



$$P(\text{Alarm} \mid \text{do}(\text{Cigar}), \text{Smoke}) \neq P(\text{Alarm} \mid \text{Cigar}, \text{Smoke})$$

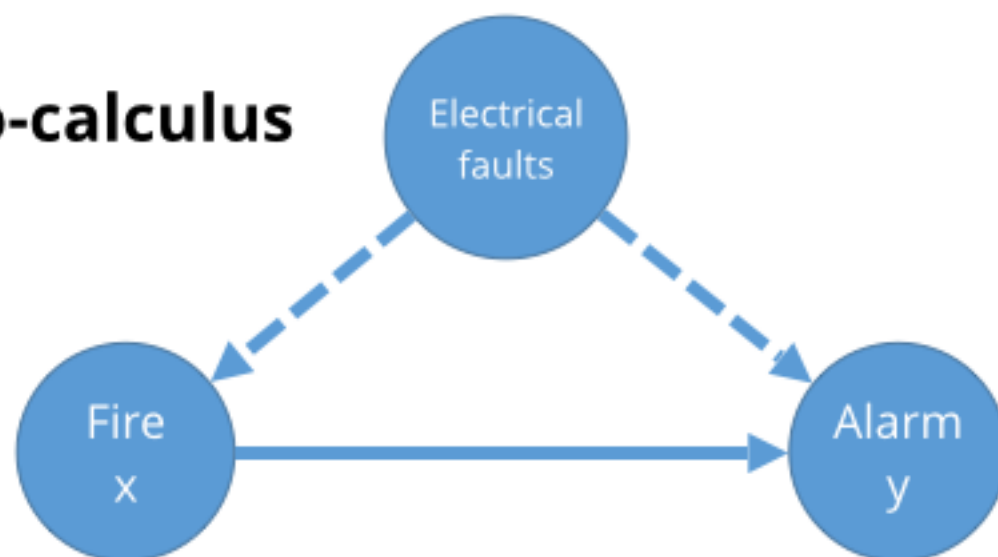
Rule 3: Delete interventions

$P(Y \mid \text{do}(X)) = P(Y)$ if there is no path from X to Y with only forward-directed arrows



$$P(\text{Alarm} \mid \text{do}(\text{Damage})) = P(\text{Alarm})$$

Applying the do-calculus



Calculation with interventions



$$\begin{aligned}
 P(y|do\{x\}) &= \sum_z P(y|do\{x\}, z)P(z|do\{x\}) \\
 &= \sum_z P(y|do\{x\}, do\{z\})P(z|do\{x\}) \\
 &= \sum_z P(y|do\{x\}, do\{z\})P(z|x) \\
 &= \sum_z P(y|do\{z\})P(z|x) \\
 &= \sum_{x'} \sum_z P(y|do\{z\}, x')P(x'|do\{z\})P(z|x) \\
 &= \sum_{x'} \sum_z P(y|z, x')P(x'|do\{z\})P(z|x) \\
 &= \sum_{x'} \sum_z P(y|z, x')P(x')P(z|x)
 \end{aligned}$$

Probability axioms

Rule 2

Rule 2

Rule 3

Probability axioms

Rule 2

Calculation with no interventions



$$= \sum_{x'} \sum_z P(y|z, x')P(x')P(z|x)$$

Rule 3

Eliminating interventions

These axioms are complete: they can discover a solution if one exists (Shpitser, 2008).

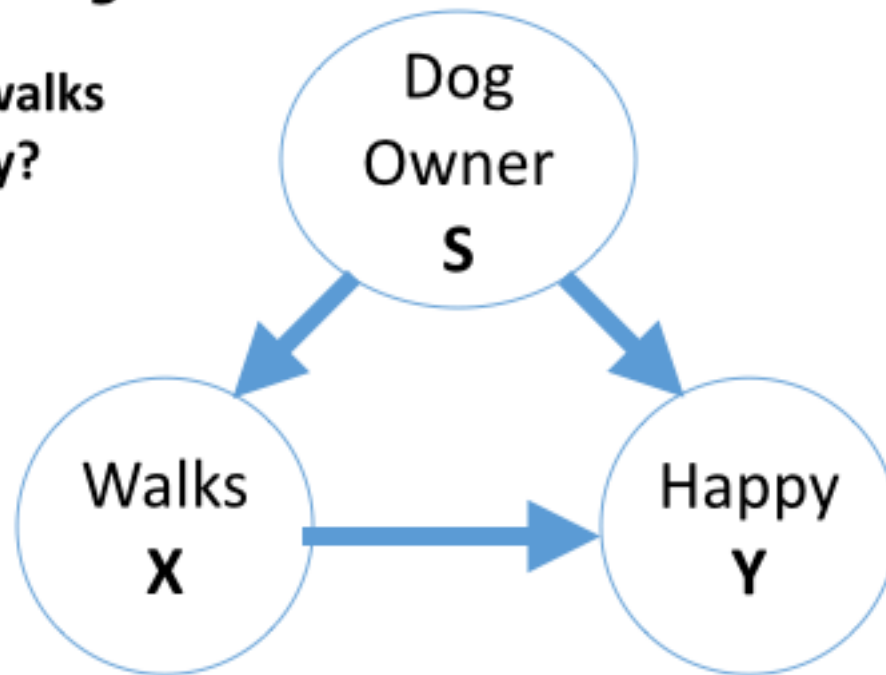
Algorithms exist to discover the solution in polynomial time (Shpitser, 2008).

If there is no solution, then we must do an experiment.

Algorithms exist to tell us what variables to experiment on, or discover other variables to manipulate if we can't manipulate the target (Bareinboim, 2012)

Back door adjustment

Does going on walks
make you happy?



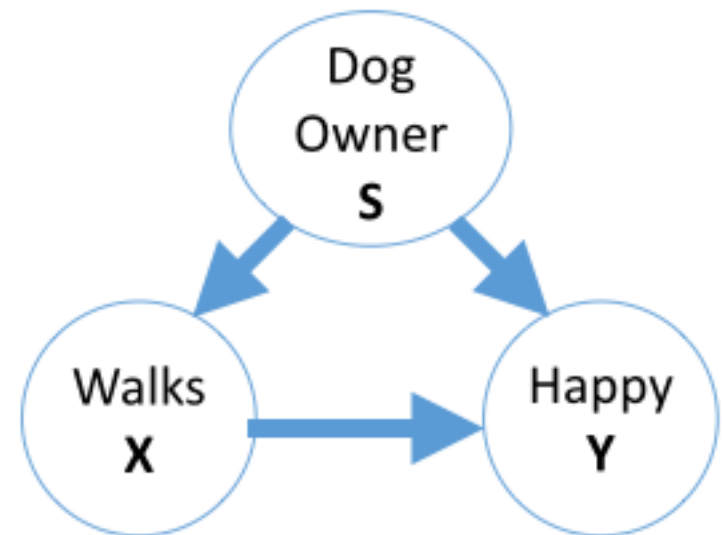
S is a set of variables that satisfies the back door criterion

$$P(Y \mid \text{do}(\mathbf{X} = x)) = \sum_s P(Y \mid \mathbf{X} = x, S = s) P(S = s)$$

