Causal discovery from “big data” (?)

Tom Heskes
Radboud University Nijmegen

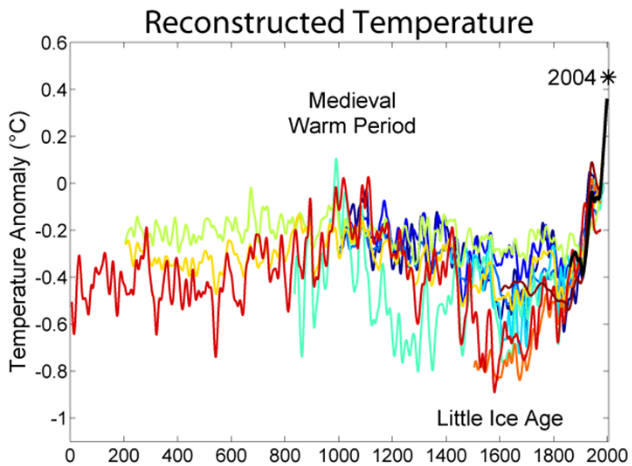
Outline

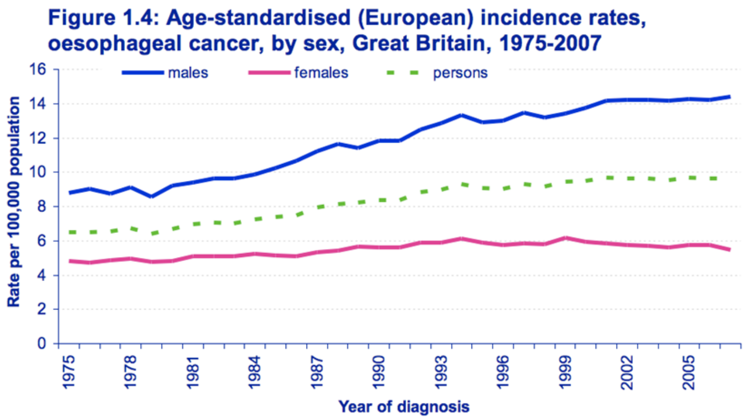
- Statistical causal discovery
- The logic of causal inference
- A Bayesian approach...
- Applications
- Current research and future goals

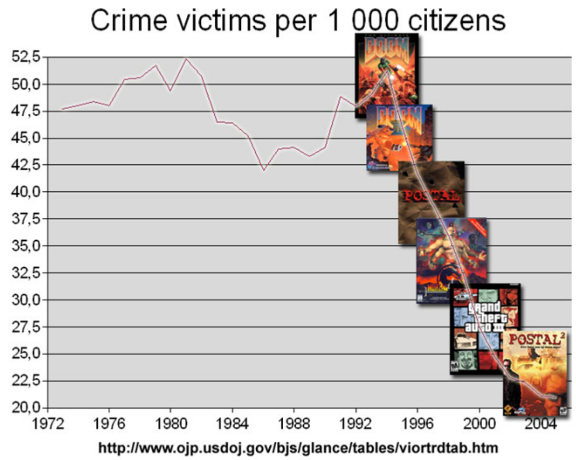
Outline

- Statistical causal discovery
  - Introduction
  - Finding causal relations
- The logic of causal inference
- A Bayesian approach...
- Applications
- Current research and future goals

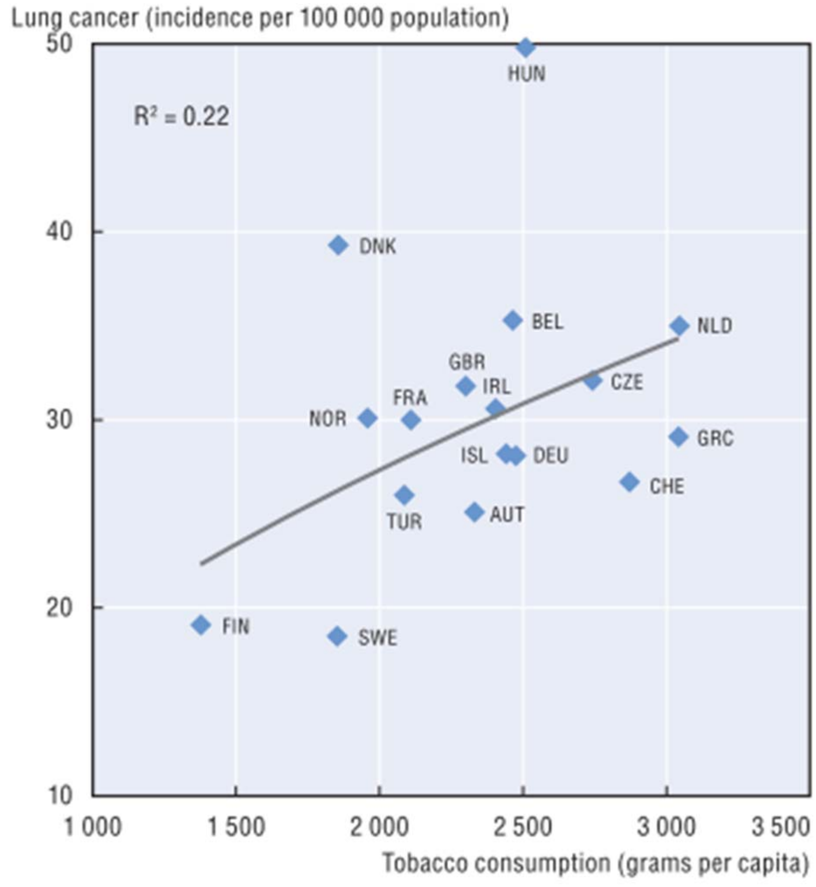
“We have discovered a link between...”







Causal discovery: smoking and lung cancer



Source: OECD Health Data 2010.

Results

- clear correlation
- strong risk factor for lung cancer

Chocolate consumption and Nobel prizes

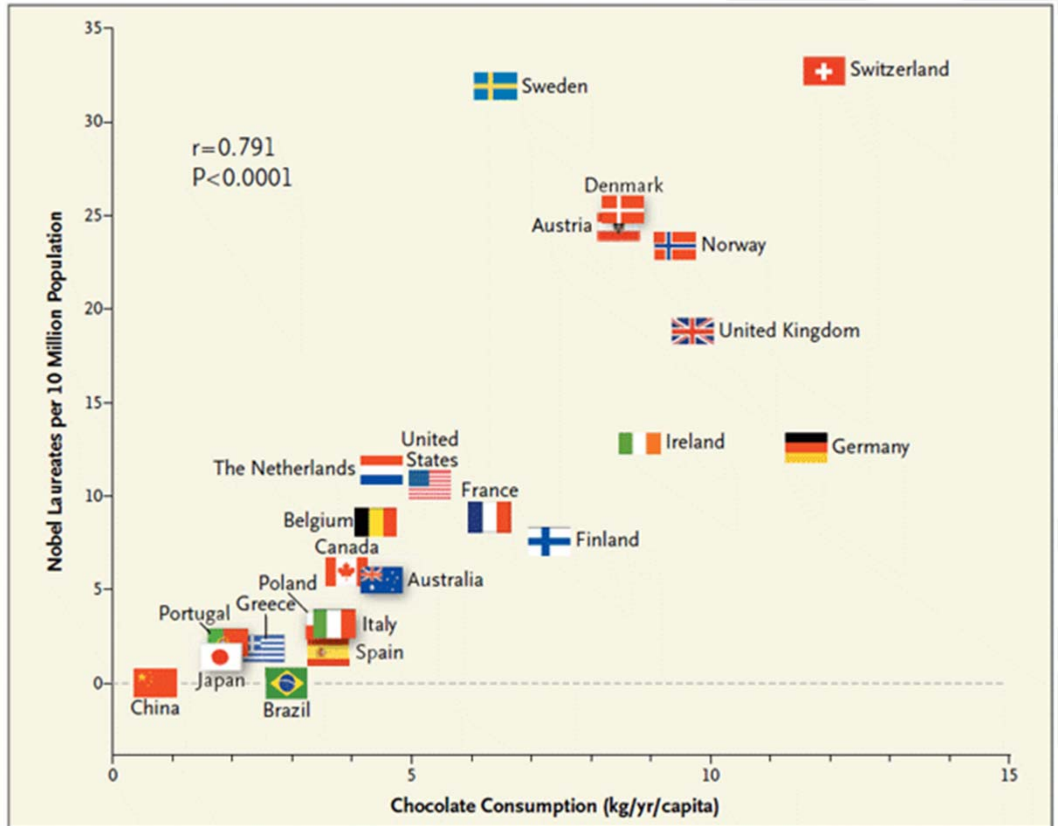


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

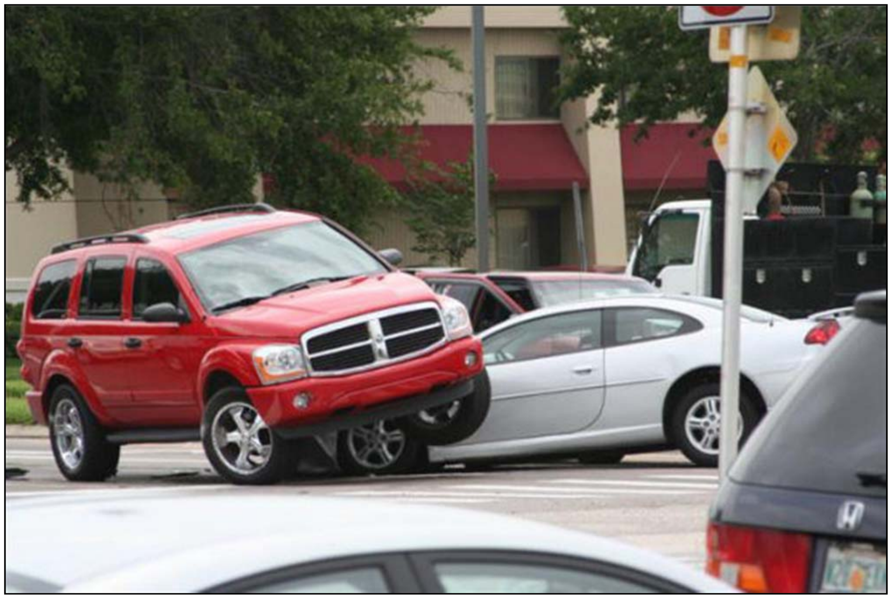
NEW ENGLAND JOURNAL OF MEDICINE

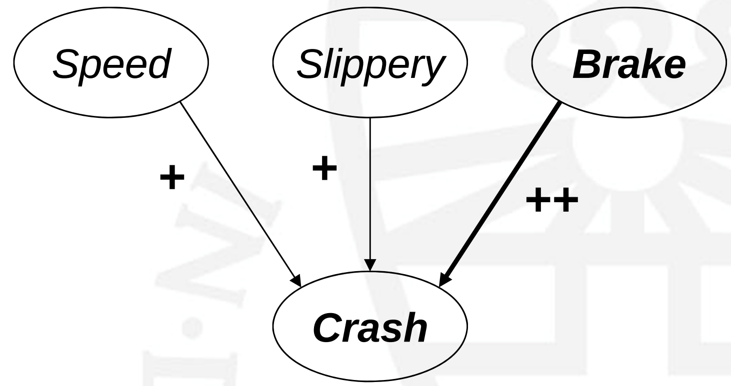
Results

- even stronger link!
- good predictor of chance on Nobel prize...

Messerli, "Chocolate Consumption, Cognitive Function, and Nobel Laureates", New England Journal of Medicine, 2012

Accident hot spots



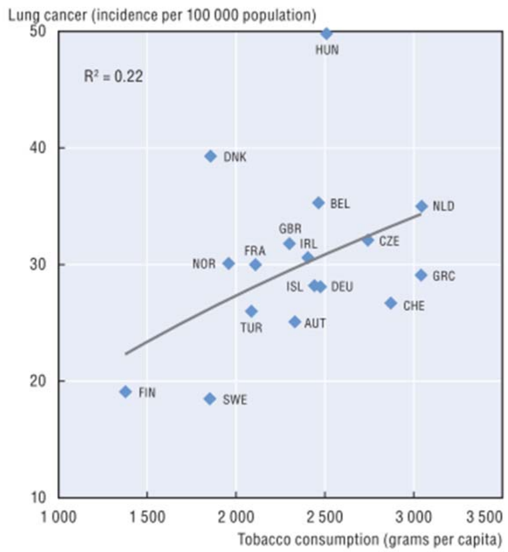


regression coefficients

Results

- strong positive correlation between *Braking heavily* and *Car Crash*?

