# CAUSAL MODELLING

# COMING FROM ONE OF THREE POSSIBLE BACKGROUNDS

**Strong statistical background, and you currently use statistical techniques frequently**: Your task – mapping existing concepts into new use cases. Reorient!

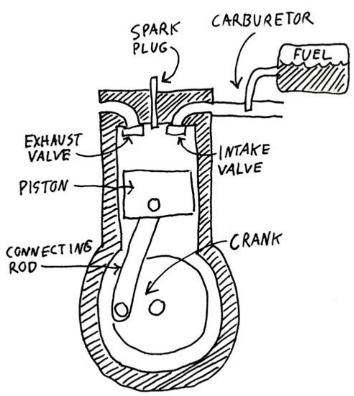
**Had a statistics background at one point, but maybe a bit rusty**: Your task – relearn possibly rusty statistical concepts in this new context of causal modelling

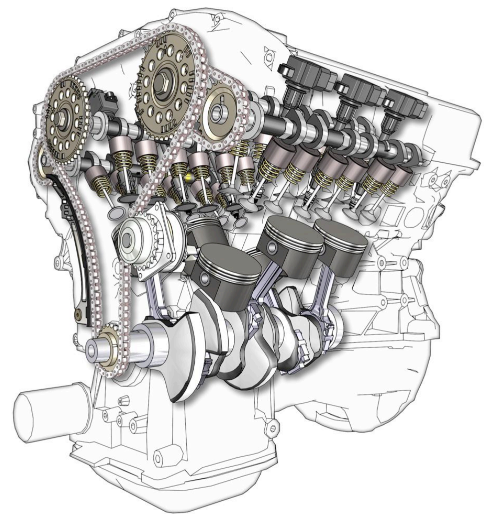
**No background in statistics**: Learn statistical concepts and in parallel learn how these statistical concepts apply to causal modelling

# COMPLEXITY AND THE WORLD

How complex is the world?

# MODELING COMPLEXITY

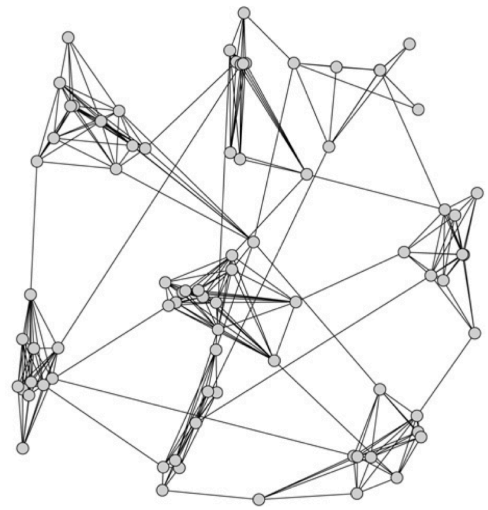


With the help of computers, we handle more complex models

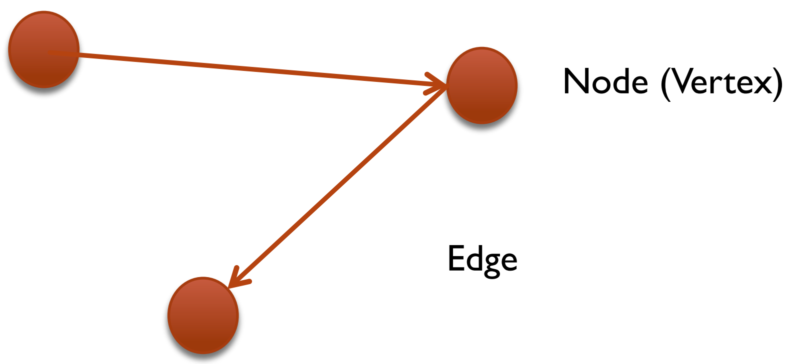
(and models of models)