Probability of Y if we intervene so
that \mathbf{X} is x

Probability of observing Y
when \mathbf{X} is x,
for each possible value of S

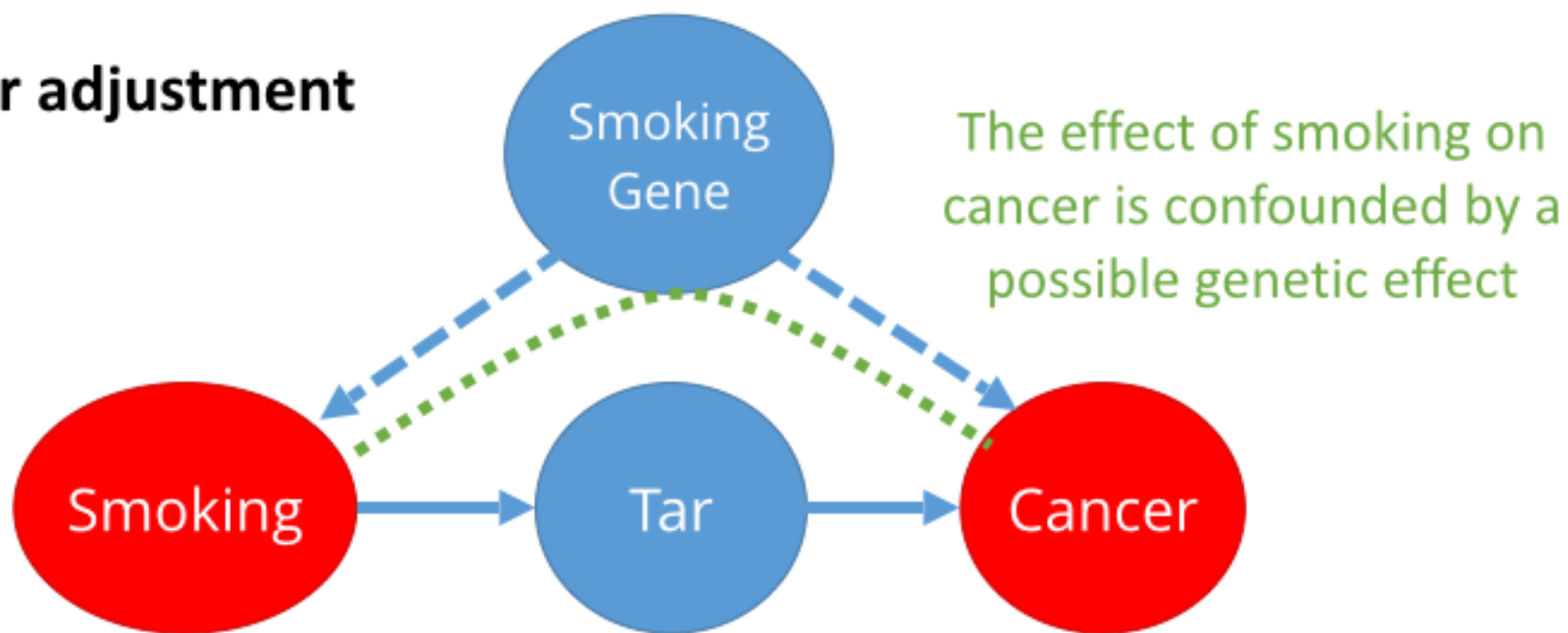
Back door adjustment

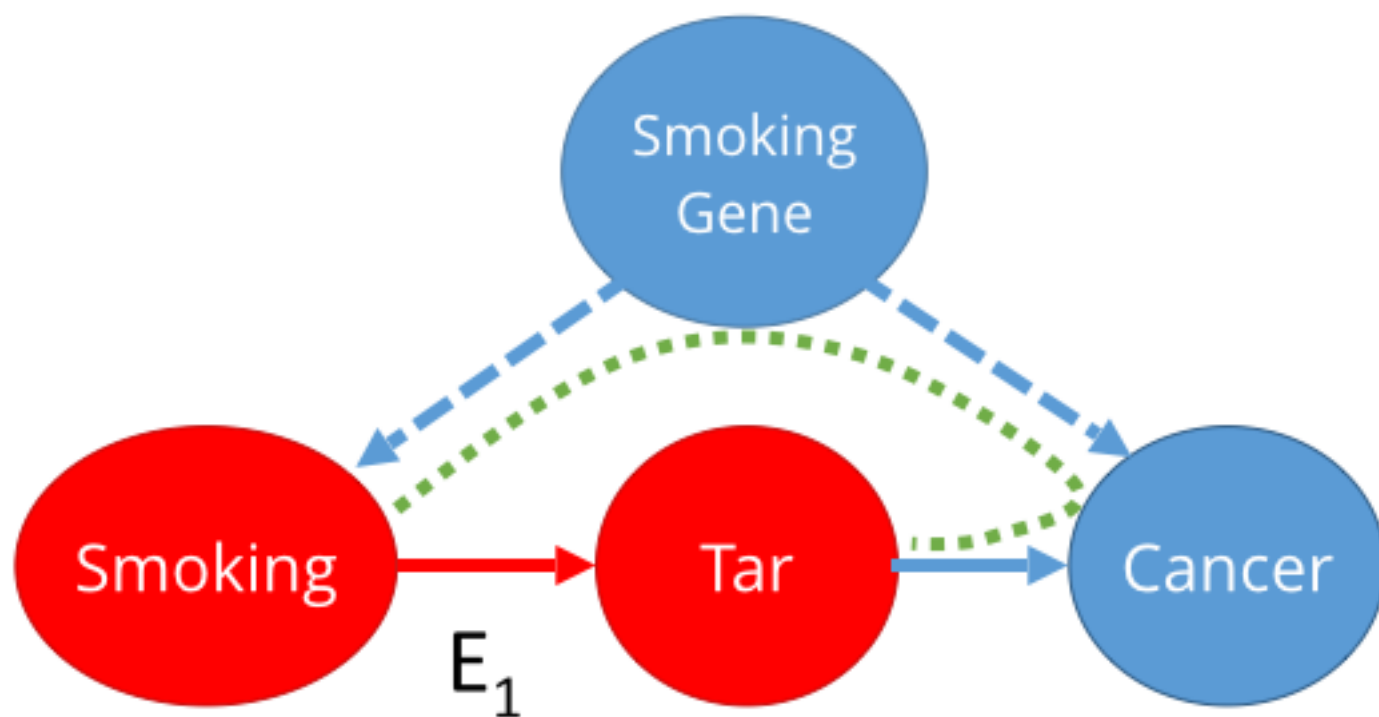


$$\begin{aligned} P(\text{Happy} \mid \text{do}(\text{Walks} = \text{Yes})) &= \\ &\sum_s P(\text{Happy} \mid \text{Walks} = \text{Yes}, \text{Dog} = s) P(\text{Dog} = s) \\ &= P(\text{Happy} \mid \text{Walks} = \text{Yes}, \text{Dog} = \text{Yes}) P(\text{Dog} = \text{Yes}) \\ &\quad + P(\text{Happy} \mid \text{Walks} = \text{Yes}, \text{Dog} = \text{No}) P(\text{Dog} = \text{No}) \end{aligned}$$

= Probability of being happy if you walk and have a dog, weighted by the probability of having a dog
+ Probability of being happy if you walk and don't have a dog, weighted by the probability of not having a dog

Front door adjustment



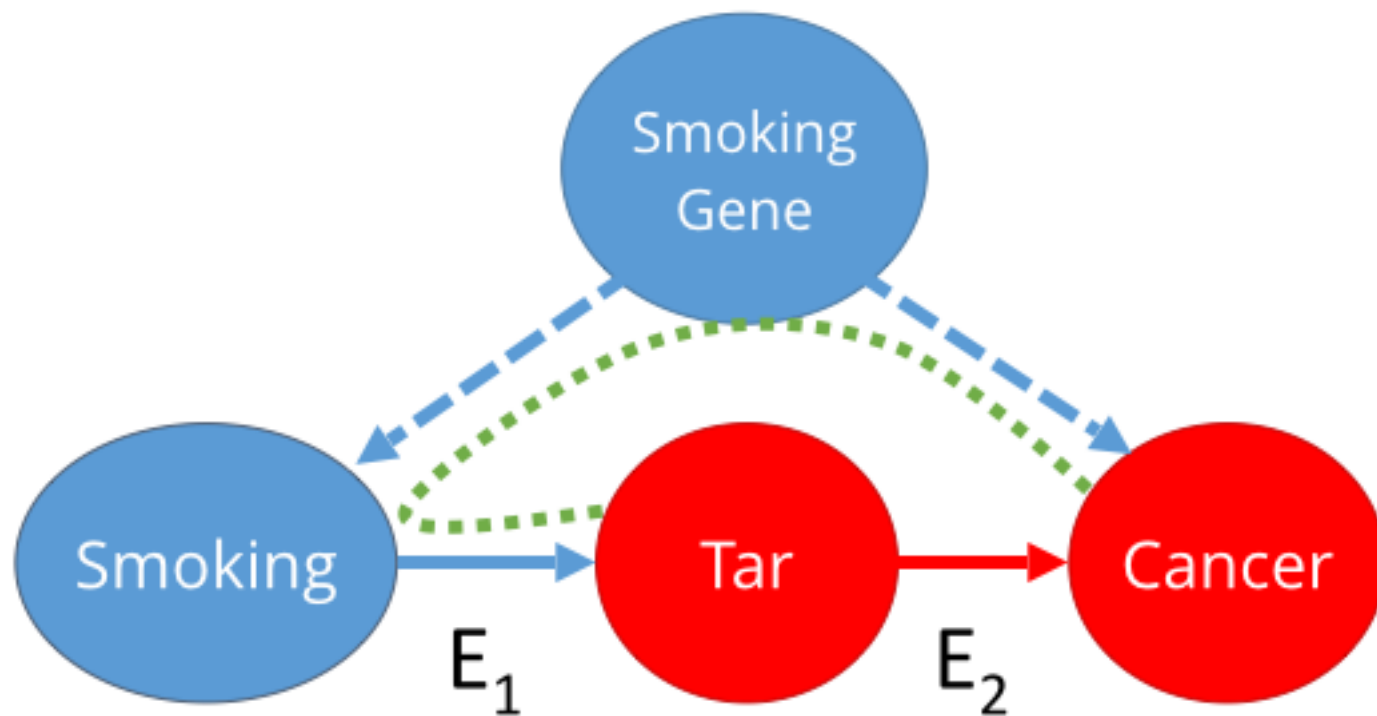


There is a back-door path from Smoking to Tar

BUT it's blocked by the collider in cancer

So we can just use the observed probabilities

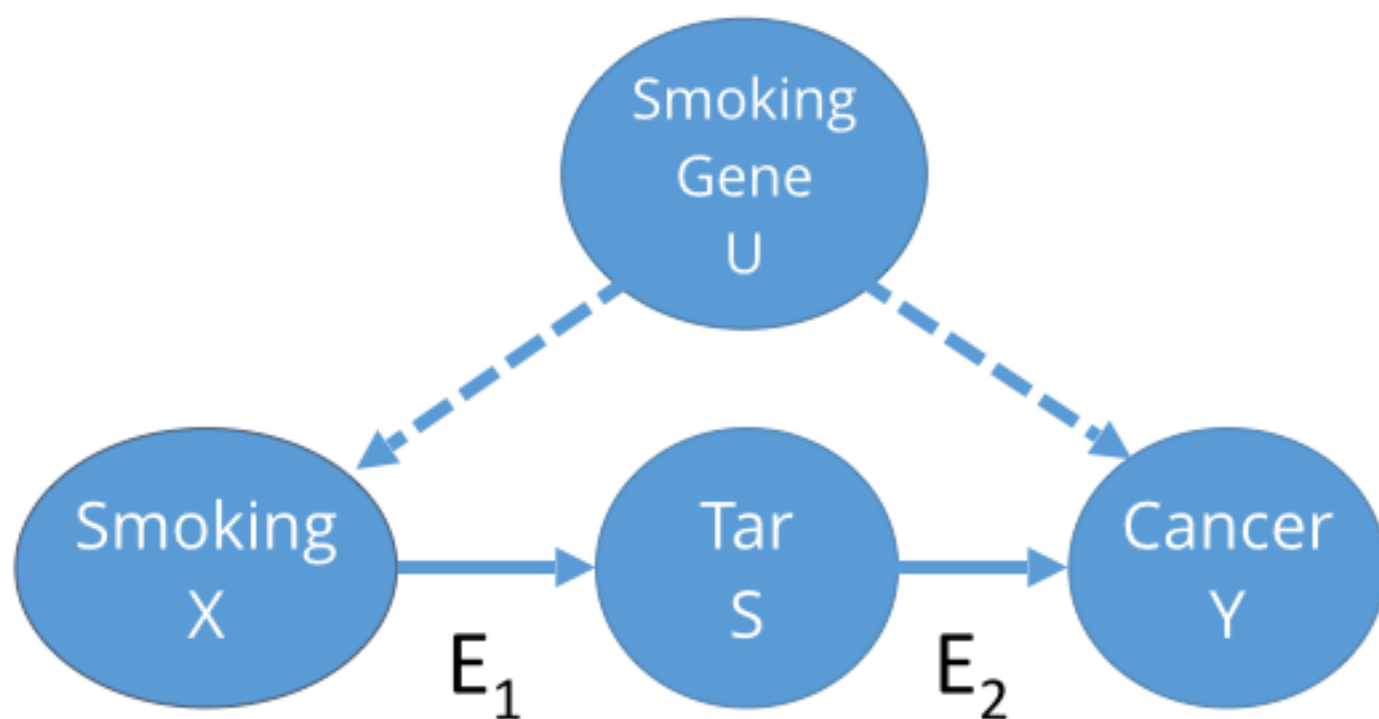
Average causal effect $E_1 = P(\text{Tar} \mid \text{Smoking}) - P(\text{Tar} \mid \text{No smoking})$



There is a back-door path from Tar to Cancer

But we can block it by controlling for smoking.