From observation to action



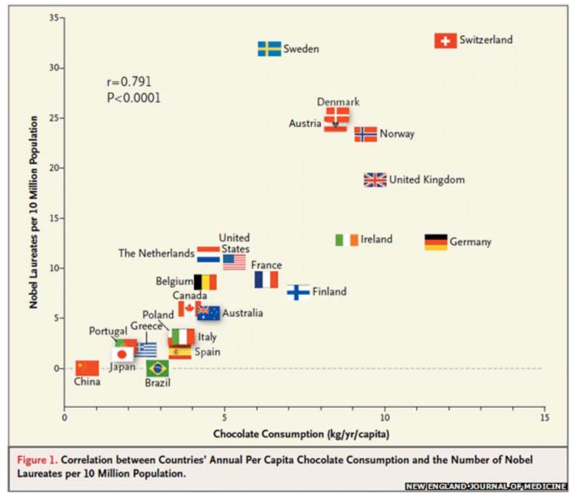
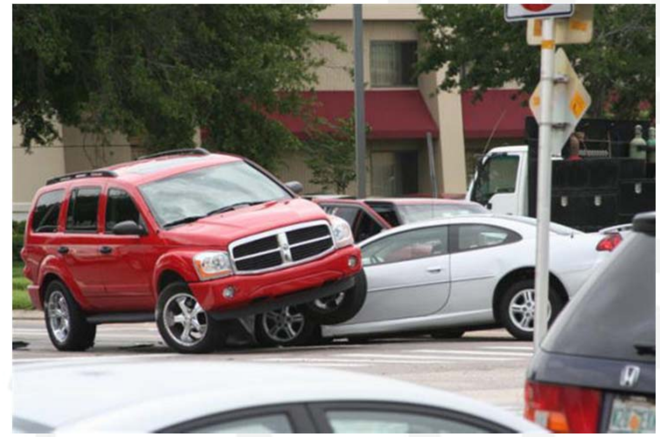


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

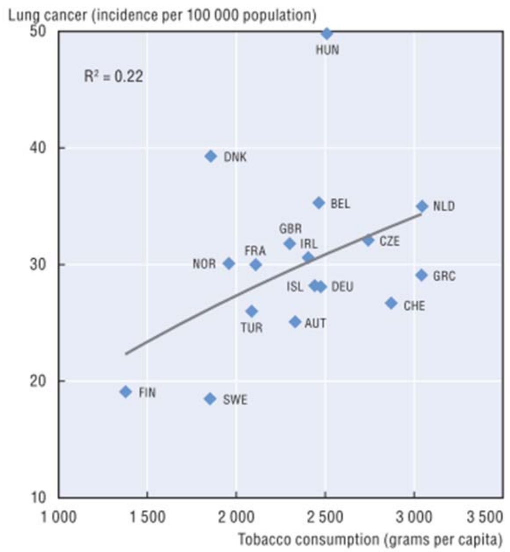
NEW ENGLAND JOURNAL OF MEDICINE

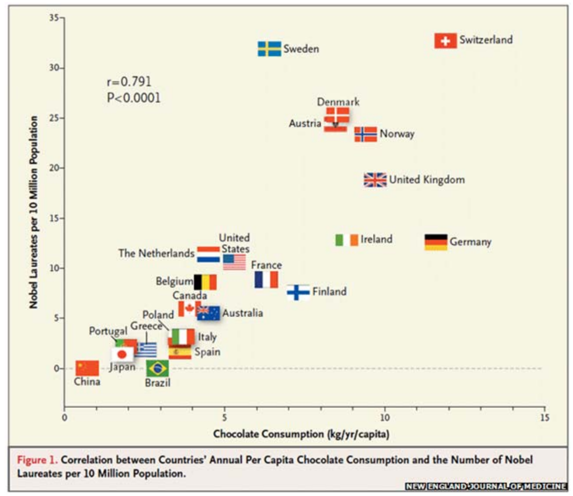


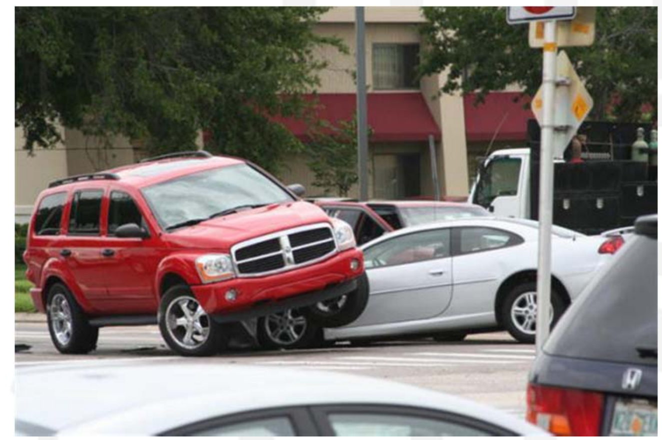
- correlations describe the world as we see it
- causal relations predict how the world will change when we intervene

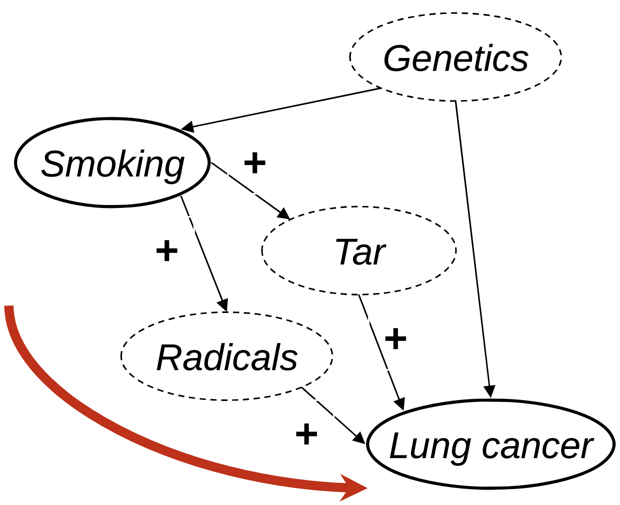
⇒ main goal of causal discovery

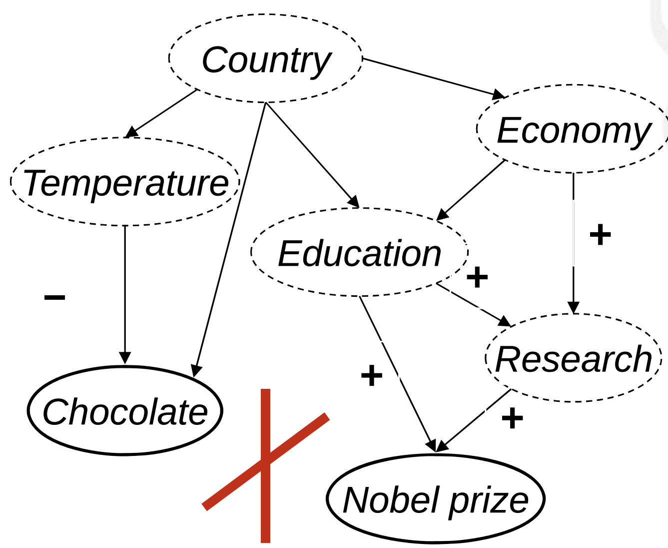
Challenge: recognize causal pathways from data

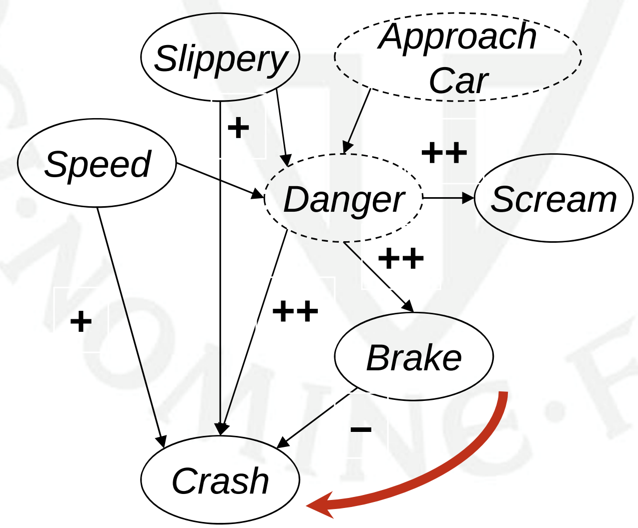








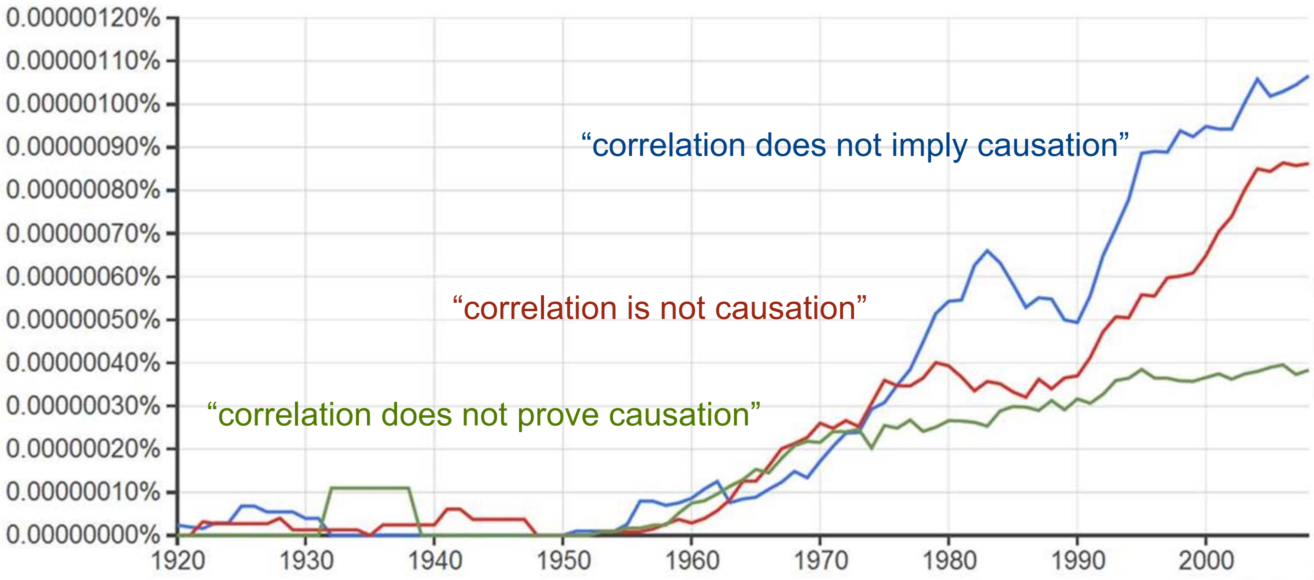


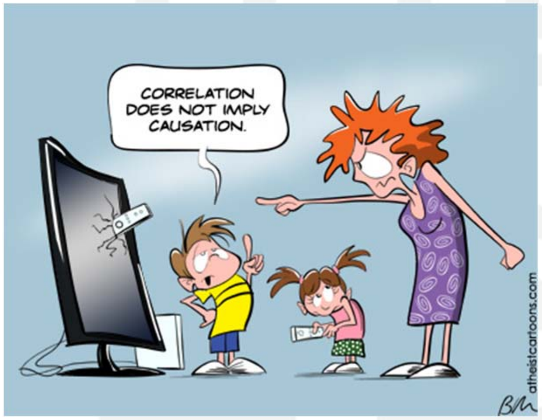


Outline

- Statistical causal discovery
  - Introduction
  - Finding causal relations
- The logic of causal inference
- A Bayesian approach...
- Applications
- Current research and future goals

A popular saying





Why do people love to say that correlation does not imply causation?

Daniel Engber: "The internet blowhard's favorite phrase"
http://www.slate.com/articles/health_and_science/science/2012/10/correlation_does_not_imply_causation_how_the_internet_fell_in_love_with_a_stats_class_clich_.html

Big data and causality

[...] society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing *why* but only *what*. This overturns centuries of established practices and challenges our most basic understanding of how to make decisions and comprehend reality.

Mayer-Schönberger & Cukier

Big data and causality

But faced with massive data, this approach to science - hypothesize, model, test - is becoming obsolete. [...] Petabytes allow us to say: ‘Correlation is enough.’ We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

Anderson (EiC Wired)

Logical fallacy
reasoning

correlation does not imply causation

thus

it is impossible to discover causal relationships from purely observational data

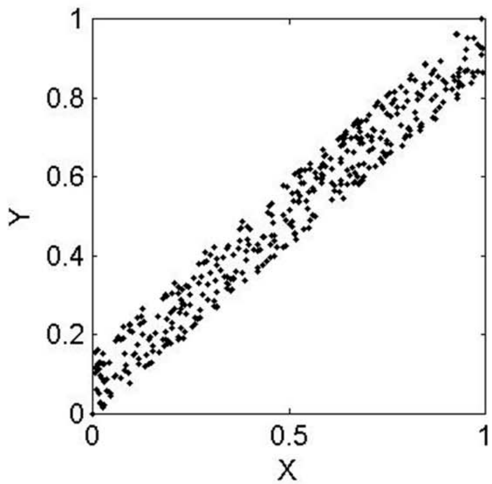
In fact

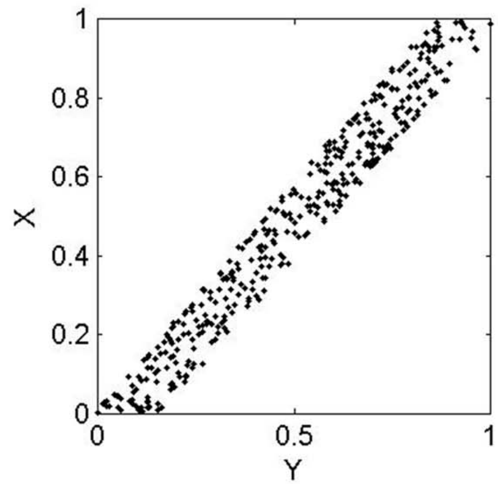
a single, simple correlation does not imply causation

yet

it is possible to discover causal relationships from purely observational data (which of course requires some assumptions, as any statistical approach)

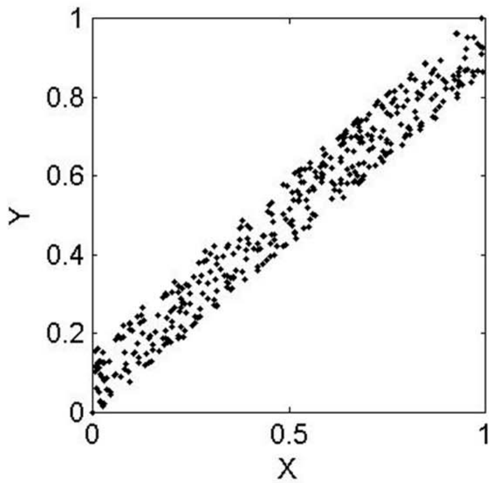
Causal direction





does X cause Y or does Y cause X?

