Title: Causal Inference in Healthcare

Speakers: Ciaran Lee

Series: Quantum Foundations

Date: February 18, 2020 - 3:30 PM

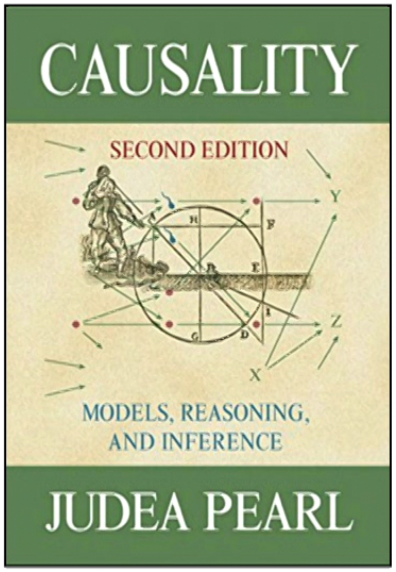
URL: http://pirsa.org/20020066

Abstract: Causal reasoning is vital for effective reasoning in science and medicine. In medical diagnosis, for example, a doctor aims to explain a patientâ€™s symptoms by determining the diseases causing them. This is because causal relations---unlike correlations---allow one to reason about the consequences of possible treatments. However, all previous approaches to machine-learning assisted diagnosis, including deep learning and model-based Bayesian approaches, learn by association and do not distinguish correlation from causation. I will show that these approaches systematically lead to incorrect diagnoses. I will outline a new diagnostic algorithm, based on counterfactual inference, which captures the causal aspect of diagnosis overlooked by previous approaches and overcomes these issues. I will additionally describe recent algorithms from my group which can discover causal relations from uncontrolled observational data and show how these can be applied to facilitate effective reasoning in medical settings such as deciding how to treat certain diseases.

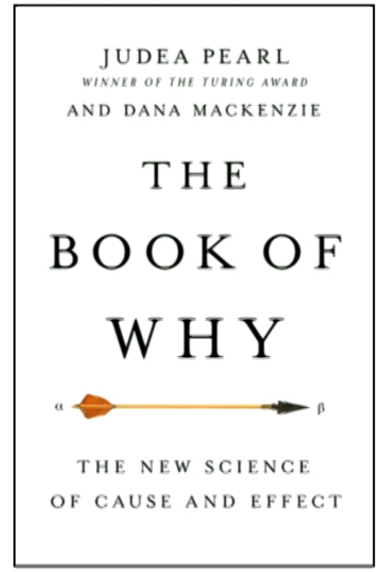
Causal Inference in Healthcare

Ciarán M. Lee
Babylon Health & University College London

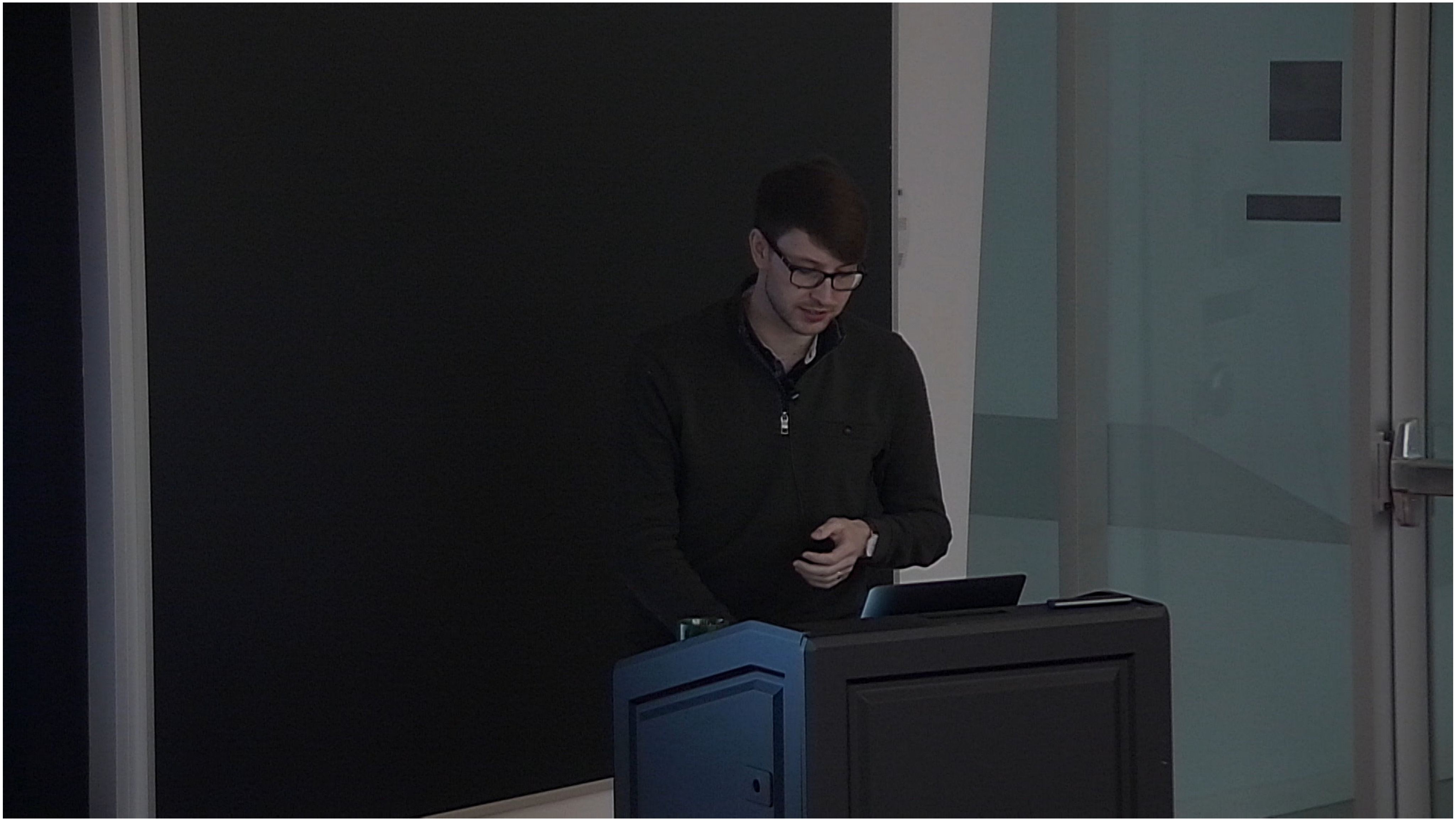
What is causal inference?



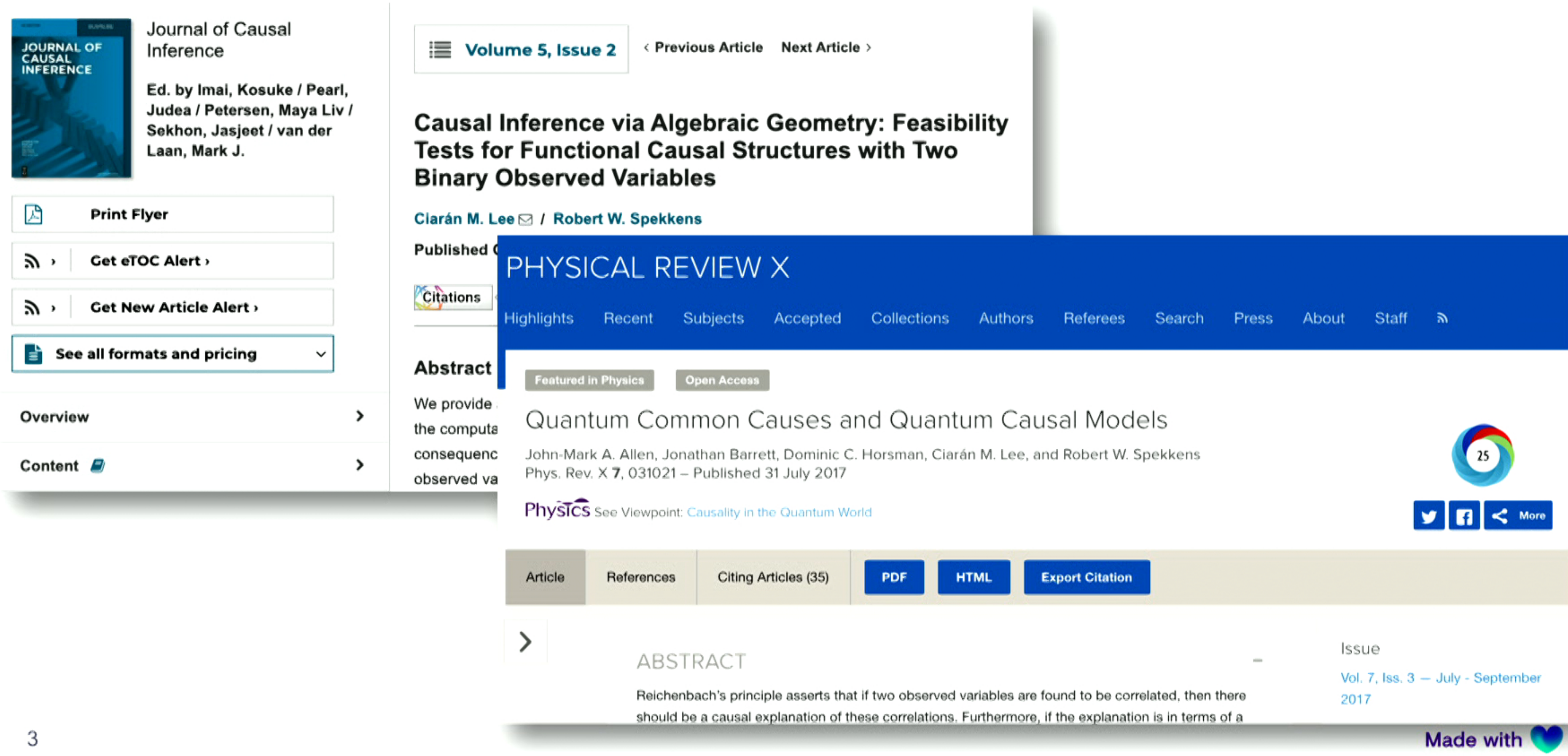
- Causal Inference provides the tools to ask and answer causal questions
- Does smoking cause lung cancer?
- Would I have cancer had I smoked?



Made with



I first got interested in causal inference as a way to better understand quantum mechanics


Journal of Causal Inference

Ed. by Imai, Kosuke / Pearl, Judea / Petersen, Maya Liv / Sekhon, Jasjeet / van der Laan, Mark J.

Print Flyer

Get eTOC Alert ›

Get New Article Alert ›

See all formats and pricing

Overview

Content

Volume 5, Issue 2

‹ Previous Article Next Article ›

Causal Inference via Algebraic Geometry: Feasibility Tests for Functional Causal Structures with Two Binary Observed Variables

Ciarán M. Lee / Robert W. Spekkens

Published

Citations

Abstract

We provide

the computa

consequenc

observed va

PHYSICAL REVIEW X

Highlights Recent Subjects Accepted Collections Authors Referees Search Press About Staff

Featured in Physics Open Access

Quantum Common Causes and Quantum Causal Models

John-Mark A. Allen, Jonathan Barrett, Dominic C. Horsman, Ciarán M. Lee, and Robert W. Spekkens

Phys. Rev. X **7**, 031021 – Published 31 July 2017

Physics See Viewpoint: Causality in the Quantum World

25

More

Article References Citing Articles (35) PDF HTML Export Citation

ABSTRACT

Reichenbach's principle asserts that if two observed variables are found to be correlated, then there should be a causal explanation of these correlations. Furthermore, if the explanation is in terms of a

Issue

Vol. 7, Iss. 3 — July - September 2017

Made with

Causal inference is useful for studying quantum cryptography

nature > npj quantum information > articles > article

npj | Quantum Information

Article | Open Access | Published: 29 April 2019

Device-independent certification of non-classical joint measurements via causal models

Ciarán M. Lee

npj Quantum Information 5

846 Accesses | **1** Citatio

This article has been u

Abstract

Quantum measuremen
rise to some of the mos

PHYSICAL REVIEW LETTERS

Highlights Recent Accepted Collections Authors Referees Search Press About

Open Access

Towards Device-Independent Information Processing on General Quantum Networks

Ciarán M. Lee and Matty J. Hoban
Phys. Rev. Lett. **120**, 020504 – Published 10 January 2018

191

More

Article | References | Citing Articles (10) | Supplemental Material | PDF | HTML | Export Citation

ABSTRACT

The violation of certain Bell inequalities allows for device-independent information processing secure against nonsignaling eavesdroppers. However, this only holds for the Bell network, in which two or

Issue

Vol. 120, Iss. 2 — 12 January 2018

Clever maths will stop hackers spying on the quantum internet

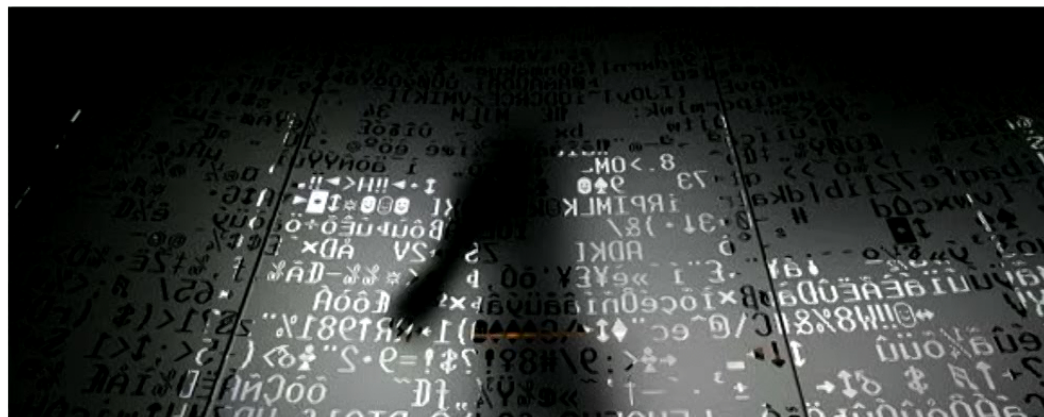
TECHNOLOGY 15 January 2018

By Jacob Aron

TECH & SCIENCE

CAN WE BUILD A HACK-PROOF INTERNET USING QUANTUM PHYSICS? NEW BREAKTHROUGH HEIGHTENS 'TECHNOLOGY ARMS RACE'

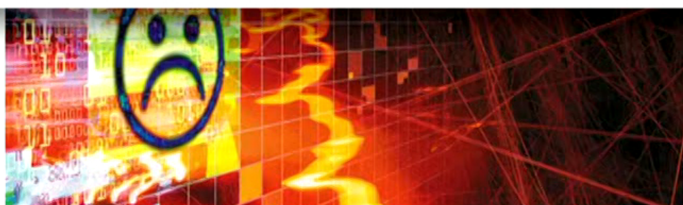
BY KASTALIA MEDRANO ON 1/12/18 AT 8:54 AM EST



Clever maths will stop hackers spying on the quantum internet

TECHNOLOGY 15 January 2018

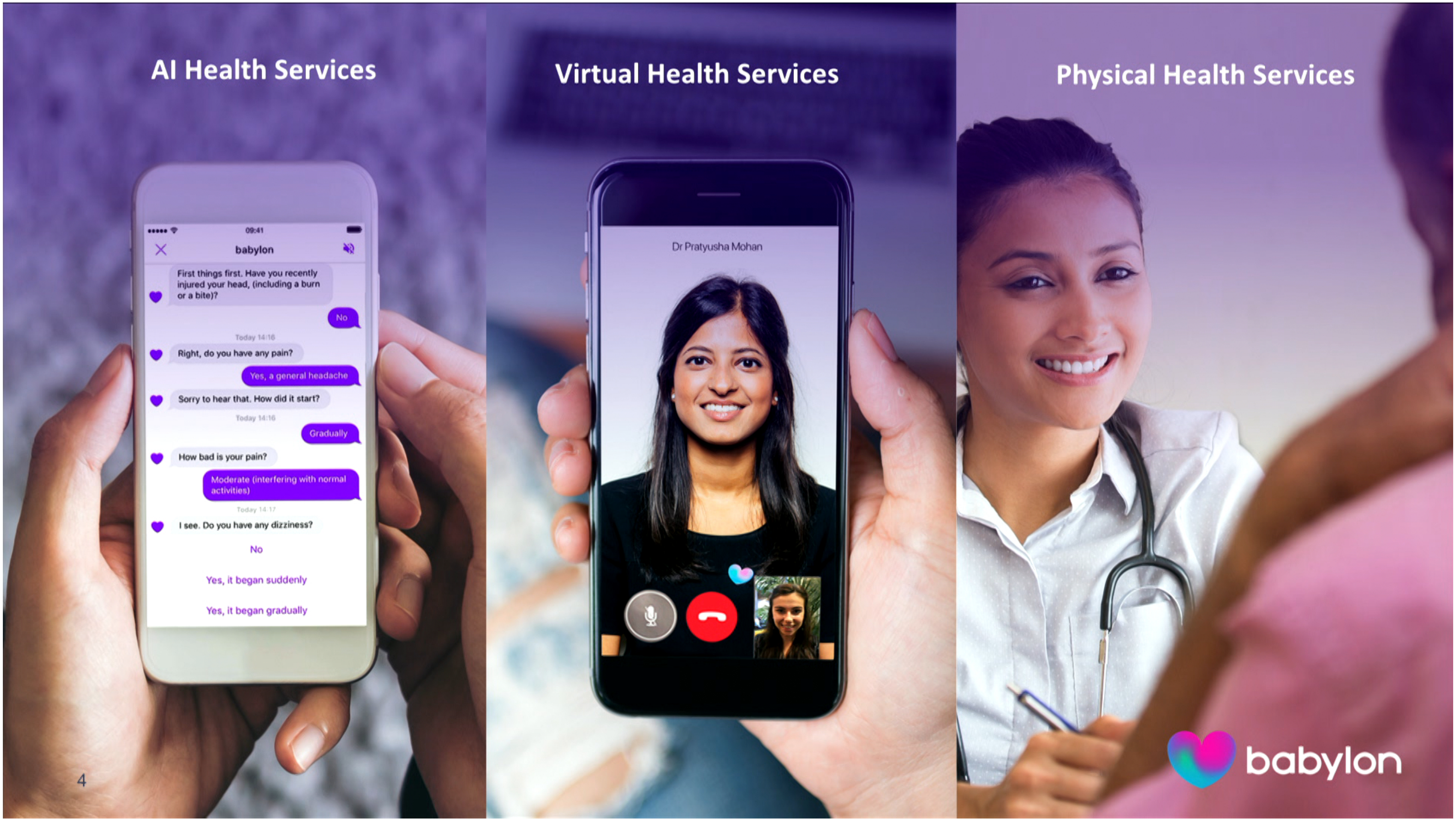
By Jacob Aron



TECH & SCIENCE

CAN WE BUILD A HACK-PROOF INTERNET USING QUANTUM PHYSICS? NEW

The pair overcome this by using a technique from machine learning called causal inference to study the structure of the network. Essentially, a computer analyses the direction of information flow between the different nodes to figure out its causal structure. For example, if node A is connected to

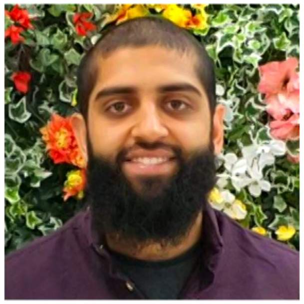

AI Health Services
Virtual Health Services
Physical Health Services
09:41
babylon
First things first. Have you recently injured your head, (including a burn or a bite)?
No
Today 14:16
Right, do you have any pain?
Yes, a general headache
Sorry to hear that. How did it start?
Today 14:16
Gradually
How bad is your pain?
Moderate (interfering with normal activities)
Today 14:17
I see. Do you have any dizziness?
No
Yes, it began suddenly
Yes, it began gradually
Dr Pratyusha Mohan
babylon
4

The causal team

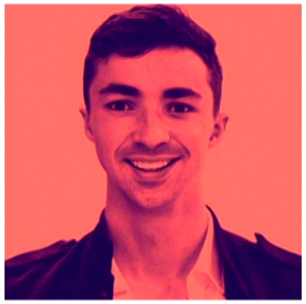
Anish Dhir
Research Scientist

Chris Hart
Research Scientist

Jon Richens
Research Scientist



Omar Jahangir
Research Intern, UCL



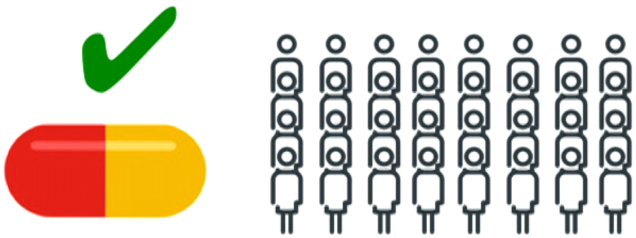
Logan Graham
Research Intern, Oxford

Ciarán Lee
Senior Research Scientist & team lead

Made with

- Causal knowledge is vital for effective reasoning in science and medicine
- In medical diagnosis, a doctor wants to determine the disease *causing* symptoms
- A direct causal relation, unlike a correlation, means treating the disease will reduce symptoms

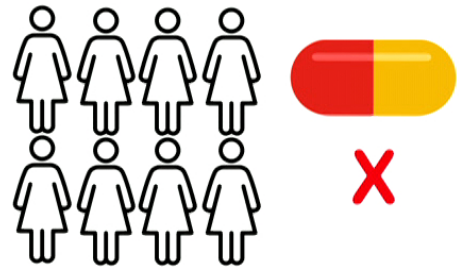
Suppose you have high cholesterol and your doctor told you a new drug has been shown to be effective when tested on the population as a whole.