Average causal effect $E_2 = P(\text{Cancer} \mid \text{do}(\text{Tar})) - P(\text{Cancer} \mid \text{do}(\text{No tar}))$

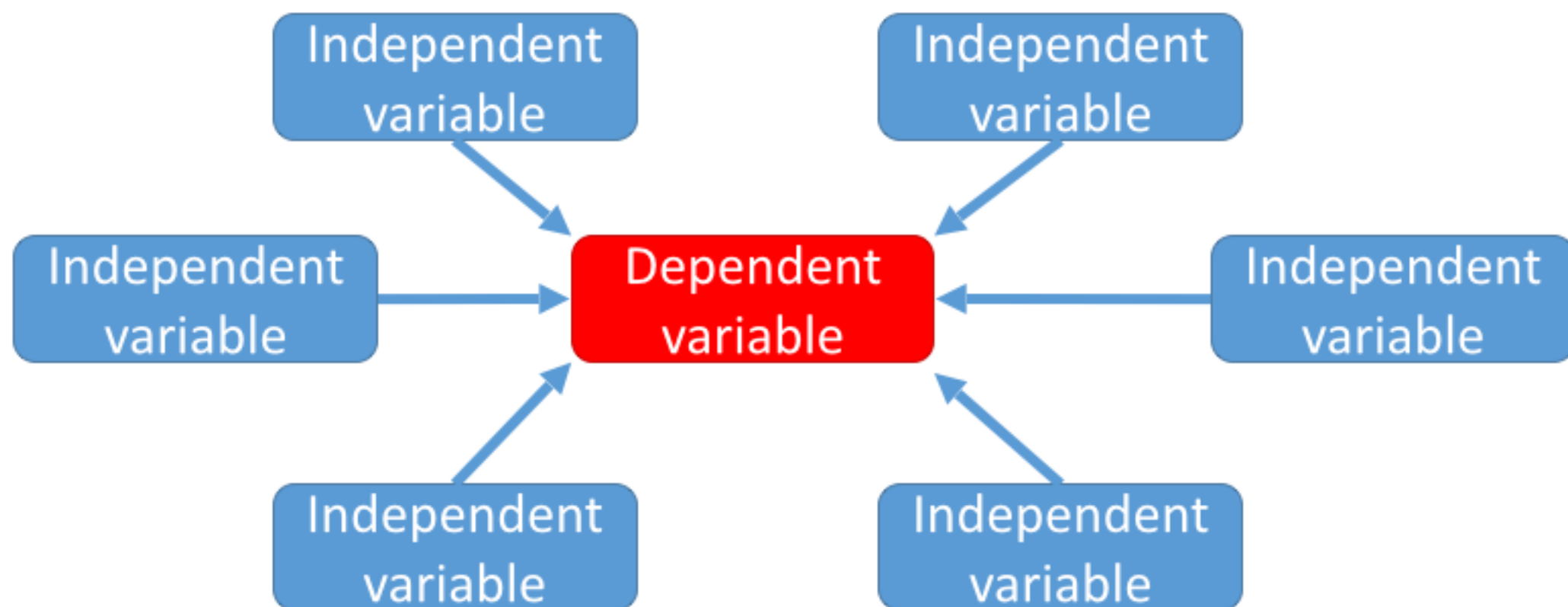


$$\Pr(Y|do(X=x)) = \sum_s \Pr(S=s|X=x) \sum_{x'} \Pr(Y|X=x', S=s) \Pr(X=x')$$

Fitting real data

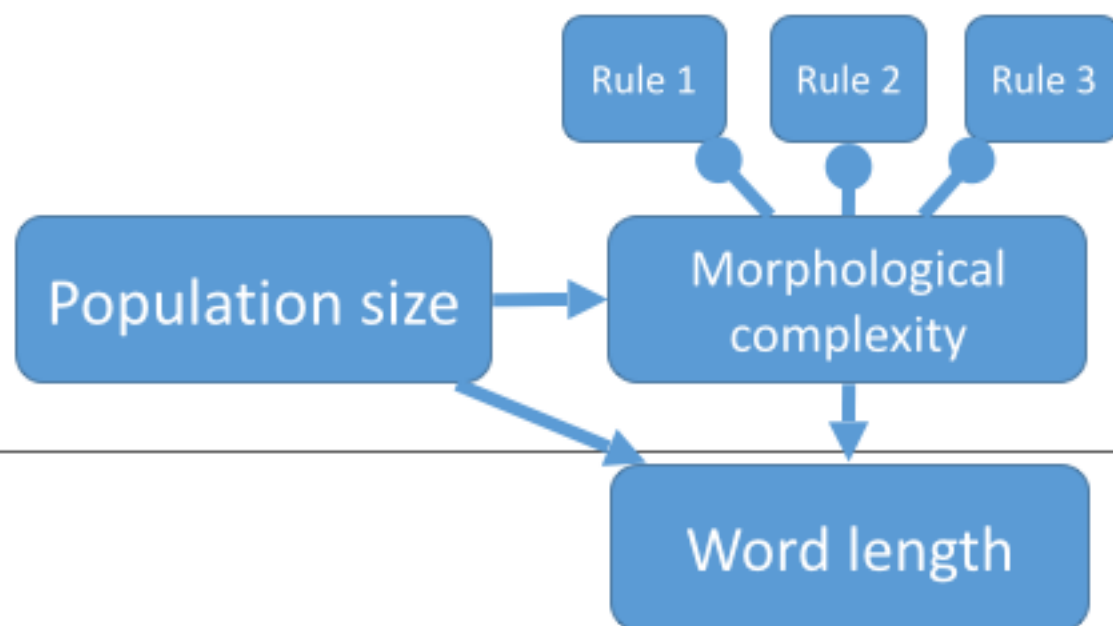
Model fitting in R

Standard assumption of linear model:



Model fitting in R

Convert a causal graph to a structural equation model



Define

```
library(lavaan)
library(semPlot)
model <- "
MorphologicalComplexity =~ Rule1 + Rule2 + Rule3
MorphologicalComplexity ~ PopulationSize
WordLength ~ MorphologicalComplexity + PopulationSize"
```

Fit

```
fit <- sem(model, data=d)
```

Statistics

```
summary(fit, standardized=TRUE)
```

Plot

```
semPaths(fit, 'std')
```

Inferring causal graphs

PC algorithm (Sprites et al., 2000; Kalisch et al., 2012)

Start with fully connected graph

For each pair of variables:

- Try to find evidence that the variables are independent:

 - no correlation,

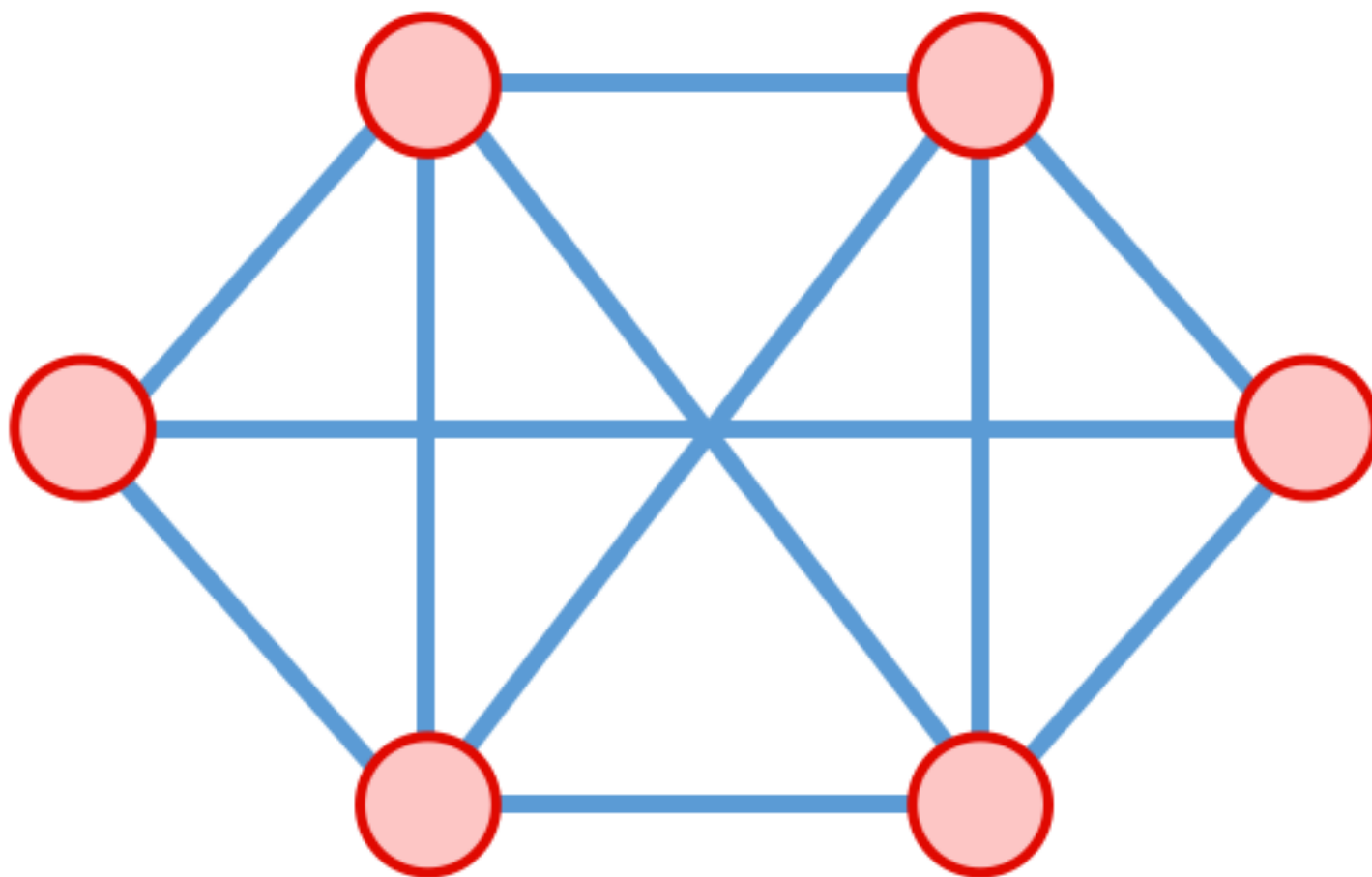
 - or correlation is explained by a set of other variables

- Any statistical test can be used (e.g. conditional independence)