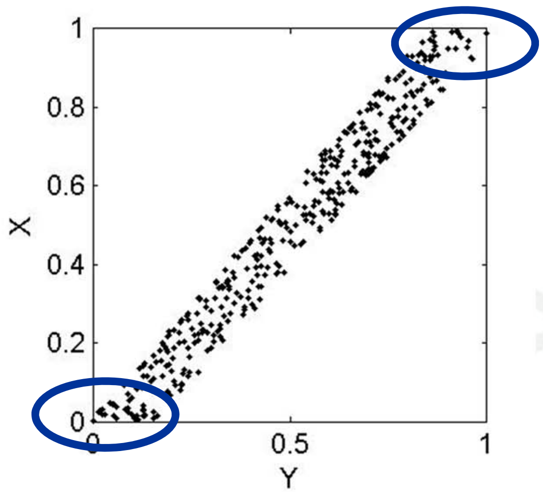
Causal direction



easy to explain as

Y = f(X) + noise

$X \rightarrow Y$



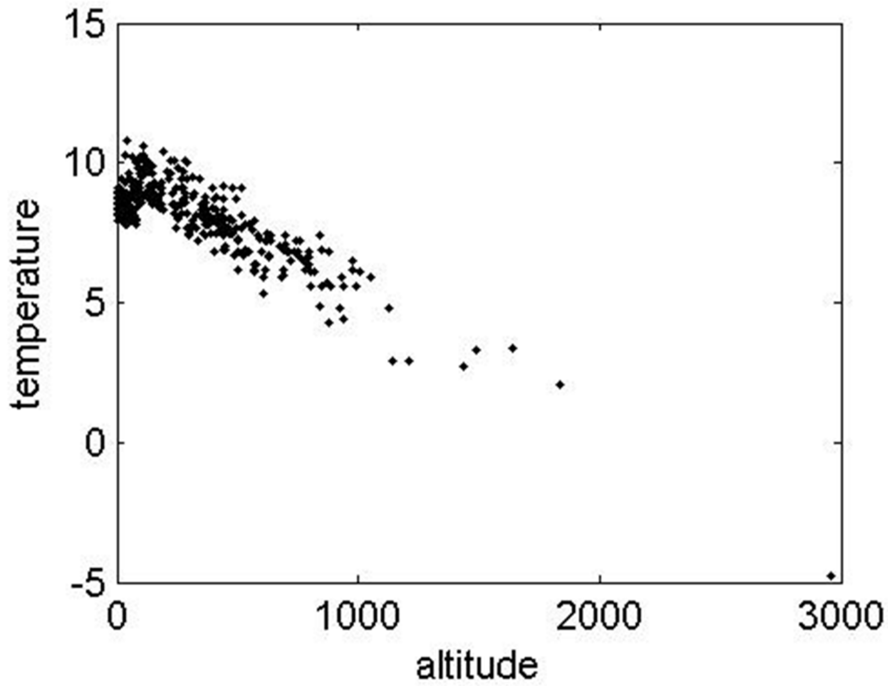
difficult to explain as

X = g(Y) + noise

$Y \not\rightarrow X$

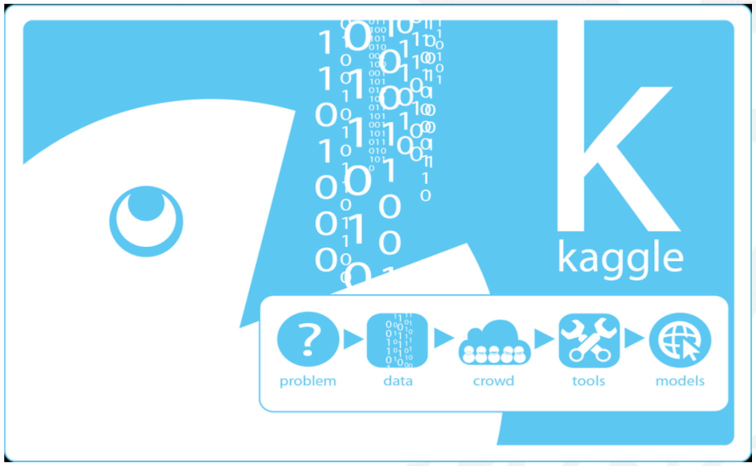
Ockham chooses a razor

Real-world cause-effect pairs



X: altitude of weather station

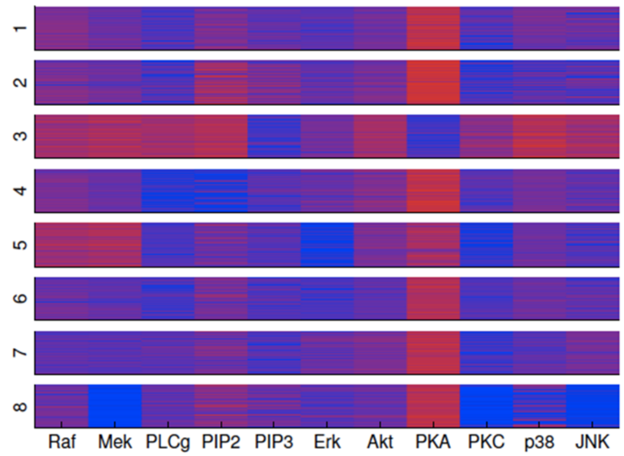
Y: temperature (average over 1961-1990)



http://webdav.tuebingen.mpg.de/cause-effect/
http://www.kaggle.com/c/cause-effect-pairs

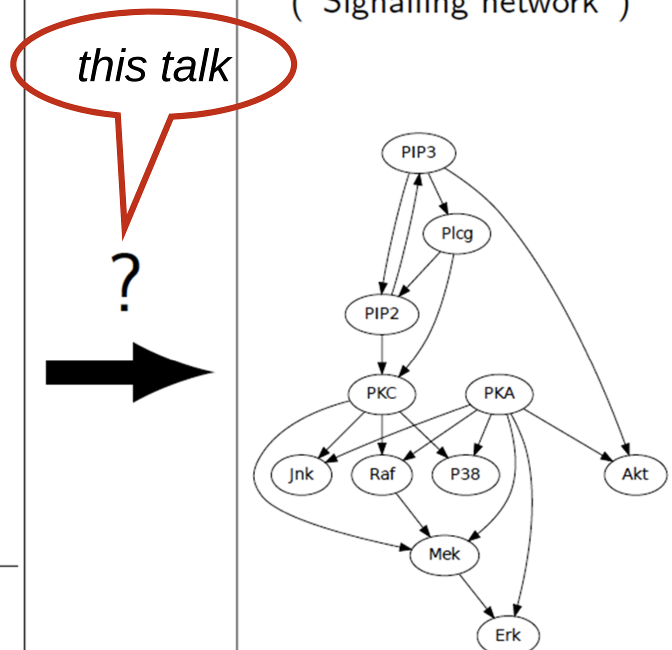
More variables: build causal model

Protein Abundance Data:
[Sachs et al., 2005]



Condition	Reagent	Intervention
1	-	observational
2	Akt-inhibitor	inhibits AKT activity
3	G0076	inhibits PKC activity
4	Psitectorigenin	inhibits PIP2 abundance
5	U0126	inhibits MEK activity
6	LY294002	inhibits PIP2/PIP3 activity
7	PMA	activates PKC activity
8	β2CAMP	activates PKA activity

Causal Mechanism:
("Signalling network")



(depicted here: "consensus" network)

Sachs et al., "Causal protein-signaling networks derived from multiparameter single-cell data", 2005

Outline

- Statistical causal discovery
- The logic of causal inference
  - Connection to structural equation models
  - Causal DAGs and constraint-based methods
  - Logical Causal Inference (LoCI)
- A Bayesian approach...
- Applications
- Current research and future goals

Structural Equation Models

Definition: SEM/SCM [Pearl, 2000; Wright, 1921]

- a set of d observed random variables $\{X_1,..,X_d\}$ and corresponding latent variables $\{E_1,..,E_d\}$,
- a set of d structural equations

'noise'

effect

$$X_i = f_i\left(\mathbf{X}_{pa(i)}, E_i\right)$$

causal mechanism

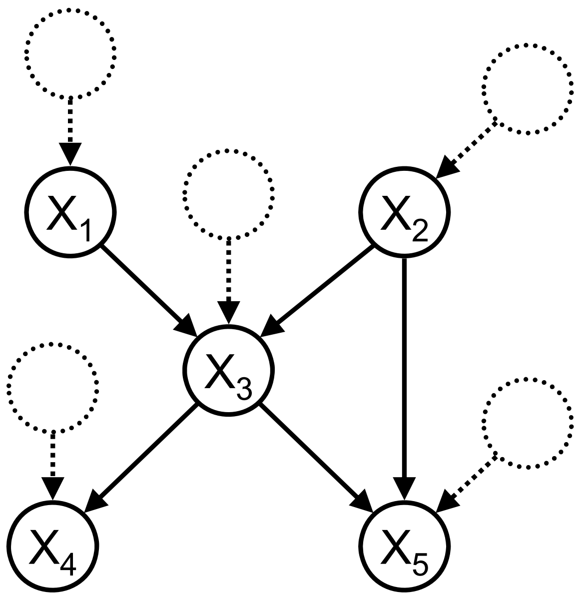
direct causes

with $pa(i)$ the observed direct causes ('parents') of X_i

- a joint probability distribution $p(E_1,..,E_d)$ on the latent variables
- inducing a joint probability distribution $p(X_1,..,X_d)$ on the observed variables

Graphical model equivalent

- variables become vertices
- direct causal mechanisms become arcs from cause to effect
- latent noise variables implicit
- *note*: SEM structure + observed probability distribution ≈ Bayesian network



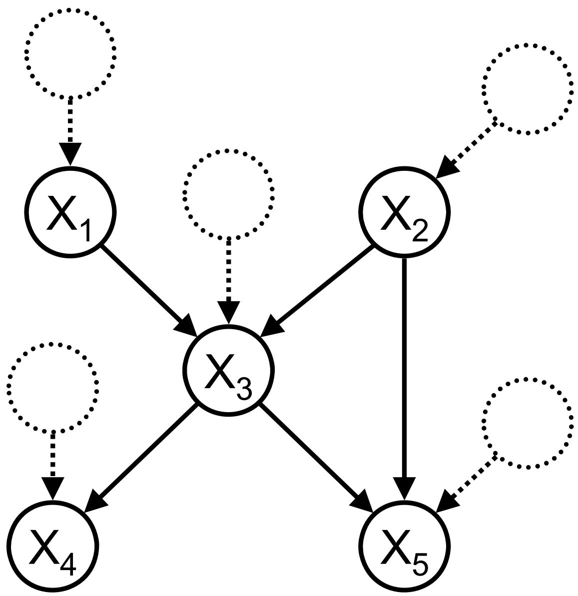
graphical representation

$$X_1 = f_1(E_1)$$
$$X_2 = f_2(E_2)$$
$$X_3 = f_3(X_1, X_2, E_3)$$
$$X_4 = f_4(X_3, E_4)$$
$$X_5 = f_5(X_2, X_3, E_5)$$

structural equation model

Interventions in a SEM

- (externally) force the value of variable X_i to a specific value / distribution
- denote: $do(X_i = \xi)$



graphical representation

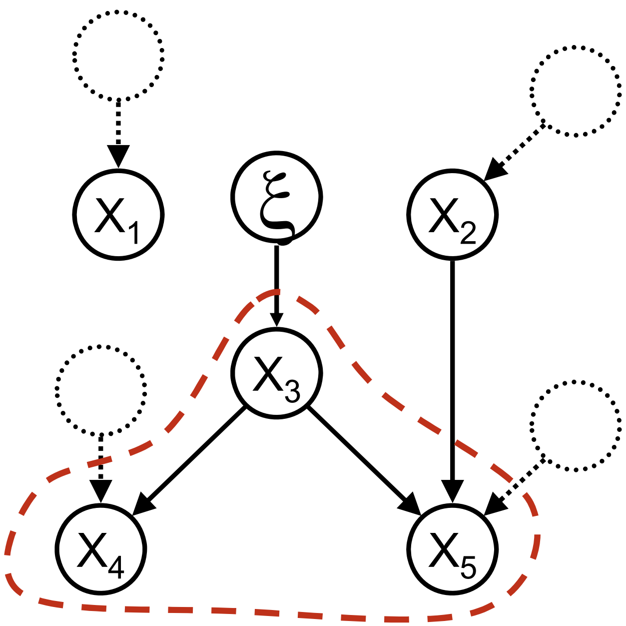
$$X_1 = f_1(E_1)$$
$$X_2 = f_2(E_2)$$
$$X_3 = f_3(X_1, X_2, E_3)$$
$$X_4 = f_4(X_3, E_4)$$
$$X_5 = f_5(X_2, X_3, E_5)$$

structural equation model

Interventions in a SEM

do(X_i = ξ)

- replaces corresponding causal mechanism
- graphical: removes incoming arcs
- only impacts on observed distribution of causal descendants



intervention on X_3

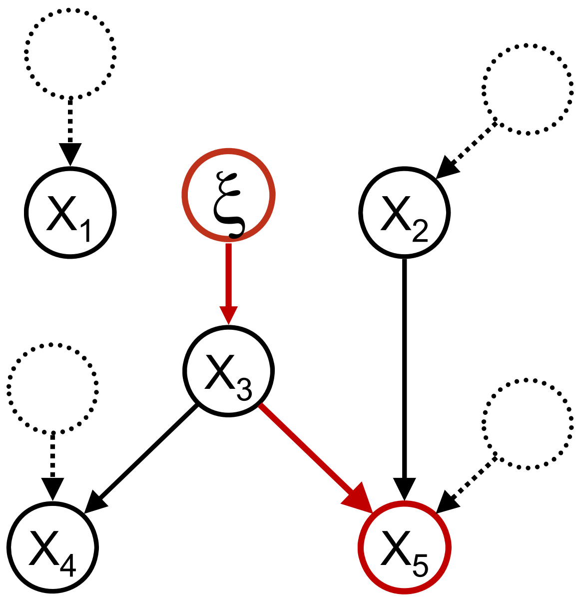
$$X_1 = f_1(E_1)$$
$$X_2 = f_2(E_2)$$
$$X_3 = \xi$$
$$X_4 = f_4(X_3, E_4)$$
$$X_5 = f_5(X_2, X_3, E_5)$$

override causal mechanism

Prediction in a SEM

- given a SEM structure with observed distribution $p(X_1,..,X_d)$
- intervention $do(X_i = \xi)$
- predict impact on distribution of other observed nodes: $p(X_j \mid do(X_i = \xi))$

- *note*: $p(X_j \mid do(X_i = \xi)) \neq p(X_j \mid X_i = \xi)$!