# MODELING USING GRAPHS



Graphs model **relationships** between objects/entities/events

# THE STAR MODEL TODAY: DIRECTED ACYCLIC GRAPHS




Directed Acyclic Graph (DAG): No loops

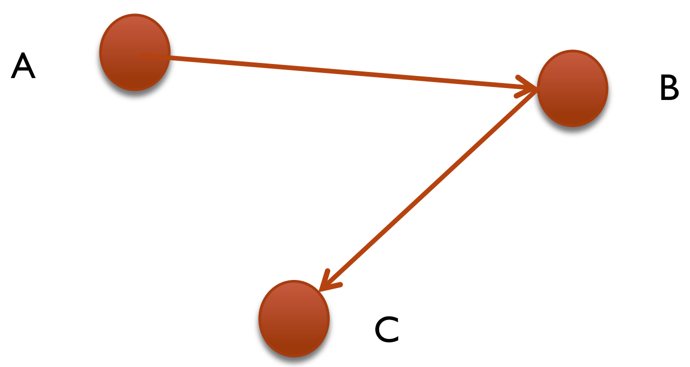
If you have three nodes, how many DAGS could you possibly make?

# WHAT ARE WE GOING TO MODEL TODAY USING DAGS?

Causation!

Causal Relationships: A causes B

We're also going to explore the relationship between causation and probability.



# SOME QUESTIONS TO GET THE WHEELS SPINNING

Does seeing a dark grey cloud cause it to rain?

Does seeing a dark grey cloud increase the probability that it will rain?

Does the presence of a dark grey cloud increase the probability it will rain?

Does the presence of a dark grey cloud cause it to rain?

# SOME QUESTIONS TO GET THE WHEELS SPINNING

Does the presence of high levels of humidity in a cloud cause it to rain?

Does the presence of high levels of humidity in a cloud increase the probability it will rain?

Does the presence of high levels of humidity in a cloud cause it to appear dark grey?

Does the presence of high levels of humidity in a cloud increase the probability it will appear dark grey?

# SOME QUESTIONS TO GET THE WHEELS SPINNING

Does the presence of symptoms CAUSE the disease?

Does the presence of symptoms increase the probability that the disease is present?

# SOME QUESTIONS TO GET THE WHEELS SPINNING

Does seeing a dark grey cloud cause it to rain?

Does seeing a dark grey cloud increase the probability that it will rain?

Does the presence of a dark grey cloud increase the probability it will rain?

Does the presence of a dark grey cloud cause it to rain?

Does the presence of high levels of humidity in a cloud cause it to rain?

Does the presence of high levels of humidity in a cloud increase the probability it will rain?

Does the presence of high levels of humidity in a cloud cause it to appear dark grey?

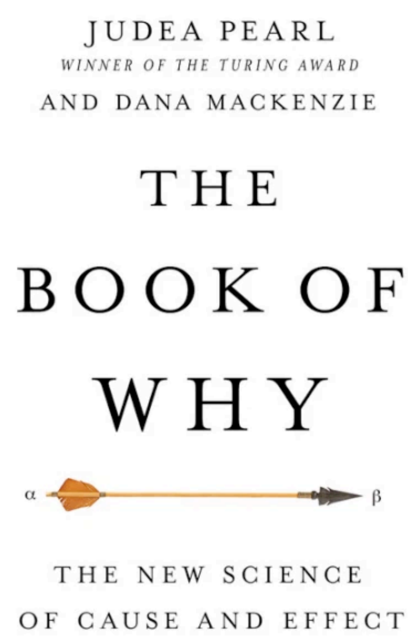
Does the presence of high levels of humidity in a cloud increase the probability it will appear dark grey?

# WHAT IS CAUSATION?

What does it mean for A to cause B?

What are examples of causation?

In the last 20 or so years academic researchers have begun to push this field forward, and have had a lot of debates on the subtle distinctions.




Published in 2020

# REAL WORLD RELEVANCE – THE EPA

"Causation is a difficult and controversial concept. Thus, any causal methodology needs a strong conceptual foundation to be useful and defensible."