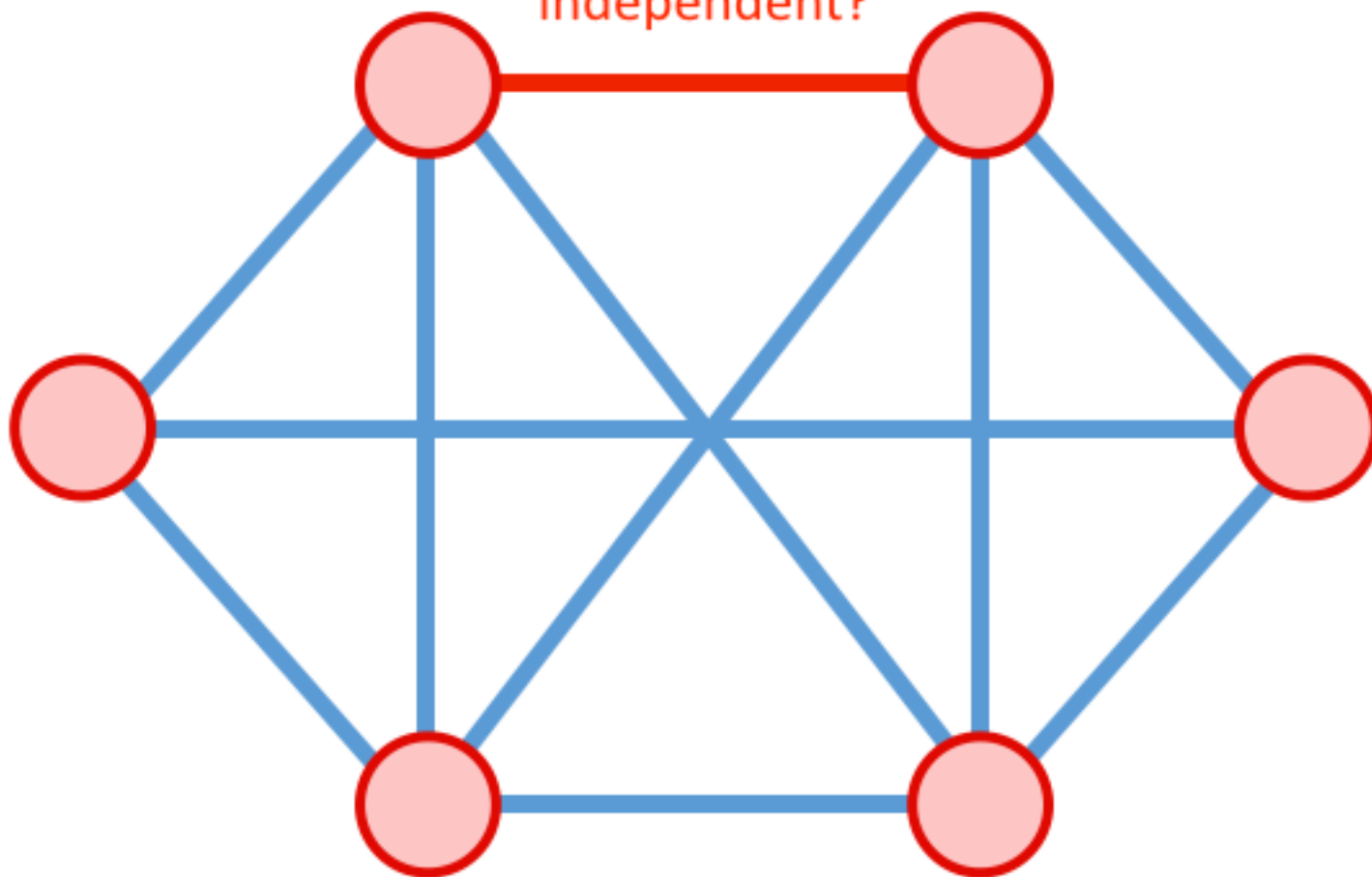
- If variables are independent, remove the edge.

The PC algorithm is an efficient way of performing only the tests which need to be done.

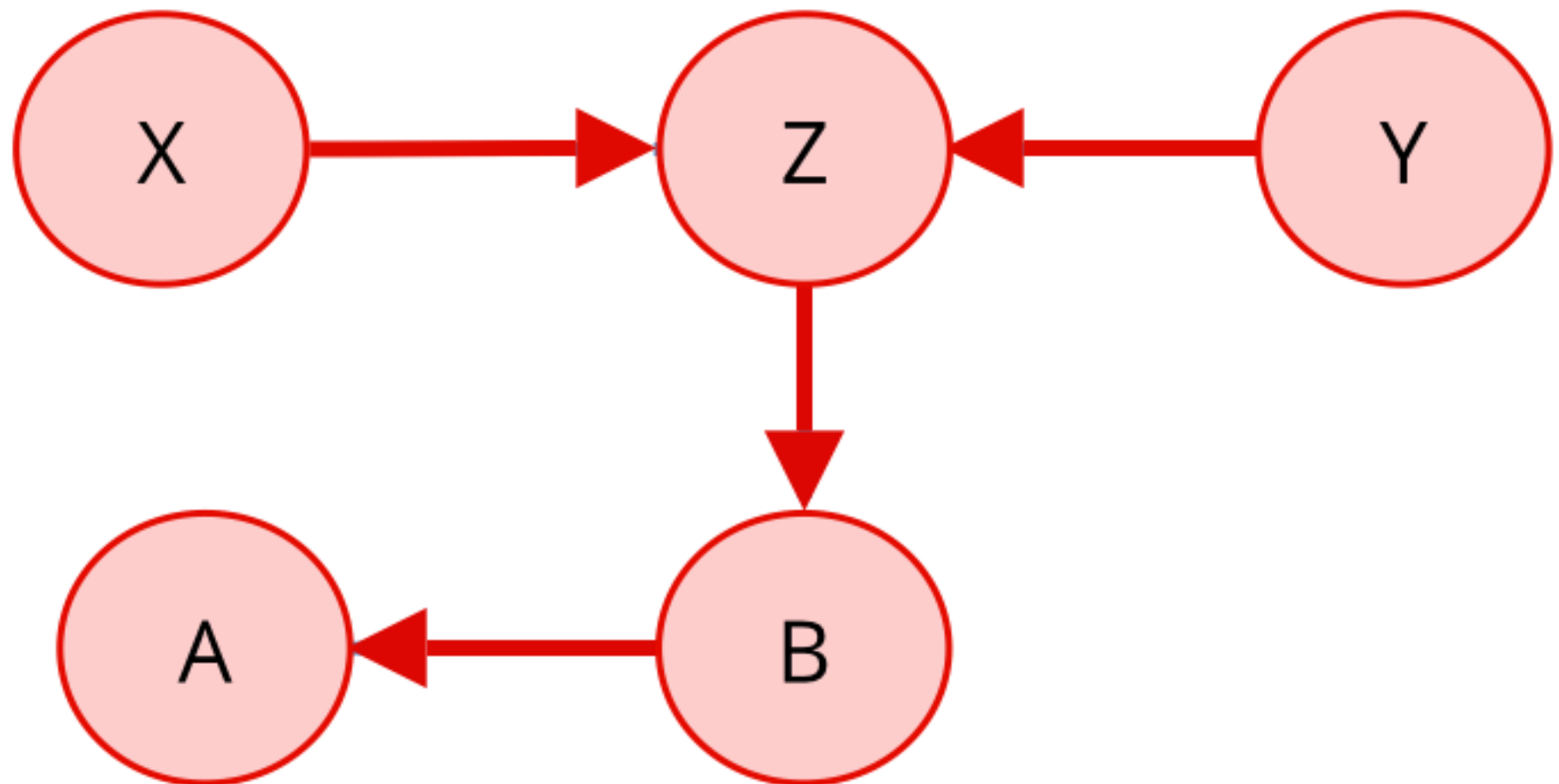
Results in a 'skeleton' graph

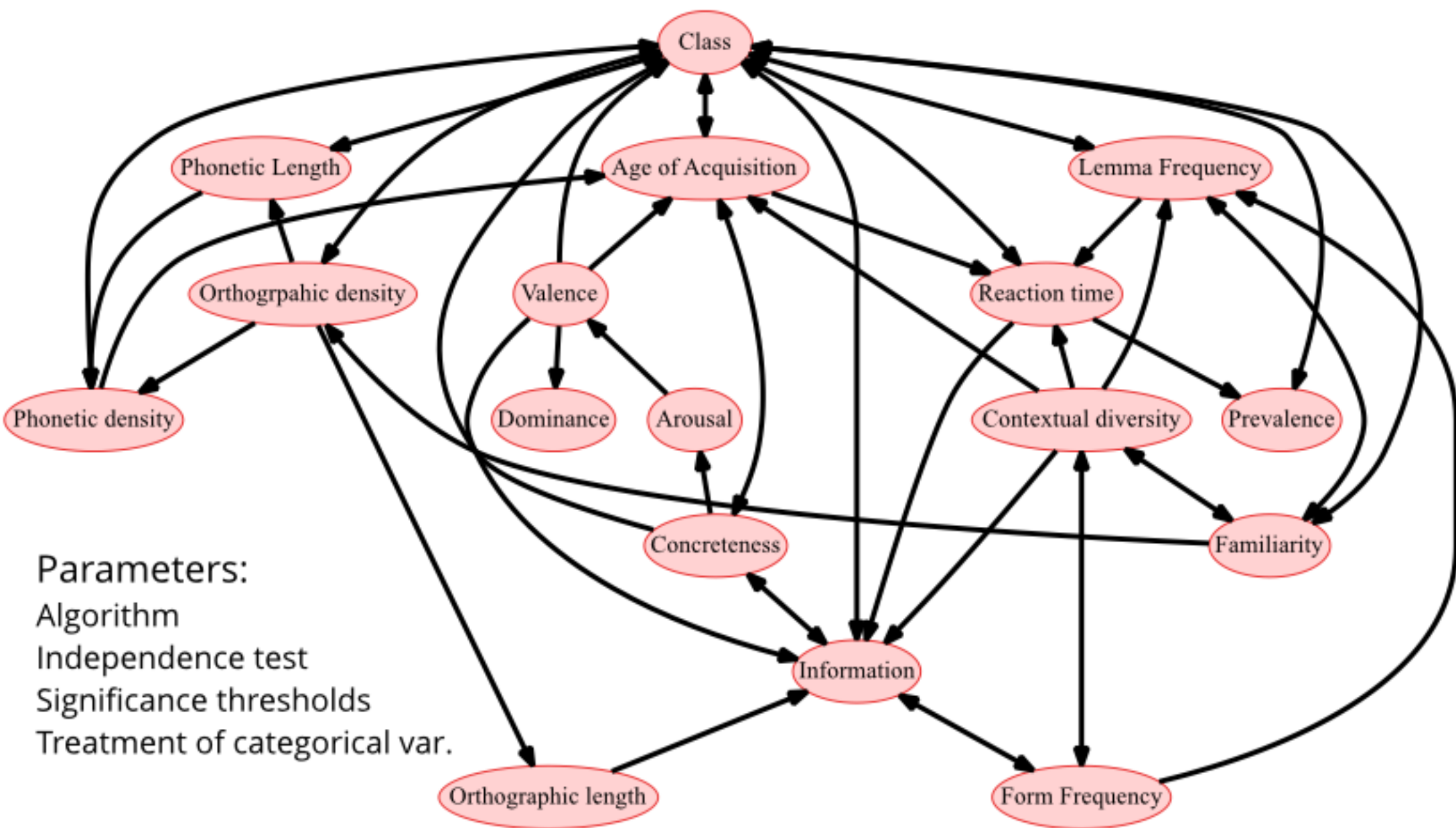


Are these variables
statistically
independent?