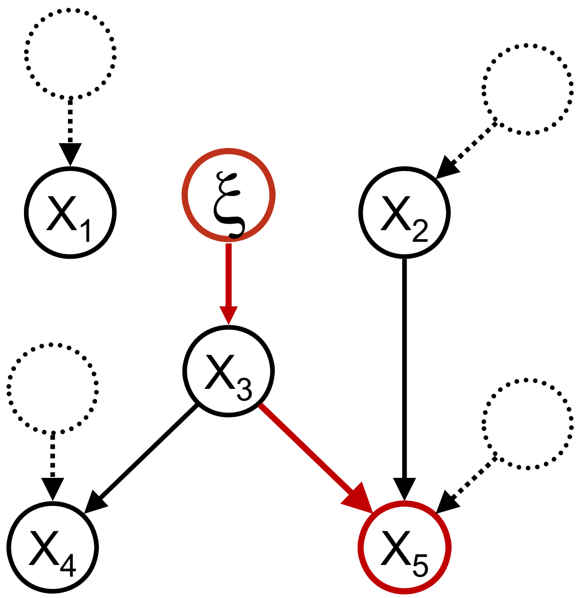


$$p\left(X_5 \middle| do\left(X_3 = \xi\right)\right) = ?$$

Prediction in a SEM

- given a SEM structure with observed distribution $p(X_1,..,X_d)$
- intervention $do(X_i = \xi)$
- predict impact on distribution of other observed nodes: $p(X_j \mid do(X_i = \xi))$

- do-calculus [Pearl, 2000]: formal method to express $p(X_j \mid do(X_i = \xi))$ in terms of $p(X_1,..,X_d)$

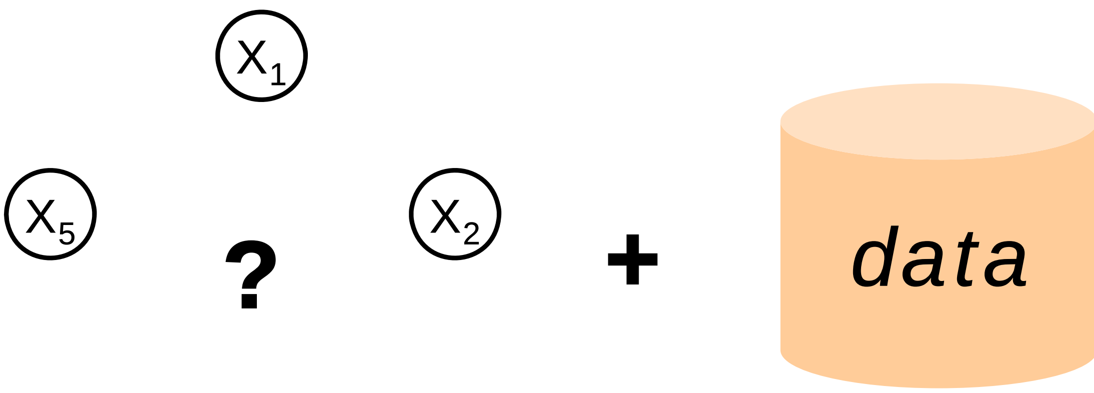


$$p\left(X_5 \middle| do\left(X_3 = \xi\right)\right) = ?$$

for example: predict the effect of treatment for an individual patient (assuming a known structure, often without confounders)

Prediction in practice

- given observed data from some distribution $p(X_1,..,X_d)$
- some reasonable assumptions,
- can we still predict $p(X_j \mid do(X_i = \xi))$?



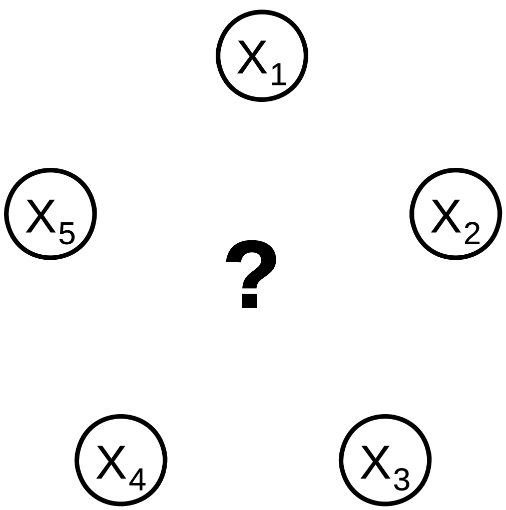
$$p\left(X_1, X_2, X_3, X_4, X_5\right)$$

$$= p\left(X_j \middle| do\left(X_i = \xi\right)\right)?$$

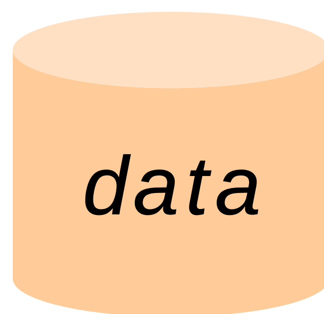
Prediction in practice

- given observed data from some distribution $p(X_1,..,X_d)$
- some reasonable assumptions,
- can we still predict $p(X_j \mid do(X_i = \xi))$?

- Yes! (sometimes): provided we can infer something about the structure…



+



$$p\left(X_1, X_2, X_3, X_4, X_5\right)$$

$$= p\left(X_j \middle| do\left(X_i = \xi\right)\right)?$$

Outline

- Statistical causal discovery
- The logic of causal inference
  - Connection to structural equation models
  - Causal DAGs and constraint-based methods
  - Logical Causal Inference (LoCI)
- A Bayesian approach...
- Applications
- Current research and future goals

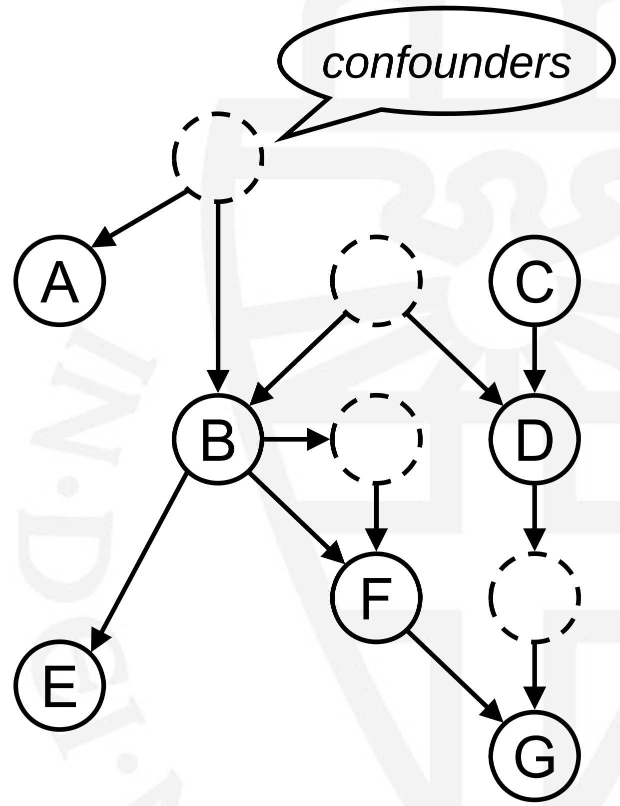
Some background theory and assumptions

Causal DAG assumption

- real-world consists of networks of causally interacting variables,
- subset of these variables observed in experiments

$$p(\mathbf{X}) = \prod_{k=1}^{K} p\left(X_k \middle| pa\left(X_k\right)\right)$$

parents of X_k in G



underlying causal DAG G
(Directed Acyclic Graph)

From causal structure to probabilities and back

Key insight:

- underlying causal structure is responsible for observed probability distribution
- identify characteristic features in the distribution to reconstruct the model

Main issues:

- what characteristics?
- how to handle latent confounders?

But also:

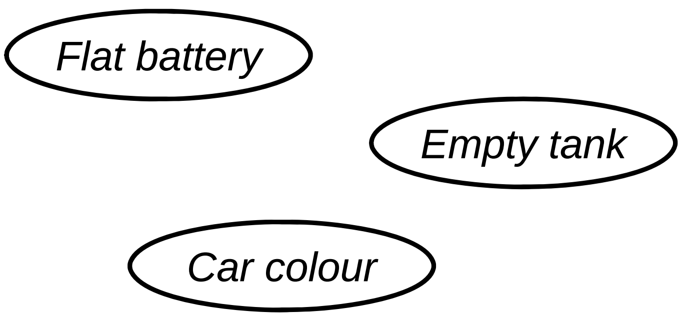
- dealing with uncertain (structural) conclusions
- complex interactions, mixed/missing data, background knowledge, etc.
- scalability to large models and/or large data sets
- ...

Some background theory and assumptions

Probabilistic independence constraints

“*X* is *independent* of *Y*”

- $X \perp\!\!\!\perp Y \quad : \quad p(X|Y) = p(X)$



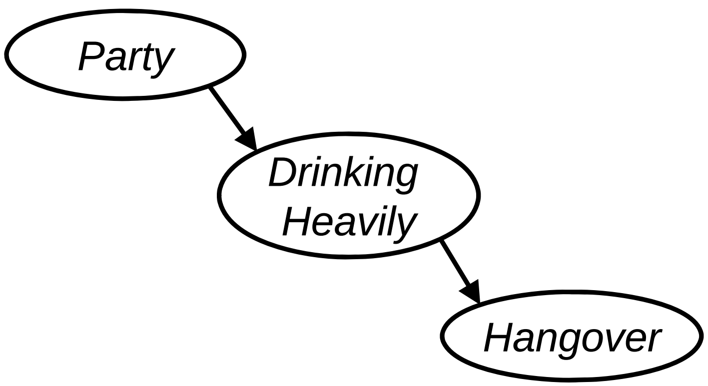
Independence

Some background theory and assumptions

Probabilistic independence constraints

- $X \perp\!\!\!\perp Y \quad : \quad p(X|Y) = p(X)$
- $X \perp\!\!\!\perp Y|Z \quad : \quad p(X|Y,Z) = p(X|Z)$

"*X* is *conditionally* independent of *Y* given *Z*"



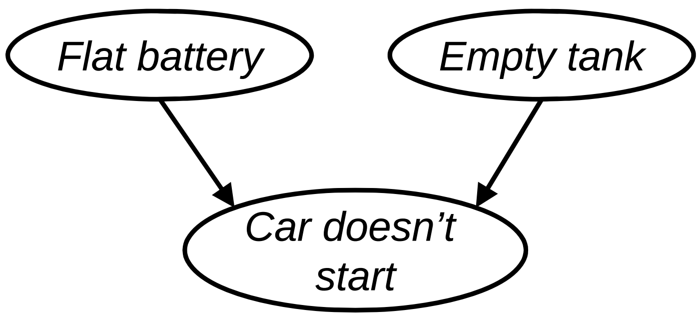
Conditional independence

Some background theory and assumptions

Probabilistic independence constraints

- $X \perp\!\!\!\perp Y \quad : \quad p(X|Y) = p(X)$
- $X \perp\!\!\!\perp Y|Z \quad : \quad p(X|Y,Z) = p(X|Z)$
- $X \not\perp\!\!\!\perp Y|Z \quad : \quad p(X|Y,Z) \neq p(X|Z)$

"*X* is (conditionally) *dependent* of *Y* given *Z*"



Conditional dependence

From causal graph to (in)dependencies and back

- Given a causal graph, we can read off all conditional (in)dependencies
- For causal inference we need to invert this and reason in the opposite direction:

Given an observed set of conditional (in)dependencies, e.g., derived from a set of data, what can we say about the underlying causal graph?

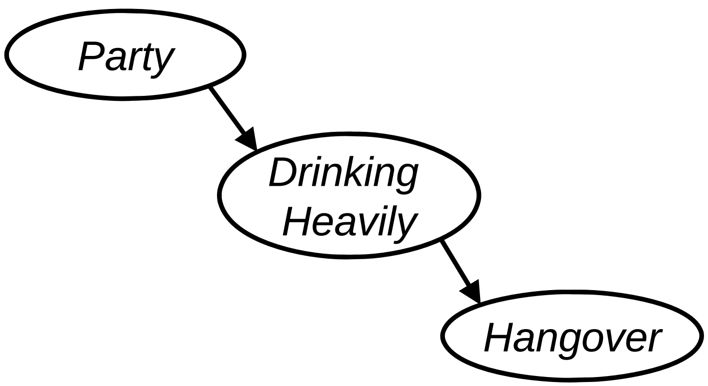
Key connection: two rules

1. $X \perp\!\!\!\perp Y \,|\, [Z] \quad : \quad (Z \Rightarrow X) \vee (Z \Rightarrow Y)$

square brackets denote 'minimal'

"is a cause of"

"if variable Z makes variables X and Y independent, then Z must have a causal relation to X and/or Y"



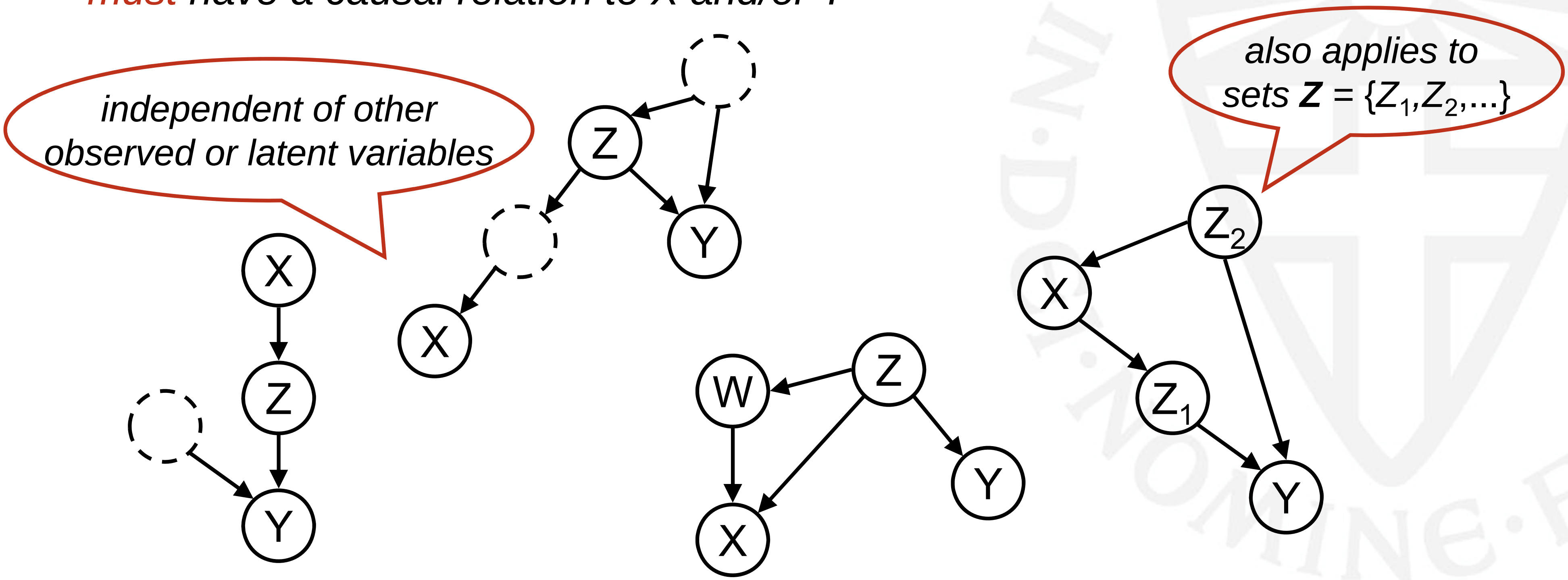
Minimal conditional independence

Reasoning:
every possible DAG in which variables *X* and *Y* are dependent when we do not condition on *Z*, yet become independent when we do condition on *Z*, has a (possibly indirect) directed path from *X* to *Z* and/or from *Y* to *Z*

Key connection: two rules

1. $X \perp\!\!\!\perp Y \,|\, [Z] \quad : \quad (Z \Rightarrow X) \vee (Z \Rightarrow Y)$

"if variable Z makes variables X and Y independent, then Z must have a causal relation to X and/or Y"