(https://www.epa.gov/caddis-vol1/about-causal-assessment)

## Causal Analysis/Diagnosis Decision Information System (CADDIS)

The Causal Analysis/Diagnosis Decision Information System, or CADDIS, is designed to help scientists and engineers in the Regions, States, and Tribes conduct causal assessments in aquatic systems. It is organized into five volumes.

### Learn About CADDIS

- Basic Information
  - How To Cite CADDIS
- Frequent Questions
- Glossary

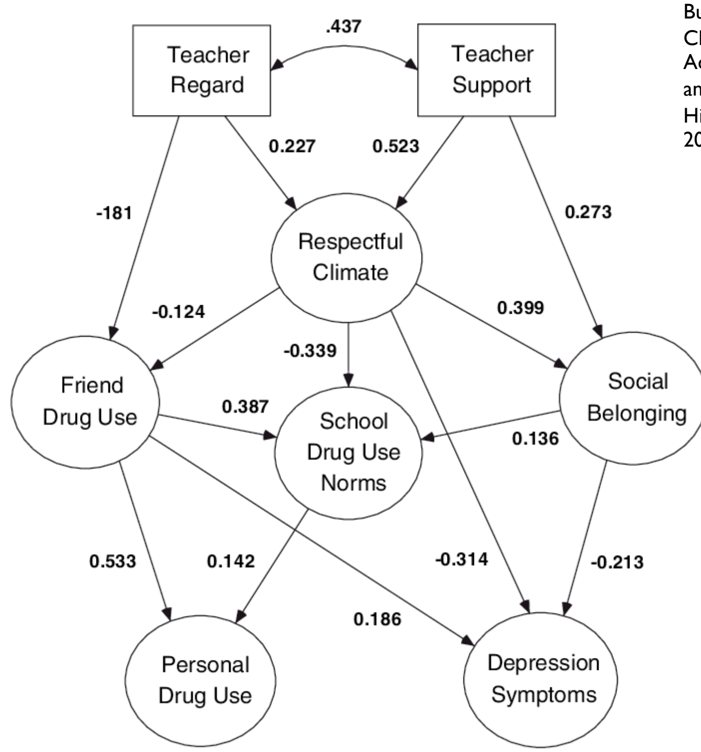
### Volume 1: Stressor Identification

- CADDIS Volume 1 Home
- Learn About Causal Assessment
- Getting Started
- Tips for Candidate Causes
- Types of Evidence Tables

# CAUSAL MODEL EXAMPLE

Causal models are very diagram oriented, and just by looking at a causal diagram you can start to interpret it.

Here's an example of a causal diagram:




Taken from: Teachers as Builders of Respectful School Climates: Implications for Adolescent Drug Use Norms and Depressive Symptoms in High School LaRusso et al 2007

# WHAT DO CAUSAL DIAGRAMS DO?

Suppose we want to predict something?

Suppose we want to ***change*** or ***affect*** something?

If I change A, will it also change X? Why? In what way? How much?

Causal diagrams help us with this!

# HOW TO MAKE CAUSAL MODELS:

From the World Health Organization (**Causal modelling** - Challenges for its development and usefulness for programme management)

1. Make comprehensive list of potential causes
2. Structure causes in layers of influence
3. Explore the plausibility of the presumed causal link (s)
4. Finalise the model
5. Prioritise relevant actions to combat the problem at hand (starting point of causal model)
6. Develop list of indicators

This will be a TEAM effort

# EXERCISE – CAUSAL MODEL SKETCH

Suggested starting points:

- Pick a situation or system of interest
- Pick an element of that system – your goal is to understand the influences on this element of the system
- List other aspects, elements, events, situations, etc. within the system that you believe could influence this element of the system.
- List aspects that in turn could influence these events
- Organize all of these elements into a network of elements that cause other elements to change.

## MOVING INTO MORE RIGOROUS TERRITORY

Causal models are gaining in popularity because they help to capture subject matter expertise in a way that does not require mathematical modelling up front.

They reveal assumptions – shared or not shared!

**Then** we can relate what is in the diagram to more rigorous mathematical and statistical concepts, and even a specific set of equations that represent a statistical model.