That is, the drug is highly correlated with recovery

Would you take it?

Suppose, for a second opinion you see a different doctor who tells you that this drug has been shown to be ineffective when considering men and women alone



That is, the drug is negatively correlated with recovery for men & woman when considered separately

Would you take it?

- By looking at different subsets of the data, associations can completely reverse! This is Simpson's paradox.

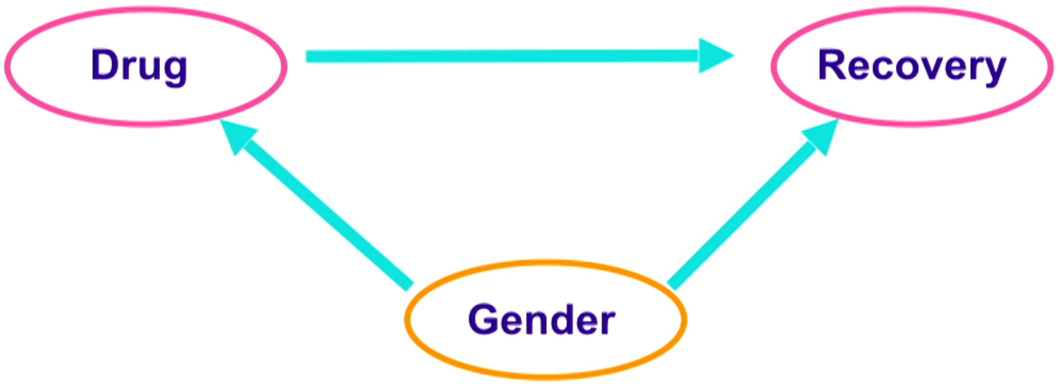
- Standard machine learning only learns patterns & correlations

- If it doesn't know your gender, it would prescribe this drug to you.

- If it knows you're a woman, it wouldn't prescribe the drug.

What's going on?

When offered the drug, men are more likely to take it. Men are also more likely to recover regardless of taking it. Gender is a confounder

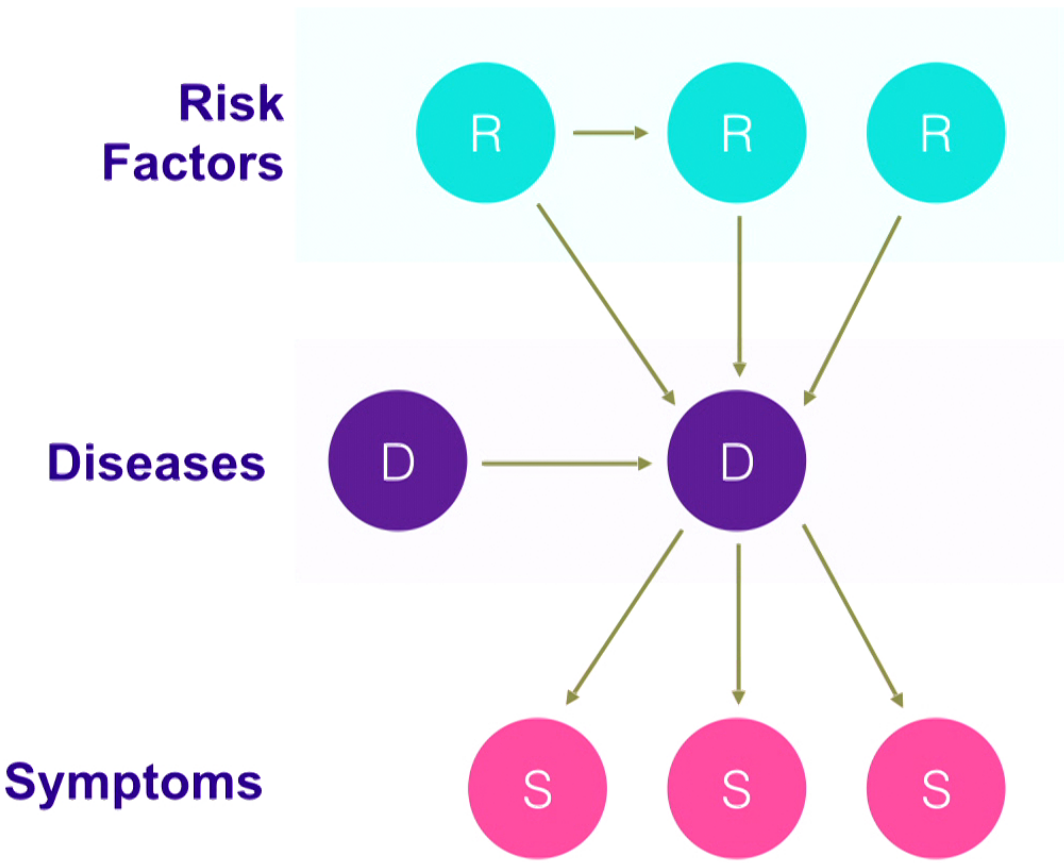


Instead of asking whether the drug is highly **correlated** with recovery, ask if it ***causes*** it

I’ll now show how the inability to disentangle correlation & causation leads to issues in an important problem in medicine:

Diagnosis

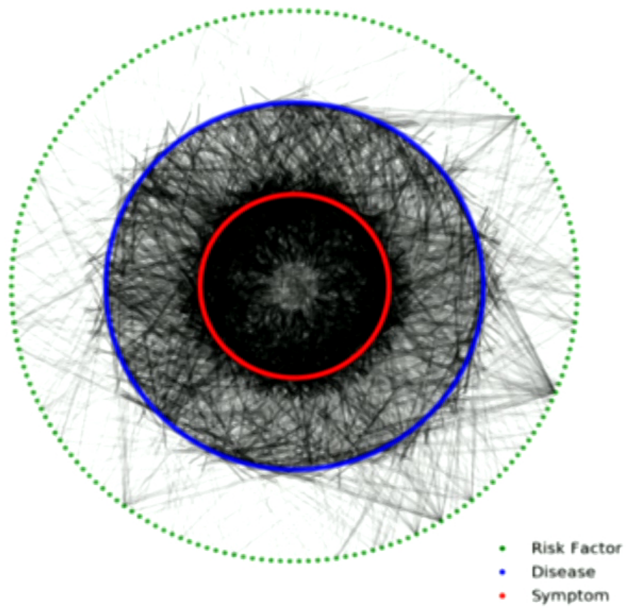
Disease Model



Probabilistic Generative Model (PGM)

- 3 layers involving risk factors (like smoking), diseases (like angina), symptoms (like chest pain).
- Causal links between nodes input by doctors and epidemiologists
- Have probabilities for each link, “How likely are you to have angina if you have chest pain and smoke”

Disease Model



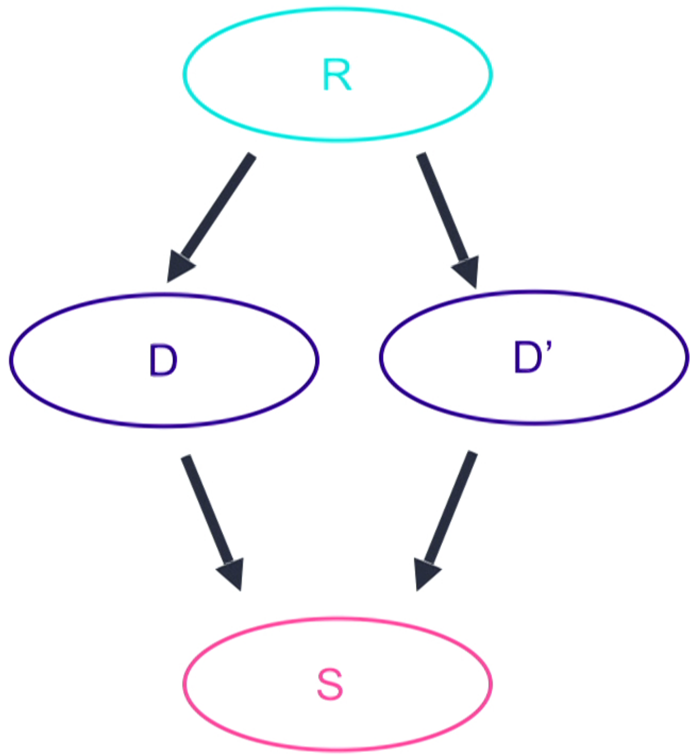
Posterior ranking approach to diagnosis

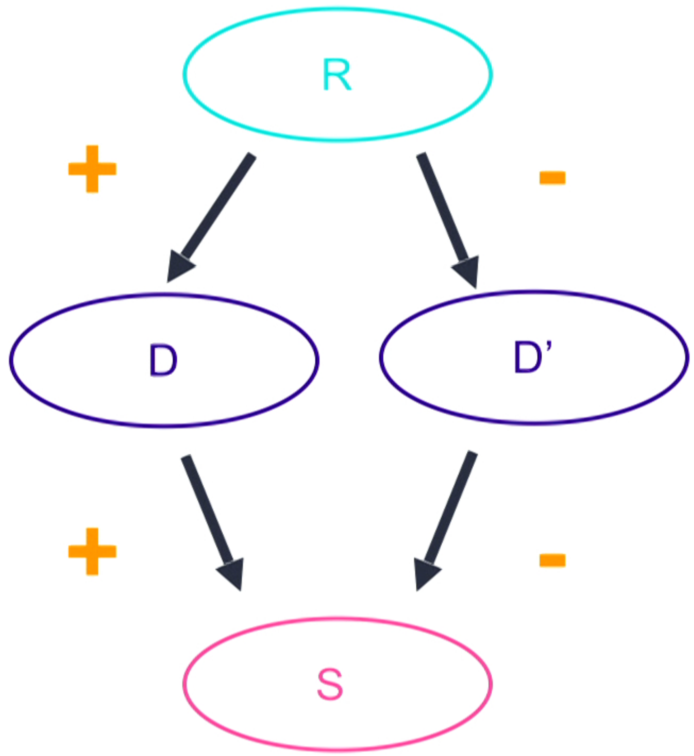
Given set of symptoms & risk factors, $\{S, R\}$, disease model is used to calculate posteriors for each disease, D:

$$P(D \mid S, R)$$

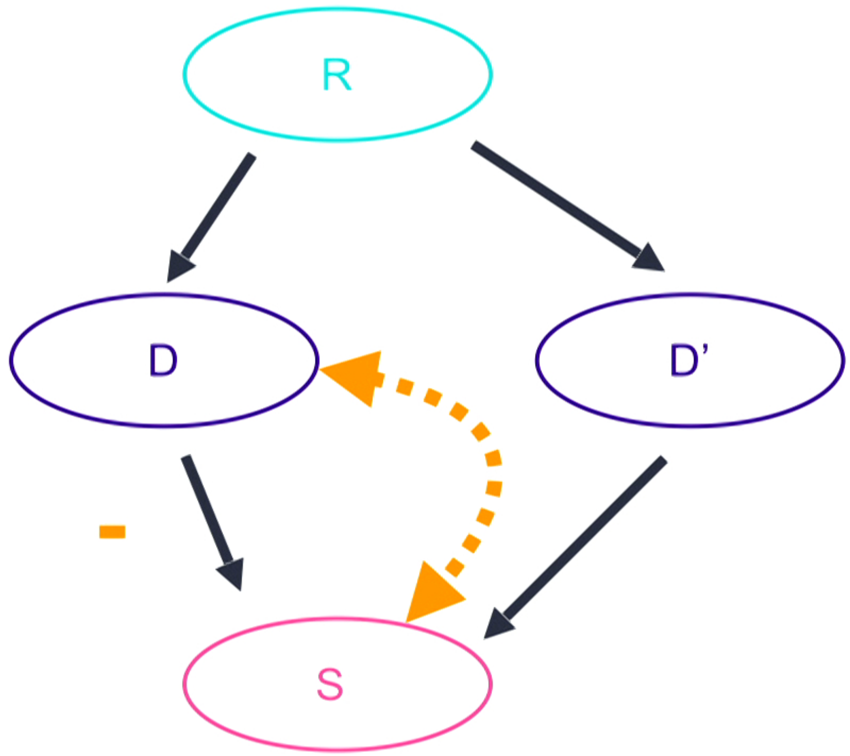
Diseases are then ranked by their posteriors, from most likely disease, to least likely

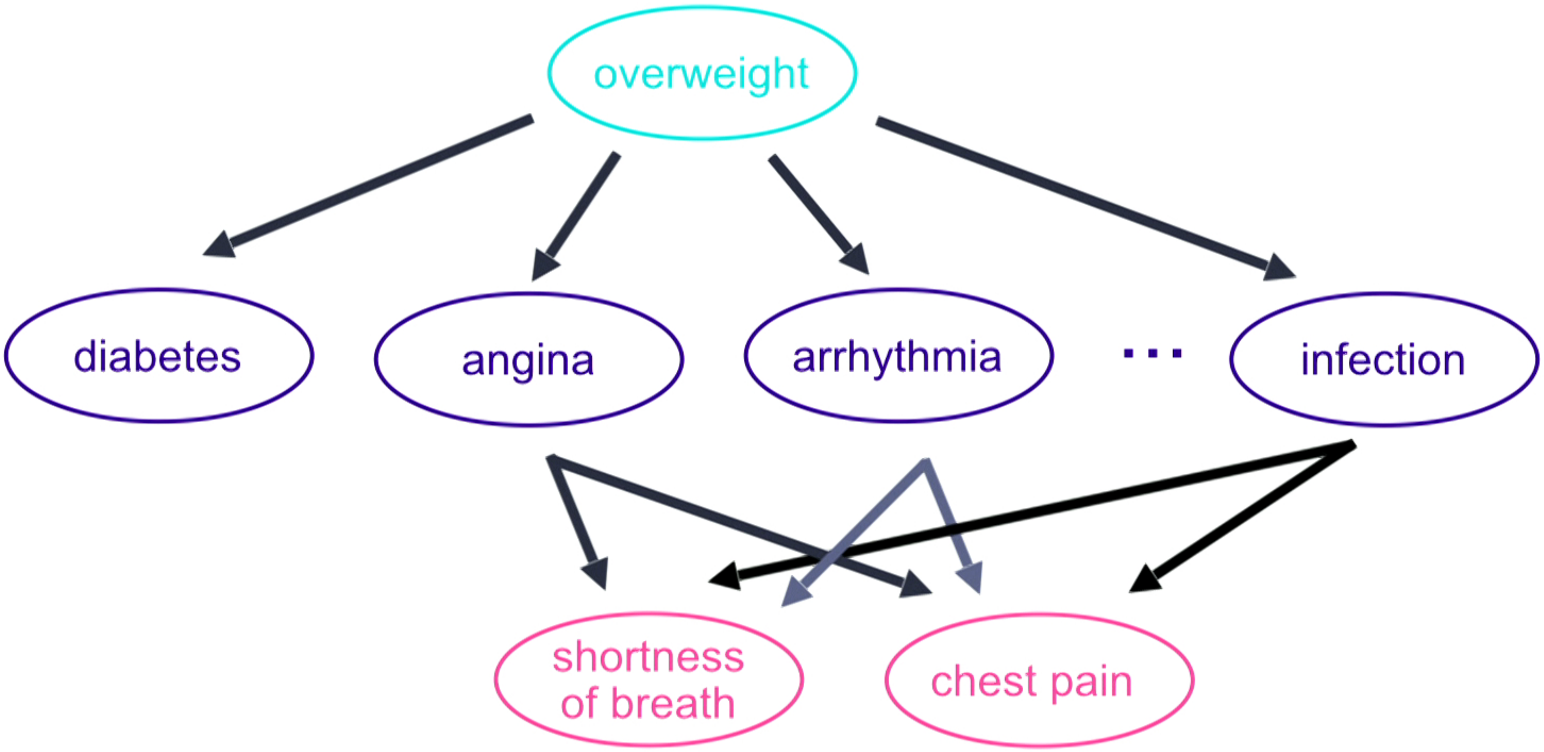
Not taking causal relations into account can lead to problems



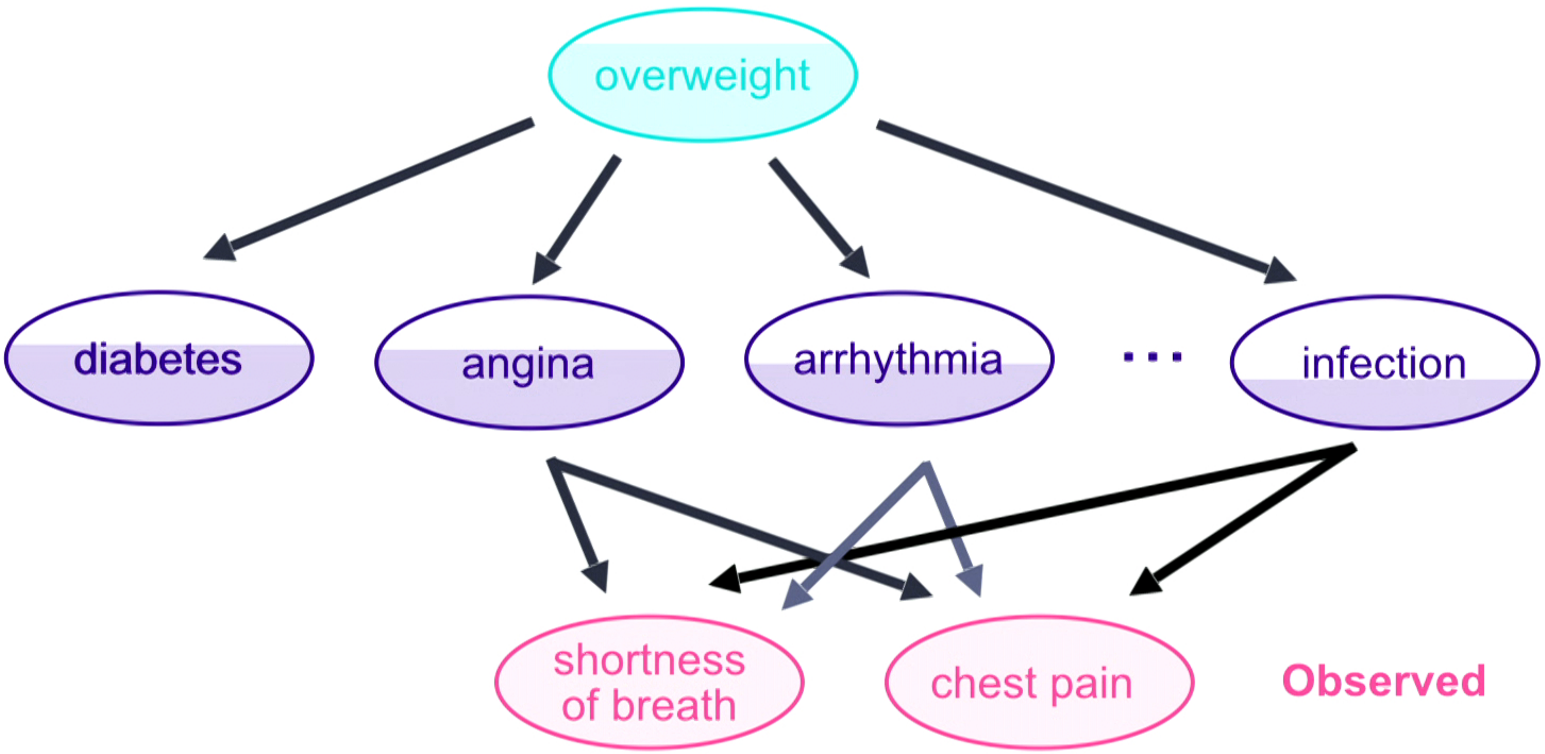

R
+
-
D
D'
+
-
S

Simpson's Paradox in action!