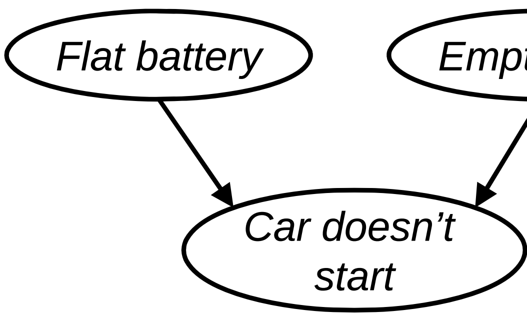


Key connection: two rules

1. $X \perp\!\!\!\perp Y \,|\, [Z] \quad : \quad (Z \Rightarrow X) \vee (Z \Rightarrow Y)$
2. $X \not\perp\!\!\!\perp Y \,|\, [Z] \quad : \quad (Z \not\Rightarrow X) \wedge (Z \not\Rightarrow Y)$

"is NOT a cause of"

"if variable Z makes variables X and Y dependent, then Z cannot have a causal relation to X and/or Y"



Reasoning:
a *DAG* in which variables *X* and *Y* are independent when we do not condition on *Z*, yet become dependent when we do condition on *Z*, cannot have a directed path from *Z* to *X*, nor from *Z* to *Y*

Minimal conditional dependence ('v-structure')

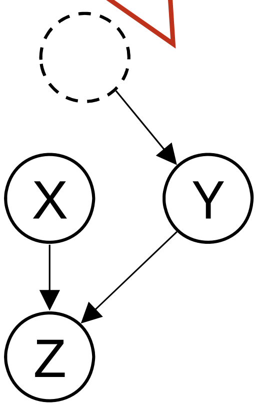
Key connection: two rules

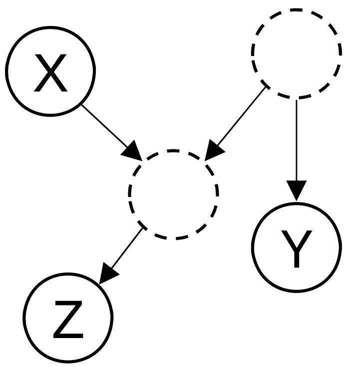
1. $X \perp\!\!\!\perp Y \mid [Z] \quad : \quad (Z \Rightarrow X) \vee (Z \Rightarrow Y)$
2. $X \not\perp\!\!\!\perp Y \mid [Z] \quad : \quad (Z \not\Rightarrow X) \wedge (Z \not\Rightarrow Y)$

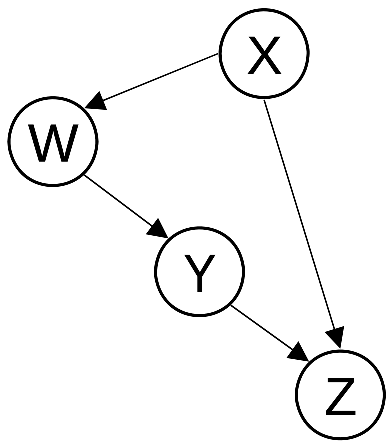
"if variable Z makes variables X and Y dependent, then Z cannot have a causal relation to X and/or Y"

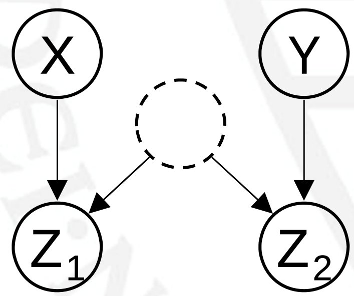
also applies to sets $\boldsymbol{Z} = \{Z_1, Z_2, \ldots\}$

independent of other observed or latent variables









Outline

- Statistical causal discovery
- The logic of causal inference
  - Connection to structural equation models
  - Causal DAGs and constraint-based methods
  - Logical Causal Inference (LoCI)
- A Bayesian approach...
- Applications
- Current research and future goals

Logical Causal Inference (LoCI)

1. $X \perp\!\!\!\perp Y \mid [Z] \quad : \quad (Z \Rightarrow X) \vee (Z \Rightarrow Y)$
2. $X \not\!\perp\!\!\!\perp Y \mid [Z] \quad : \quad (Z \not\Rightarrow X) \wedge (Z \not\Rightarrow Y)$
3. [something slightly more complicated, needed for completeness]

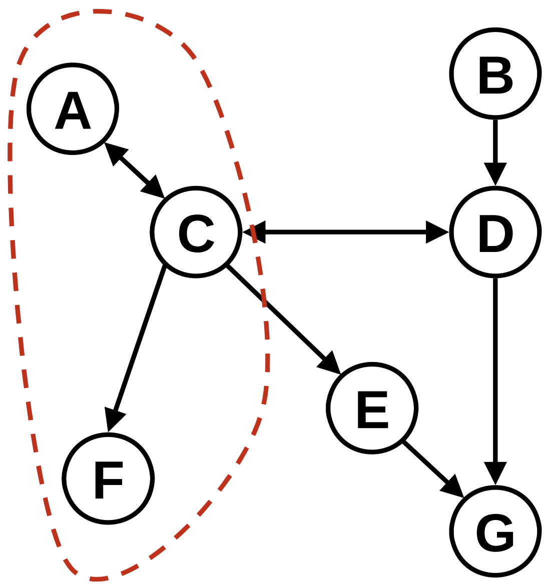
\+ subsequent logical deduction on standard causal properties

- transitivity $\quad (X \Rightarrow Y) \wedge (Y \Rightarrow Z) \quad : \quad (X \Rightarrow Z)$
- acyclicity $\quad (X \Rightarrow Y) \quad : \quad (Y \not\Rightarrow X)$

Theorem: "LoCI rules are sound and complete for causal discovery in the presence of latent confounders and selection bias." [Claassen & Heskes, 2011]

Example – infer causal relation

- introduce efficient search strategy over subsets

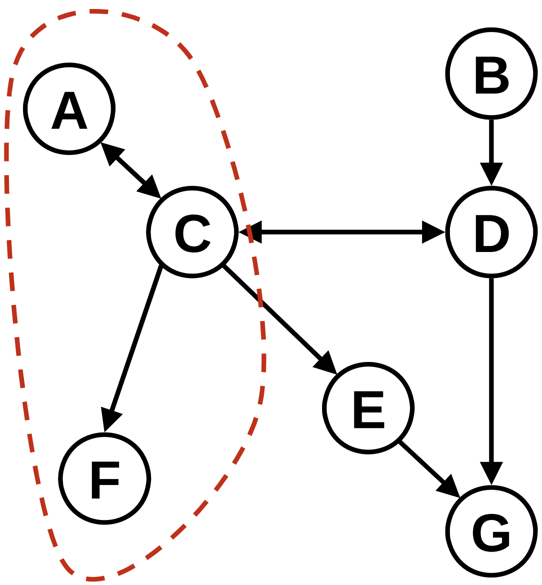


underlying causal structure G

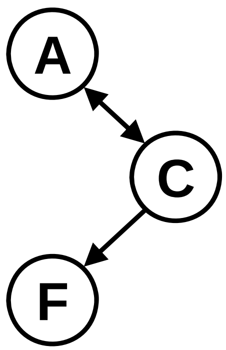
Example – infer causal relation

- introduce efficient search strategy over subsets
- identify minimal in/dependencies in subset

pick subset



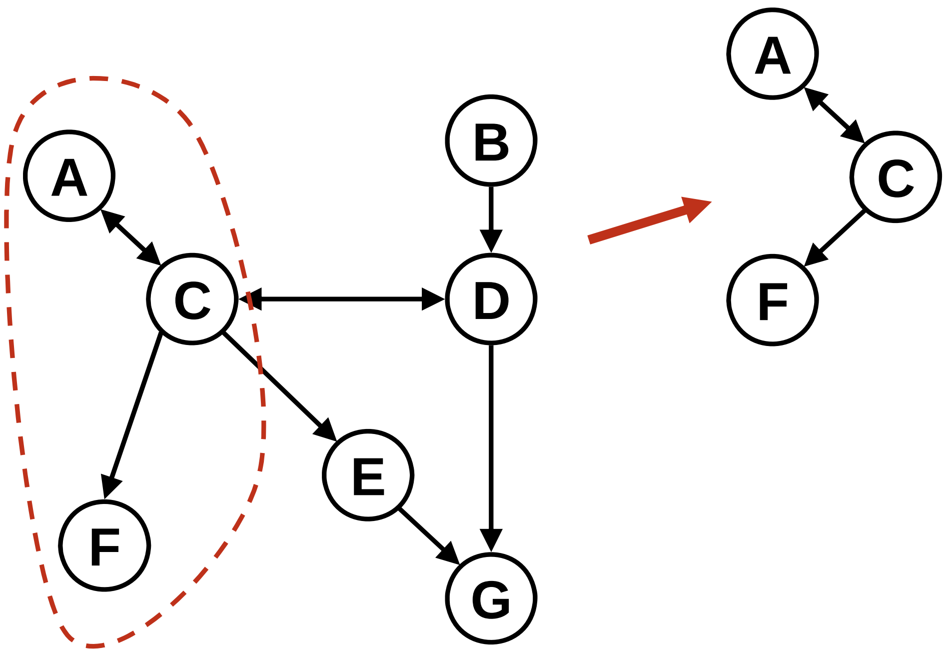
underlying causal structure G



$$A \perp\!\!\!\perp F \,|\, [C] \;\; : \;\; (C \Rightarrow A) \vee (C \Rightarrow F)$$

Example – infer causal relation

- introduce efficient search strategy over subsets
- identify minimal in/dependencies in subset
- collect implied causal information in list



underlying causal structure G

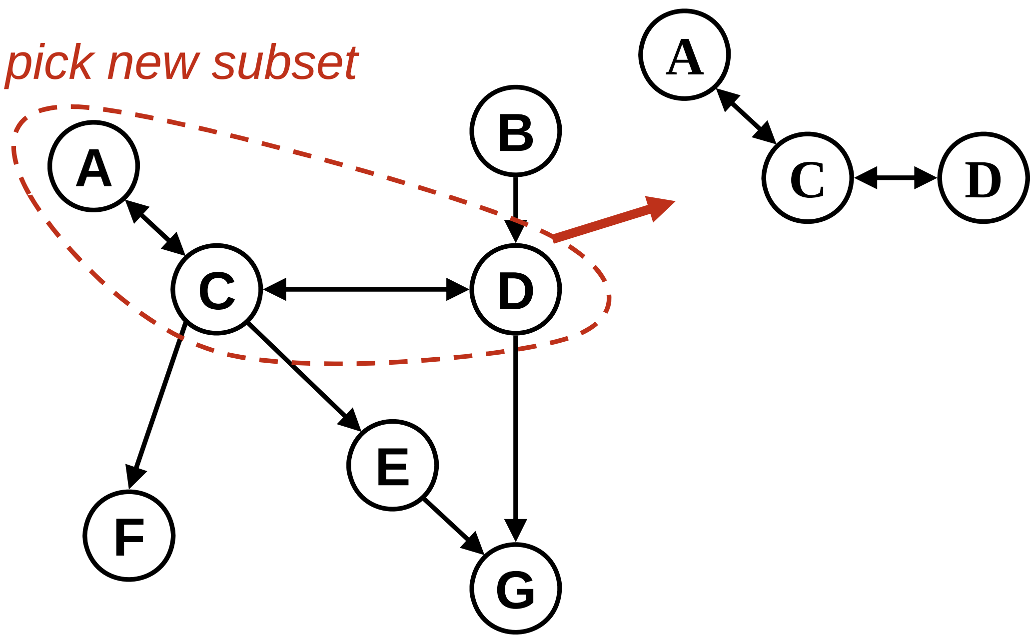
$$A \perp\!\!\!\perp F \mid [C] \;:\; (C \Rightarrow A) \vee (C \Rightarrow F)$$

$$\mathcal{L}(\mathcal{G}) = \{(C \Rightarrow A) \vee (C \Rightarrow F)\}$$

collect in list

Example – infer causal relation

- introduce efficient search strategy over subsets
- identify minimal in/dependencies in subset
- collect implied causal information in list
- repeat...



underlying causal structure G

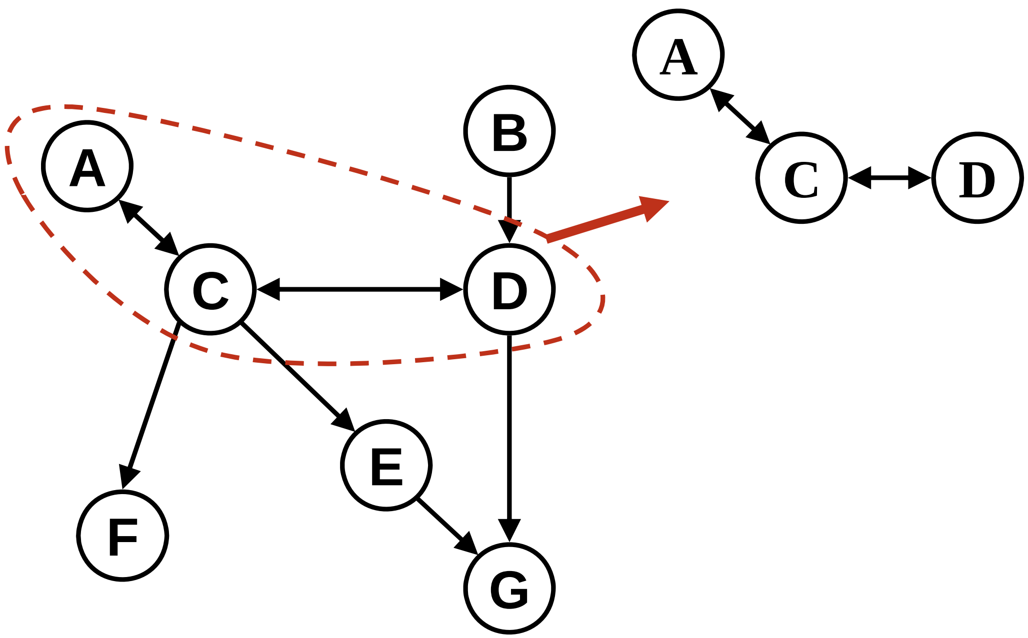
$$A \not\perp\!\!\!\perp D \mid [C] \;:\; (C \not\Rightarrow A) \wedge (C \not\Rightarrow D)$$

$$\mathcal{L}(\mathcal{G}) = \left\{ \begin{array}{l} (C \Rightarrow A) \vee (C \Rightarrow F) \\ (C \not\Rightarrow A) \\ (C \not\Rightarrow D) \end{array} \right\}$$

add to list

Example – infer causal relation

- introduce efficient search strategy over subsets
- identify minimal in/dependencies in subset
- collect implied causal information in list
- find new causal information through logical deduction



underlying causal structure G

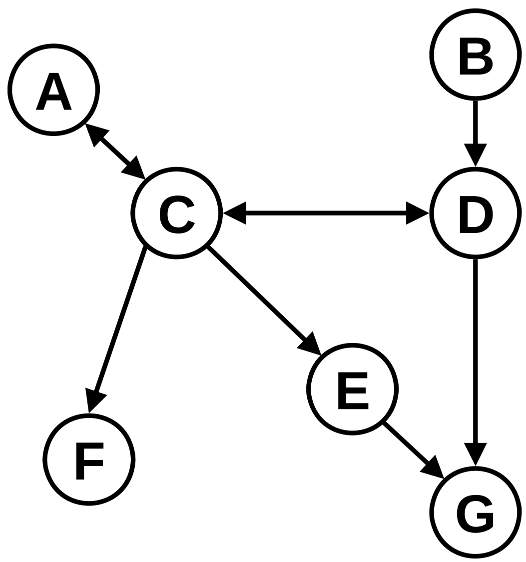
$$A \not\perp D \mid [C] \;:\; (C \not\Rightarrow A) \wedge (C \not\Rightarrow D)$$

$$\mathcal{L}(\mathcal{G}) = \begin{Bmatrix} (C \Rightarrow A) \vee (C \Rightarrow F) \\ (C \not\Rightarrow A) \\ (C \not\Rightarrow D) \\ (C \Rightarrow F) \end{Bmatrix}$$