How do we go from one to the other?

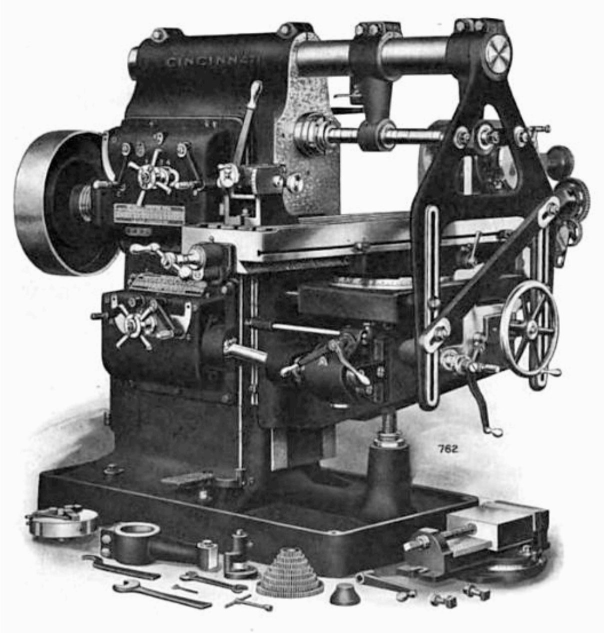
# THE MECHANICS OF CAUSATION

Billiard balls and milling machines: a clear mechanical physical process we can trace.

But that's sometimes too much: we want to say that eating cake is a partial contributor to feeling happy, without worrying about the exact physical processes underlying it.

What about statistical models? They seem to model relationships between elements without getting into the mechanics.

What can they tell us about causality?



# CORRELATION DOES NOT EQUAL CAUSATION… BUT…

Statisticians have somewhat used this truth as a shield - but people have been demanding more: "statistical regularities are the 'symptoms' of causal relations " Russo (2011)

Four possible meanings if something is correlated (e.g. using Pearson's correlation coefficient)

1. A causes B
2. B causes A
3. C causes A and B
4. It's spurious correlation - there's no causation involved.

# RELATIONSHIP BETWEEN PROBABILITY AND CAUSATION - MORE DETAILS

We know that probability and causality are not the same thing, but what exactly is the relationship? This is the topic of 'Probabilistic Causation'

Perspective on this question has changed over time:

1950s – 2000s: Causes raise the probability of their effects. End of story?

2000s – today: We CAN infer causality from probabilities, but it's complicated! Causal structure <u>constrains</u> the values of probabilities. Also, we like Bayes.

Side note: specific causality vs general causality

(https://plato.stanford.edu/entries/causation-probabilistic/#ProbRaisTheoCaus)

# DAGGITY

A free web-app for"creating, editing, and analyzing causal diagrams (also known as directed acyclic graphs or causal Bayesian networks)"

http://www.dagitty.net

This site has many resources, and in the process of learning how to use the app you will encounter many of the more rigorous and detailed elements of creating causal models.

# KEY CONCEPT: COUNTERFACTUAL

In causal modelling, we are exploring counterfactuals:

- If A were to occur B would also occur
- If A had NOT occurred, B would NOT have occurred
- If A were to change, B would change

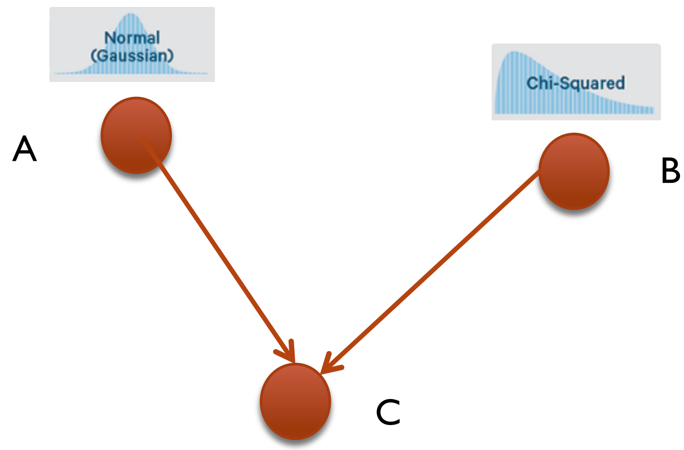
Here we are stating a specific direction of cause and effect.

