Title: Colloquium - Causal and counterfactual inference and what they're good for

Speakers: CiarÃ¡n Lee

Series: Quantum Foundations, Quantum Information

Date: September 18, 2024 - 2:00 PM

URL: https://pirsa.org/24090092

Abstract: Causal reasoning is vital for effective reasoning in many domains, from healthcare to economics. 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 and to answer counterfactual queries. In this talk I will present some recent work done with my collaborators about how one can learn and reason with counterfactual distributions, and why this is importantly for decision making. In all cases I will strive to motivate and contextualise the results with real word examples.

Some applications of Causal Inference in the real world

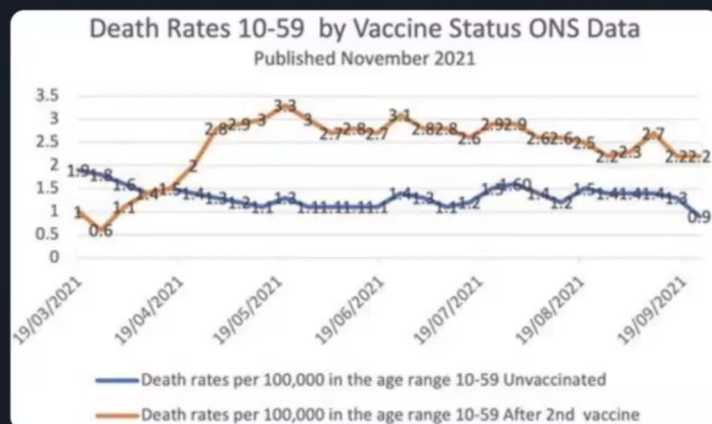
Ciarán Gilligan-Lee

Spotify & University College London

Dr Anthony Hinton
@TonyHinton2016

Thanks to @abirballan
orange line= vaccinated 10-59 y old
blue line= unvaccinated 10-59 y old
Orange line is ABOVE blue line
=> MORE deaths in vaccinated than unvaccinated?

Is the right policy to vaccinate low-risk populations?
ANY explanation @CMO_England @sajidjavid @ONS ???



08:58 · 20/11/2021 · Twitter for iPad

1,109 Retweets **148** Quote Tweets **1,706** Likes

Motivating example

- The findings were very weird indeed, flying in the face of medical knowledge and confounding experts

- Yet the finding was irrefutable: death rates for vaccinated people are higher than for unvaccinated people

Do vaccines work?

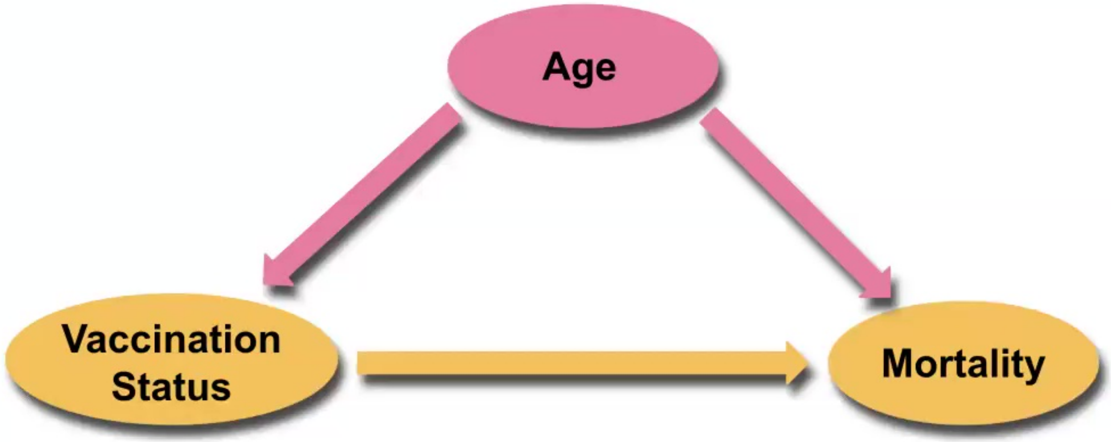
Let’s look at a picture

The average unvaccinated person is much younger than the average vaccinated person.

Therefore they have a lower mortality rate. Any benefit from the vaccines is swamped by the increase in mortality with age!

Age is a **confounder** between Vaccine Status and Mortality

But when we control for age, vaccinations are shown to reduce mortality rate.



Motivating example

- Any action or policy change based on these correlations—such as whether to vaccinate—would not increase patient survival.

Take home: Relying on correlations extracted from observational data can lead to embarrassing, costly, and dangerous mistakes.

- To overcome this, we need to understand cause and effect

Why is this important for Spotify?

Why is this important for Spotify?

- Usually randomised controlled trials or A/B tests tell us about cause & effect.

- But sometimes A/B tests can't be performed. They could be too damaging to user experience, or technically too hard to implement:

  - "Do app crashes cause churn?"
  - "Does podcast or audiobook consumption cause retention?"

- **Causal Inference** provides a set of methods and tools for learning and quantifying cause and effect, even without A/B tests – given some assumptions. Understanding causality is vital for actionable decision-making.

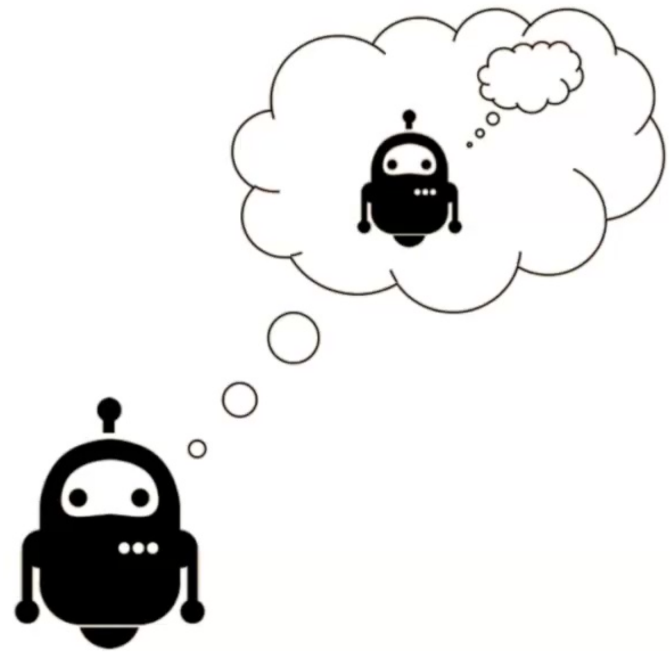
What types of causal questions can we answer?