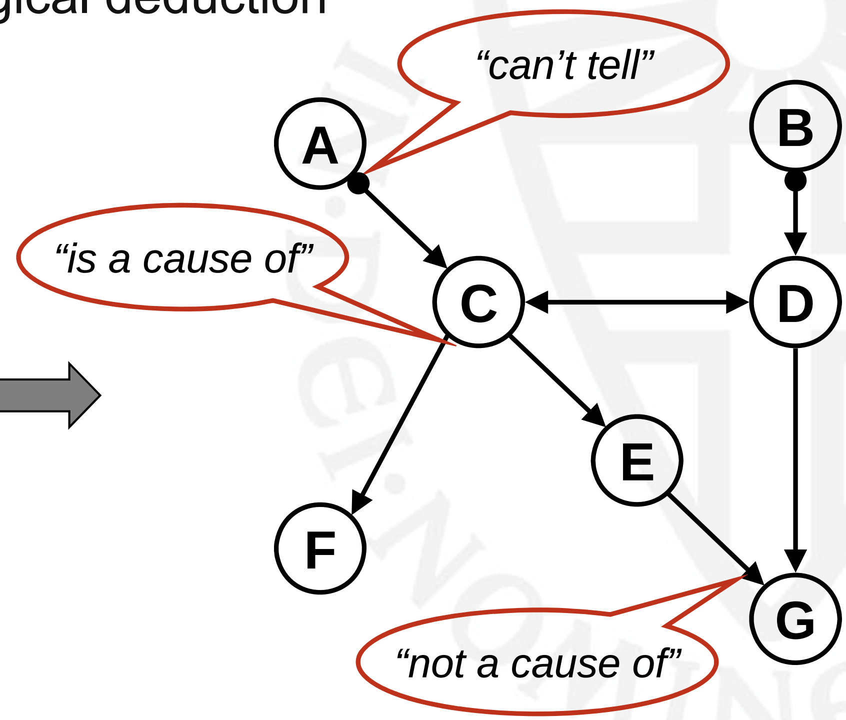
causal relation!

Example – infer causal relation

- introduce efficient search strategy over subsets
- identify minimal in/dependencies in subset
- collect implied causal information in list
- find new causal information through logical deduction
- finally: output causal model



underlying causal structure G



inferred causal model P

Outline

- Statistical causal discovery
- The logic of causal inference
  - Connection to structural equation models
  - Causal DAGs and constraint-based methods
  - Logical Causal Inference (LoCI)
- A Bayesian approach...
- Applications
- Current research and future goals

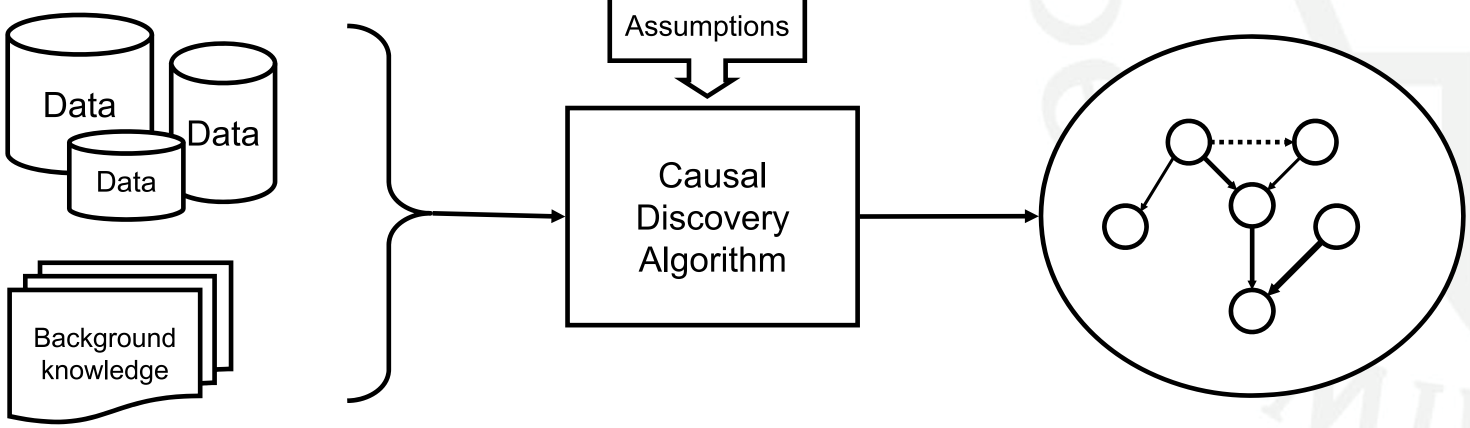
Improving reliability

- categorical decisions based on finite data are not robust
- mistakes propagate through the model
- impact of insecure decisions not visible in output

- Idea: distinguish between reliable and 'marginal' conclusions

- Goal:

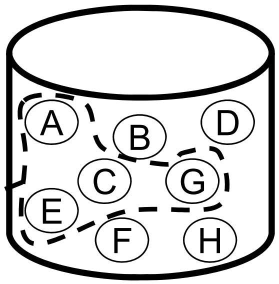
$\boxed{p(\text{“}X \Rightarrow Y\text{”} | \mathbf{D}, \mathrm{I})}$



Bayesian Constraint-based Causal Discovery

Claassen & Heskes,
best paper award UAI 2012

$\{A, C, E, G\}$



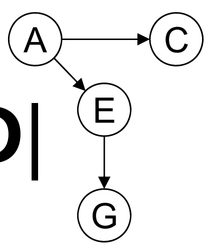
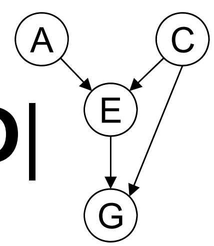
repeat until done

$$p(L|\mathbf{D}) \propto \sum_{\mathcal{G} \to L} p(\mathbf{D}|\mathcal{G})\, p(\mathcal{G})$$

1: select (new) subset of variables from **D**

3: translate into logical causal statements

4: collect in global list

$p(\mathbf{D}|$  $, I), p(\mathbf{D}|$  $, I), \ldots$

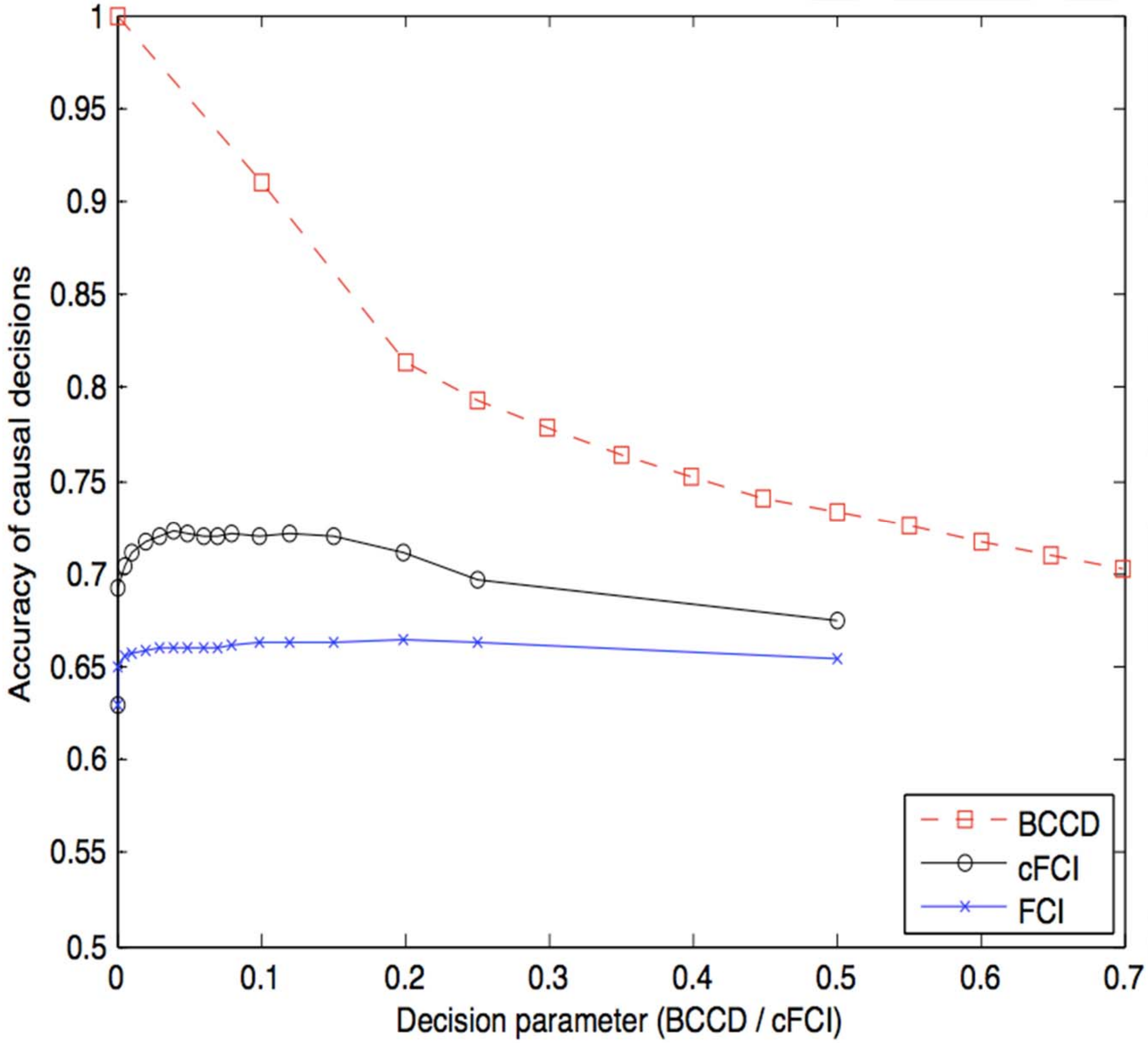
$$\mathcal{L}: \begin{cases} p(C \Rightarrow A \vee G) = 0.82 \\ p(B \Rightarrow F) = 0.78 \\ p(C \not\Rightarrow A) = 0.67 \\ \ldots \end{cases}$$

2: compute Bayesian likelihoods for **all** marginal structures G over selected subset

5: rank and process into causal model

Probability of a causal relation

- BCCD accuracy can be 'tuned' by changing the threshold
- competitors such as (conservative) FCI shift the balance between (in)dependence decisions, but cannot tune accuracy of causal statements
- good (slightly conservative) estimate of $p(\text{“}X \Rightarrow Y\text{”} \mid \mathbf{D})$

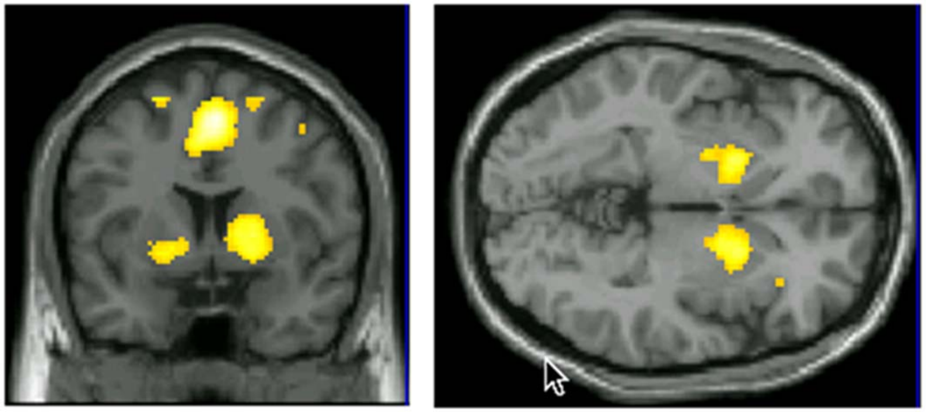


Outline

- Statistical causal discovery
- The logic of causal inference
  - Connection to structural equation models
  - Causal DAGs and constraint-based methods
  - Logical Causal Inference (LoCI)
- A Bayesian approach...
- Applications
- Current research and future goals