overweight
diabetes
angina
arrhythmia
...
infection
shortness of breath
chest pain

Made with


overweight
diabetes
angina
arrhythmia
...
infection
shortness of breath
chest pain
Observed

Made with

Diagnostic Desiderata

- **Consistency**: The likelihood that a disease is causing symptoms should be proportional to the posterior likelihood of that disease
- **Causality**: Any disease D that cannot cause any of the patient's observed symptoms should not be included in a diagnosis
- **Simplicity**: Diseases that explain a greater number of the patient's symptoms should be more likely

Diagnostic Desiderata

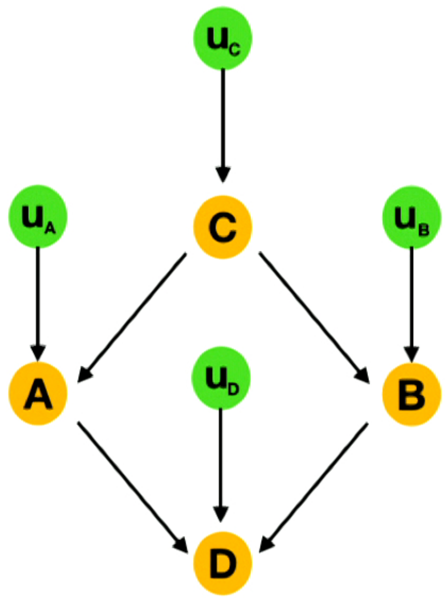
Posterior ranking

- **Consistency**: The likelihood that a disease is causing symptoms should be proportional to the posterior likelihood of that disease
- ~~**Causality**: Any disease D that cannot cause any of the patient's observed symptoms should not be included in a diagnosis~~
- ~~**Simplicity**: Diseases that explain a greater number of the patient's symptoms should be more likely~~

29

Made with

Causal Models



- Observed terms are deterministic function of parents and latent "noise"
- Noise terms are distributed according to latent distribution
- $A = f(C, u_A)$, $u_A \sim p(u_A)$
- These jointly generate $P(A|C)$

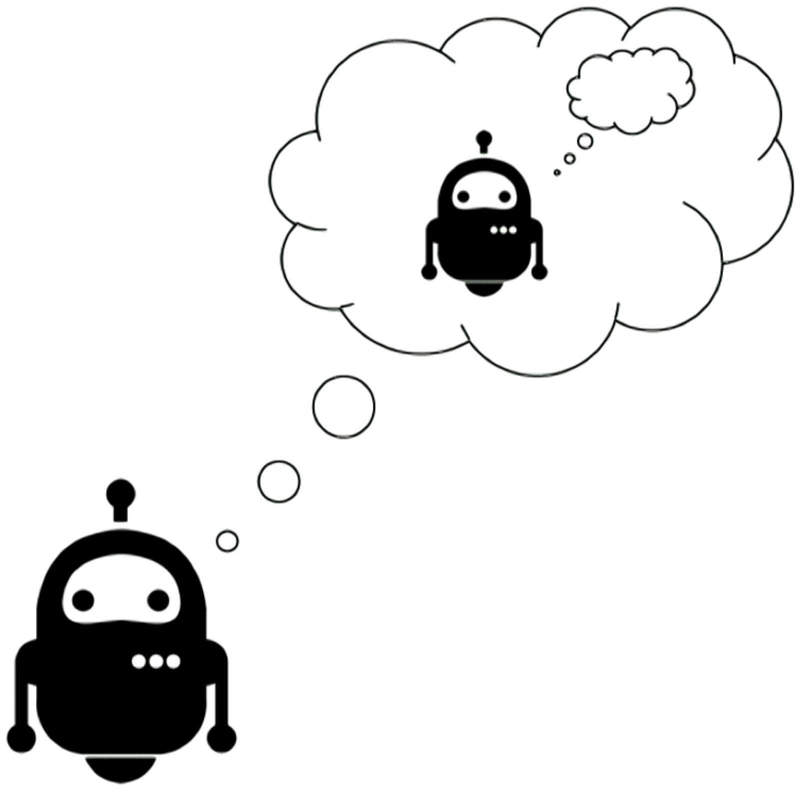
Made with

What can we do with them

Inference Hierarchy

Complexity

- Seeing (observations)
- Acting (interventions)
- Imagining (counterfactuals)

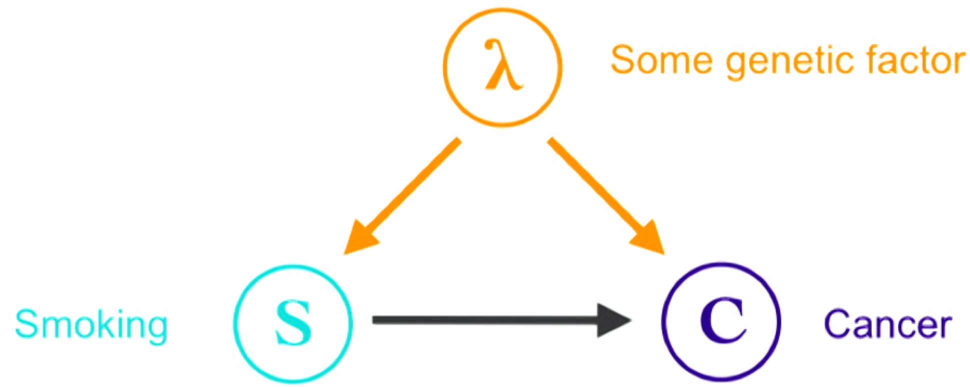


Made with

Seeing (observations)

Acting (interventions)

Imagining (counterfactuals)



given

Probability of cancer

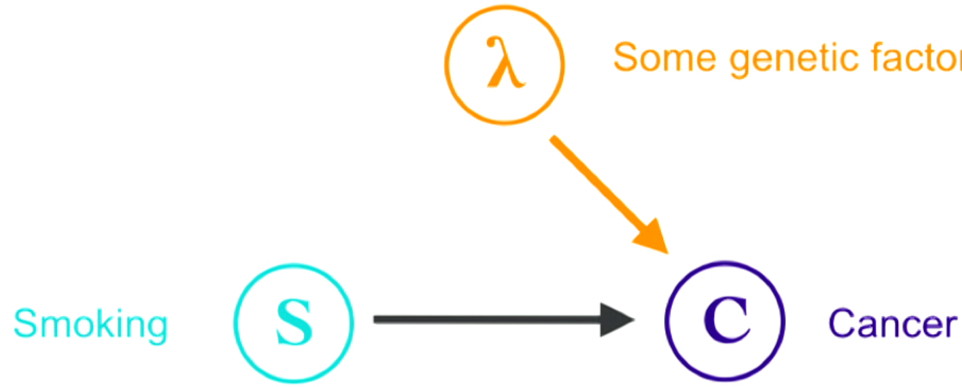
Smoking has been observed

$$P(C = T|S = T)$$

Seeing (observations)

Acting (interventions)

Imagining (counterfactuals)

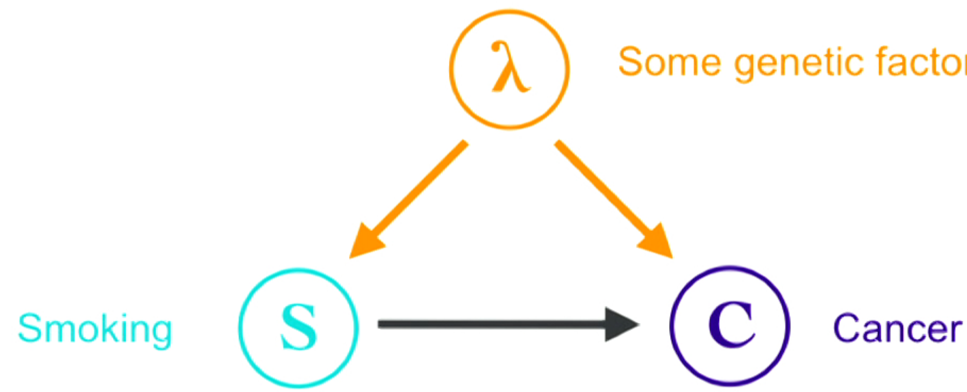


$$\overbrace{P(C=T}^{\text{Probability of cancer}}\overset{\text{given}}{\Big|}\overbrace{\mathrm{do}(S=T)}^{\text{subject was made to smoke}})$$

Seeing (observations)

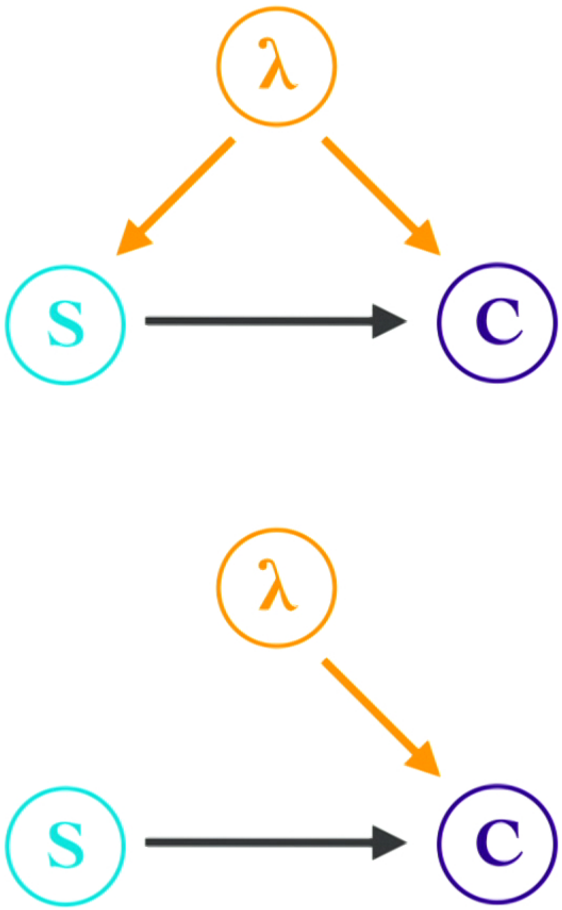
Acting (interventions)

Imagining (counterfactuals)



$$\overbrace{P(C = F}^{\text{Probability of subject not having cancer}} \overset{\text{given}}{|} \overbrace{C = T}^{\text{subject has cancer}}, \overbrace{\mathrm{do}(S = F)}^{\text{and if subject was made to not smoke}})$$

Made with



Counterfactual Inference compute P(C=F | C=T, S=T, do(S=F)):

1. **Abduction**: update P(λ) to P(λ | S=T, C=T)

2. **Action**: Apply do(.) operator to force S=F

3. **Predict**: Compute P(C=F) in model with do(S=F) & P(λ | S=T, C=T)

Posterior ranking versus **Counterfactual Inference**

$P(D=T \mid S=T, R)$ versus $P(S=F \mid S=T, R, do(D=F))$

"What is most likely disease, given evidence?"

"Given symptoms are present, would they not be, had disease be cured?"

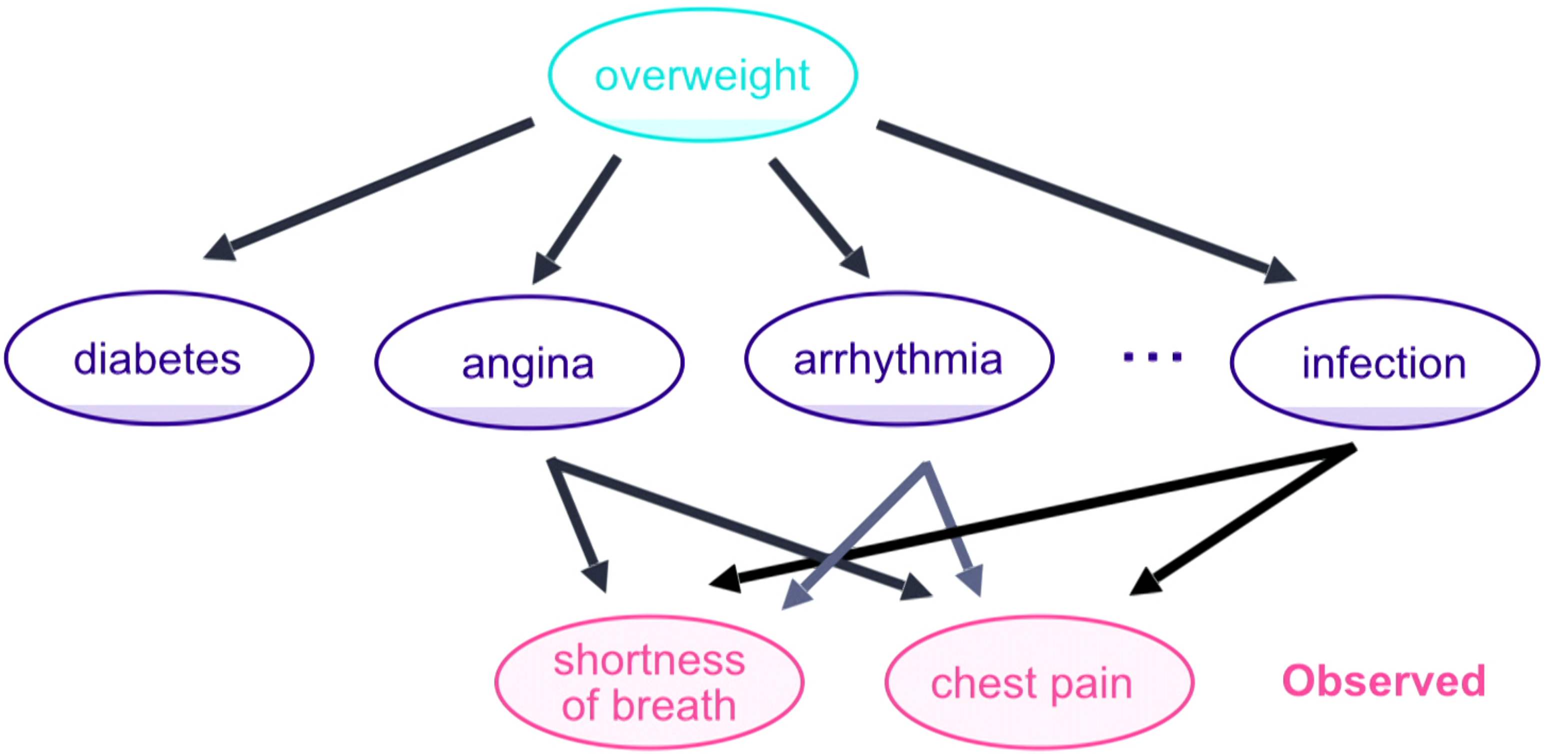
Posterior ranking versus **Counterfactual Inference**

$P(\, D{=}T \mid S{=}T, R\,)$ versus $P(S{=}F \mid S{=}T, R, do(D{=}F)\,)$

Probability of Disablement

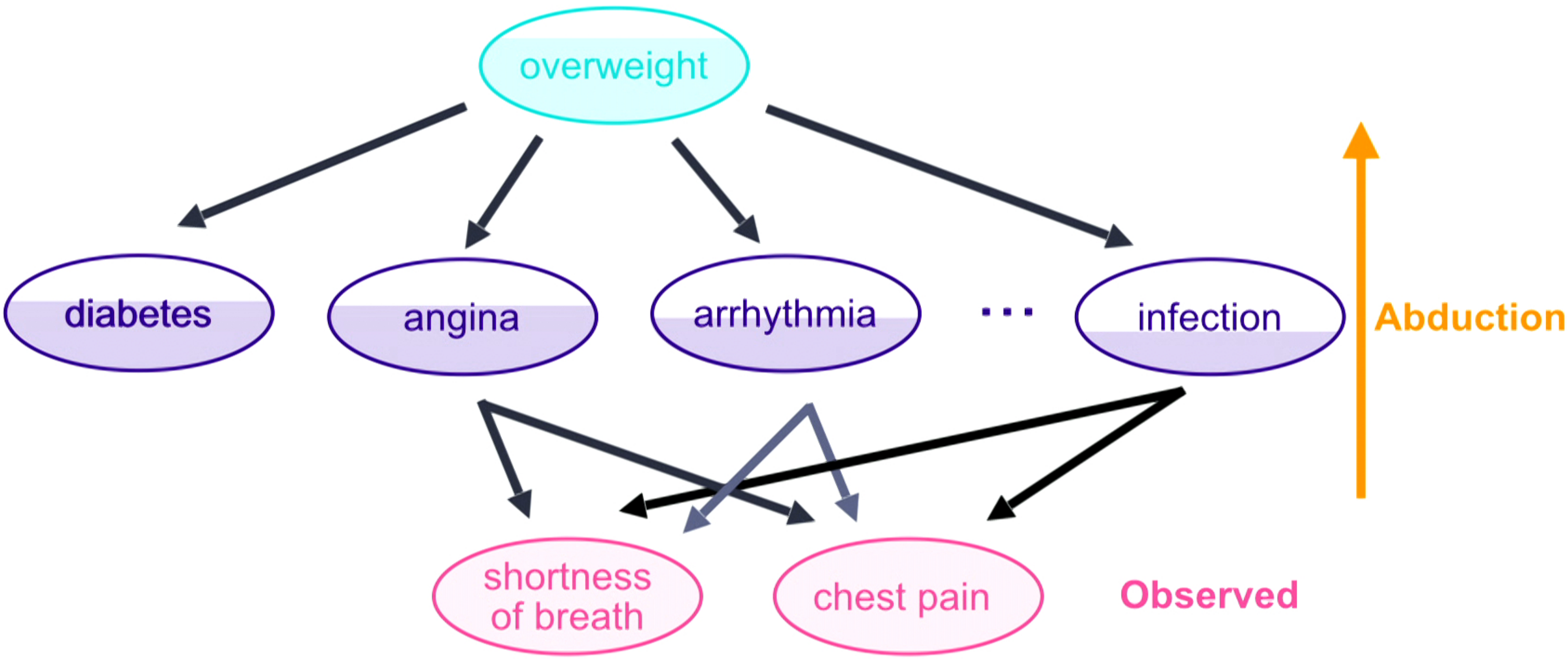
Made with

Disablement



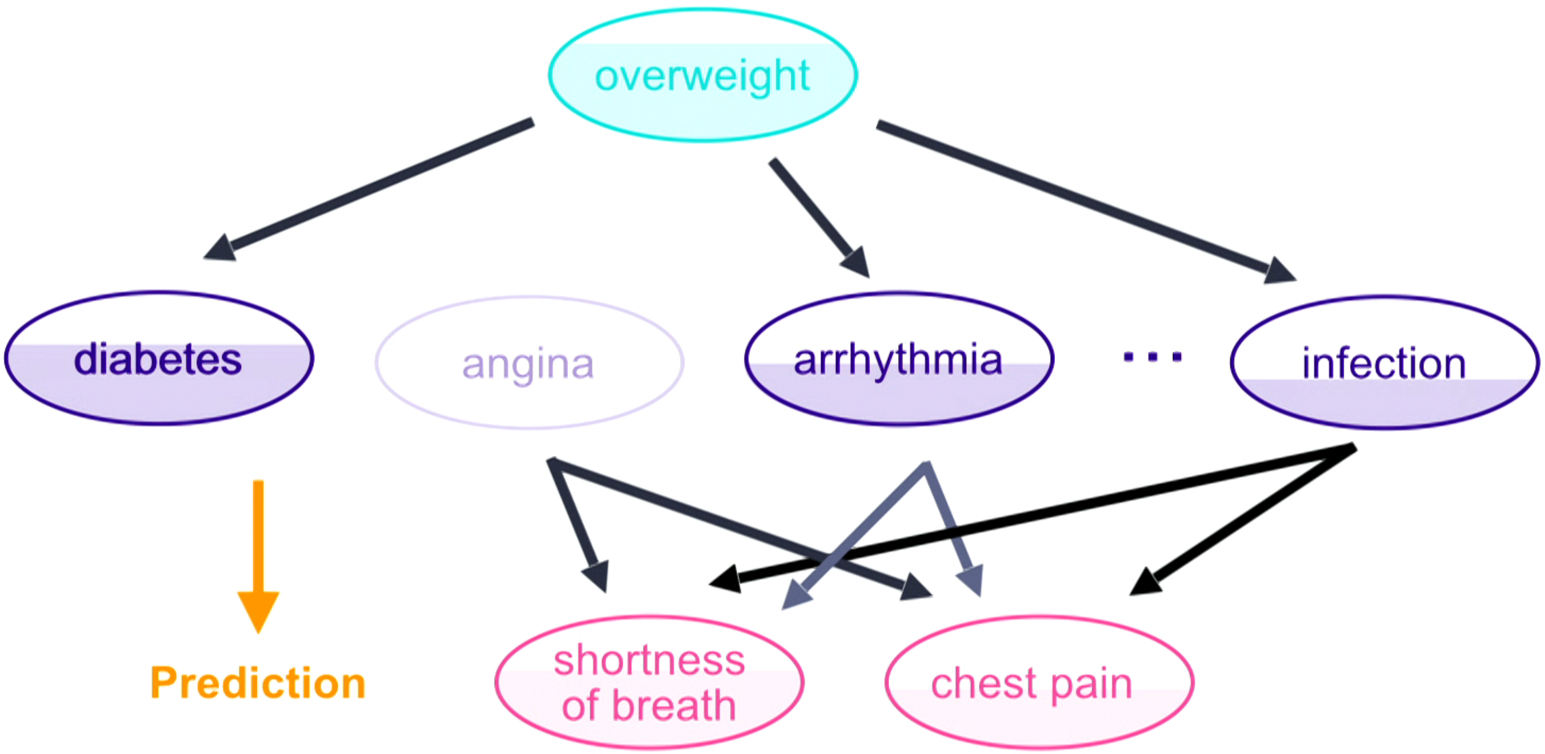
Made with

Disablement



Made with

Disablement



Made with

Expected Disablement

$$\mathbb{E}_{dis}(D_k, \mathcal{E}) := \sum_{\mathcal{S}'} \left| \mathcal{S}_+ \setminus \mathcal{S}'_+ \right| p(\mathcal{S}' | \mathcal{E}, do(D_k = 0))$$