CAUSAL HIERARCHY - PEARL'S HIERARCHY

01 Seeing - Observations

02 Acting - Interventions

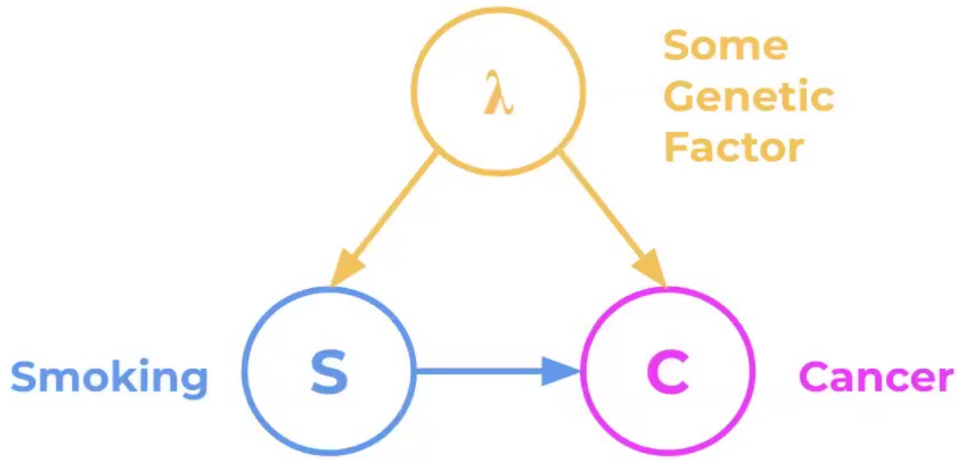
03 Imagining - Counterfactuals



CAUSAL HIERARCHY

01

Seeing - Observations



Probability of Cancer

Smoking has been observed

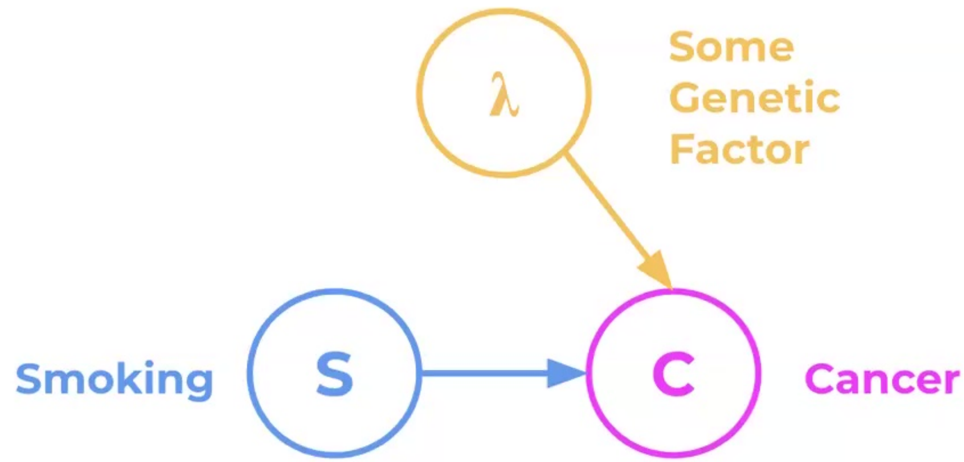
P(C=T | S=T)

given

CAUSAL HIERARCHY

02 Acting - Interventions

Simulate randomised controlled trial



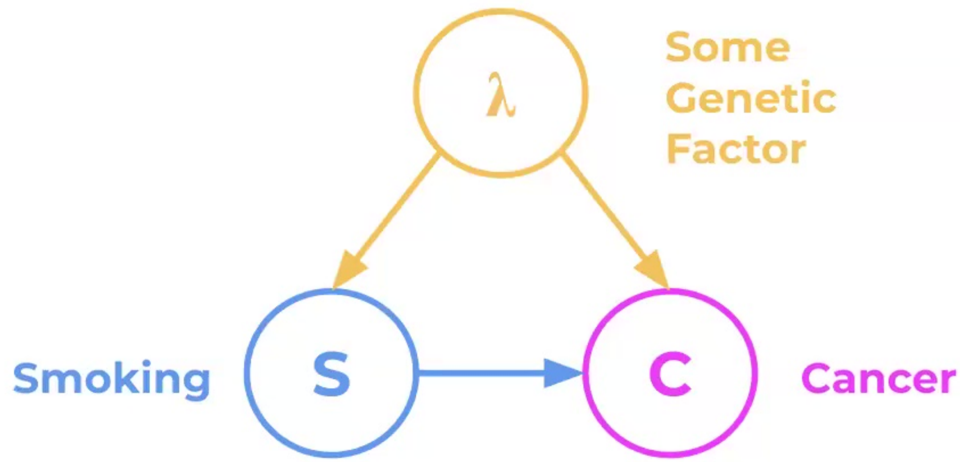
Probability of Cancer

Subject was MADE to Smoke

$$P(\,C_{S=T} = T)$$

CAUSAL HIERARCHY

03 Imagining - Counterfactuals



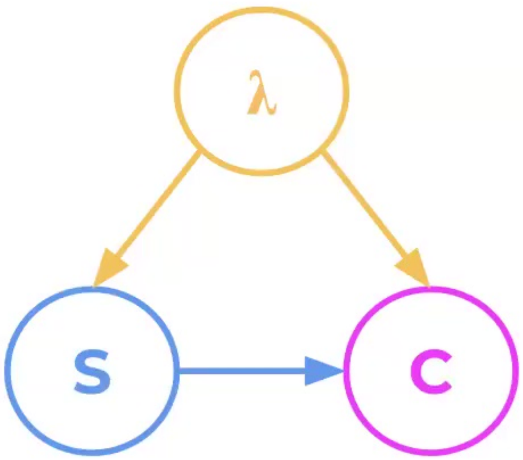
Counterfactuals ask **What If** questions

"Given subject is a smoker and has cancer, what is the chance they wouldn't if they didn't smoke

Probability of No Cancer	Subject has Cancer	Subject was MADE not to Smoke

$$P(\, C_{S=F} = F \mid C=T, S=T)$$

HOW TO ANWER COUNTERFACTUALS?



Counterfactual Inference compute:

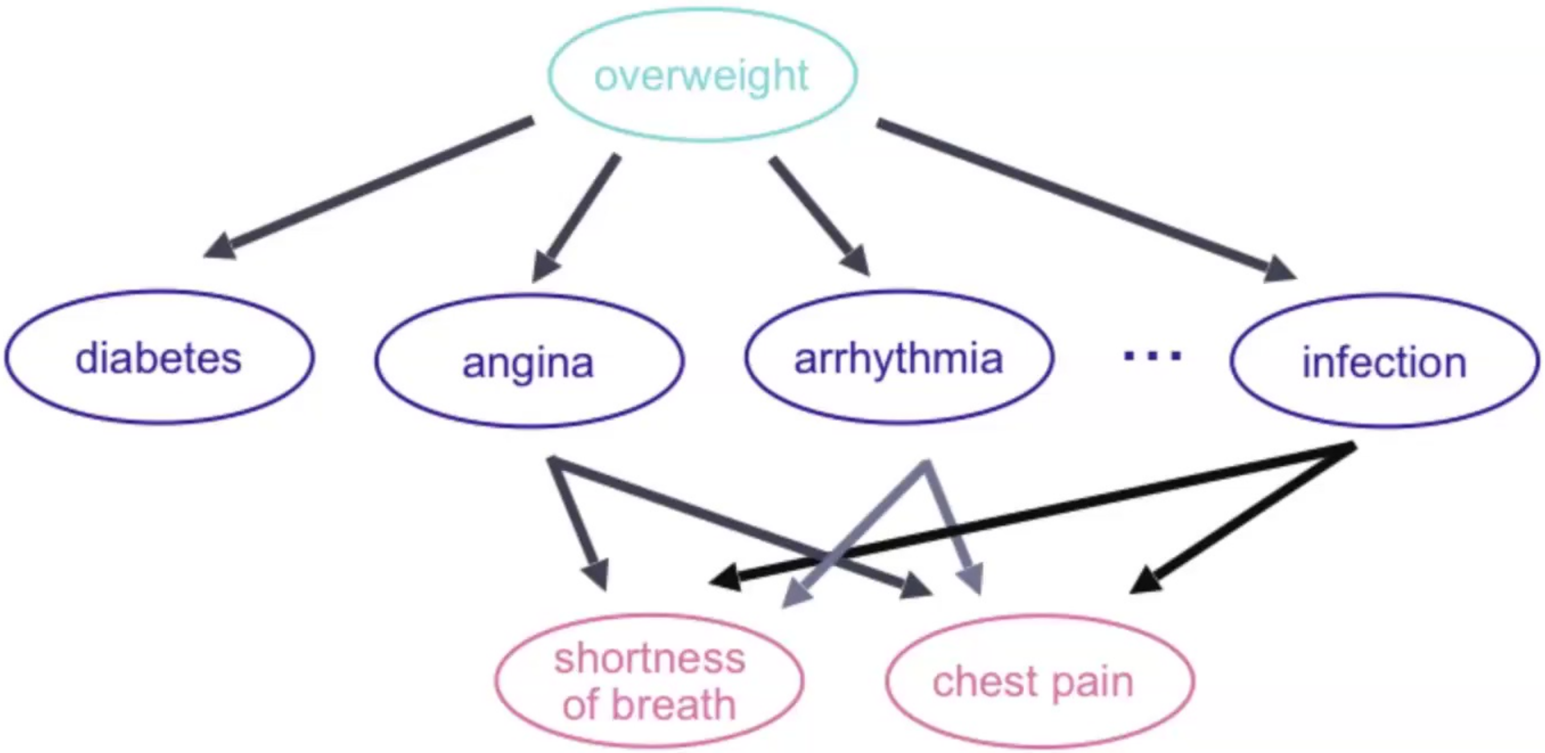
$$P(C_{S=F}=F \mid C=T, S=T)$$

1. **Abduction:**
   Update P(λ) to : $P(\lambda \mid C = T, S = T)$

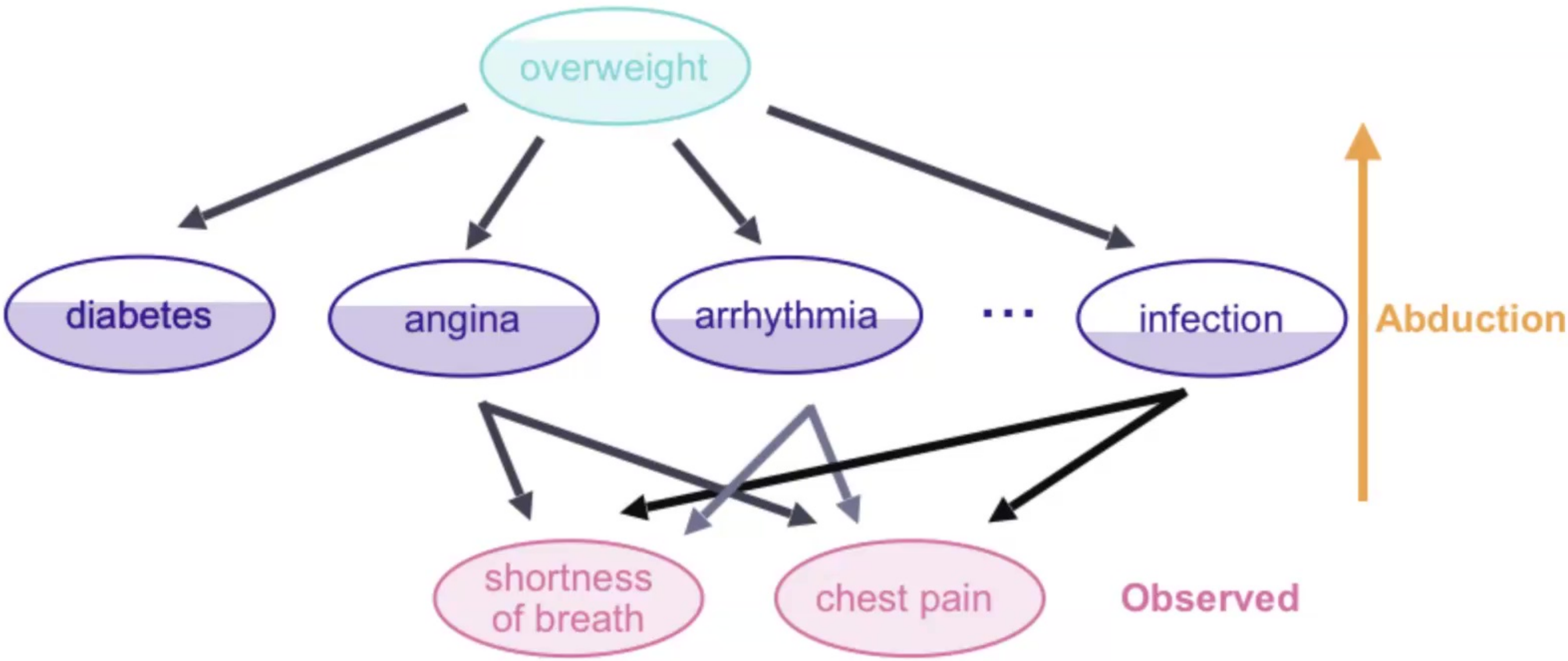
2. **Action:**
   Apply *do(.)* operator to force intervention ***S=F***

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

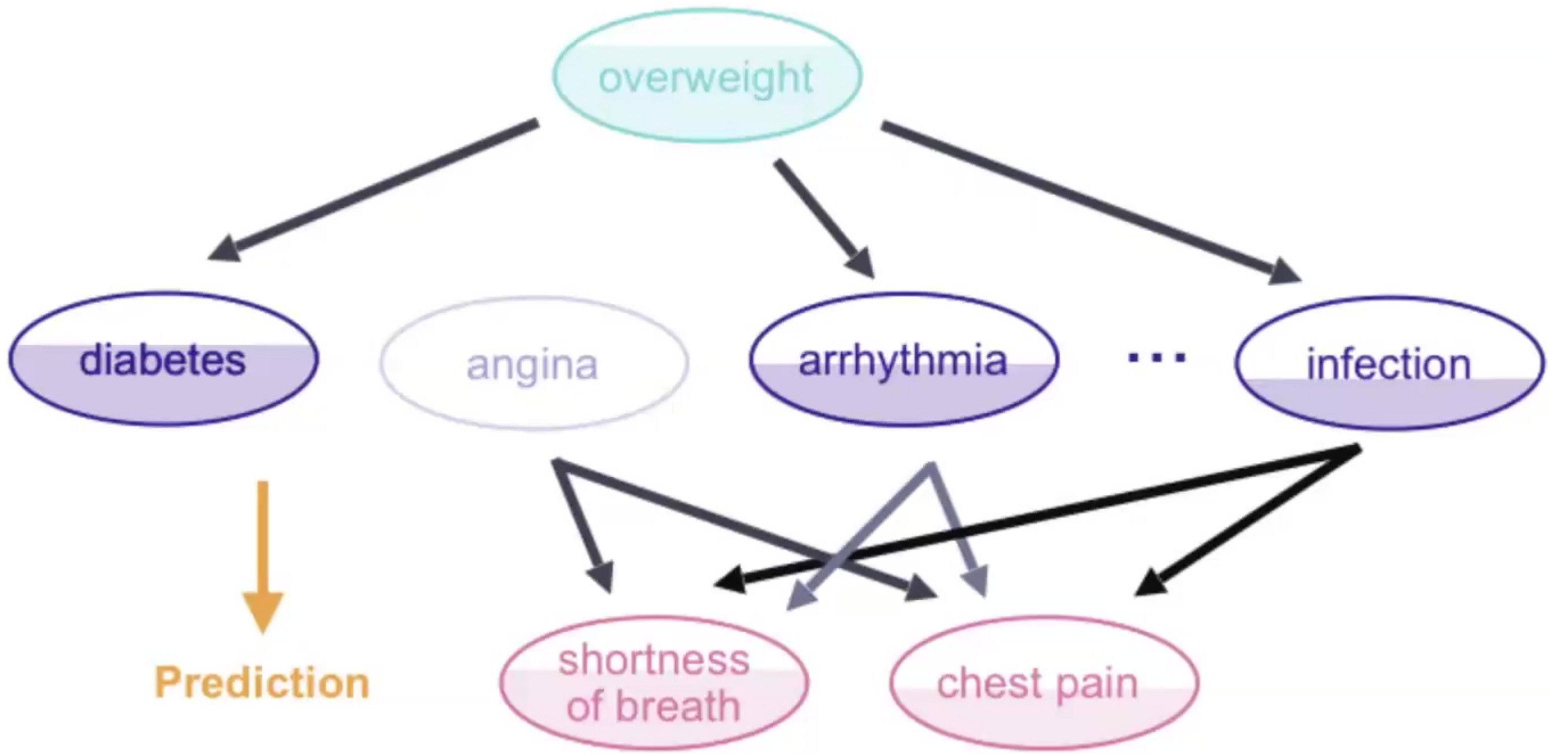
Example: $\mathbf{P(Symptoms_{Disease=off} = reduce \mid Observe\ Symptoms)}$