Orienting the edges





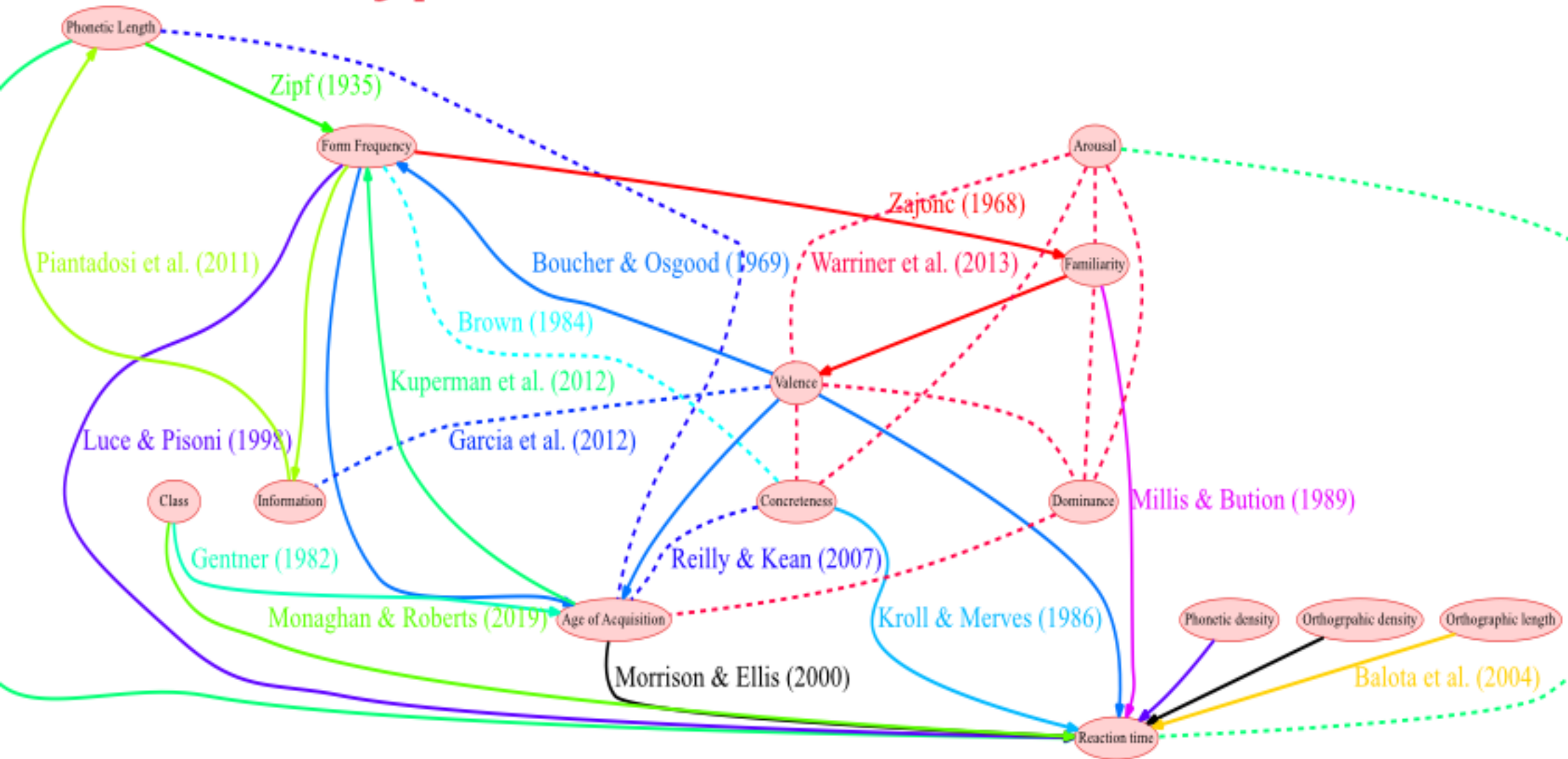
DANGER

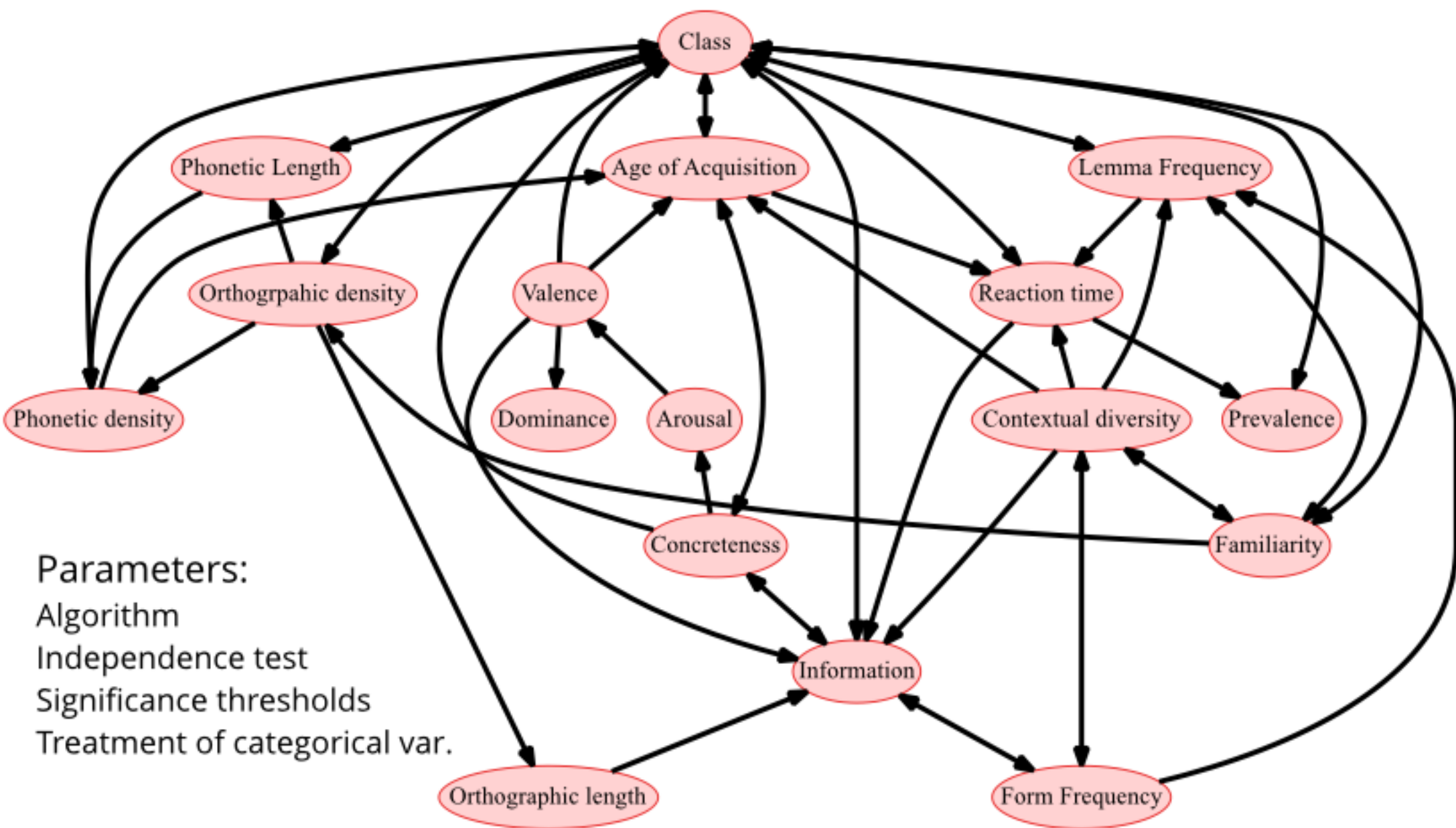
Causal discovery is tricky

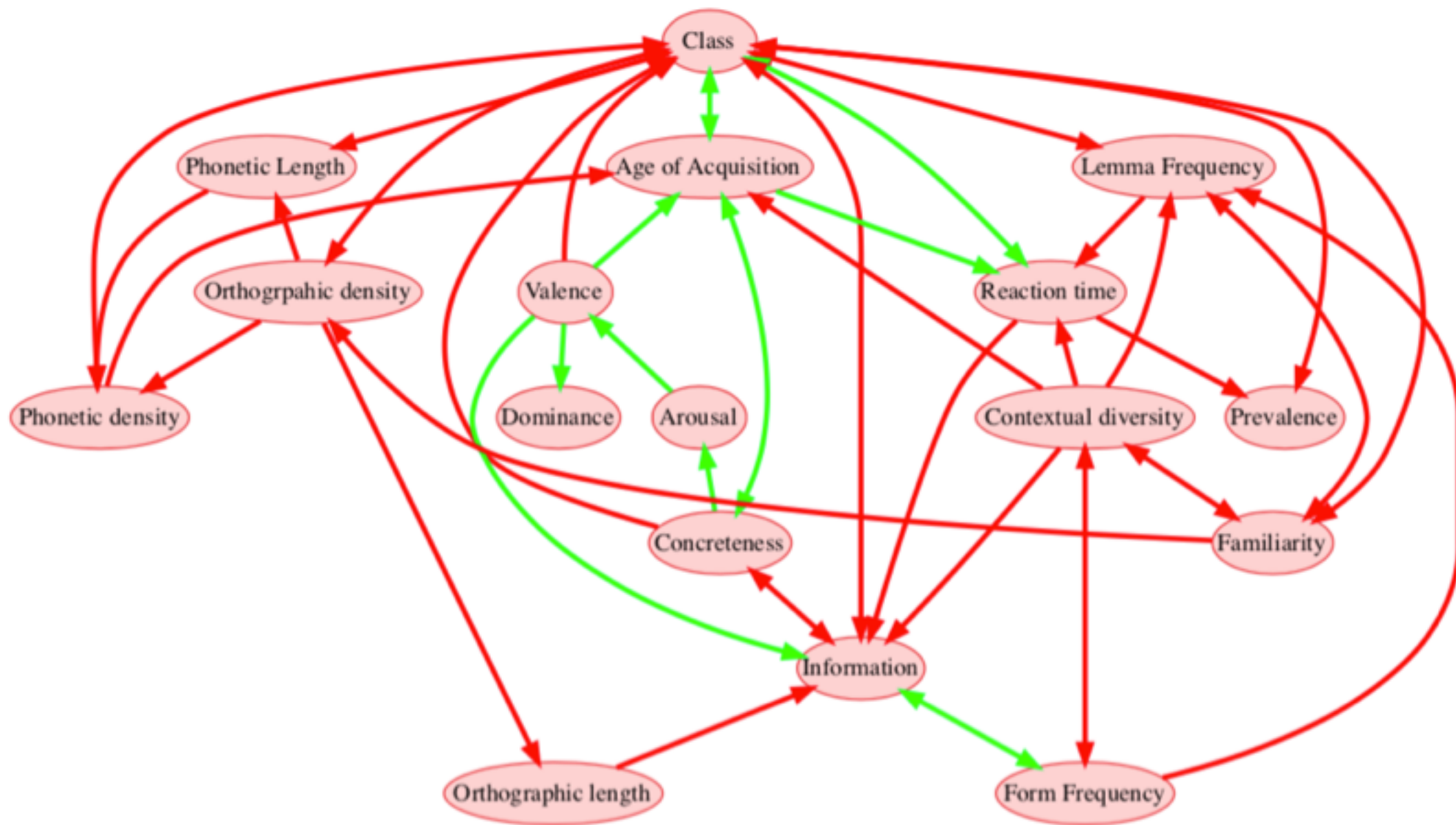
- Requires a lot of data
- Orientation of edges is not robust
- It's easy to justify patterns

How can you protect yourself from ad-hoc storytelling?

Hypotheses about the lexicon