Derives from notion of necessary cause and measures how how well a single disease explains presented symptoms

Expected Sufficiency

$$\mathbb{E}_{suff}(D_k, \mathcal{E}) := \sum_{\mathcal{S}'} \left| \mathcal{S}'_+ \right| p(\mathcal{S}' | \mathcal{E}, do(\mathsf{Pa}(\mathcal{S}_+) \setminus D_k = 0))$$

Derives from notion of sufficient cause and measures how many symptoms we expect to be caused by a disease

Made with

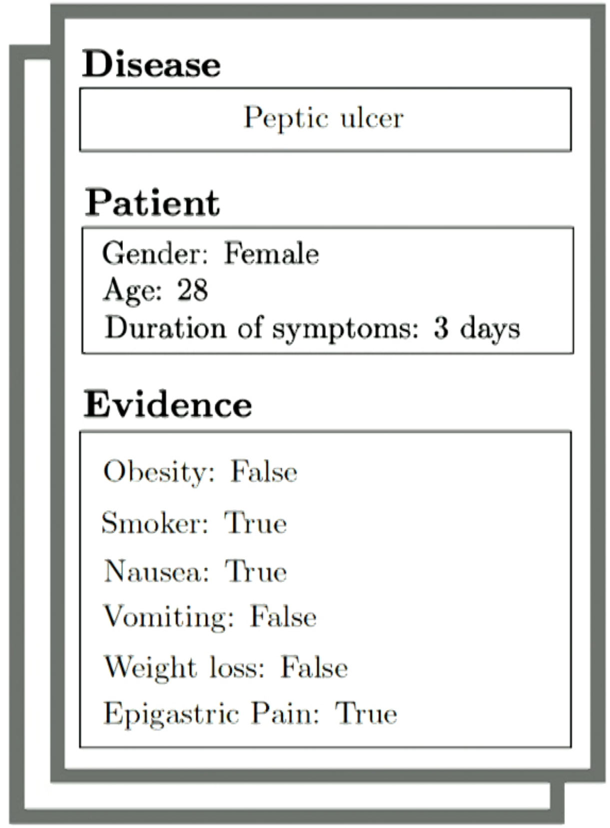
Theorem: Expected Disablement and Expected Sufficiency satisfy three desiderata of Consistency, Causality, and Simplicity

Comparing posterior ranking and counterfactual ranking

We test using 1700 medical cases prepared by a panel of doctors

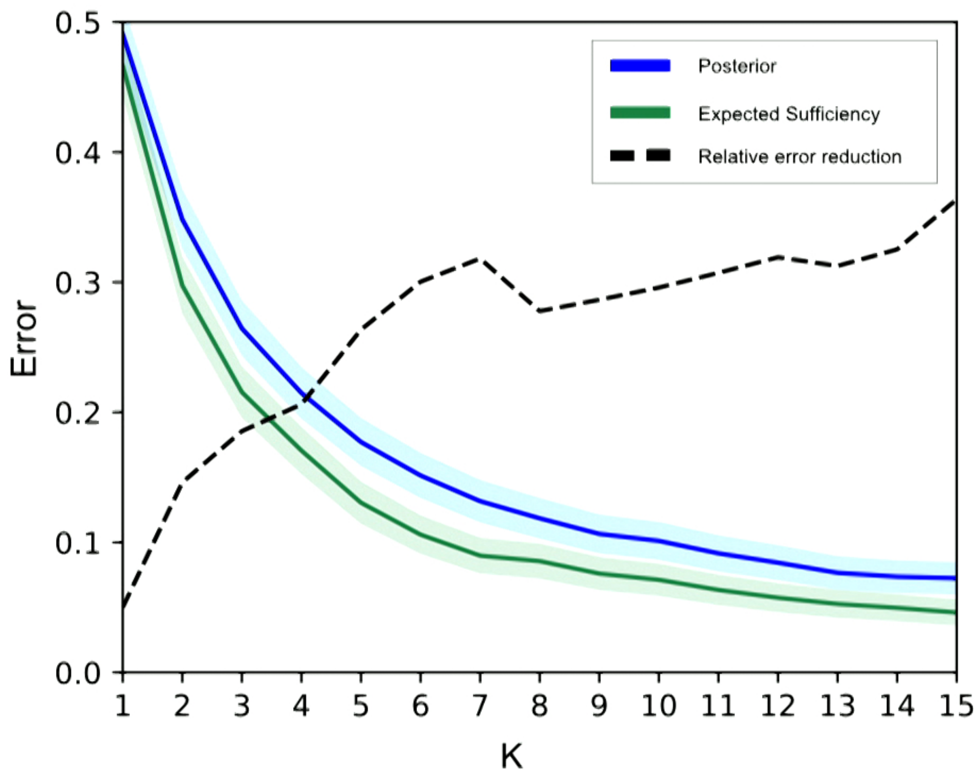
From symptoms & medical history from case, diagnose disease

Output ranked list of diseases, compare probability of case disease in top k

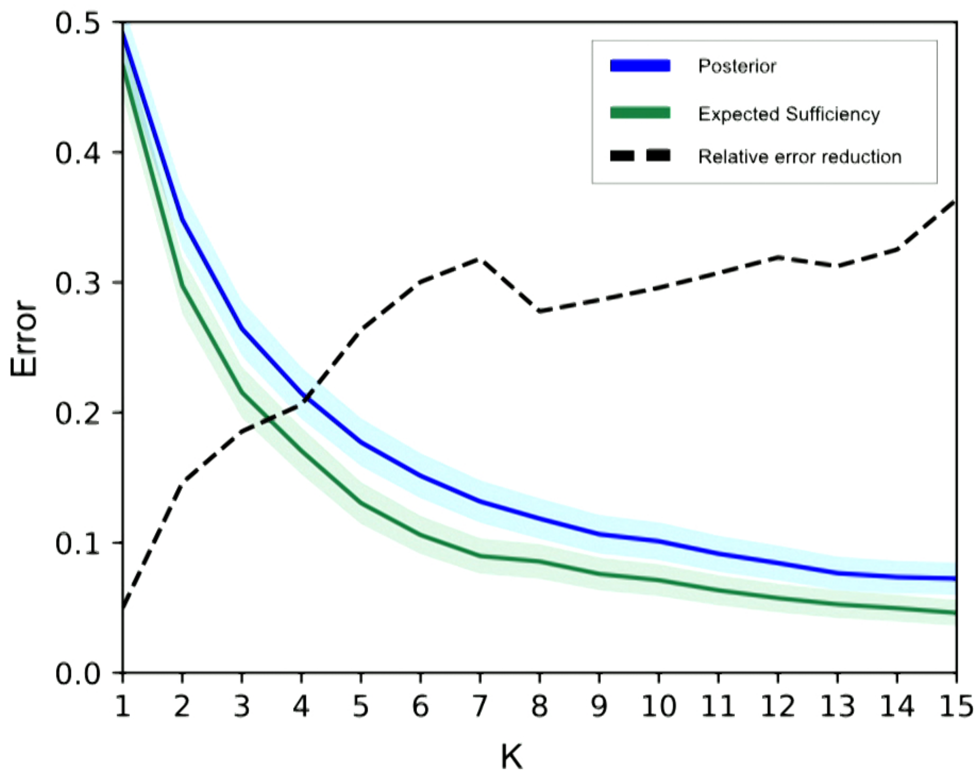


Made with

$$\mathbb{E}_{suff}(D_k, \mathcal{E}) := \sum_{\mathcal{S}'} \left|\mathcal{S}'_+\right| p(\mathcal{S}'|\mathcal{E}, do(\mathsf{Pa}(\mathcal{S}_+) \setminus D_k = 0))$$

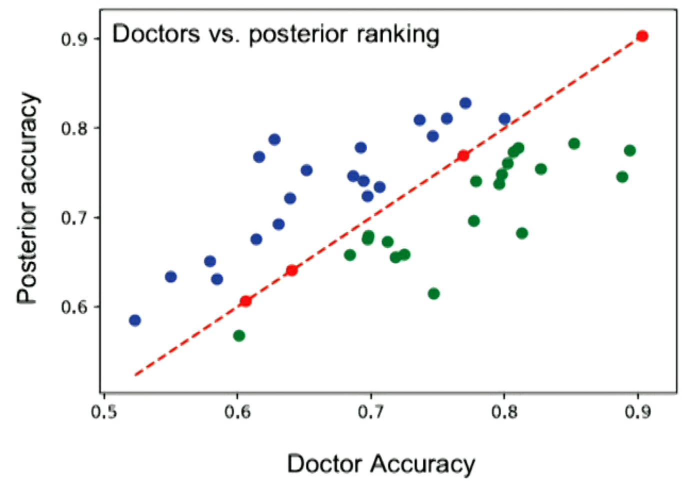


$$\mathbb{E}_{suff}(D_k, \mathcal{E}) := \sum_{\mathcal{S}'} \left|\mathcal{S}'_+\right| p(\mathcal{S}'|\mathcal{E}, do(\mathsf{Pa}(\mathcal{S}_+) \setminus D_k = 0))$$

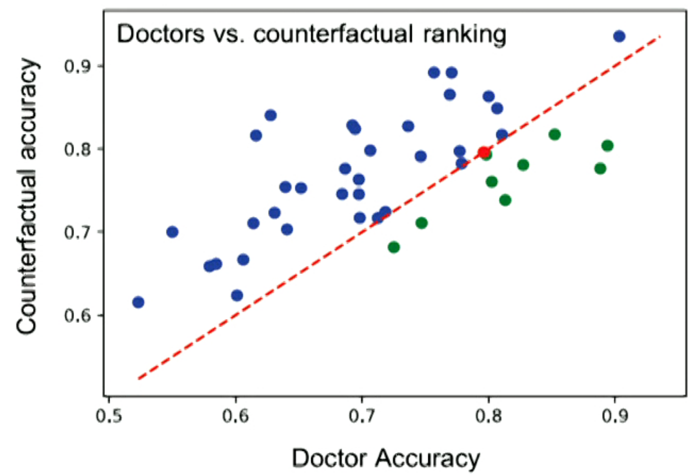


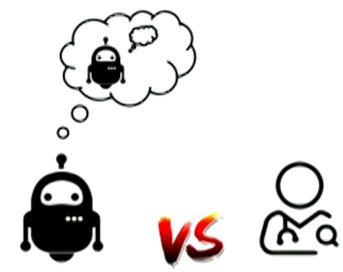
Compare to 44 doctors not involved in creating medical cases

Standard Bayesian updating places in top 48% of doctors, achieving average clinical accuracy



Counterfactual inference places in top 25% of doctors, achieving expert clinical accuracy





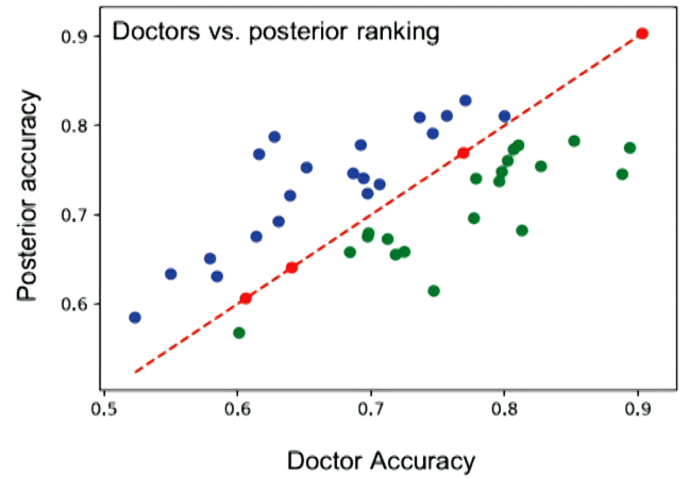
49

Made with

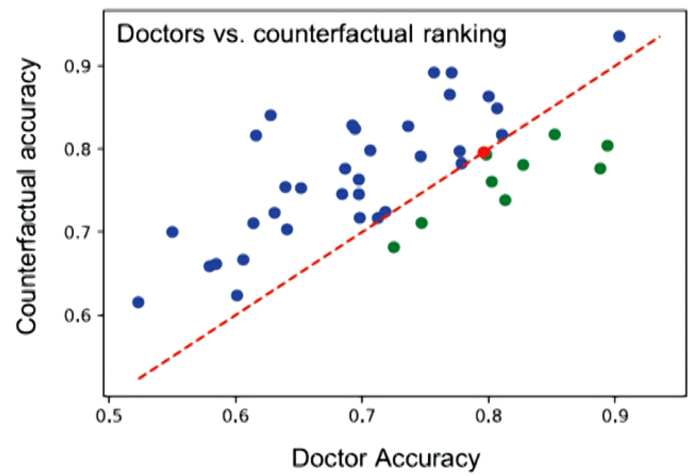
	Vignettes					
	All	VCommon	Common	Uncommon	Rare	VRare
N	1671	131	413	546	353	210
Mean (A)	3.81	2.85	2.71	3.72	4.35	5.45
Mean (C)	3.16	2.5	2.32	3.01	3.72	4.38
Wins (A)	31	2	7	9	9	4
Wins (C)	412	20	80	135	103	69
Draws	1228	131	326	402	241	137

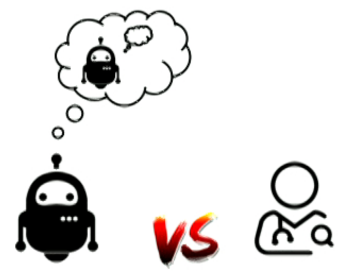
Compare to 44 doctors not involved in creating medical cases

Standard Bayesian updating places in top 48% of doctors, achieving average clinical accuracy



Counterfactual inference places in top 25% of doctors, achieving expert clinical accuracy





Made with

Accepted to:

1. ***Frontiers of AI-assisted Care symposium***
2. ***Causal Machine Learning workshop at NeurIPS 2019 (selected as Spotlight)***

arXiv: 1910.06772

Counterfactual diagnosis

Jonathan G. Richens,[1] Ciarán M. Lee,[1, 2] and Saurabh Johri[1]

[1] *Babylon Health, London, United Kingdom**

[2] *University College London, United Kingdom*

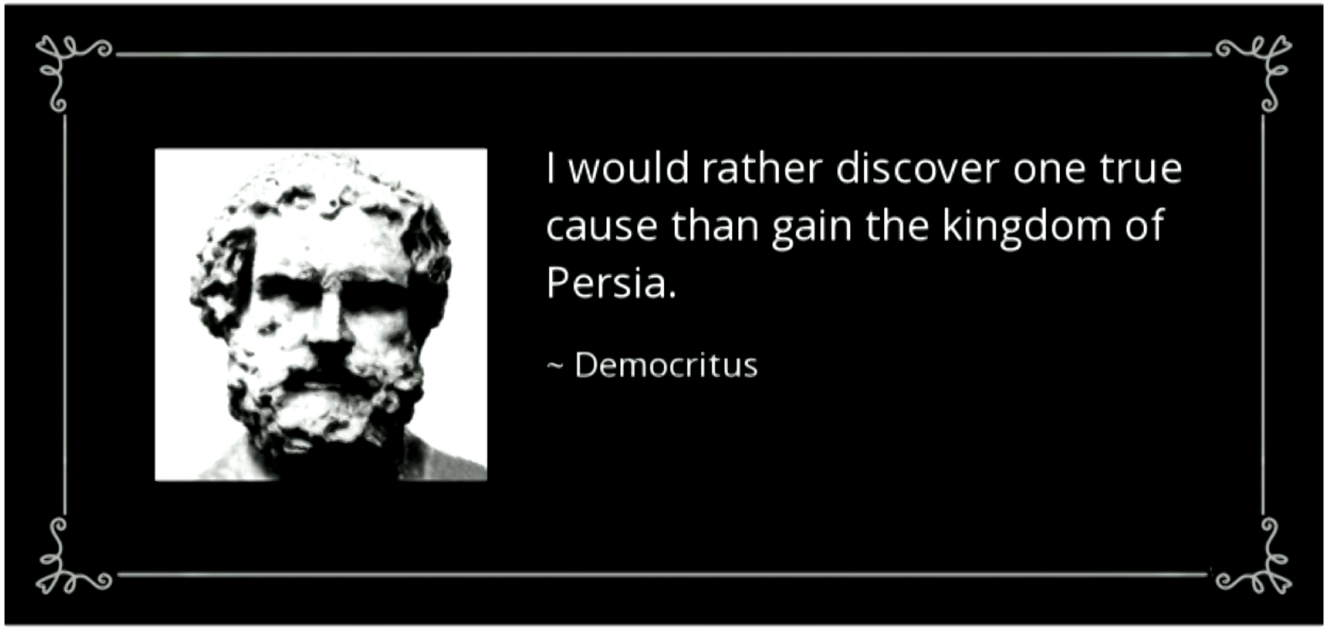
Causal knowledge is vital for effective reasoning in science and medicine. In medical diagnosis for example, a doctor aims to explain a patients symptoms by determining the diseases *causing* them. However, all previous approaches to Machine Learning assisted diagnosis, including Deep Learning and model-based Bayesian approaches, do not distinguish correlation from causation. Here, we propose a new diagnostic algorithm based on counterfactual inference which captures the causal aspect of diagnosis overlooked by previous approaches. Using a statistical disease model, which describes the relations between hundreds of diseases, symptoms and risk factors, we compare our counterfactual algorithm to the standard Bayesian diagnostic algorithm, and test these against a cohort of 44 doctors. We use 1763 medical cases created by a separate expert panel of doctors to benchmark performance. Each medical case provides a non-exhaustive list of symptoms and medical history simulating an instance of a single disease. The algorithms and doctors are tasked with

Made with

Knowing causal structure was crucial in the above.
How can we learn causal structure?

Made with

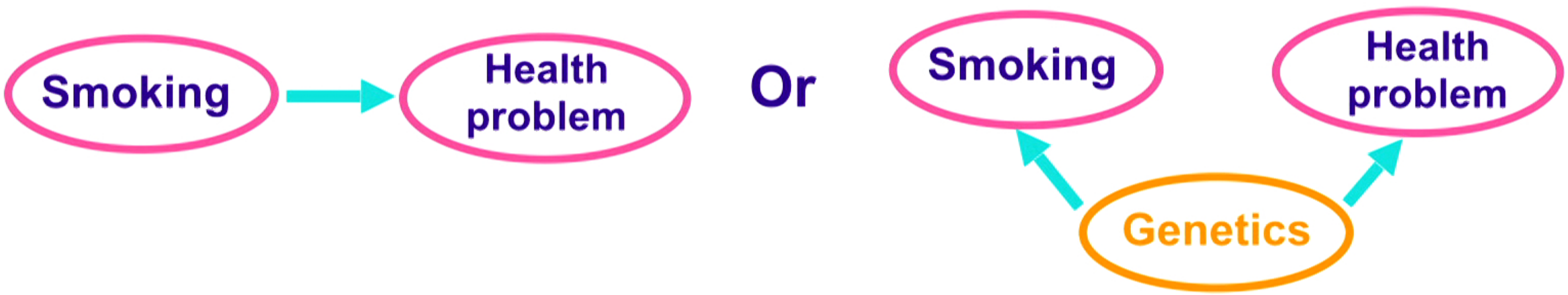
Learning causal relations between a set of variables is an incredibly important problem in science, medicine, economics



How do we learn causal relationships?