Example: $\mathbf{P(Symptoms_{Disease=off} = reduce|\ Observe\ Symptoms)}$



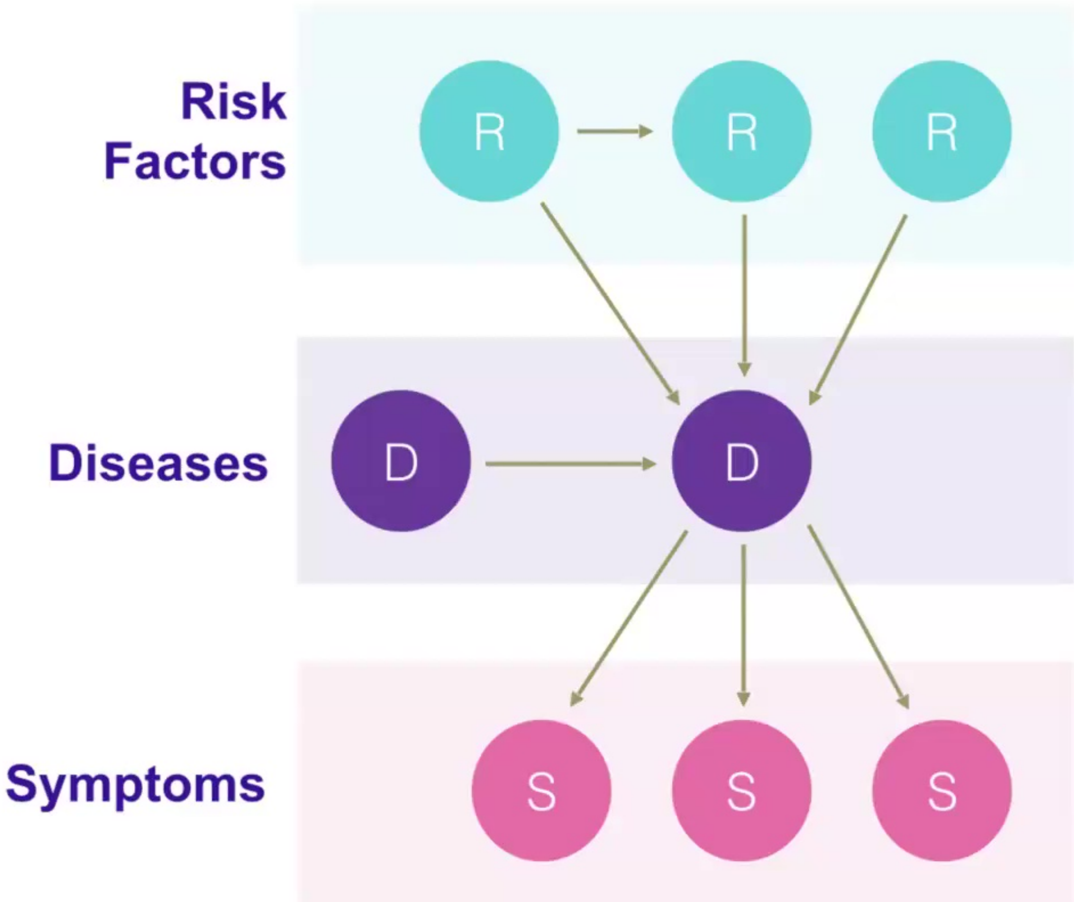
Example: **P(Symptoms$_{\text{Disease=off}}$ = reduce| Observe Symptoms)**



ARE COUNTERFACTUALS USEFUL?

- But are counterfactuals useful for anything?
- Is just quantifying interventions enough for deciding what actions to take?
- Let's explore them in the context of medical 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”

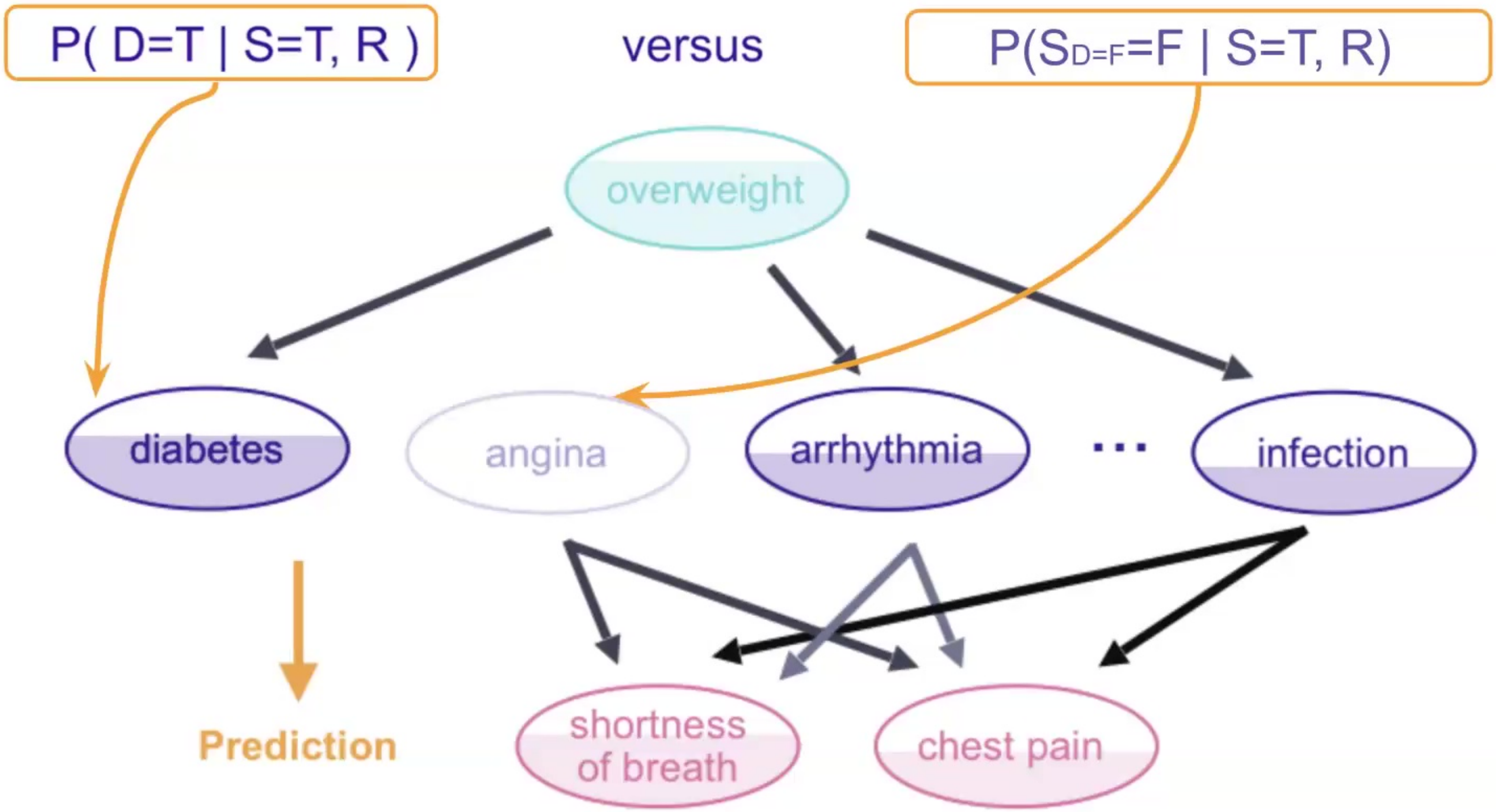
Posterior ranking versus **Counterfactual Inference**

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

"What is most likely disease, given evidence?"

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

Why might we expect this to work?