This is different from: The data shows that when A changes, B changes

# KEY CONCEPT: DISTRIBUTION OF VARIABLE

Probability of each range of values of the variable, represented by an equation.

If each node in our causal diagram represents a variable, there is an equation associated with each node



## KEY CONCEPT: CONDITIONAL PROBABILITY

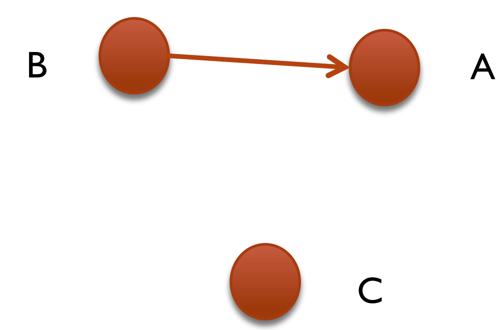
Variable A might be *dependent on* variable B, but not on variable C

The causal diagram here shows this.

Here, P(A and B and C) = P(B) * P(A|B) * P(C)

This diagram could mean: B influences/causes A, but B doesn't cause C and A doesn't cause B or C and C doesn't cause A or B

What you *don't* connect is *as meaningful* as what you do connect



# CONDITIONAL INDEPENDENCE

Variables A and B are conditionally independent from each other if, *once we know the value of another variable C*, neither A nor B affects the other.

The patterns of edges in a causal diagram (d-separation, colliders) encode conditional independence

Once we have created the diagram, we can use techniques to confirm these relationships.

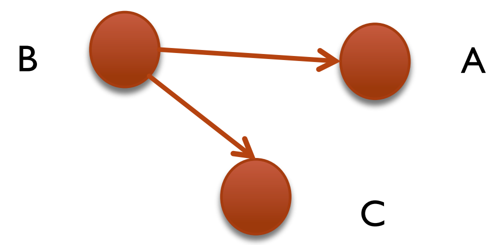
Note: If we say that variable A is **conditioned on** variable B, that means we assume that variable B represents a state, condition or event that has occurred.

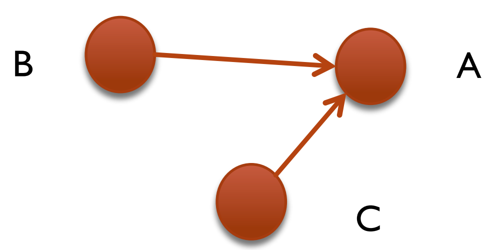
# KEY CONCEPTS: CONFOUNDER, COLLIDER, MEDIATOR

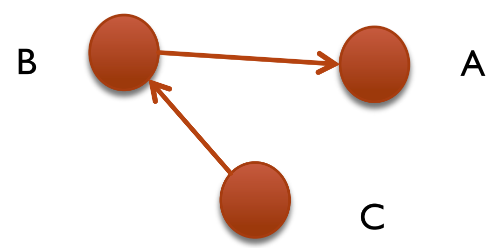
Confounder: Common cause of X and Y

Collider: Common Effect of X and Y

Mediator: A variable in the middle of a path



# KEY CONCEPTS: D-SEPARATION, D-CONNECTION

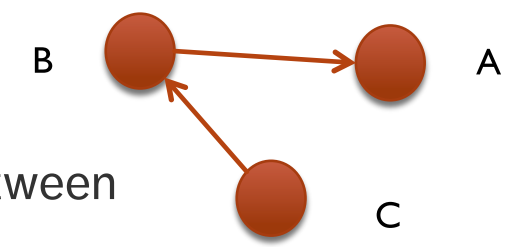
We want to understand change! What is the *effect* of changing X?

d-Separation: helps you to determine if a **set** of variables X is independent from another **set** of variables Y (given certain conditions)

Variables that are *not connected* in the diagram are independent

*x* and *y* are *d*-connected if there is an *unblocked path* between them

There are a number of ways that a variable C can *block* the path between other variables



# CAUSAL MODEL COUSINS

**Structural Equation Models (SEMs)**: Comes from traditional statistics. Sets of equations relating exogenous and endogenous variables to effect (dependent) variables

**Bayesian Networks/Belief Networks**: Comes from the newer Bayesian tradition as well as Pearl's tradition. People argue that SEMs are essentially the same as these and vice versa.