- Gold standard are randomised controlled trials (RCTs)

- Asking people if they smoke and have health problems isn't enough to conclude smoking causes these problems



- Need to force some to smoke and some not to, and look at relative number of health problems between groups

53

Made with

- Much of the time RCTs are **unethical**

- They can also be **expensive**–such as in drug trials–or technologically **unfeasable**–such as with astronomical bodies

- If we can't perform RCTs, what do we do?

Made with

- Causal Discovery algorithms provide an elegant approach to this problem

- They employ assumptions about what it means for one variable to cause another, aiming to capture the essence of the “asymmetry” between cause and effect

Made with

- Rough idea is that **causal** mechanism:

  P(effect | cause),

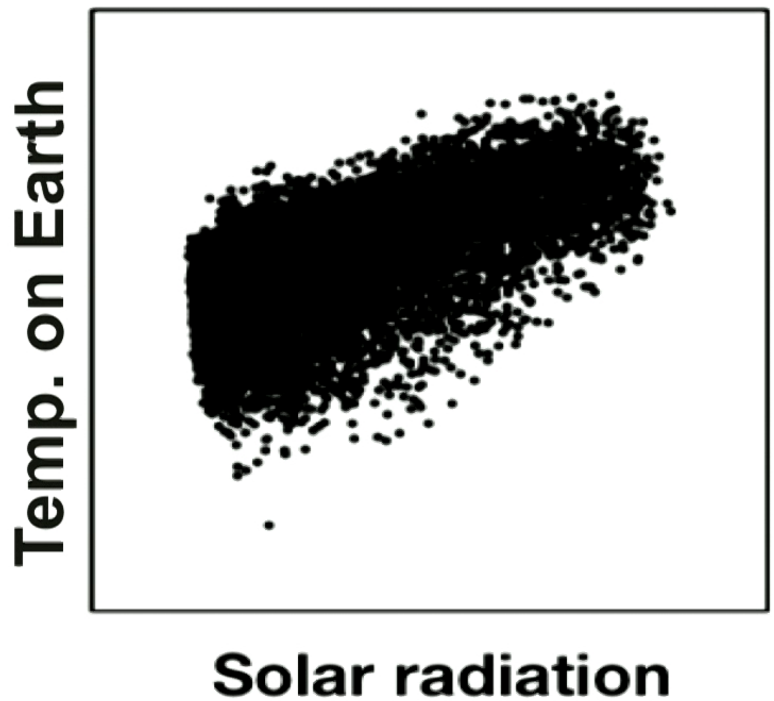
  is “simpler” to describe than the **acasual** one:

  P(cause | effect).

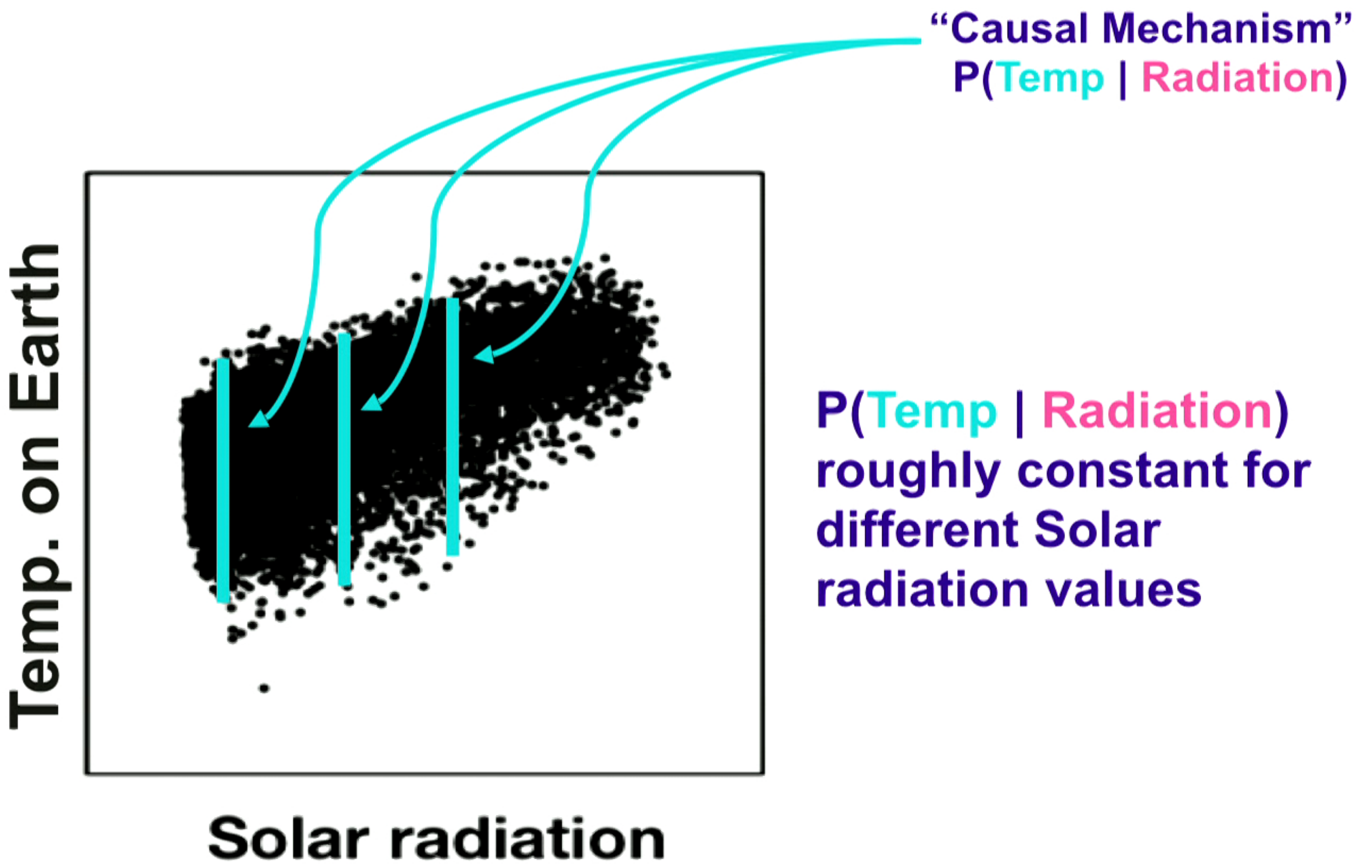
- “Easier to smash a cup than to un-smash it,”

- Let’s take a look at a simple example

Brief example

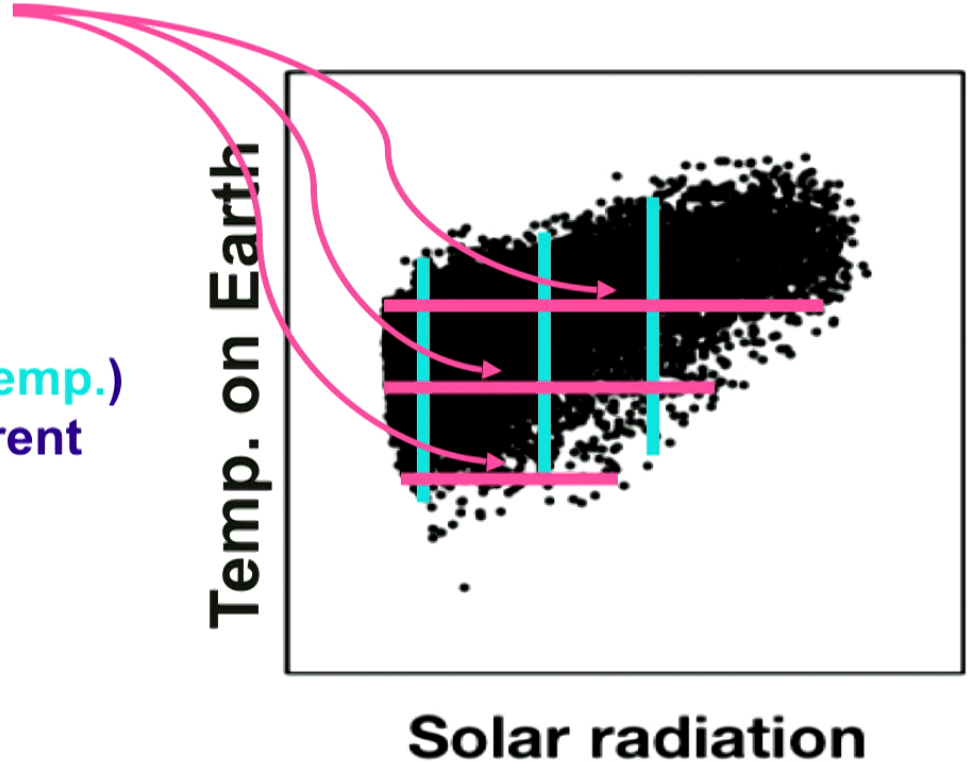


Made with



P(Temp | Radiation) roughly constant for different Solar radiation values

"Acausal mechanism"
P(Radiation | Temp.)



P(Radiation | Temp.) varies for different Temp. values

P(Temp | Radiation) "simpler" to describe than P(Radiation | Temp), so algorithm outputs:

Rad. → Temp.

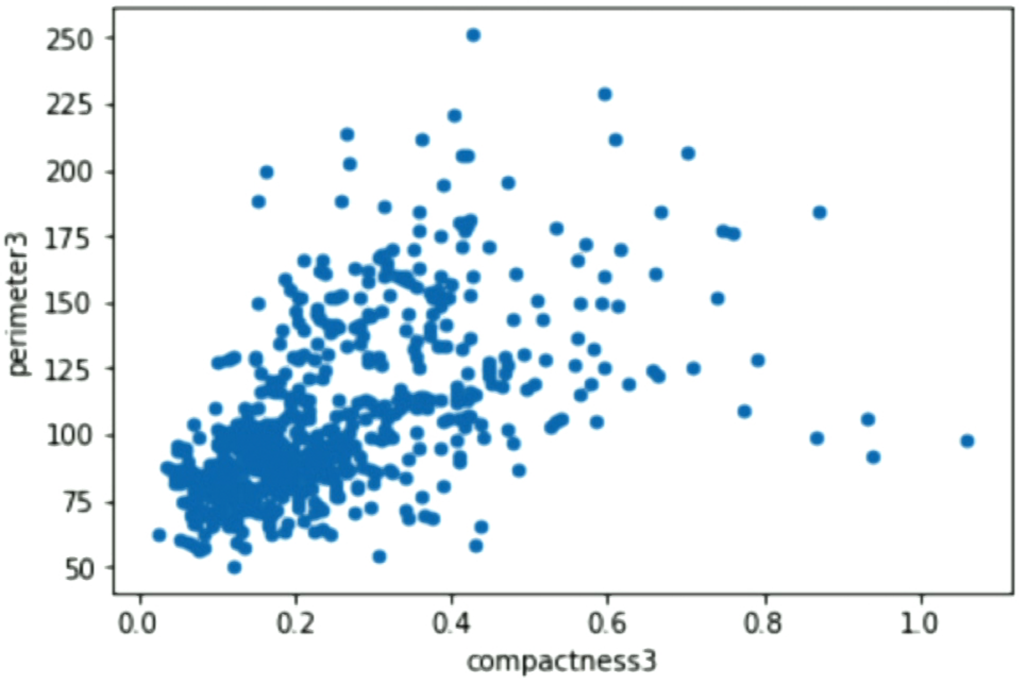
Different causal discovery algorithms quantify "simplicity" differently

(e.g. conditional kernel mean embeddings define norm over conditional distributions, variance quantifies simplicity)

61

Made with

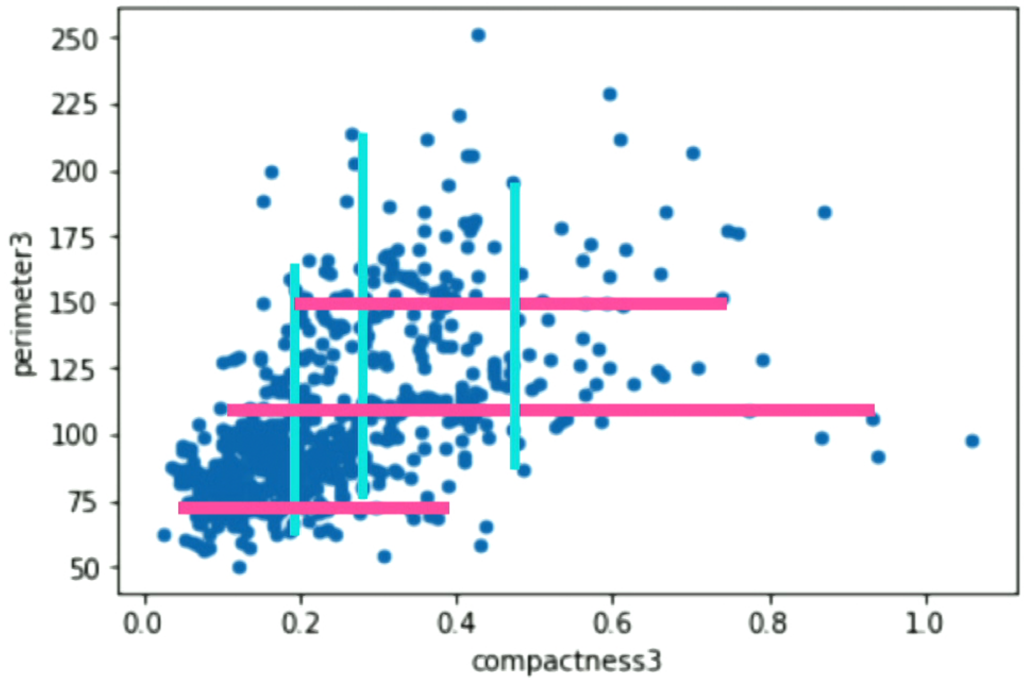
What’s the causal direction?



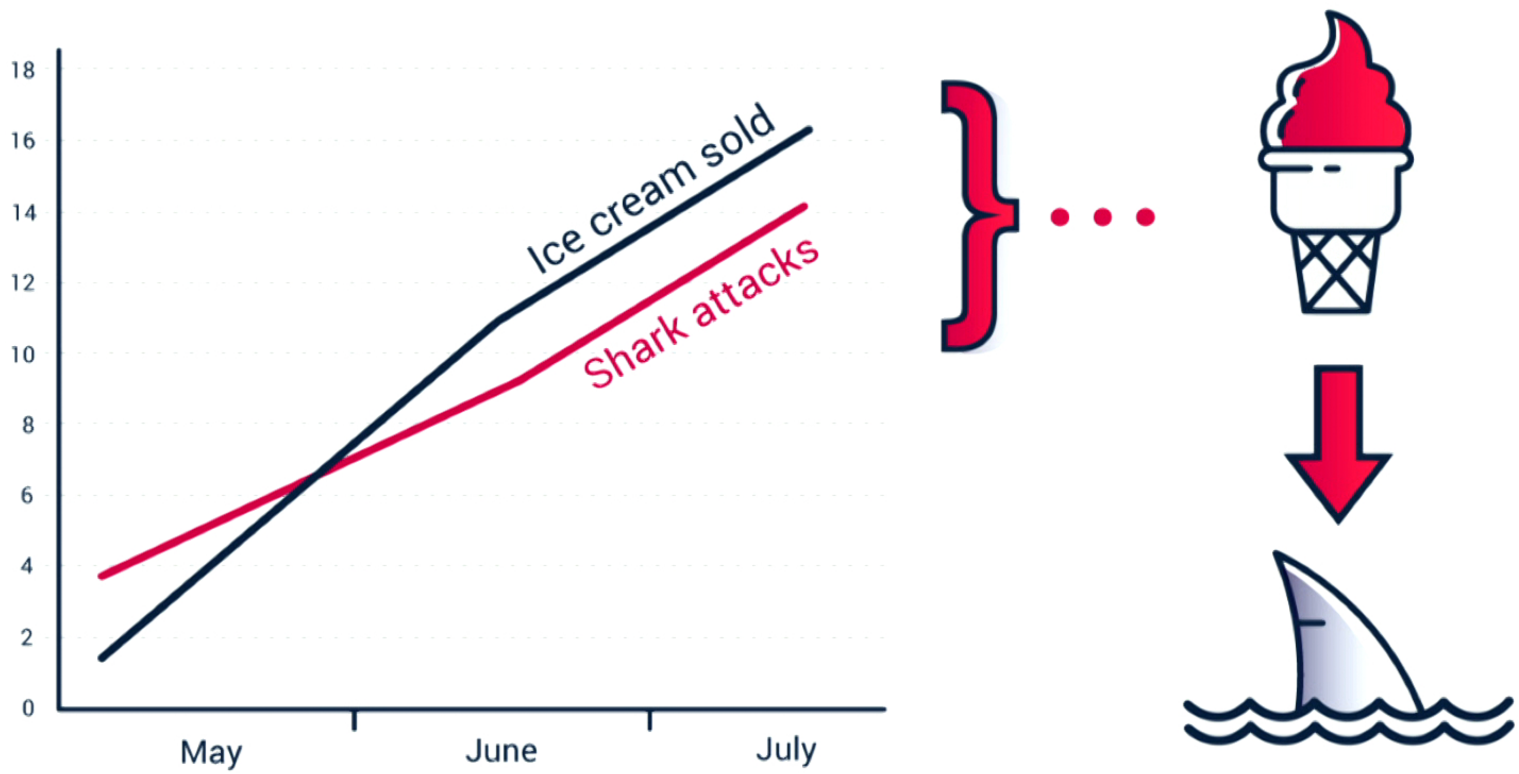
62

Made with

What's the causal direction?



Made with


18
16
14
12
10
8
6
4
2
0
May
June
July
Ice cream sold
Shark attacks

64

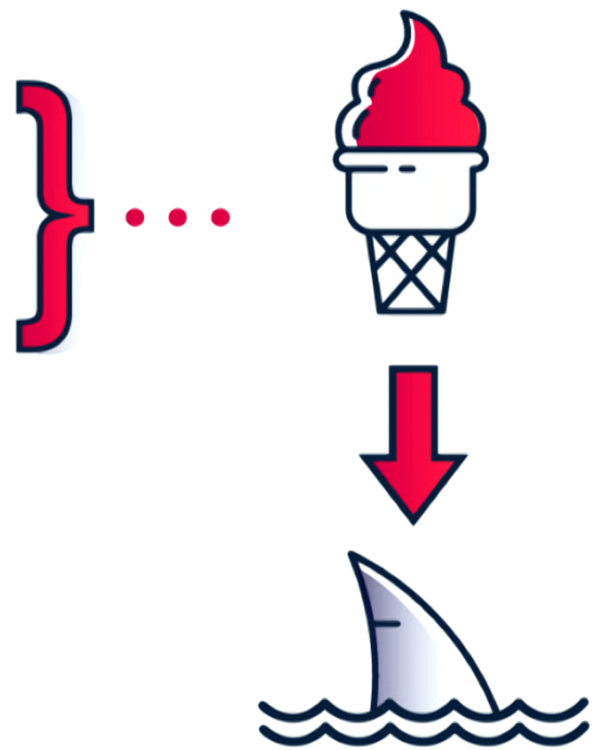
Made with

Soda Totally Turns Teens Into Killers

Anna North
10/25/11 6:00PM • Filed to: HEALTH

9.5K 25 Save

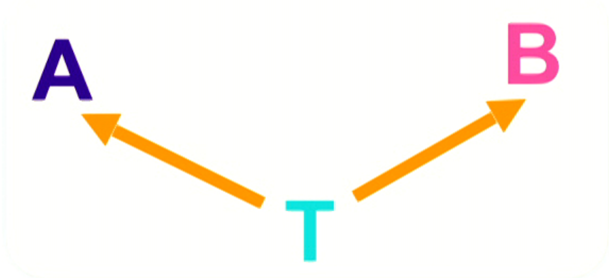
Turns out soda isn't just bad for your teeth: if you're a teenager, it could make you more likely to knife someone.