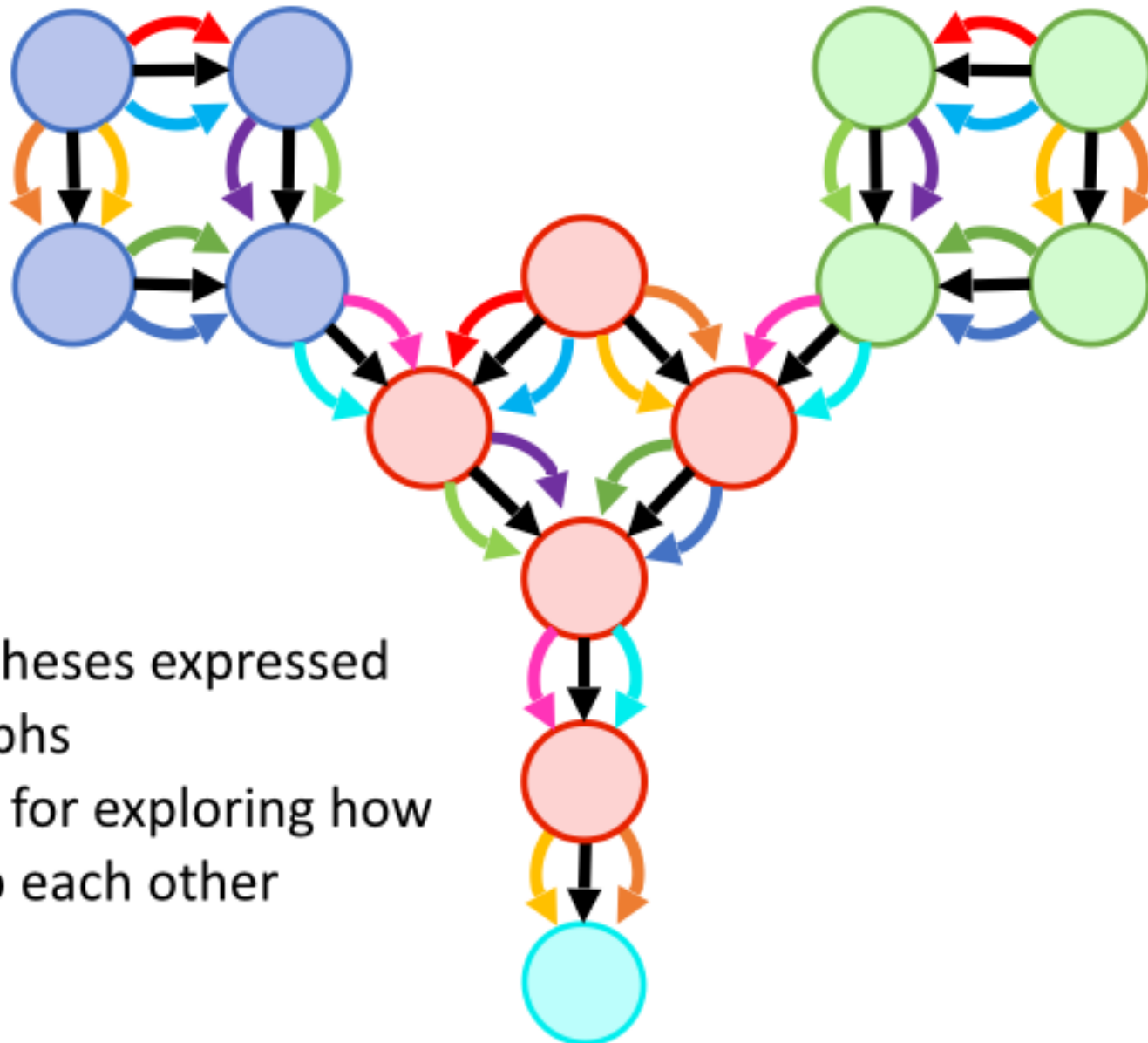


CHIELD

**Causal Hypotheses in Evolutionary
Linguistics Database**



<https://chield.excd.org>

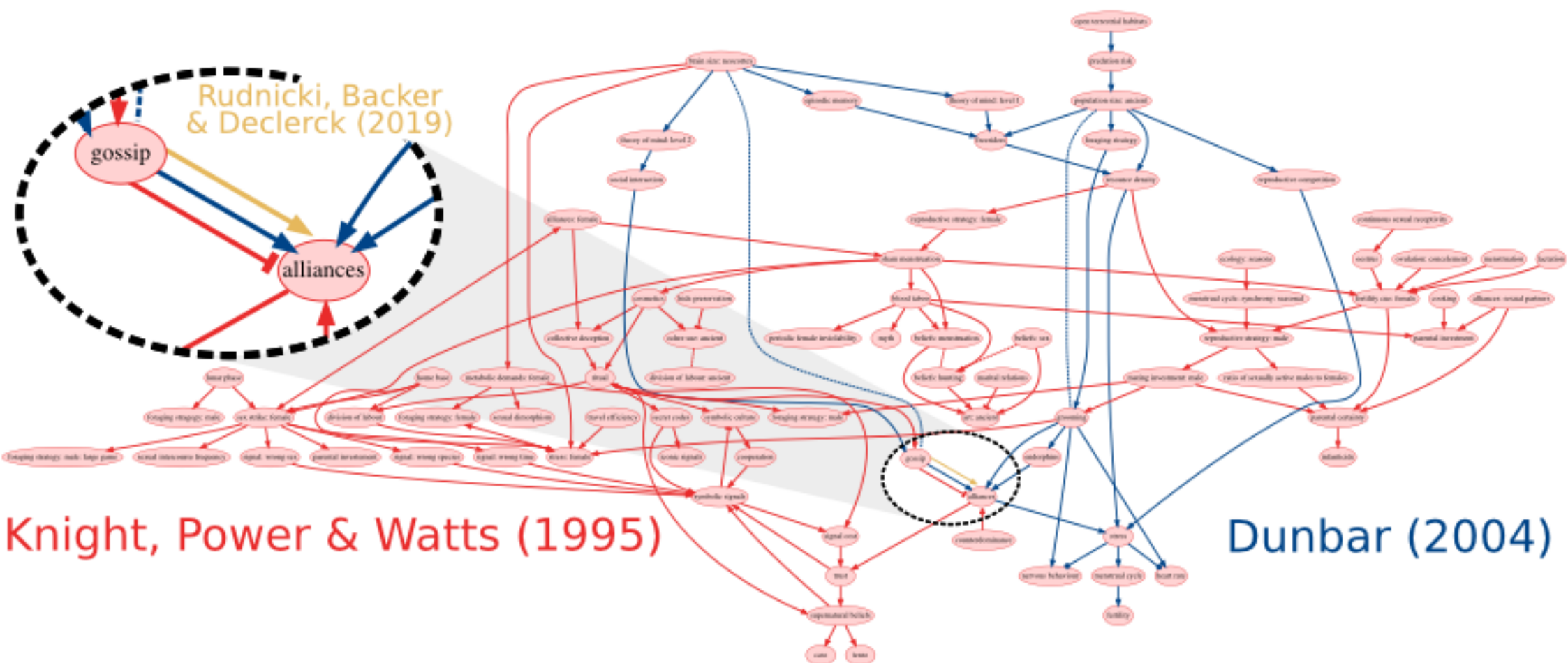


Aims:

- Collect hypotheses expressed as causal graphs
- Provide tools for exploring how they relate to each other

<https://chield.excd.org>

Compare theories, find critical differences and tests of those critical differences