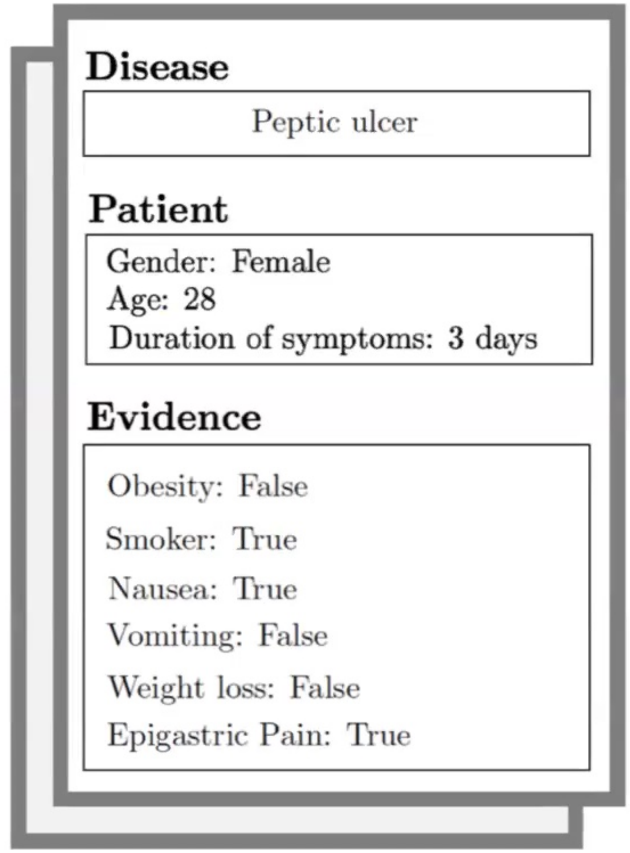


P(D=T | S=T, R)
versus
P(SD=F=F | S=T, R)
overweight
diabetes
angina
arrhythmia
...
infection
Prediction
shortness of breath
chest pain

Comparing posterior ranking and counterfactual inference

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

From symptoms & medical history from
case, diagnose disease



Disease

Peptic ulcer

Patient

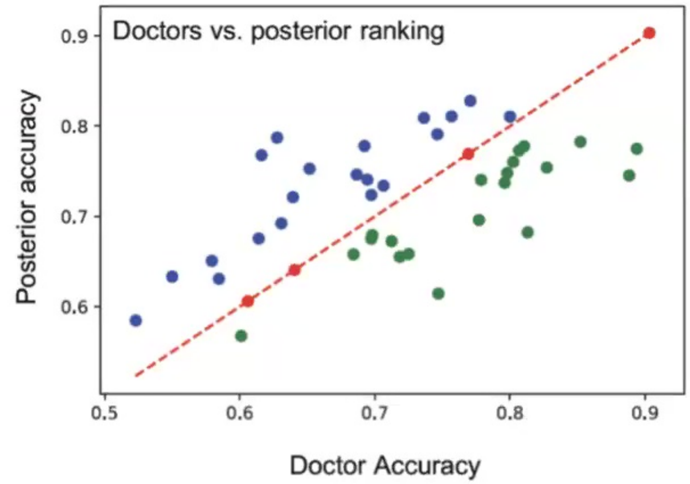
Gender: Female
Age: 28
Duration of symptoms: 3 days

Evidence

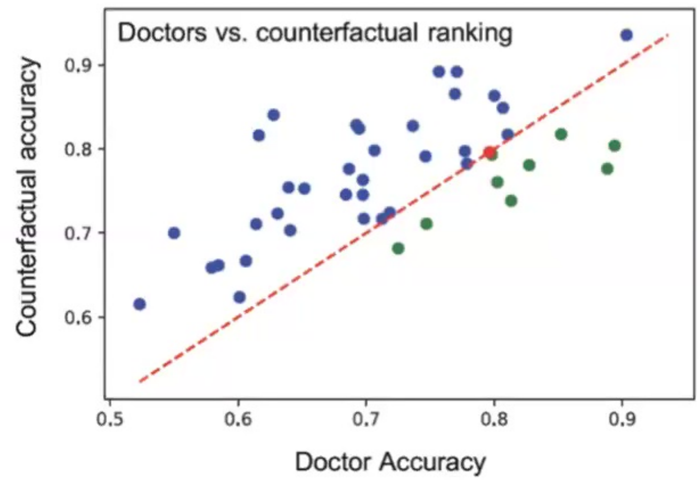
Obesity: False
Smoker: True
Nausea: True
Vomiting: False
Weight loss: False
Epigastric Pain: True

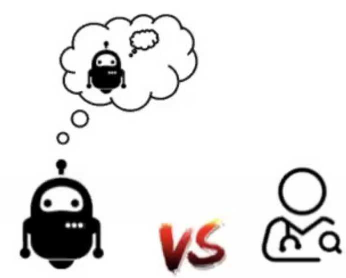
Compare to 44 doctors not involved in creating medical cases

Associative places in top 48% of doctors, achieving average clinical accuracy

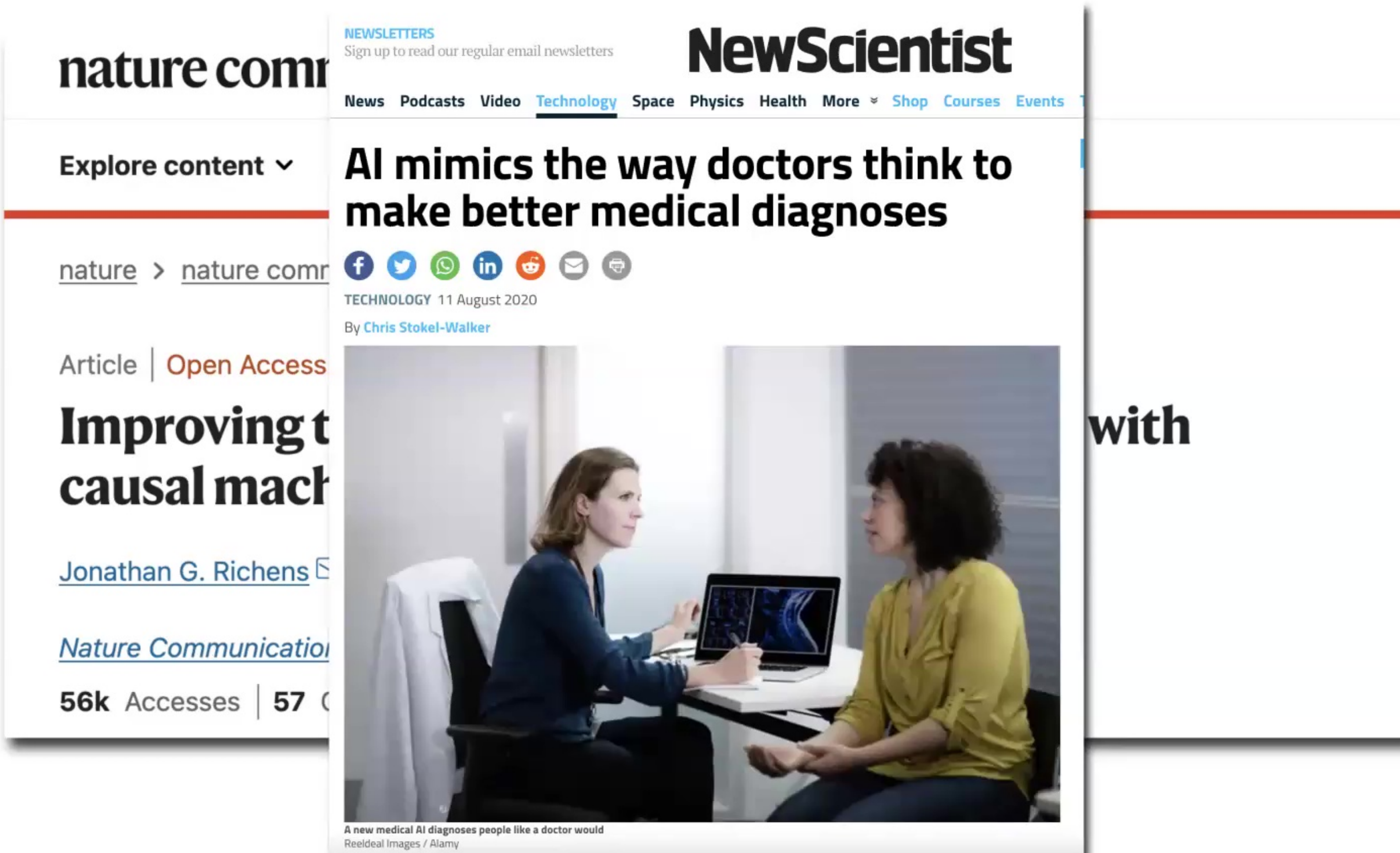


Counterfactual inference places in top 25% of doctors, achieving expert clinical accuracy