65

Made with

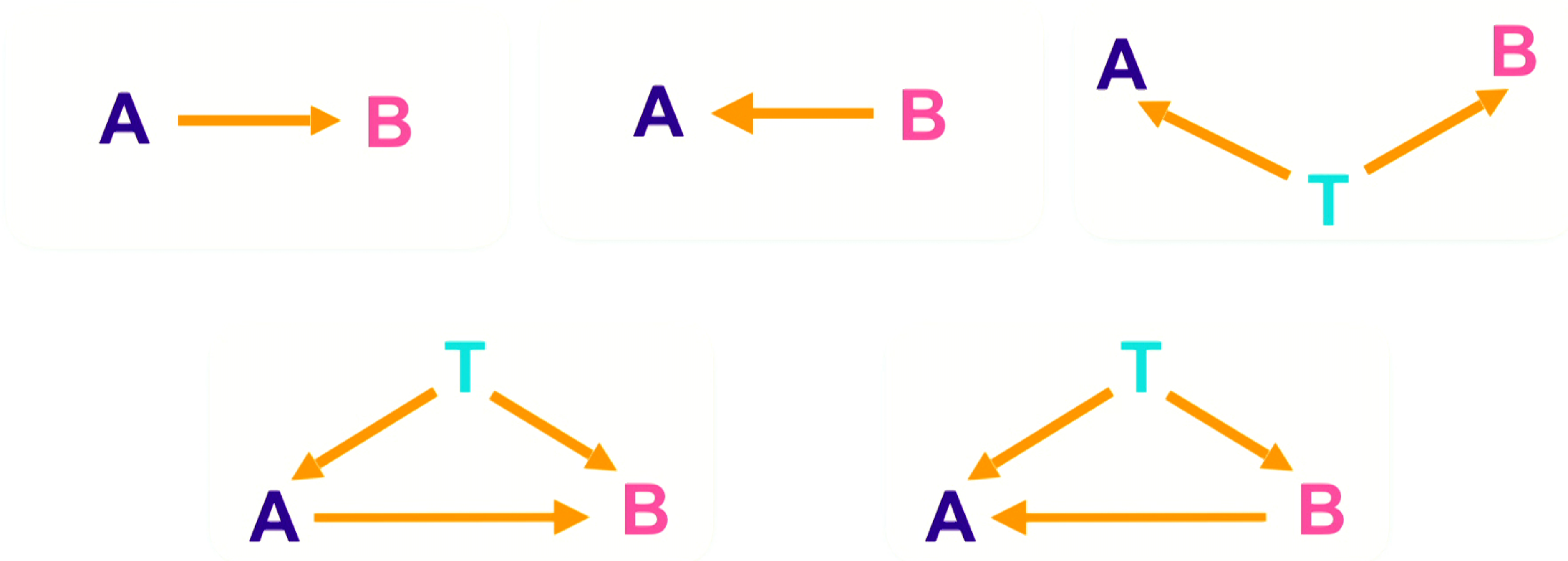
It turns out neither variable is a cause of the other



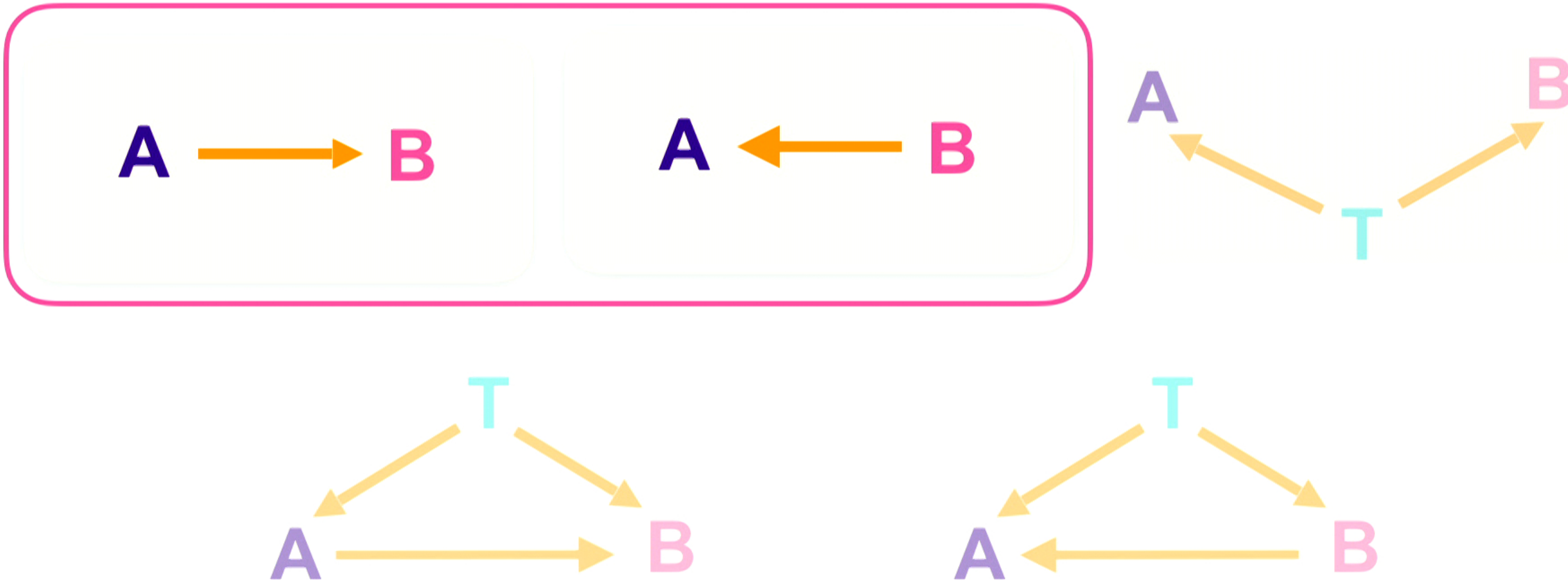
- A, B could be diseases and T a risk factor
- This causal relationship is of a different kind than the “purely directed” relations
- Treating one observed variable will not “cure” the other
- Discovering causal structure important for diagnosis and treatment

Made with

Actually there are 5 different causal structures between 2 correlated variables



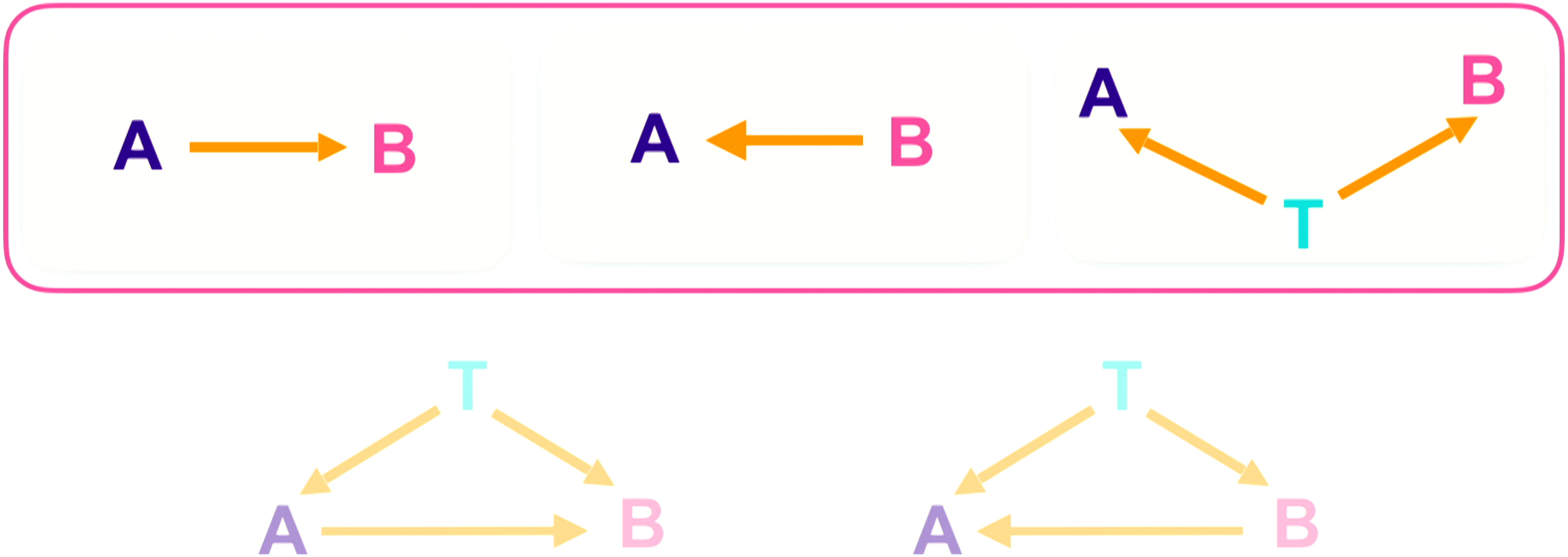
Actually there are 5 different causal structures between 2 correlated variables



68

Made with

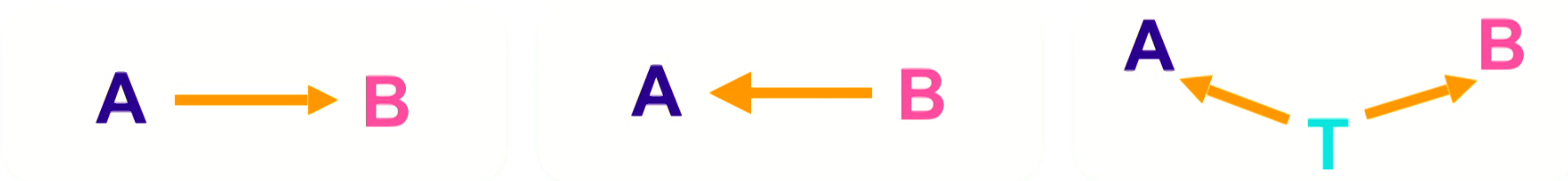
Actually there are 5 different causal structures between 2 correlated variables



Main Result

Given algorithm for distinguishing purely directed causal structures

method turns it into one that can distinguish



While maintaining original accuracy in distinguishing purely directed causal structures

Common Cause				
Exp.	Algorithm	Normal	Uniform	Expon.
1	modKCDC	96%	95%	97%
	modIGCI	99%	96%	99%
2	modKCDC	98%	95%	96%
	modIGCI	100%	100%	100%
	CAN	80%	66%	100%
3	modKCDC	94%	99%	95%
	modIGCI	98%	96%	97%
4	modKCDC	95%	96%	96%
	modIGCI	96%	96%	97%
5	modKCDC	97%	100%	95%
	modIGCI	95%	100%	94%
6	modKCDC	96%	95%	96%
	modIGCI	94%	96%	93%

Additive noise:

$$(1)\ A = \sin(10T) + e^{3T} + n_A$$
$$B = \log(T+10) + T^6 + n_B$$
$$(2)\ A = \log(T+10) + T^6 + n_A$$
$$B = T^2 + T^6 + n_B.$$

Multiplicative noise:

$$(3)\ A = \big(\sin(10T) + e^{3T}\big)e^{n_A}$$
$$B = \big(T^2 + T^6\big)e^{n_B}$$
$$(4)\ A = \big(\sin(10T) + e^{3T}\big)e^{n_A}$$
$$B = \big(\log(T+10) + T^6\big)e^{n_y}$$

Additive and Mulitplicative noise:

$$(5)\ A = \log(T+10) + T^6 + n_A$$
$$B = \big(T^2 + T^6\big)e^{n_B}$$
$$(6)\ A = \sin(10T) + e^{3T} + n_A$$
$$B = \big(T^2 + T^6\big)e^{n_B}$$

Common Cause				
Exp.	Algorithm	Normal	Uniform	Expon.
1	modKCDC	96%	95%	97%
	modIGCI	99%	96%	99%
2	modKCDC	98%	95%	96%
	modIGCI	100%	100%	100%
	CAN	80%	66%	100%
3	modKCDC	94%	99%	95%
	modIGCI	98%	96%	97%
4	modKCDC	95%	96%	96%
	modIGCI	96%	96%	97%
5	modKCDC	97%	100%	95%
	modIGCI	95%	100%	94%
6	modKCDC	96%	95%	96%
	modIGCI	94%	96%	93%

Additive noise:

$$(1)\ A = \sin(10T) + e^{3T} + n_A$$
$$B = \log(T+10) + T^6 + n_B$$
$$(2)\ A = \log(T+10) + T^6 + n_A$$
$$B = T^2 + T^6 + n_B.$$

Multiplicative noise:

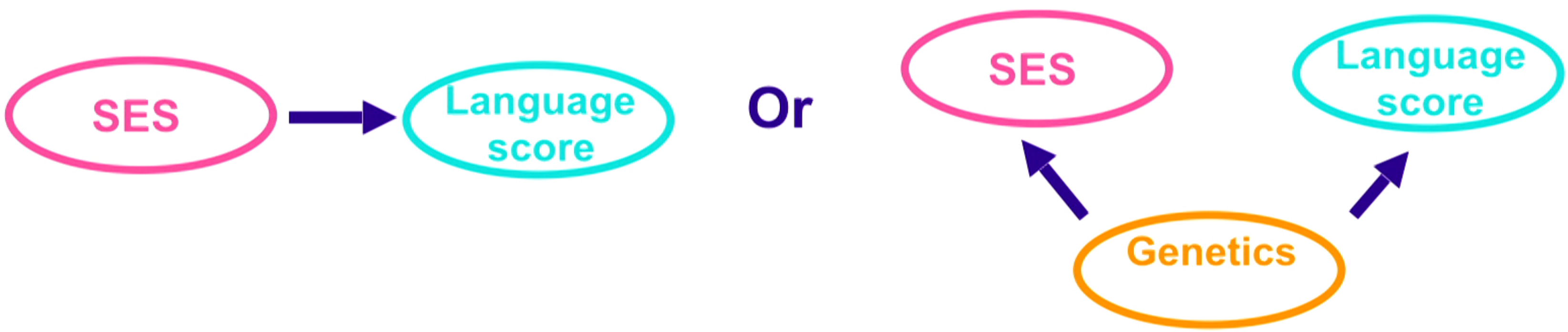
$$(3)\ A = \left(\sin(10T) + e^{3T}\right)e^{n_A}$$
$$B = \left(T^2 + T^6\right)e^{n_B}$$
$$(4)\ A = \left(\sin(10T) + e^{3T}\right)e^{n_A}$$
$$B = \left(\log(T+10) + T^6\right)e^{n_y}$$

Additive and Mulitplicative noise:

$$(5)\ A = \log(T+10) + T^6 + n_A$$
$$B = \left(T^2 + T^6\right)e^{n_B}$$
$$(6)\ A = \sin(10T) + e^{3T} + n_A$$
$$B = \left(T^2 + T^6\right)e^{n_B}$$

Data from 2287 eighth-grade pupils (aged about 11) in 132 classes in 131 schools in the Netherlands tested **language score** & **socioeconomic status (SES)**

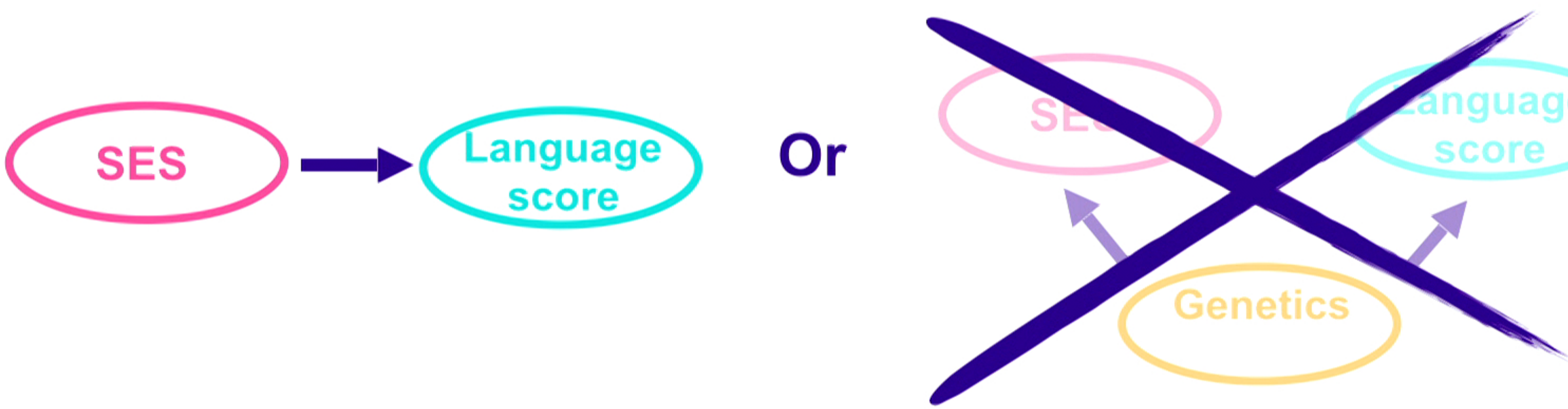
Our algorithms can be used to find the true causal structure



Made with

Data from 2287 eighth-grade pupils (aged about 11) in 132 classes in 131 schools in the Netherlands tested **language score** & **socioeconomic status (SES)**

Our algorithms can be used to find the true causal structure



73

Made with

Our work resulted in a new algorithm for detecting common causes

*We presented our work at **Frontiers of AI-assisted Care symposium**, which was held at Stanford in late 2019*

arXiv: 1910.10174

Leveraging directed causal discovery to detect latent common causes

Ciarán M. Lee*
Babylon Health &
University College London

Chris Hart
Babylon Health

Jonathan G. Richens
Babylon Health

Saurabh Johri
Babylon Health

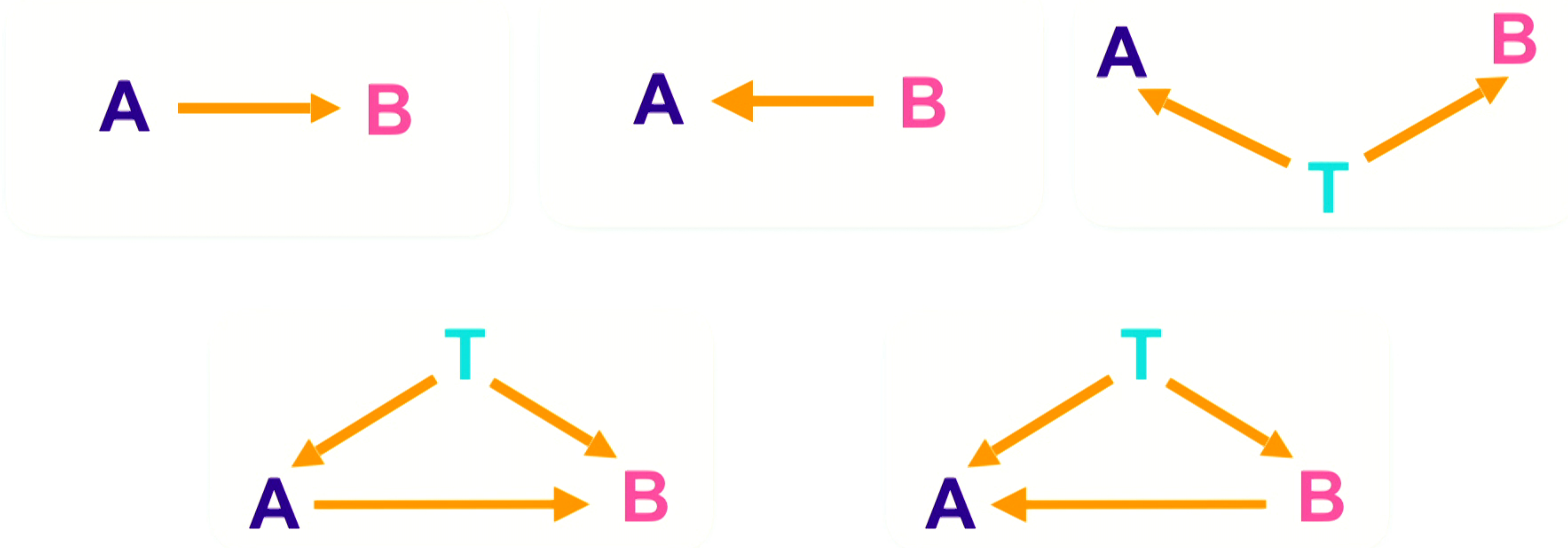
Abstract

Causal knowledge is crucial to our understanding of the world; it is a prerequisite to reasoning about the effects of interventions and ascertaining the truth of counterfactuals. As such, the discovery of causal relationships is a fundamental problem in science. In recent years, many elegant approaches to discovering causal relationships between two variables

al. 2009; Shimizu et al. 2006; Janzing et al. 2012b; Mitrovic, Sejdinovic, and Teh 2018; Louizos et al. 2017; Janzing et al. 2012a; Goudet et al. 2017; Zhang and Hyvärinen 2009; Fonollosa 2016; Lopez-Paz et al. 2015). However, most of these approaches deal only with purely directed causal relationships and cannot detect latent common causes. That is, given two variables A and B, these algo-

Made with

Some algorithms which can distinguish all 5, but these make strong assumptions about causal models



- Many medical studies only measure variables pertinent to the study, due to ethical reasons.
- This results in many datasets measuring overlapping but not exactly coinciding variables.
- Can we extract causal information from non-jointly measured variables?