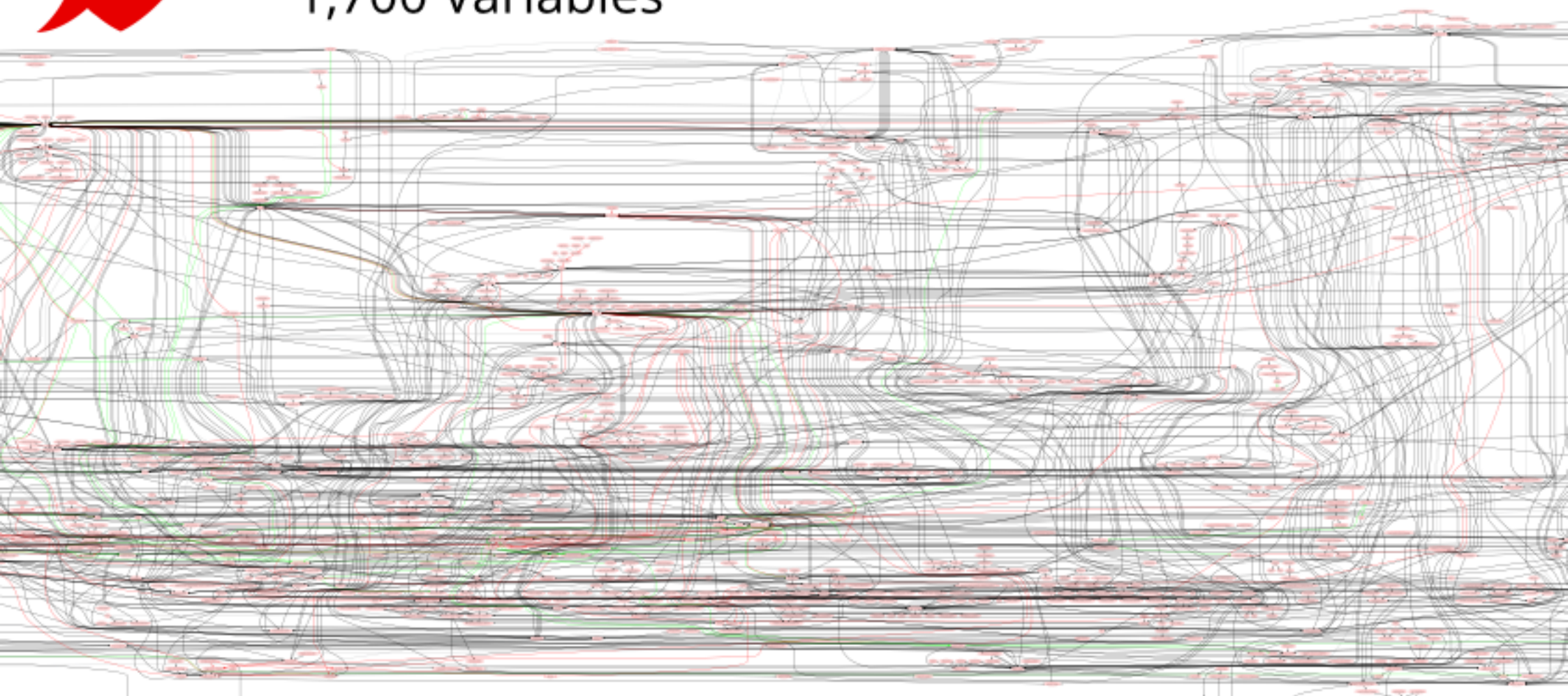


CHIELD Documents			
University of Bristol (GB) https://chield.excd.org/documents.html 120%			
CHIELD Documents Causal Links Variables Authors Explore Add Data Downloads About Help			
Authors ↑↓	Year ↑↓	Title ↑↓	Contributor ↑↓
<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>
Alexandre Celma-Miralles; Juan M Toro	2018	Beat perception in a non-vocal learner: rats can identify isochronous beats	Jasmine Calladine
Andreas Baumann	2018	Linguistic stability increases with population size, but only in stable learning environments	Stephen Mann
Andrew Feeney	2018	Dual-processing and the representational hypothesis: accounting for the emergence of language	Angarika Deb
Curdin Derungs; Martina Köhl; Robert Weibel; Balthasar Bickel	2018	Environmental factors drive language density more in food-producing than in hunter-gatherer	Lindell Bromham



400 Documents
3,400 Links
1,700 Variables

Database is live
<https://chield.excd.org>



Exporting



GraphViz



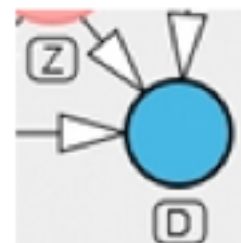
Gephi



phylopath



CSV



DAGitty

Contribute, Discuss, Edit

CHIELD Add Data

University of Bristol (GB) <https://chield.excd.org/addData.html?document=thomas2014a&f>

C.H.I.E.L.D. Documents Causal Links Variables Topics Explore Add Data Downloads About Help

Contributor Reference Causal Links Submit

Upload csv Download csv

Var1	Relation	Var2	Cor	Topic	Stage	Type	Confirmed	Notes
	>							
traditional transmissio	>	iterated learning	pos	transmission	preadaptation			
social cognition	>	iterated learning	pos	transmission	preadaptation			
self-domestication	>	cognition: learn new s	pos	learning	preadaptation			
self-domestication	>	cognition: infer comm	pos	learning	preadaptation			

Code Issues Pull requests Projects Wiki Insights

Question about Evans (2003) #322

Closed rutherford opened this issue 11 days ago · 5 comments

rutherford commented 11 days ago · edited +

Document: [evans2003comment](#)
Contributors: [@angerkuebb](#)

Is this bit "linguistic practices > linguistic ideologies" meant to say that Evans (2003) hypothesises that certain language practices cause certain language ideologies? I'm not sure if Evans (2003) makes this claim. This is a bit controversial, for two reasons. Firstly the causal link between language practices and language ideologies would usually be understood to go the other way (language ideologies > language practices), although some now argue it's both ways <=>. Also, from the link is usually understood as somewhat indirect, because people often don't practice what they preach. Also the phrase "language ideologies" is conventionally used, not "linguistic ideologies", so it might be a better name for the variable. Similarly for language/"linguistic/practices".

1

seanm2 commented 11 days ago Contributor

Thank you for these comments - they are very helpful! We're not experts in this area, so we are grateful for feedback. We'll review the document and make some changes. We'll definitely make the variable name changes you suggest. (You are also welcome to make the edits yourself by pressing the 'edit' button, and you will be credited if you do).

We're interested in capturing theories about language maintenance and change in CHIELD. It would be super interesting to see your 2016 paper on the maintenance of multilingualism as a causal graph. As some motivation, anyone who adds or edits 5 documents will become a co-author on the paper we're writing about this database. (and promotional stuff)

rutherford commented 8 days ago

Hi Sean, I'm glad that those comments were useful. I would be happy to edit the entry for Evans (2003) and contribute some more papers, maybe even 5! I'm reading papers on language evolution at the moment as I'm working on a fellowship proposal on multilingualism and linguistic diversity that includes a language evolution component. I've started reading your PhD thesis, which Matt Spike suggested, it's

<https://chield.excd.org>

Problems

Controlling for some things removes spurious correlations
Controlling for other things **creates** spurious correlations

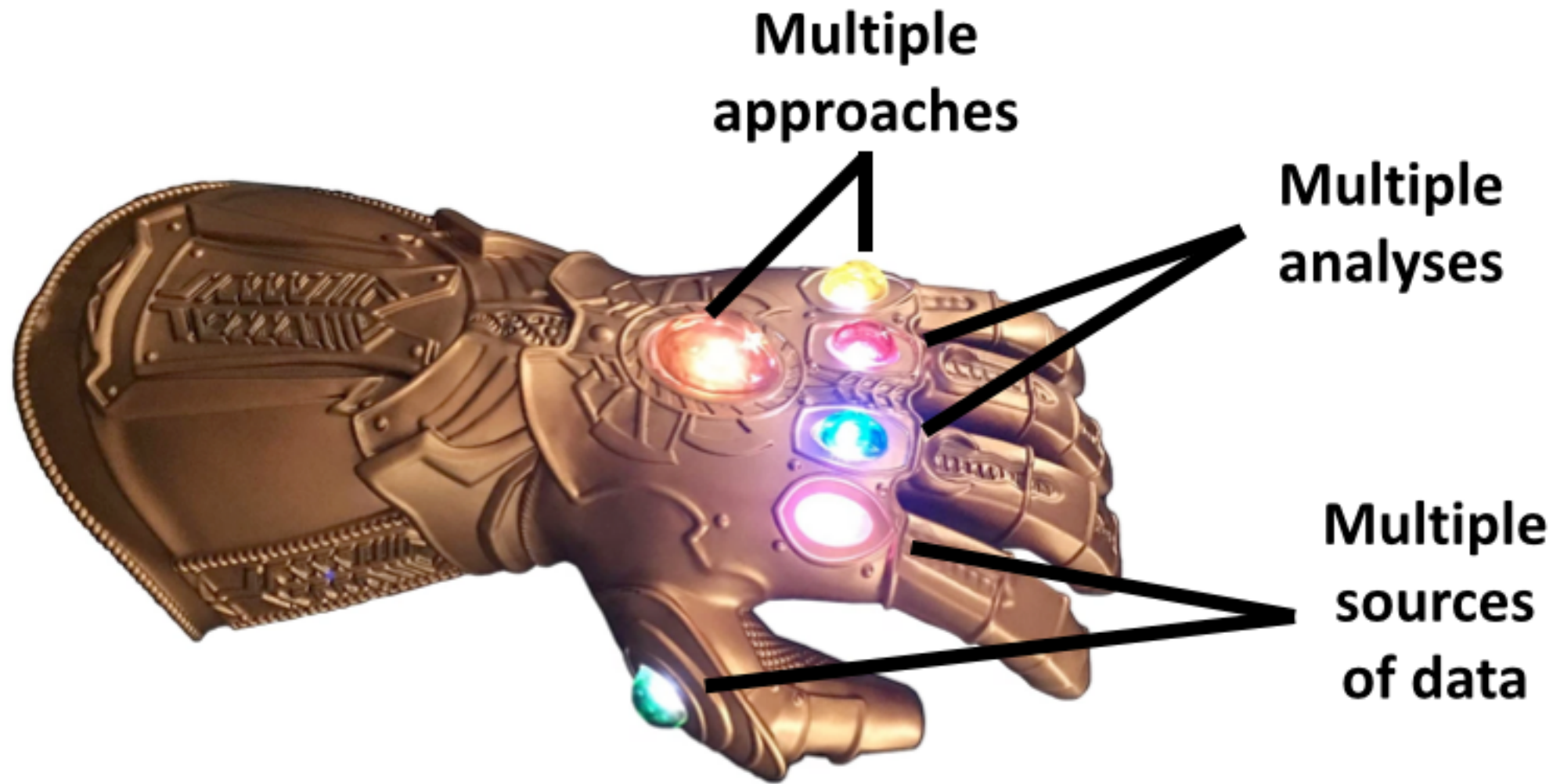
Solution: Build good causal models

Solution 1: Collaboration

- Best data
- Best methods
- Best theories
- More fun



Solution 2: Robust approach to explanation



Solution 3: Incremental approach

- Modest goals for papers
- Modest interpretations



Solution 4:

Causal graphs

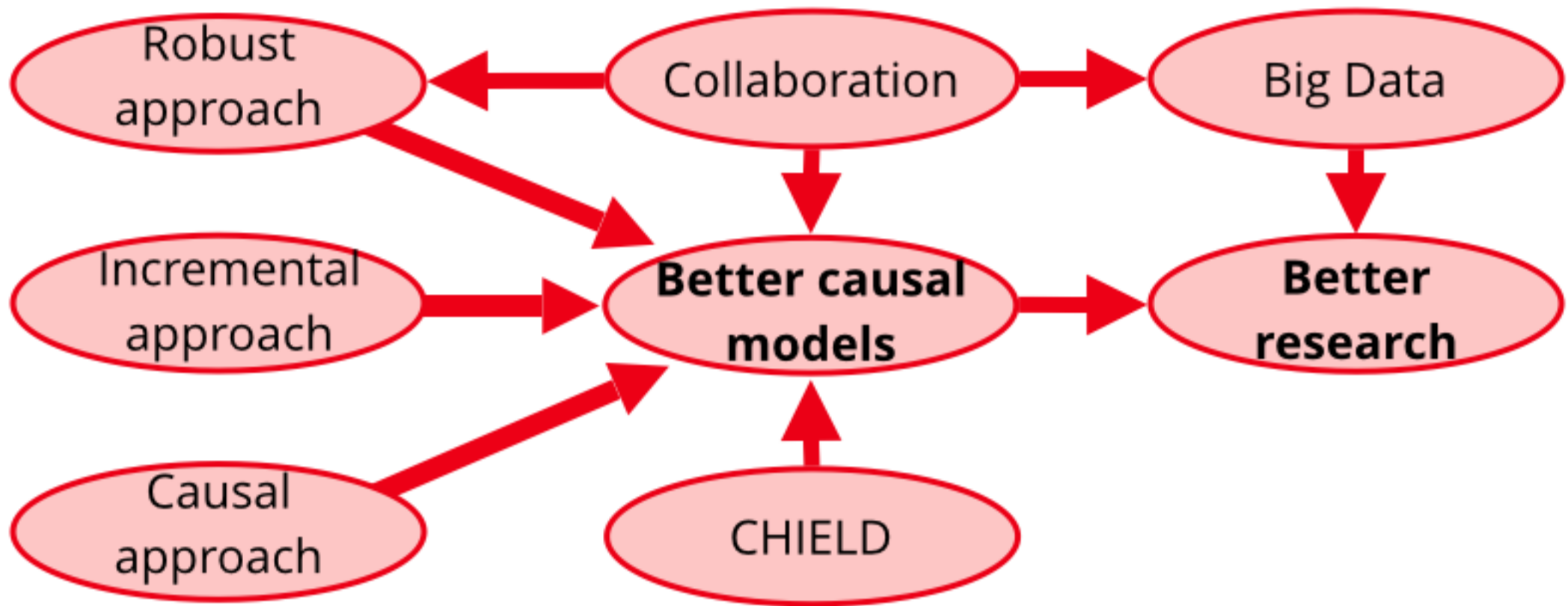
Causal graphs help us to:

- Be more explicit about our hypotheses
- Identify confounds
- Focus on critical differences between hypotheses
- Identify connections between theories

Conclusion

Big data brings many opportunities and challenges

Meeting them will require building better causal models



Recommended reading

- Rohrer, J. M. (2017). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data.
- Pearl, J., & Mackenzie, D. (2018). The book of why: the new science of cause and effect. Basic Books.
- Roberts, S. (2018). Robust, causal and incremental approaches to investigating linguistic adaptation. *Frontiers in Psychology*, 9, 166.

References

Bareinboim, E., & Pearl, J. (2012). Transportability of causal effects: Completeness results. In Twenty-Sixth AAAI Conference on Artificial Intelligence.

Pearl, J., & Mackenzie, D. (2018). The book of why: the new science of cause and effect. Basic Books.

Shrier, I. and Platt, R. W. (2008) . Reducing bias through directed acyclic graphs. BMC Medical Research Methodology, 8(70), 2008.

Shpitser, I. (2008). Complete identification methods for causal inference (Doctoral dissertation, UCLA).

Shpitser, I., & Pearl, J. (2008). Complete identification methods for the causal hierarchy. Journal of Machine Learning Research, 9(Sep), 1941-1979.

Textor, J., Benito van der Zander, Mark K. Gilthorpe, Maciej Liskiewicz, George T.H. Ellison.

[Robust causal inference using directed acyclic graphs: the R package 'dagitty'.](#)

International Journal of Epidemiology 45(6):1887-1894, 2016.

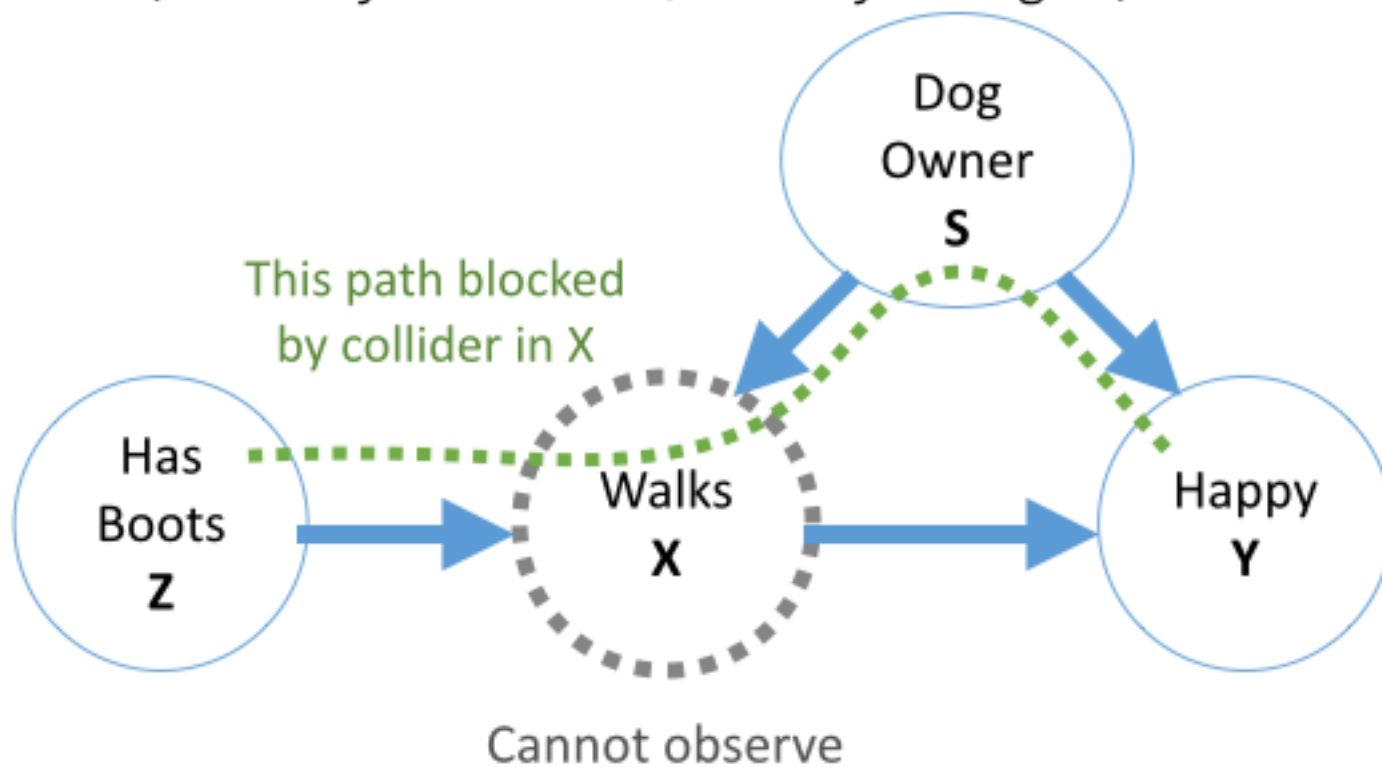
Instrumental variables

Variable Z is an instrument of X for a dependent variable Y if

Z causally influences X,

There are no unblocked alternative paths from Z to other determiners of Y (S)

(Z causally influences Y, but only through X)

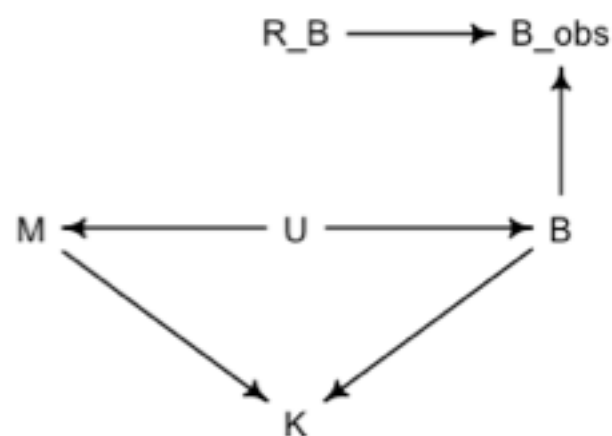


Z is correlated with Y,
but the only source of the
correlation is through X

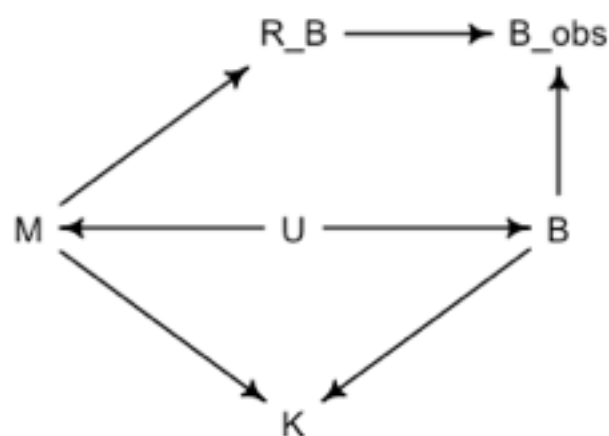
Z simulates an intervention

Data imputation

MISSING COMPLETELY
AT RANDOM



MISSING AT RANDOM



MISSING NOT
AT RANDOM