Heritability factors in adult ADHD

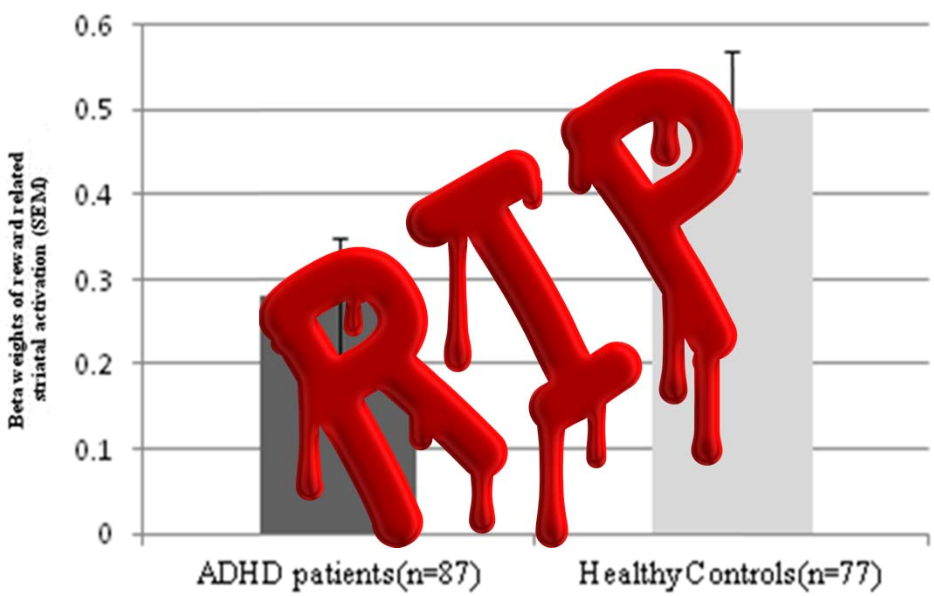
- ADHD - Attention Deficit Hyperactivity Disorder
- Two types of symptoms:
  - Hyperactivity / Impulsivity
  - Inattention / concentration problems
- Highly heritable
- DAT1 gene related to brain reward / motivation functioning, and associated with ADHD in adulthood(1)

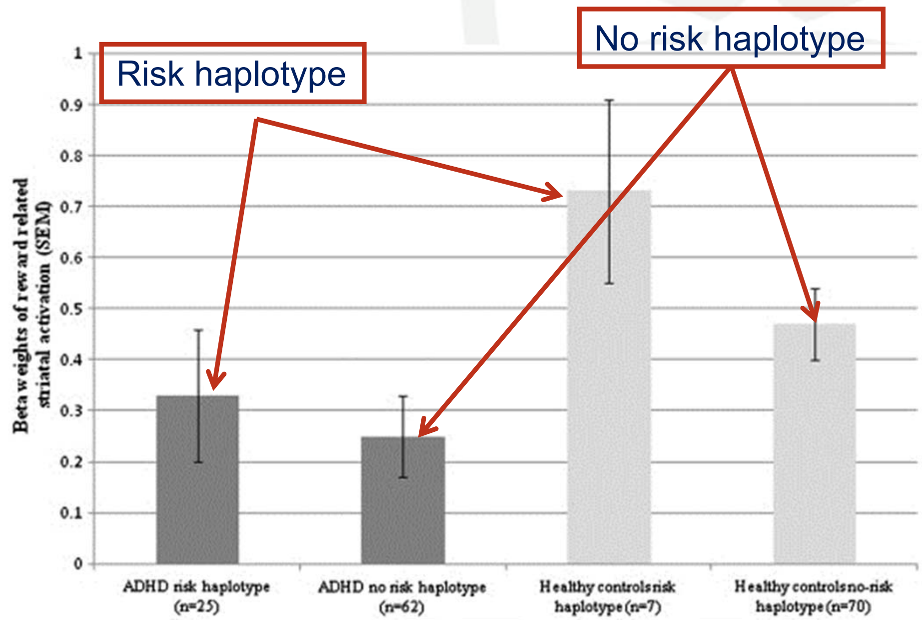


M. Hoogman et al., "The dopamine transporter haplotype and reward-related striatal responses in adult ADHD", European Neuropsychopharmacology (2012)

Previous fMRI results

- Risk haplotype is strong risk factor for ADHD
- Significant link between reward related brain activation and ADHD
- Weak dependency between haplotype and activation?





- Relevant? How to interpret?? Need to understand the causal interactions

BCCD on IMpACT data

- Sample size =164 (patients = 87, controls=77)
- probabilities on presence/absence of cause-effect relations, both direct and indirect
- includes background knowledge that nothing can causes *risk haplotype* and diagnosis *patient/control* cannot cause *hyperactivity* and *inattention*

	Activation	Smoking	Hyperactivity	Inattention	Patient/Control	Medication	Risk haplotype
Activation			50%	50%	50%		100%
Smoking			66%	66%	66%		100%
Hyperactivity							100%
Inattention	50%	69%	86%		94%	92%	100%
Patient/Control	50%	66%	100%	100%		89%	100%
Medication			89%	89%	89%		100%
Risk haplotype							

A causes B:

75%-100%
50%-75%
0%-50%

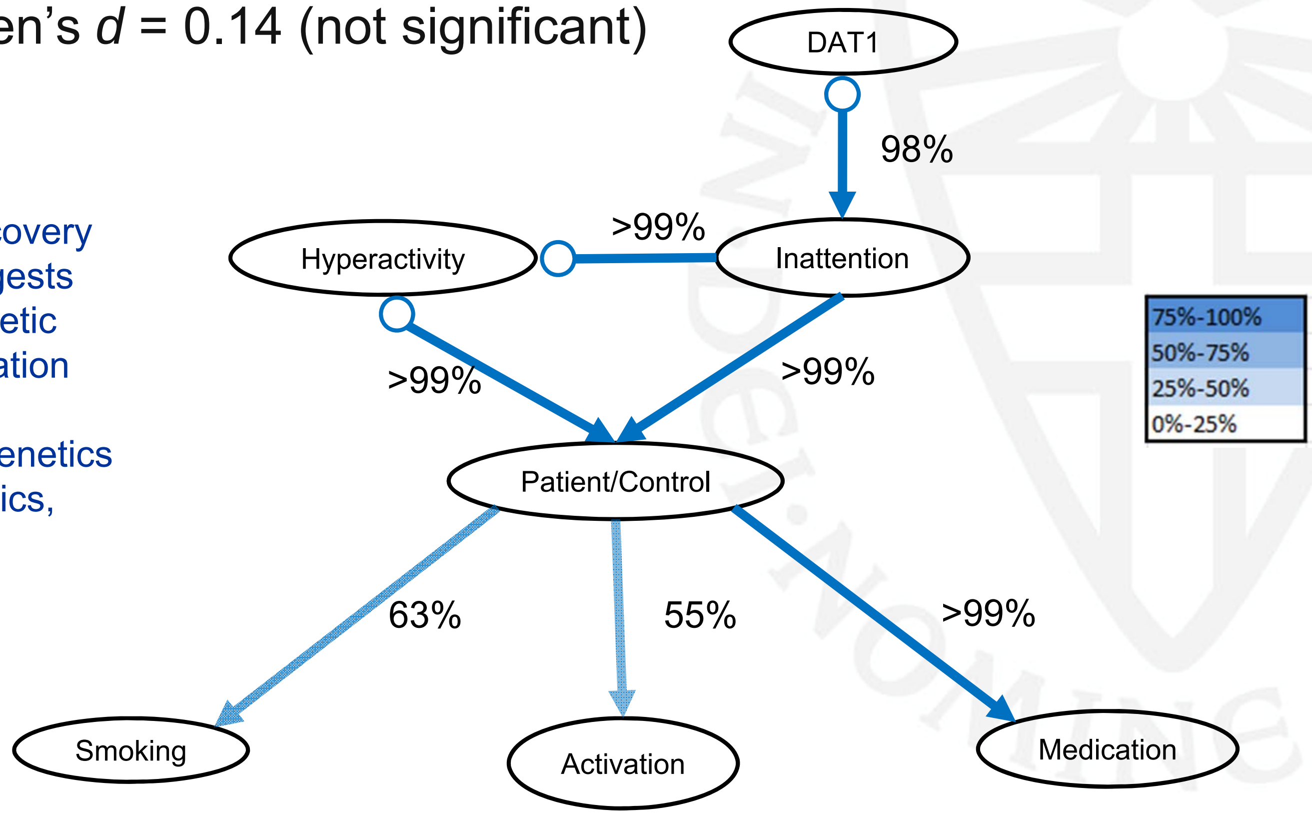
A does not cause B:

75%-100%
50%-75%
0%-50%

BCCD on IMpACT study

- global model for ADHD
- *risk haplotype* does appear to affect *(striatal response) activation*, but only via *inattention*
- total effect size: Cohen's *d* = 0.14 (not significant)

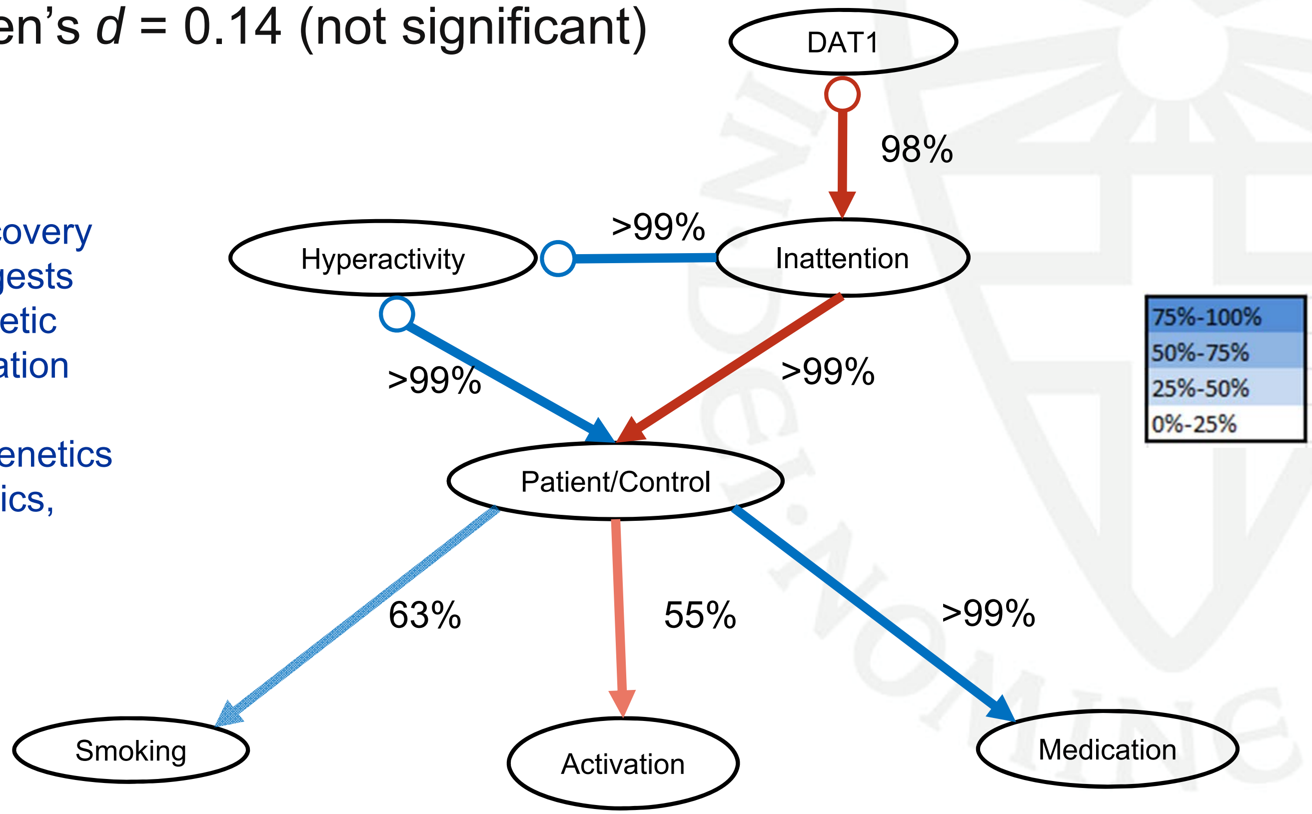
E. Sokolova et al., "Causal discovery in an adult ADHD data set suggests indirect link between DAT1 genetic variants and striatal brain activation during reward processing", American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2015



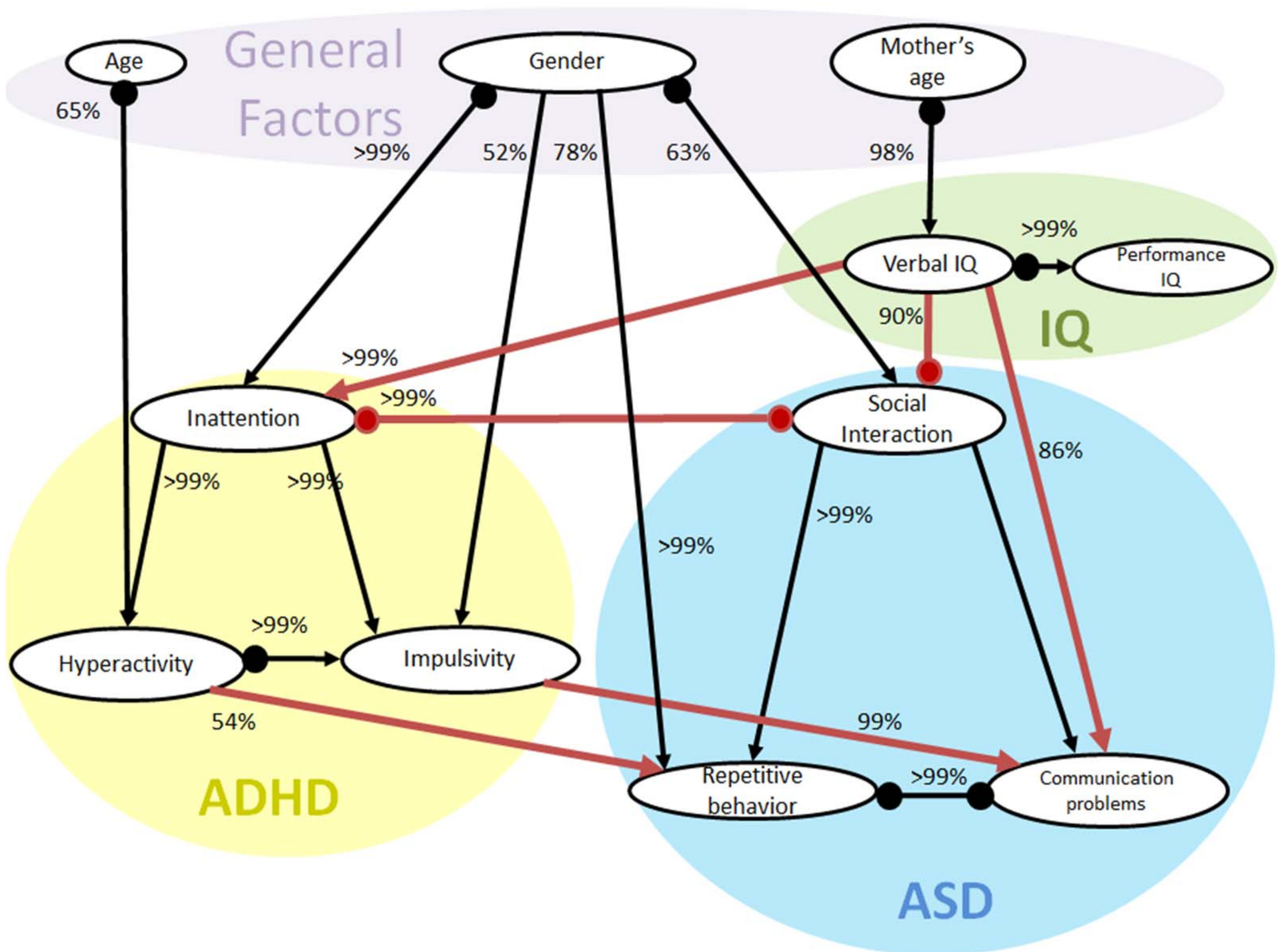
BCCD on IMpACT study

- global model for ADHD
- *risk haplotype* does appear to affect *(striatal response) activation*, but only via *inattention*
- total effect size: Cohen's d = 0.14 (not significant)

E. Sokolova et al., “Causal discovery in an adult ADHD data set suggests indirect link between DAT1 genetic variants and striatal brain activation during reward processing”, American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2015



Comorbidity between autism and ADHD



E. Sokolova et al., A causal and mediation analysis of the comorbidity between attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD), Journal of Autism and Developmental Disorders, 2017

Outline

- Statistical causal discovery
- The logic of causal inference
  - Connection to structural equation models
  - Causal DAGs and constraint-based methods
  - Logical Causal Inference (LoCI)
- A Bayesian approach...
- Applications
- Current research and future goals

Big data

- many applications typically contain thousands of variables (e.g. genetics): large *p*

- learning optimal sparse Bayesian networks is NP-hard [Chickering,1995]

⇒ high-dimensional 'big data sets' not suitable for causal discovery?