Example: If we have a study that shows relation between vitamin D & obesity, and heart risk & obesity, can we learn if low vitamin D contributes to heart risk?

Dataset 1

X	Y
3	8
1	5
7	9
5	6

Overlap

Dataset 2

Y	Z
1	3
6	9
0.7	1
0.3	0.9

X and Z are never jointly measured

Overlap

Dataset 1

X	Y
3	8
1	5
7	9
5	6

Causal Discovery

$X \longrightarrow Y$

Dataset 2

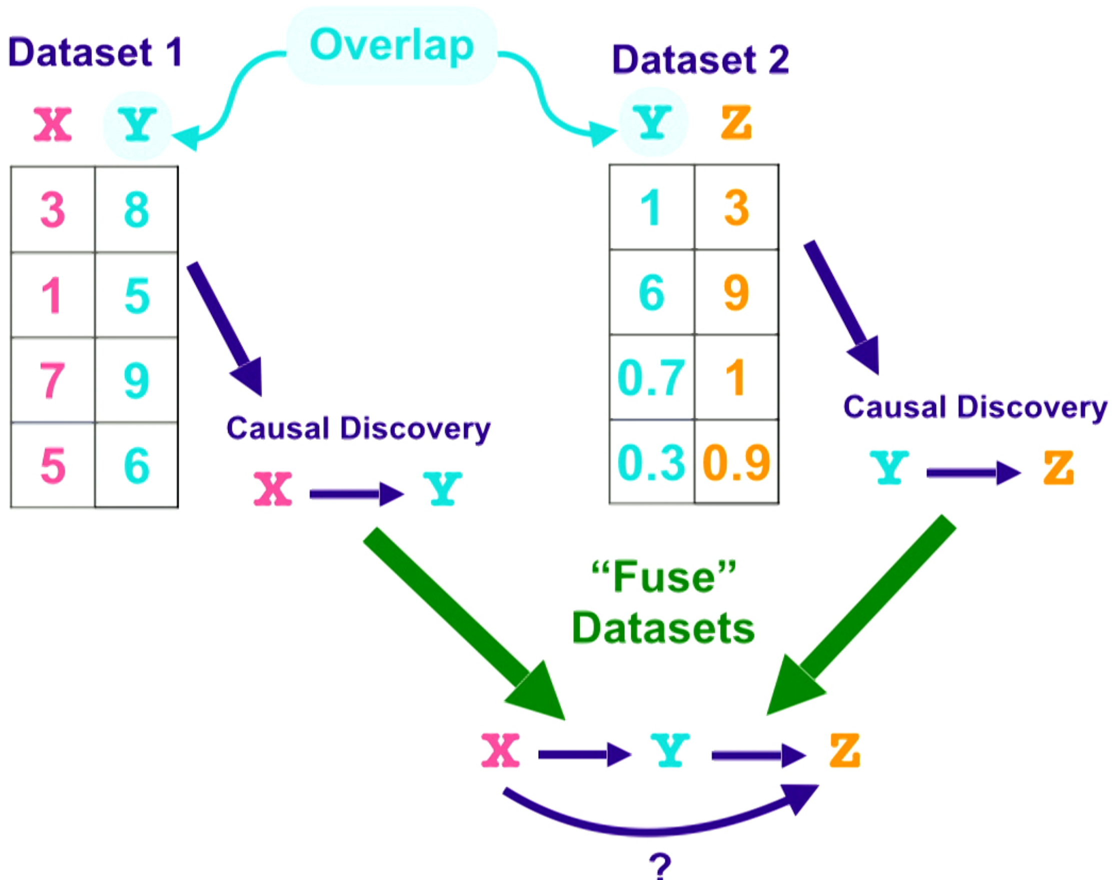
Y	Z
1	3
6	9
0.7	1
0.3	0.9

Causal Discovery

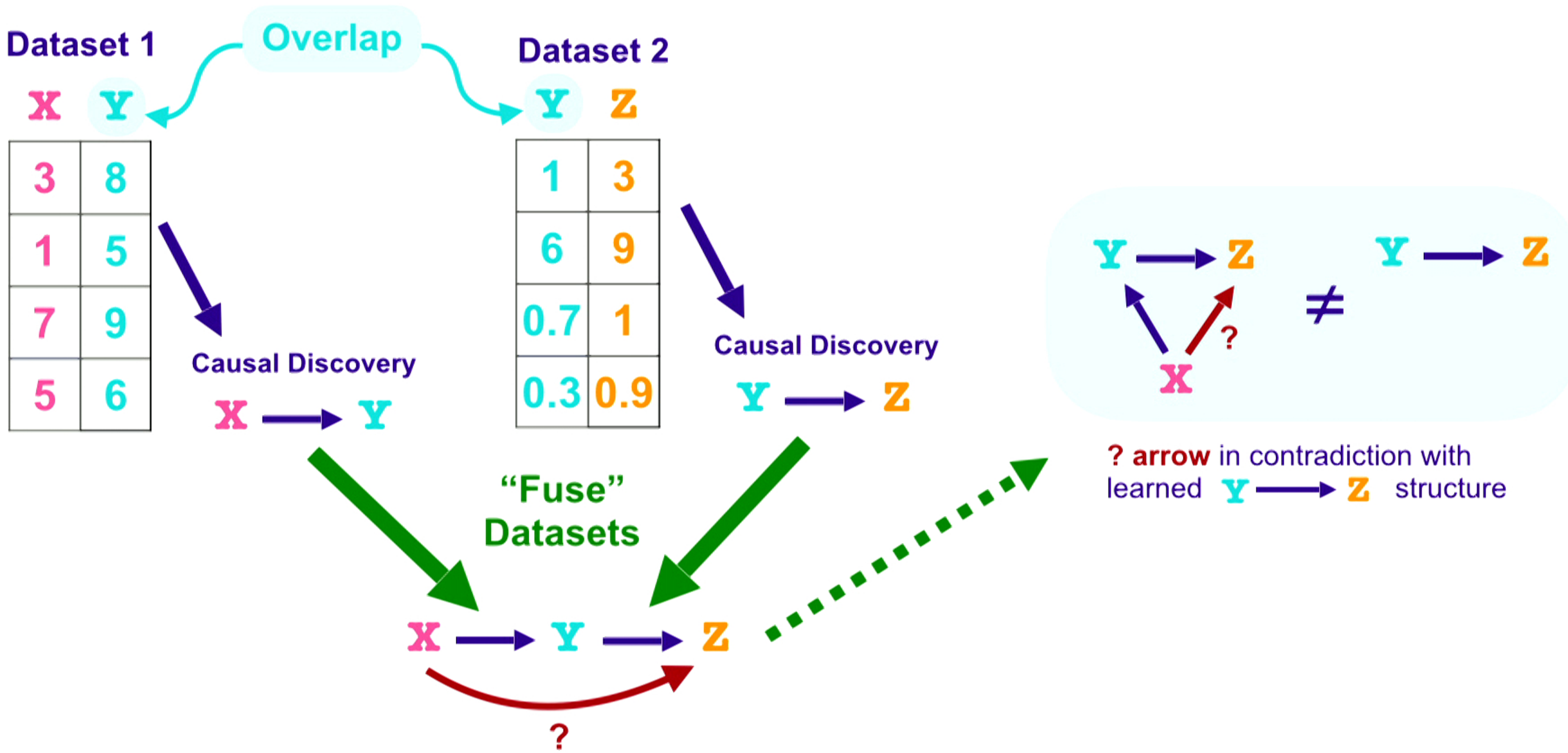
$Y \longrightarrow Z$

78

Made with


Dataset 1
Overlap
Dataset 2
X Y
3 8
1 5
7 9
5 6
Y Z
1 3
6 9
0.7 1
0.3 0.9
Causal Discovery
X → Y
Causal Discovery
Y → Z
“Fuse” Datasets
X → Y → Z
?

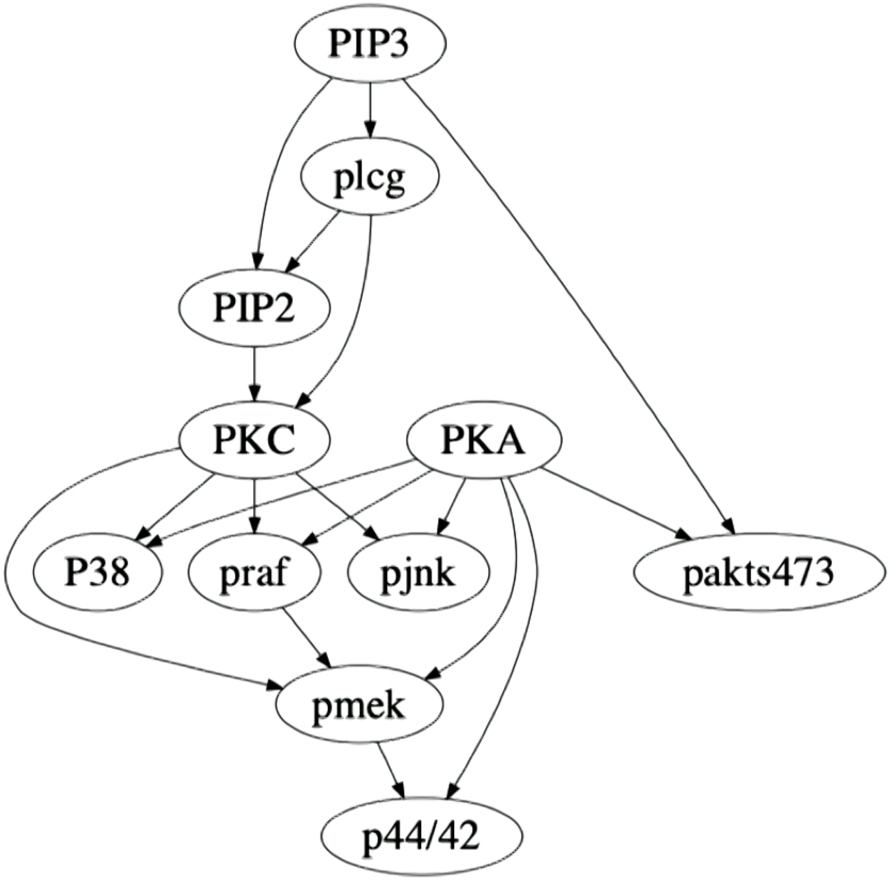
Made with


Dataset 1
Overlap
Dataset 2
X Y
3 8
1 5
7 9
5 6
Y Z
1 3
6 9
0.7 1
0.3 0.9
Causal Discovery
X → Y
Causal Discovery
Y → Z
"Fuse" Datasets
X → Y → Z
?
Y → Z
X
?
≠
Y → Z
? arrow in contradiction with learned Y → Z structure

MIT Technology Review

An algorithm that can spot cause and effect could supercharge medical AI

The technique, inspired by quantum cryptography, would allow large medical databases to be tapped for causal links



Split into two and gave to algorithms:

Previous SOTA: 61,740
Fraction correct edges: 40%

Our algorithm: 3
Fraction correct edges: 50%

How efficient is counterfactual inference?

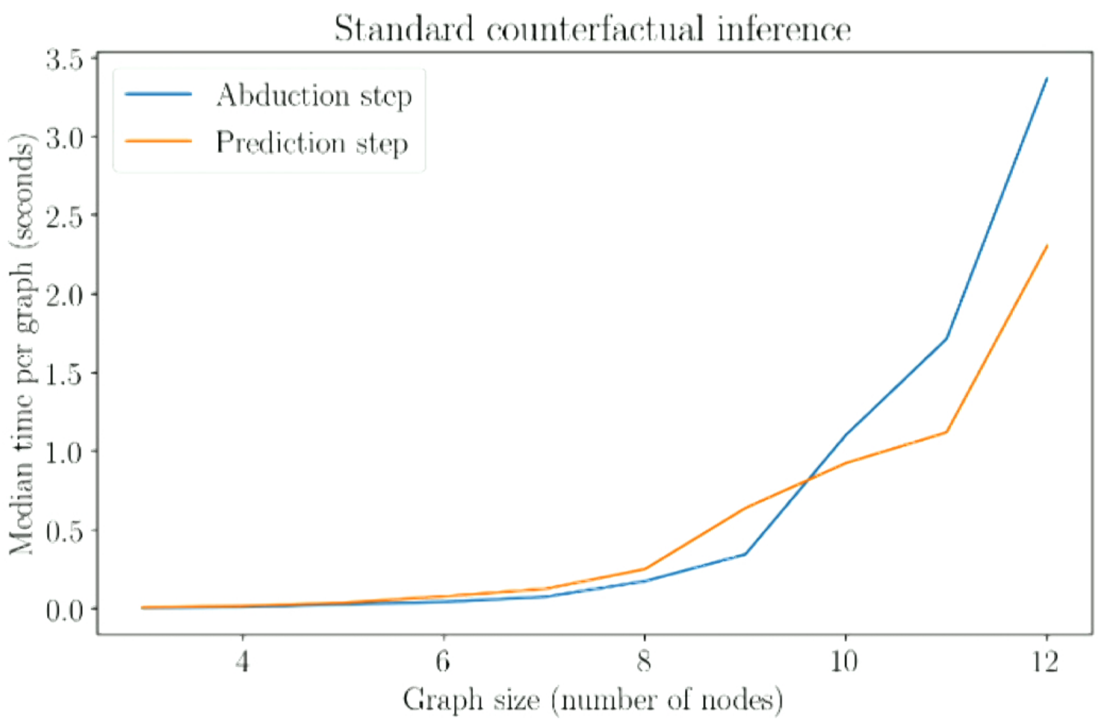
86

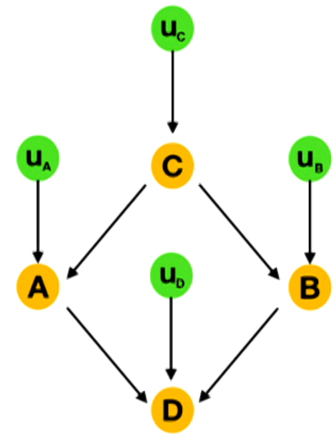
Made with

How efficient is counterfactual inference?

Made with

How efficient is counterfactual inference?

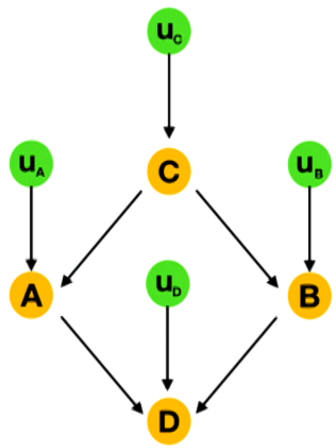


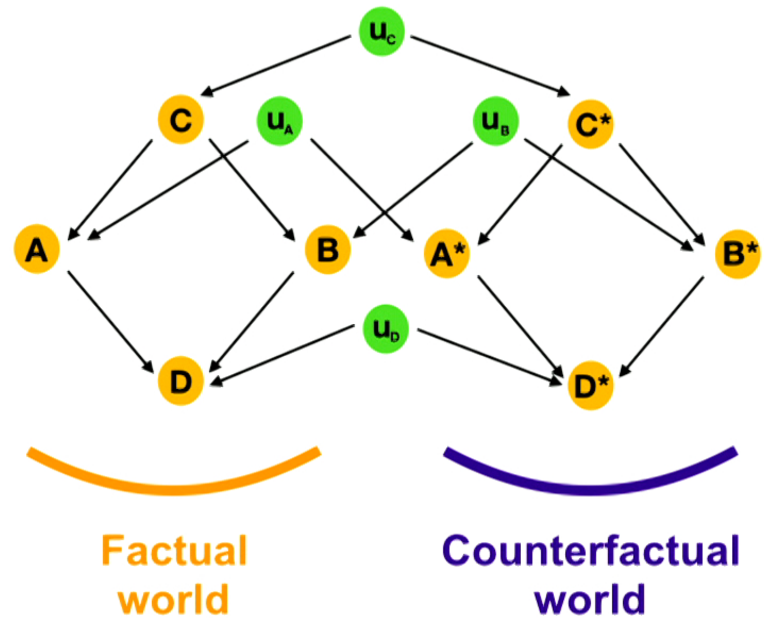


Abduction: update $P(u_A, u_B, u_C, u_D)$ given evidence

Made with

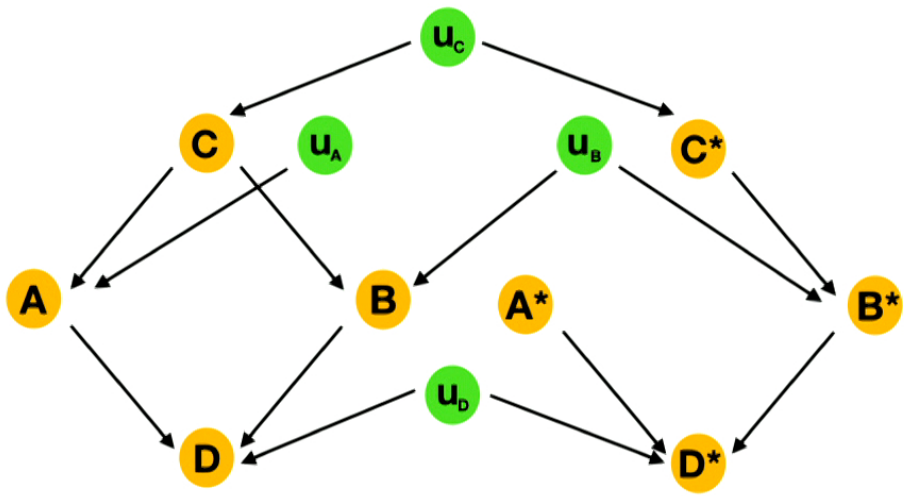
Efficient counterfactual inference with Twin Networks





Made with

Efficient counterfactual inference with Twin Networks



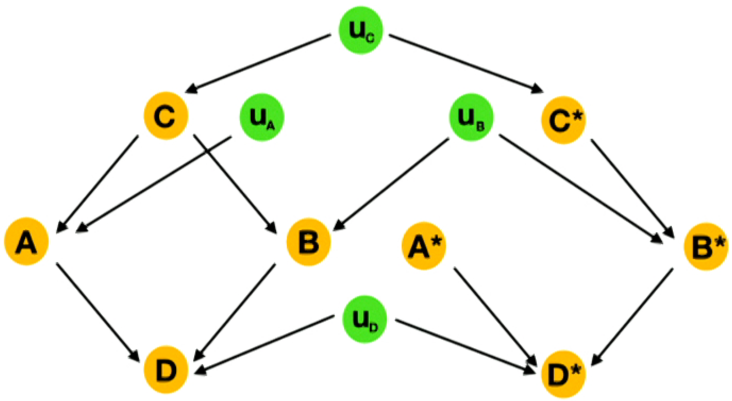
Compute counterfactual

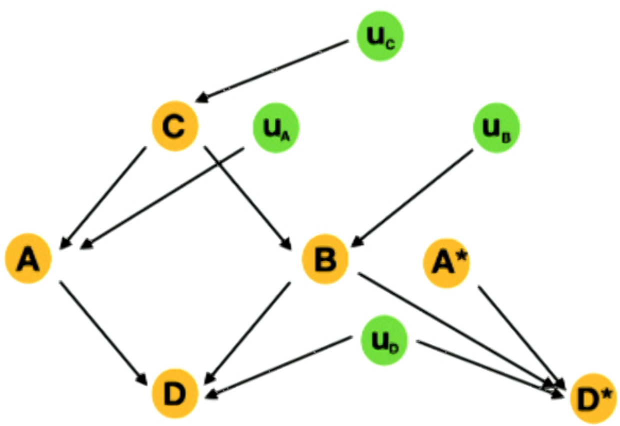
Standard: P(D | D=T, do(A=F))
1. Abduction
2. Action
3. Prediction