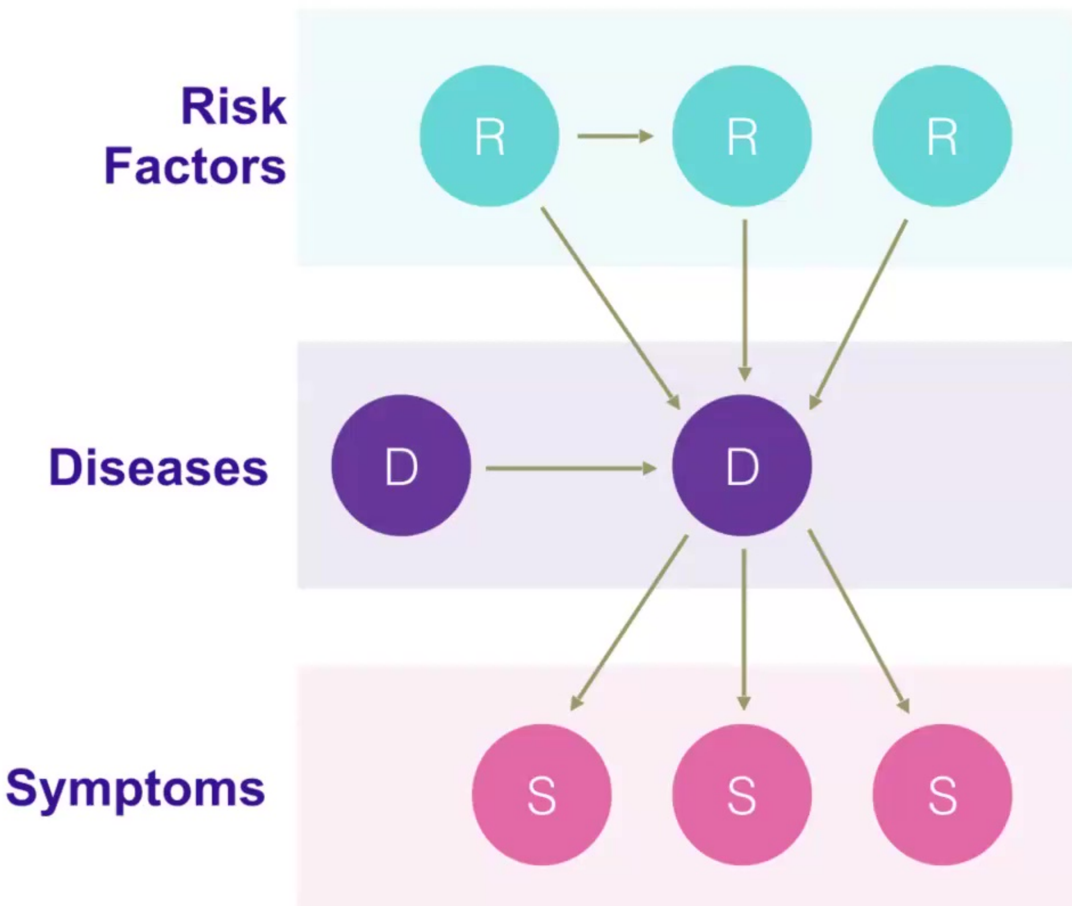




We even got some media coverage in New Scientist!



So counterfactual inference is useful, great!
What else can we do with it?



- In this example we used a parametric model to understand the relationship between causes and effects – the Noisy-OR model
- Can we answer counterfactuals questions from data **without parametric assumptions**?
- **What types of real-world problems would this help with?**

Motivating example II

- Diarrheal diseases are a leading cause of disease and mortality in the developing world
- To reduce diarrheal diseases in children in the Busia district of Western Kenya, a local NGO built protective cement structures around a randomly selected group of springs

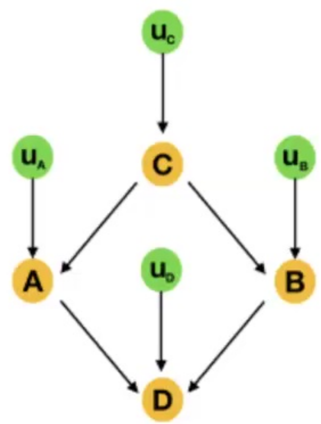
Motivating example II

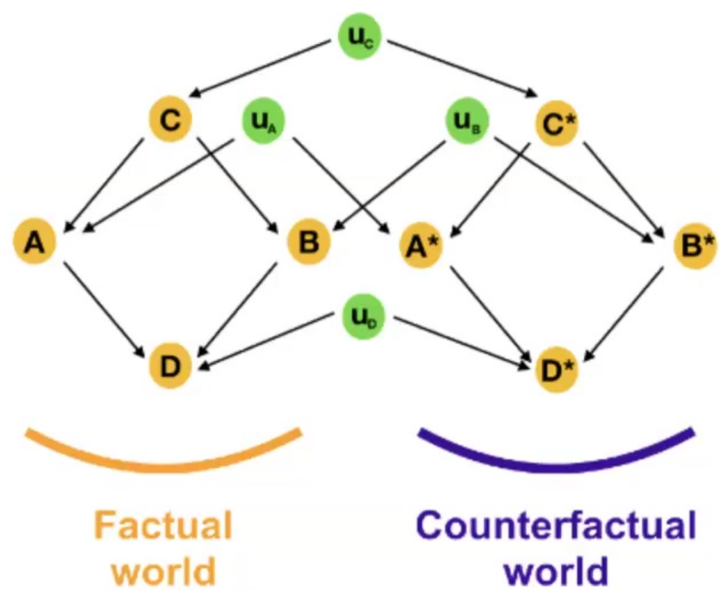
- Working with the Poverty Action Lab, they found that spring protection significantly reduces diarrhea for children under age three by 25%
- Should we scale this intervention up? Interventions are expensive, need to be certain it will help as expected
- To really answer this, need to answer the following *counterfactual* question:

Question: "Given a child developed diarrhea after drinking from an unprotected spring, would they have still developed it if they drank from a protected spring?"

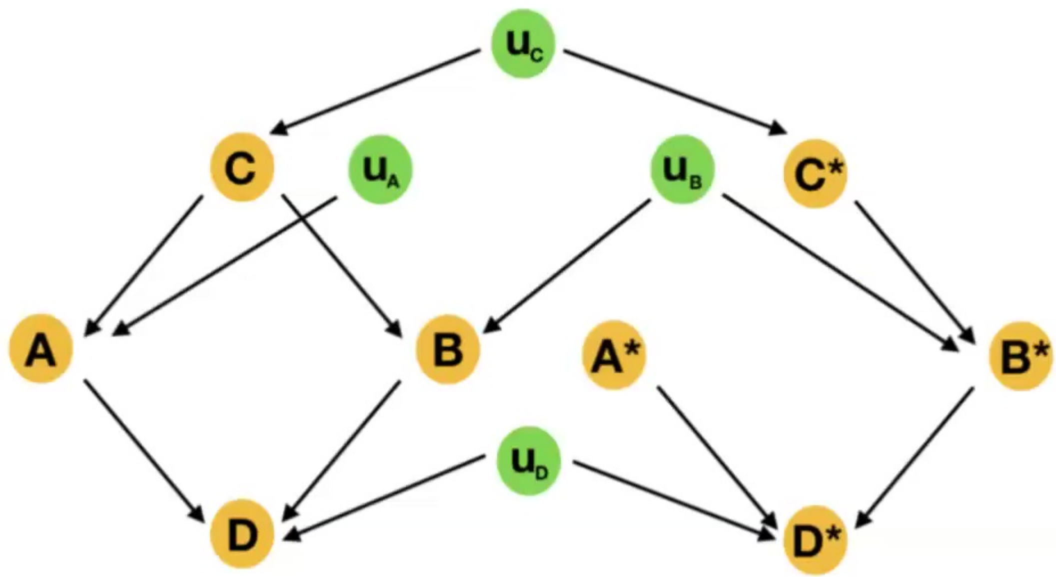
$$P(\text{Diarrhoea}_{\text{Protect Springs}=\textbf{Yes}} = \textbf{No} \mid \text{Diarrhoea} = \textbf{Yes}, \text{Protect Springs} = \textbf{No}, Z)$$