**Markov Chains**: Typically used to represent memory-less state changes for a single variable, but conceptually quite similar, in an abstract sense

Many other cousins: (e.g. Dynamic Bayesian Networks, Fuzzy Cognitive Maps)

# MODELS PLUS MONTE CARLO (CRYSTAL BALL)

Crystal ball allows you to create Monte Carlo (MC) simulations

What is a Monte Carlo simulation?

- You have probability distributions for variables
- You have a model connecting these variables
- The MC simulation generates specific values for the variables based on the distributions, and determine the specific outcome.
- By doing this repeatedly, you can generate synthetic data

# REPRESENTING GRAPH MODELS

You can use the Monte Carlo method to run simulations using models investigating probability relationships and causality (e.g. Markov Chains and Bayesian Networks)

# EXAMPLE CASE STUDY

**A Bayesian Approach to Integrated Ecological and Human Health Risk Assessment for the South River, Virginia Mercury-Contaminated Site**

**Meagan J. Harris, Jonah Stinson, and Wayne G. Landis***

Used Crystal Ball + Bayesian Network to carry out a risk assessment

We conducted a regional-scale integrated ecological and human health risk assessment by applying the relative risk model with Bayesian networks (BN-RRM) to a case study of the South River, Virginia mercury-contaminated site. Risk to four ecological services of the South River (human health, water quality, recreation, and the recreational fishery) was evaluated using a multiple stressor–multiple endpoint approach. These four ecological services were selected as endpoints based on stakeholder feedback and prioritized management goals for the river. The BN-RRM approach allowed for the calculation of relative risk to 14 biotic, human health, recreation, and water quality endpoints from chemical and ecological stressors in five risk regions of the South River. Results indicated that water quality and the recreational fishery were the ecological services at highest risk in the South River. Human health risk for users of the South River was low relative to the risk to other endpoints. Risk to recreation in the South River was moderate with little spatial variability among the five risk regions. Sensitivity and uncertainty analysis identified stressors and other parameters that influence risk for each endpoint in each risk region. This research demonstrates a probabilistic approach to integrated ecological and human health risk assessment that considers the effects of chemical and ecological stressors across the landscape.

**KEY WORDS:** Bayesian network relative risk model; ecological risk assessment; ecological services; human health risk assessment; mercury

# WHAT'S THE POINT?

So what can you do with these causal diagrams? What's the point?

Making explicit what was implicit - now I can draw my diagram and you can draw yours and we can see if they match! If they don't…

Design good studies - how will your study take into account all of the identified variables and hypothesized causal relationships between variables

Does your data match up with your picture?

# READ TO LEARN MORE

**A Technical Primer On Causality**

(https://medium.com/@akelleh/a-technical-primer-on-causality-181db2575e41)

**A Brief Introduction to Graphical Models and Bayesian Networks**

(https://www.cs.ubc.ca/~murphyk/Bayes/bayes_tutorial.pdf)

**The Metaphysics of Causation** (Stanford Encyclopedia of Philosophy)

(https://plato.stanford.edu/entries/causation-metaphysics/)

**Counterfactual Theories of Causation** (Stanford Encyclopedia of Philosophy)

(https://plato.stanford.edu/entries/causation-counterfactual/)

**Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data**

(https://journals.sagepub.com/doi/10.1177/2515245917745629)

**Causal Networks, blocking, and d-separation**

(http://www.sachaepskamp.com/files/NA2014/d-separation2013.pdf)