Twin: P(D* | D=T, A*=F)
Bayesian Inference on **Twin network**

Made with

Optimising Twin Networks





Node merging

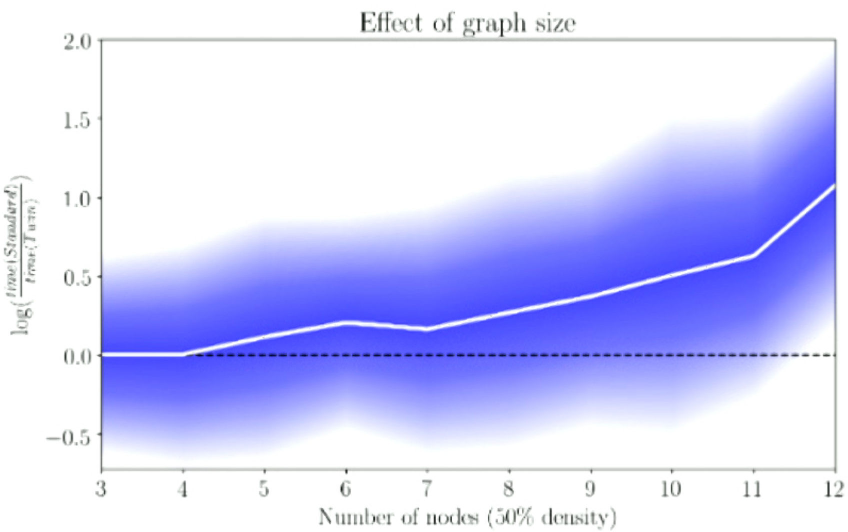
Not a priori clear that standard Bayesian inference on network twice the size of original is more efficient

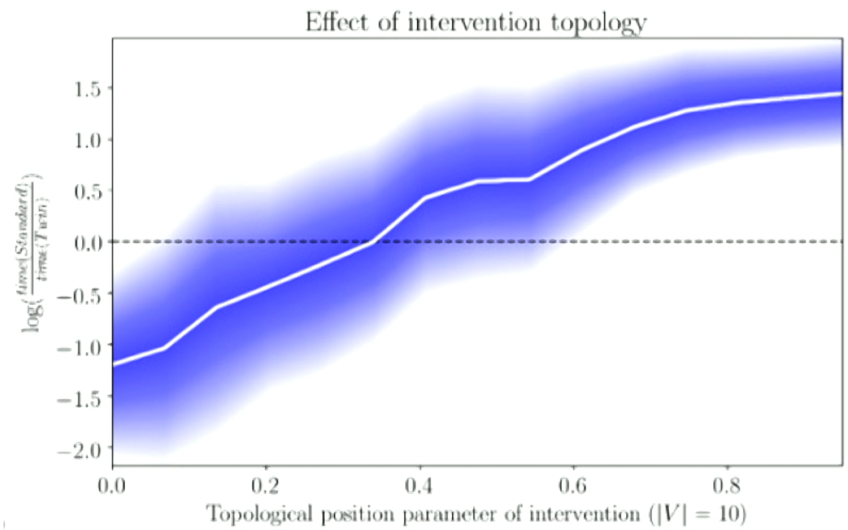
Nodes in counterfactual network that are not descendants of an intervention are exact copies of the original node

9(

Made with

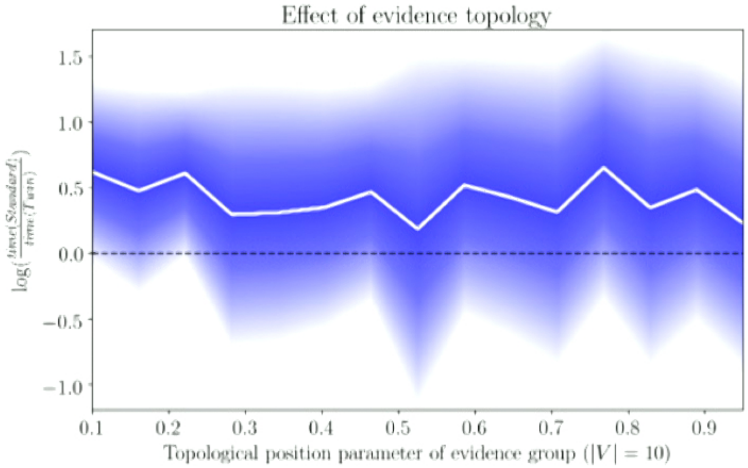
Computational advantage of Twin Networks

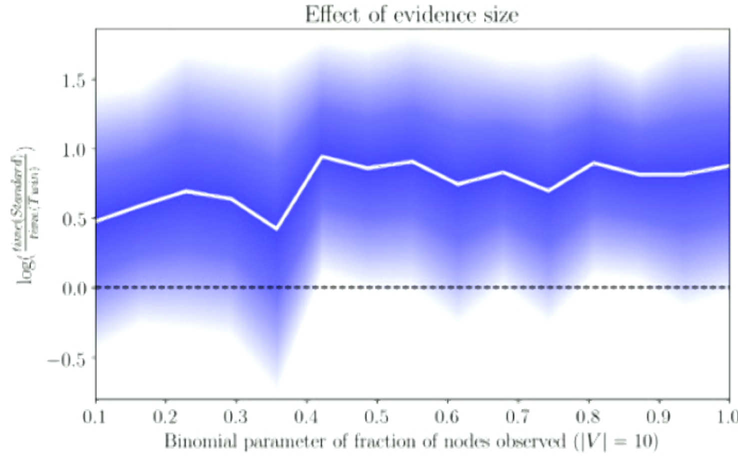


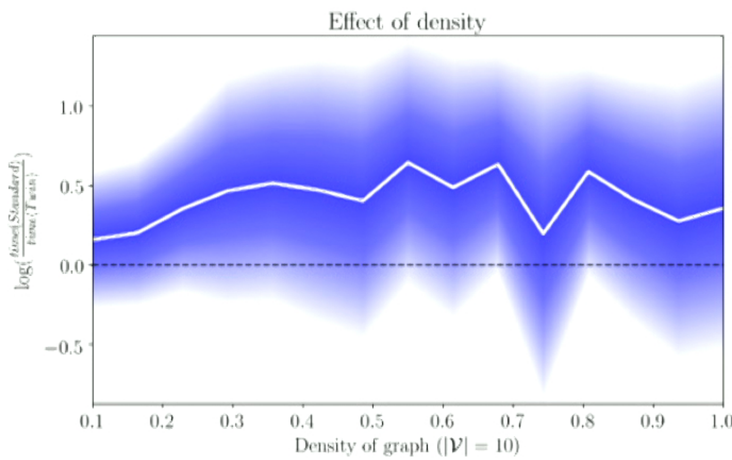


Made with

No sustained improvement in some cases



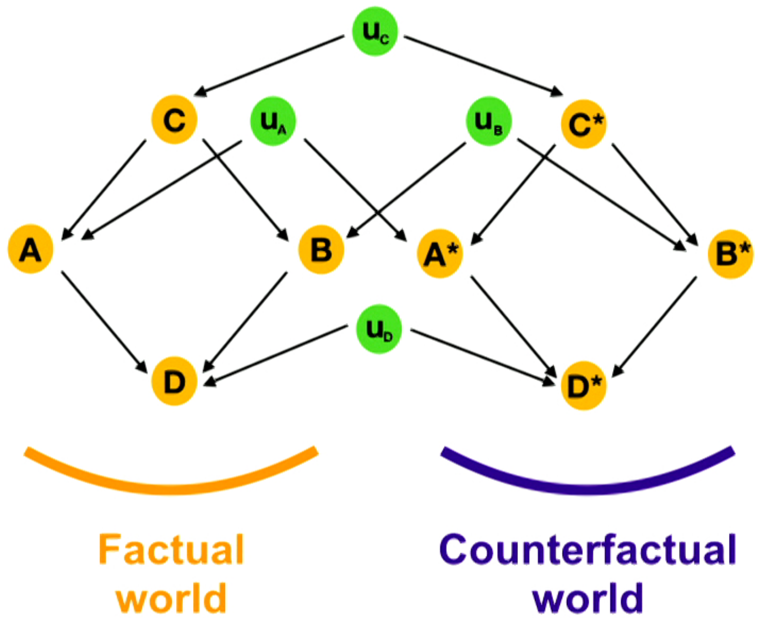


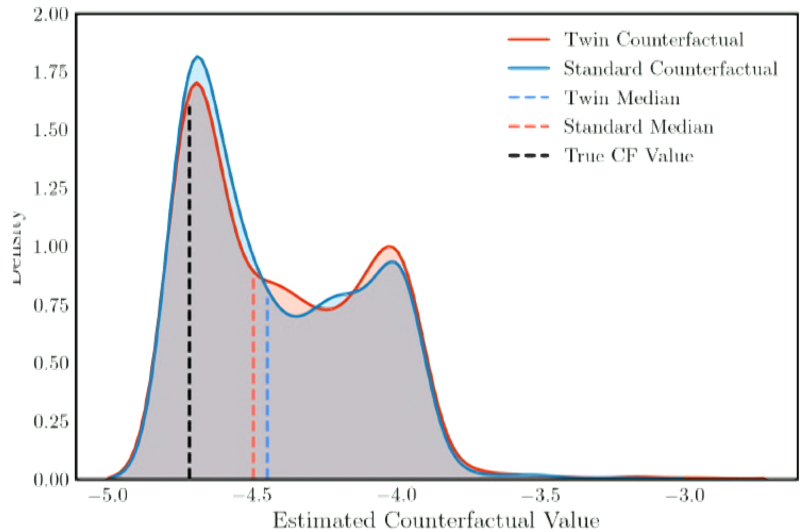


93

Made with

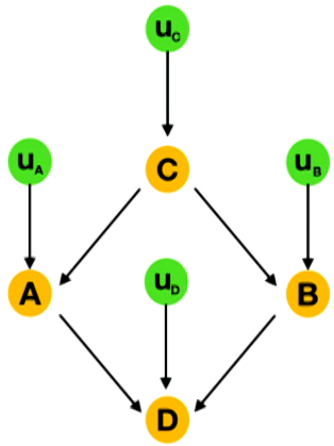
No massive improvement in approximate inference cases

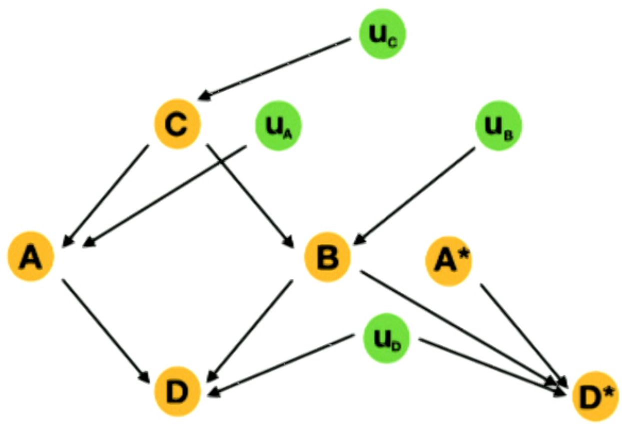


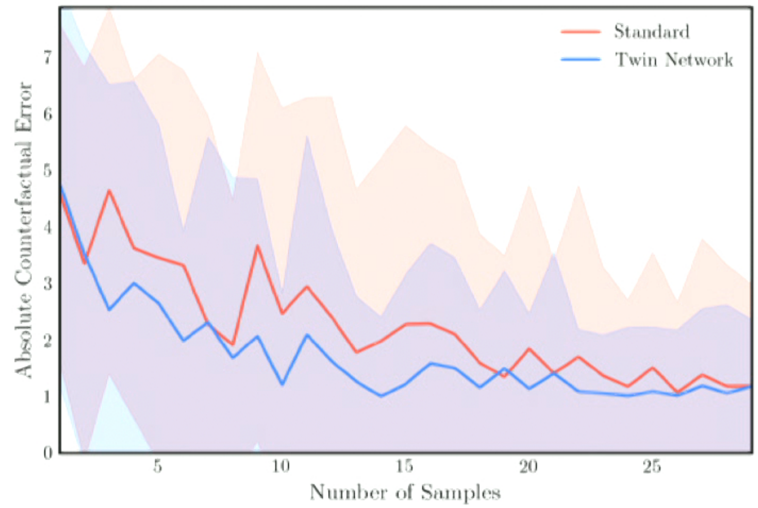


Made with

Slight improvement with merging

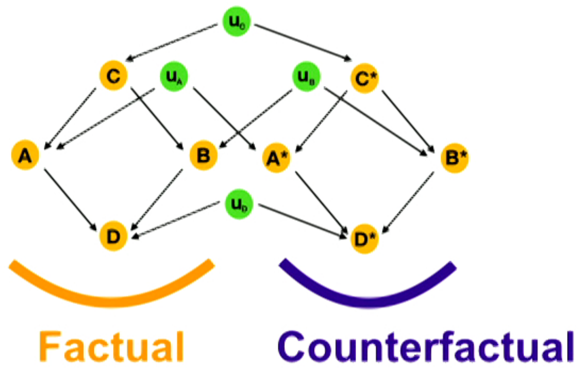


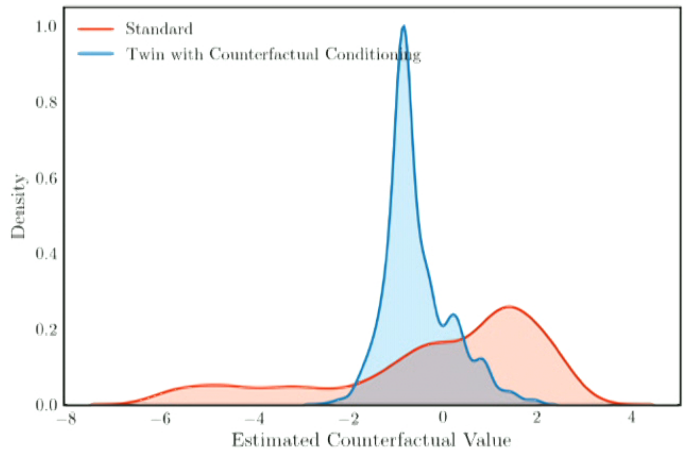


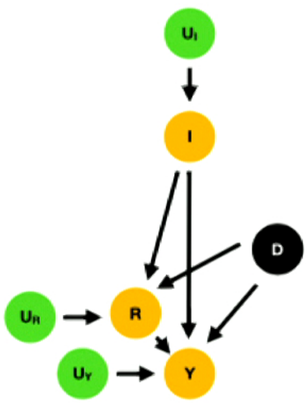


Made with

Improvement when conditioning in counterfactual world

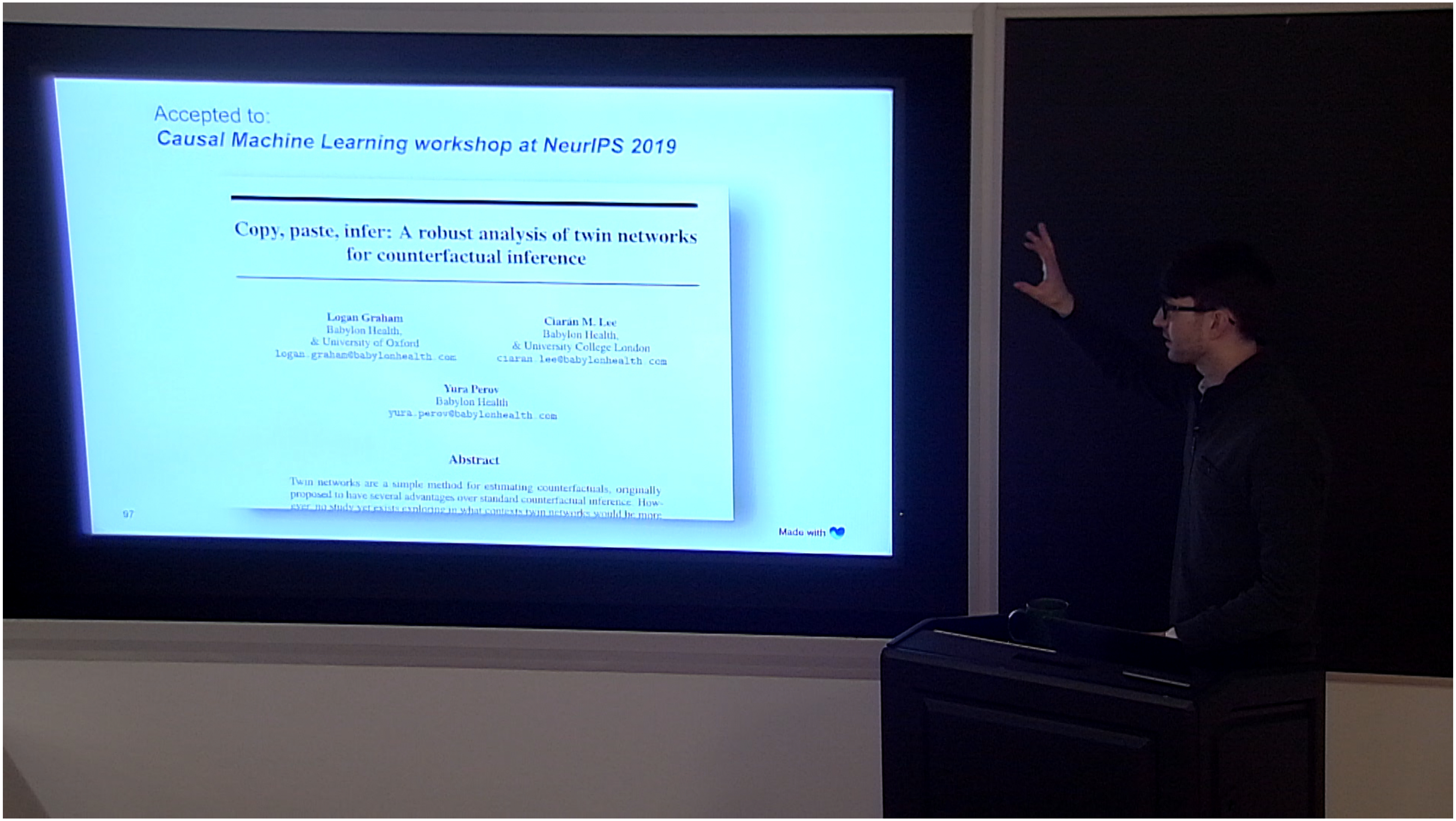






"What would happen to the patient's inflammation I* if I gave the patient drug D* and they had a large response to it"

Made with


Accepted to:
Causal Machine Learning workshop at NeurIPS 2019
Copy, paste, infer: A robust analysis of twin networks for counterfactual inference
Logan Graham
Babylon Health,
& University of Oxford
logan.graham@babylonhealth.com
Ciarán M. Lee
Babylon Health,
& University College London
ciaran.lee@babylonhealth.com
Yura Perov
Babylon Health
yura.perov@babylonhealth.com
Abstract
Twin networks are a simple method for estimating counterfactuals, originally
proposed to have several advantages over standard counterfactual inference. How-
97
Made with

arXiv: 1910.08091 & Accepted to:

1. ***Advances in Approximate Bayesian Inference 2019***
2. ***ProbProg, Probabilistic Programming Conference 2020***

MultiVerse: Causal Reasoning using Importance Sampling in Probabilistic Programming

Yura Perov[*, †] YURA.PEROV@BABYLONHEALTH.COM
Logan Graham[*, †, ‡] LOGAN.GRAHAM@BABYLONHEALTH.COM
Kostis Gourgoulias[†] KOSTIS.GOURGOULIAS@BABYLONHEALTH.COM
Jonathan G. Richens[†] JONATHAN.RICHENS@BABYLONHEALTH.COM
Ciarán M. Lee[†, §] CIARAN.LEE@BABYLONHEALTH.COM
Adam Baker[†] ADAM.BAKER@BABYLONHEALTH.COM
Saurabh Johri[†] SAURABH.JOHRI@BABYLONHEALTH.COM

Abstract

We elaborate on using importance sampling for causal reasoning, in particular for counterfactual inference. We show how this can be implemented natively in probabilistic programming. By considering the structure of the counterfactual query, one can significantly optimise the inference process. We also consider design choices to enable further optimisations. We introduce MultiVerse, a probabilistic programming prototype engine for approximate causal reasoning. We provide experimental results and compare with Pyro, an existing probabilistic programming framework with some of causal reasoning tools.

Conclusions and next steps

- Developed new causal learning algorithms
- **Next steps**: scale them up & start mining data for causal relations
- Showed that counterfactual reasoning is useful in healthcare & devised efficient ways of performing it
- **Next steps**: start using this for healthcare simulations

babylon

Thank you
for your attention!