TWIN NETWORKS





TWIN NETWORKS



Counterfactual Inference compute:

$P(C_{S=F}=F | C=T, S=T)$

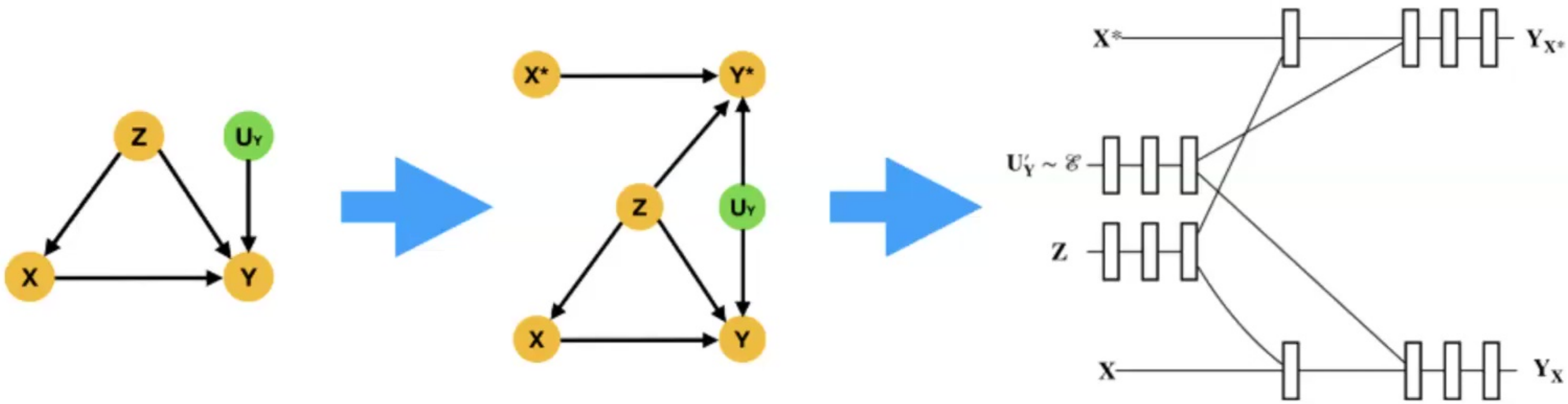
Standard:

1. **Abduction**
2. **Action**
3. **Prediction**

Twin network:

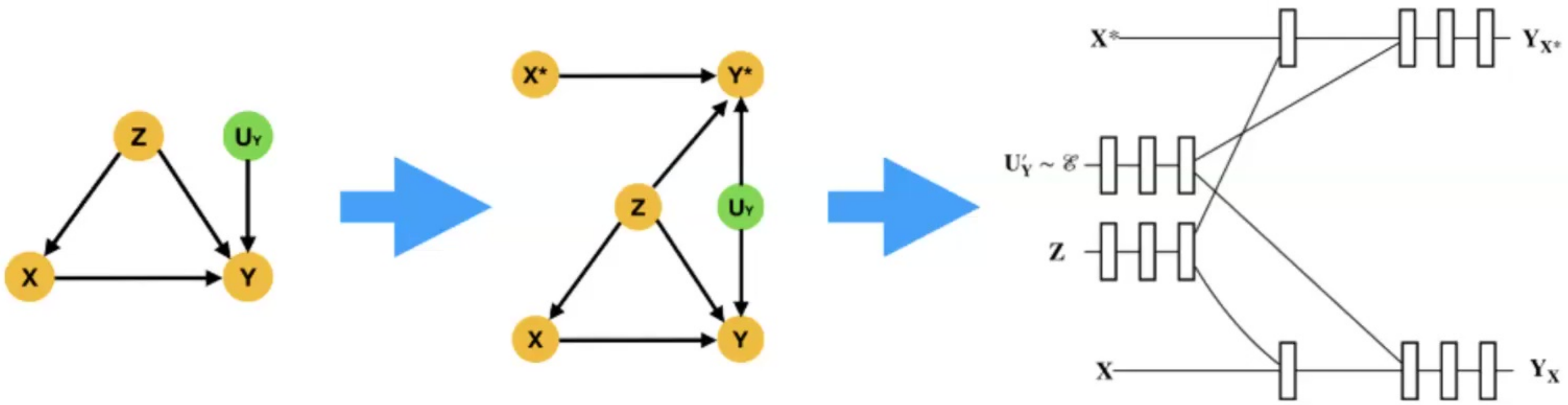
Bayesian inference on Twin Network with S* set to F: P(C*=F | C=T, S=T)

DEEP TWIN NETWORKS



- ➔ Graphical nature of twin networks makes them very amenable to deep learning
- ➔ As Neural network architecture matches twin network structure, counterfactual inference is just standard inference on network heads

DEEP TWIN NETWORKS



- ➔ Graphical nature of twin networks makes them very amenable to deep learning
- ➔ As Neural network architecture matches twin network structure, counterfactual inference is just standard inference on network heads

Should we protect water springs in this manner?

Method	P(N)	P(S)	P(N&S)
KW Median Child *Cuellar et al. 2020*	0.12 ± 0.01	-	-
KW TN Median Child	0.13598 ± 0.049	0.09811 ± 0.031	0.31778 ± 0.012
KW TN Test Set	0.06273 ± 0.020	0.03914 ± 0.016	0.08521 ± 0.034

$$P(\text{Sufficiency}) = P(\text{Diarrhoea}_{\text{Protect Springs=Yes}} = \text{Yes} \mid \text{Diarrhoea} = \text{No}, \text{Protect Springs=No}, Z)$$

- ➔ P(Sufficiency) is small, hence children would would likely have still had diarrhoea even if the springs were protected.
- ➔ This provides evidence that protecting water springs in this manner has little effect on the development of diarrhoea in children in these populations

Check out our paper to dive deeper arXiv:2109.01904

nature machine intelligence

Article

https://doi.org/10.1038/s42256-023-00611-x

Estimating categorical counterfactuals via deep twin networks

Received: 23 May 2022

Accepted: 4 January 2023

Published online: 20 February 2023

Check for updates

Athanasios Vlontzos[1,5], Bernhard Kainz[1,2] & Ciarán M. Gilligan-Lee[3,4]

Counterfactual inference is a powerful tool, capable of solving challenging problems in high-profile sectors. To perform counterfactual inference, we require knowledge of the underlying causal mechanisms. However, causal mechanisms cannot be uniquely determined from observations and interventions alone. This raises the question of how to choose the causal mechanisms so that the resulting counterfactual inference is trustworthy in a given domain. This question has been addressed in causal models with binary variables, but for the case of categorical variables, it remains unanswered. We address this challenge by introducing for causal models with categorical variables the notion of counterfactual ordering, a principle positing desirable properties that causal mechanisms should possess and prove that it is equivalent to specific functional constraints on the causal mechanisms. To learn causal mechanisms satisfying these constraints, and perform counterfactual inference with them, we introduce deep twin networks. These are deep neural networks that, when trained, are capable of twin network counterfactual inference—an alternative to the abduction–action–prediction method. We empirically test our approach on diverse real-world and semisynthetic data from medicine, epidemiology and finance, reporting accurate estimation of counterfactual probabilities while demonstrating the issues that arise with counterfactual reasoning when counterfactual ordering is not enforced

Judea Pearl @yudapearl

1/ This paper tells me that I was wrong in dismissing twin-networks as no-longer useful once you establish conditional ignorability. It shows them to be computational instruments, especially useful in Bayesian analysis. The paper will also be useful for readers working on

> **Athanasios Vlontzos** @vlontzos · Sep 7
> 🚨 New Paper !
> In our most recent work with @BernhardKainz1 and @quantumciaran we evaluate probabilities of causation and perform counterfactual-level causal inf by combining @yudapearl's Twin Networks and Deep NNs
> arxiv.org/abs/2109.01904 #CausalML
> #MachineLearning
> Show this thread

9:04 AM · Sep 11, 2021 · Twitter Web App

17 Retweets **1** Quote Tweet **109** Likes

Judea Pearl @yudapearl · Sep 11
Replying to @yudapearl
2/ explanation which, IMO, must estimate probabilities of causation, both necessary and sufficient, a task rarely tackled in the "explainable-AI" industry.