Ongoing NWO Top Grant with
Aad van der Vaart

NWO
Netherlands Organisation for Scientific Research

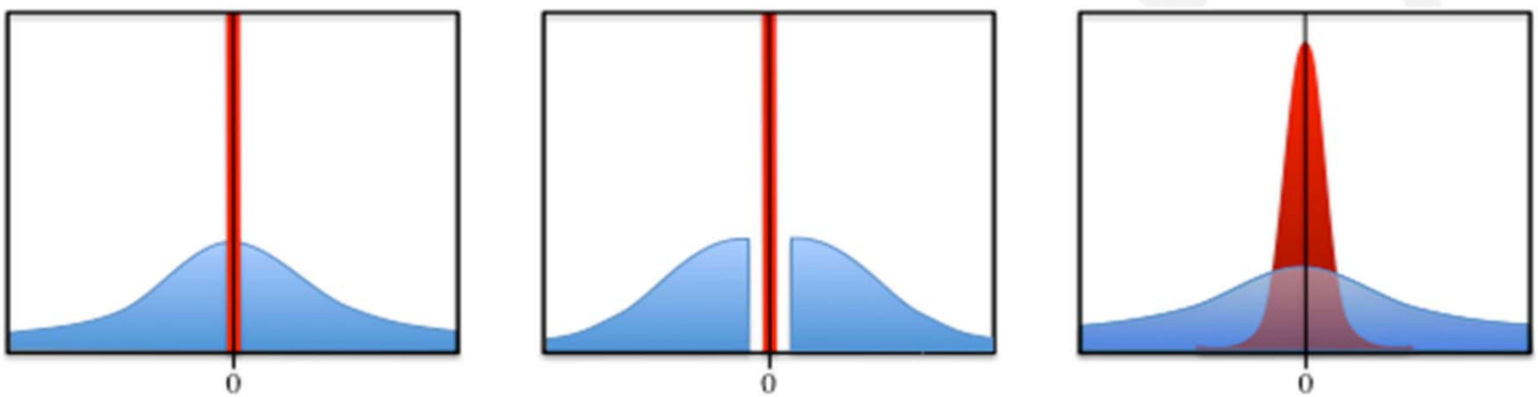
Big data

Answer(-ish)

- learning sparse causal models is not NP-hard! [Claassen, Mooij, Heskes, 2013]
- modular approach: split up in (many....) overlapping subproblems
- for sparse models feasible up to thousand nodes
- parallelize algorithms to utilize GPU power [Fabian Gieseke, tbd]

Big data

- in theory: more data = more reliable output causal model
- in practice too much data, large *N*, can hurt! (weak dependencies)
  $\Rightarrow$ 'everything is connected to everything else, but we have no clue how'
- large (p,N): standard faithfulness insufficient for uniform consistency: theoretical analyses typically based on strong faithfulness assumptions



'default' faithfulness *'strong' faithfulness* *'weak' faithfulness*

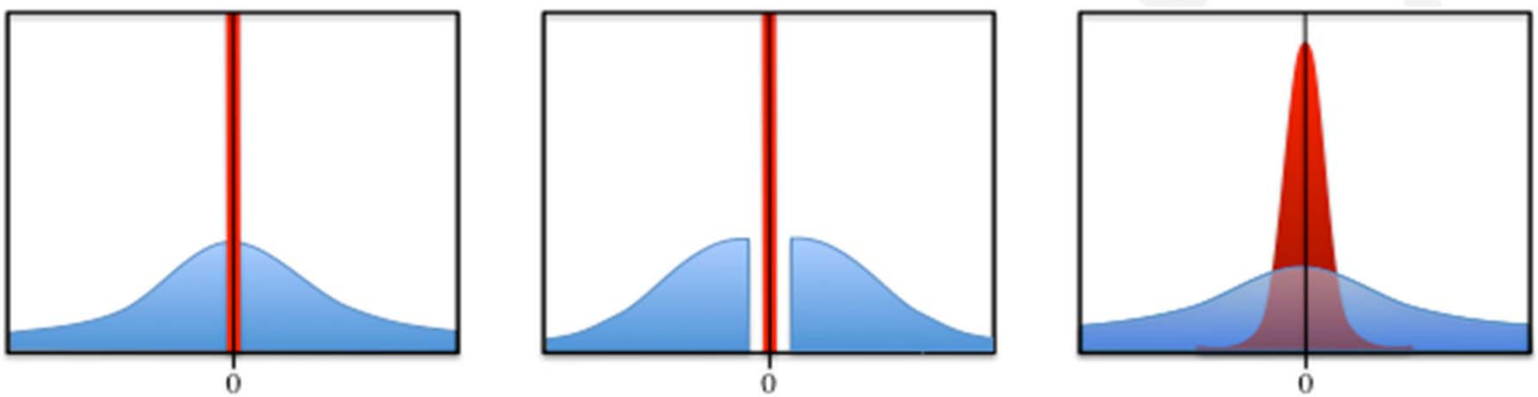
Big data

G. Bucur et al., Robust causal estimation in the large-sample limit without strict faithfulness, AISTATS, 2017

Possible approach

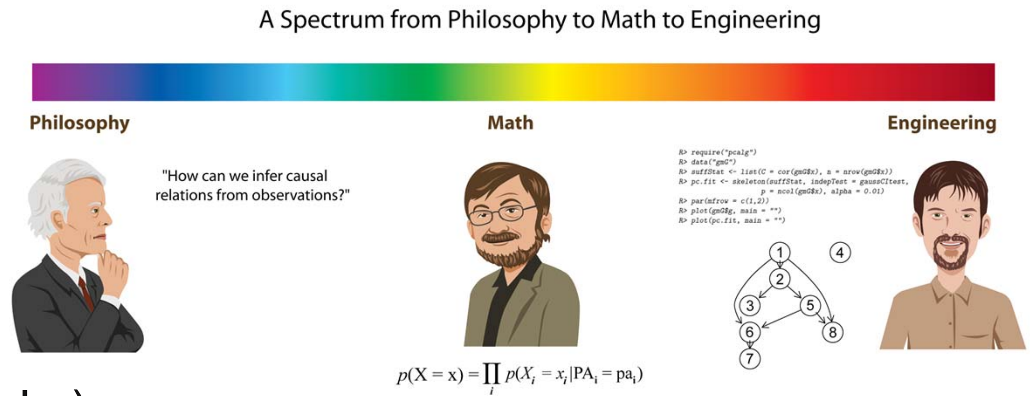
no 'accidental' causal cancellations

- forget about faithfulness
- change focus: complete model ⇒ all 'relevant' causal relations
- similar (but simpler) problems, e.g., needle in a haystack, have been tackled under weaker assumptions (weak ℓq-balls)



'default' faithfulness *'strong' faithfulness* *'weak' faithfulness*

Lots of improvements



Other challenges

- allow for complicated models (feedback in gene-regulatory networks)
- handle mixed data
- overlapping data sets (multiple experiments)
- longitudinal data sets
- joint estimation of structure and (treatment) effects

R. Cui et al., Copula PC algorithm for causal discovery from mixed data, ECML/PKDD, 2016

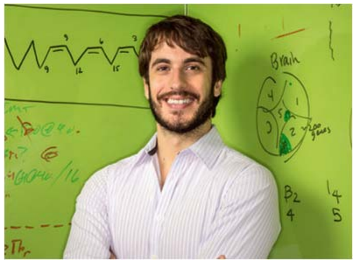
R. Rahmadi et al., Causality on longitudinal data: Stable specification search in constrained structural equation modeling. Statistical Methods in Medical Research, 2017

Ultimate goal

- principled causal discovery methods usable for mainstream scientific research and data analysis
- available software implementations
- results reported in terms of standard ‘causal confidence measures’

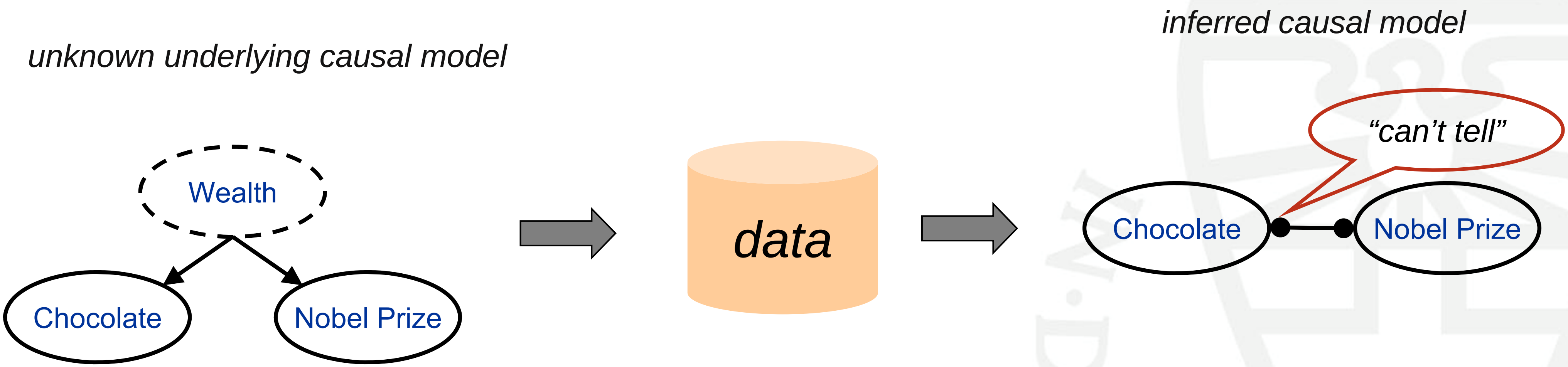
Big data and causality

Even in the last 20 to 30 years there has been a pretty big evolution in the statistical tools that we have at our disposal for actually inferring causality in an observational study [...] When I talk to my old colleagues at Facebook, they're spending a lot of time thinking about this problem. If you become increasingly skeptical of the results of your data analysis, you're going to become increasingly reliant on these tools for causal inference in observational studies. So I think that the world is actually moving in the direction of removing the opacity of the models that it generates.



Jeff Hammerbacher
(Cloudera)

Take-home message



Correlation does not imply causation

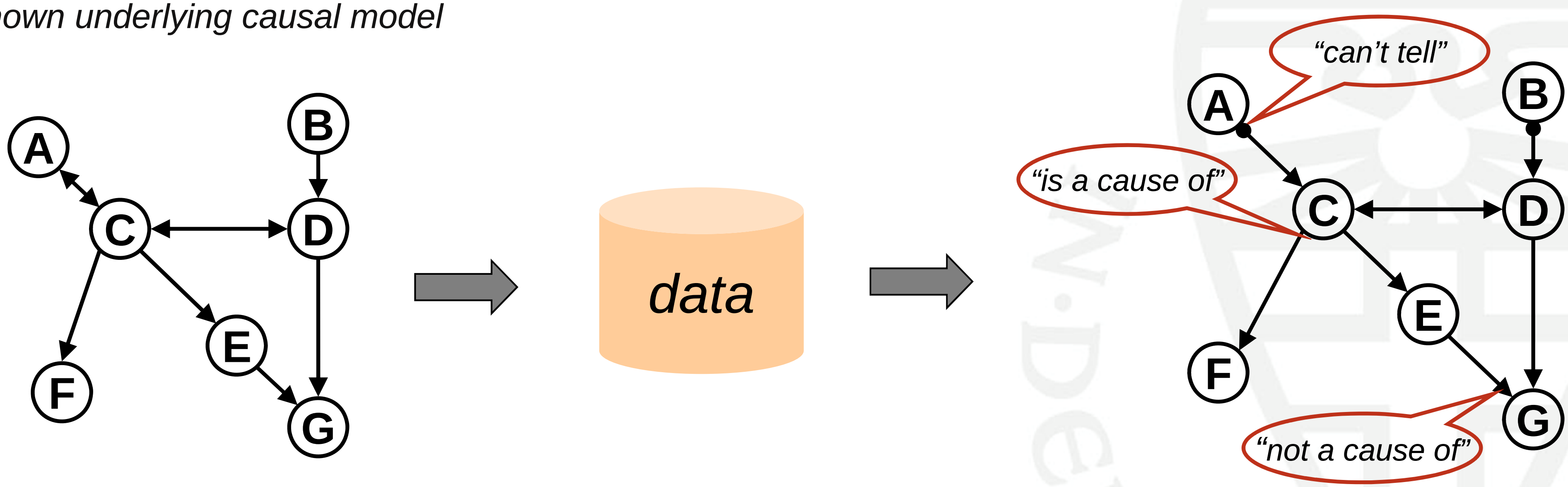
just a pair of variables
just a single symmetric number summarizing their dependence

BORING!

Take-home message

unknown underlying causal model

inferred causal model



Causal discovery from big data

challenging multi-disciplinary research
exciting opportunities

Many thanks to:
Tom Claassen, Joris Mooij, Elena Sokolova, Perry Groot, Ridho Rahmadi, Gabriel Bucur, Ruifei Cui