17

Examples of counterfactuals at Spotify

- **Which users should be sent a promotion:** which users Z "need" to be shown a promotion to engage with the content?

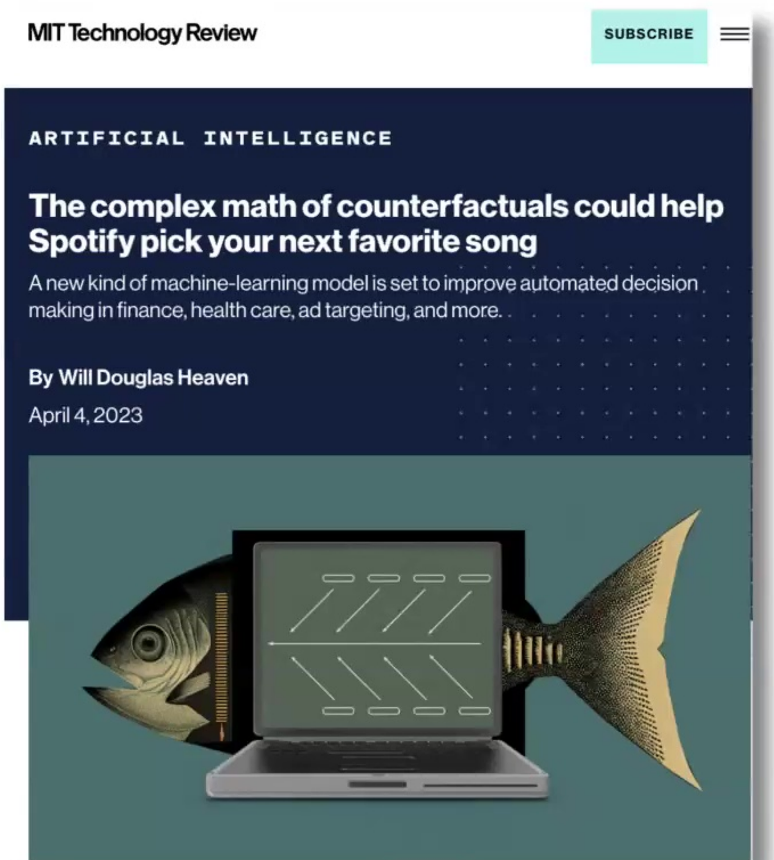
$$P(Y_{X=\text{promotion}} = \text{engaged}, Y_{X=\text{no promotion}} = \text{not engaged} \mid Z)$$

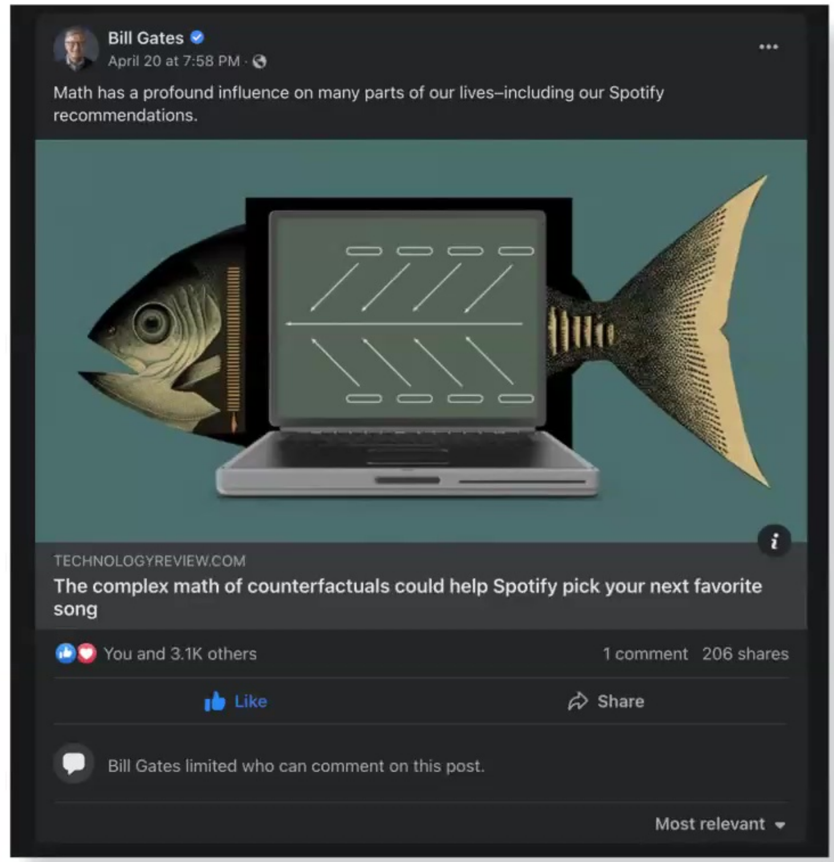
- **New content to enjoy:** If user Z listened to specific content and enjoyed it, which other content would they also have enjoyed?

$$P(Y_{X=\text{new content}} = \text{engage} \mid Y = \text{engage}, X = \text{current content}, Z)$$

And many more….

Our work received coverage from MIT Tech Review
(and was even shared by Bill Gates!)

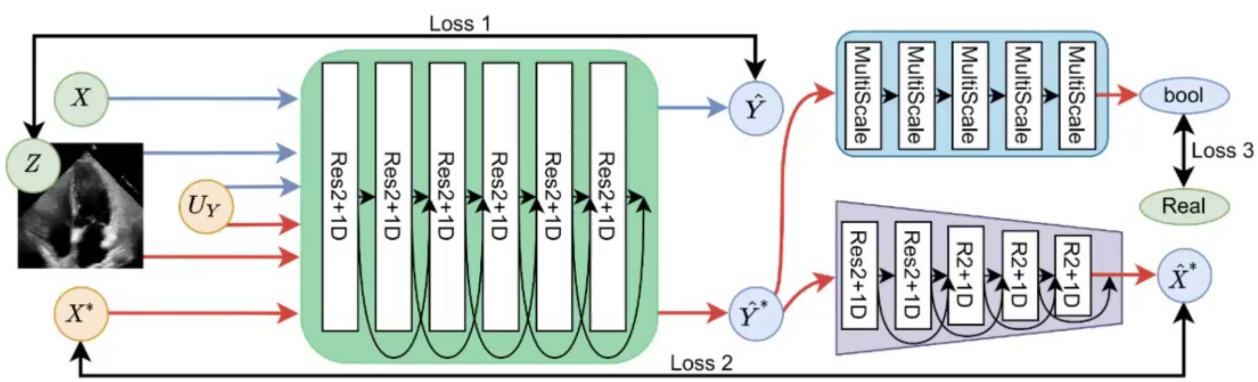


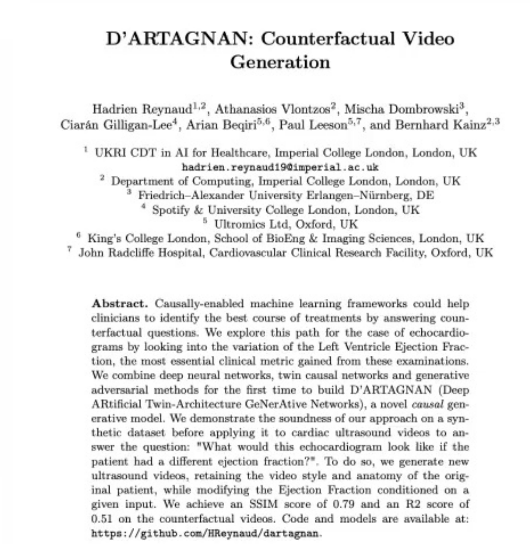


DEEP TWIN NETWORKS LET US WORK WITH HIGH DIMENSIONAL DATA, LIKE VIDEOS

arXiv:2206.01651v2

On cardiac ultrasound videos we answer the question: "What would this echocardiogram look like if the patient had a different ejection fraction?



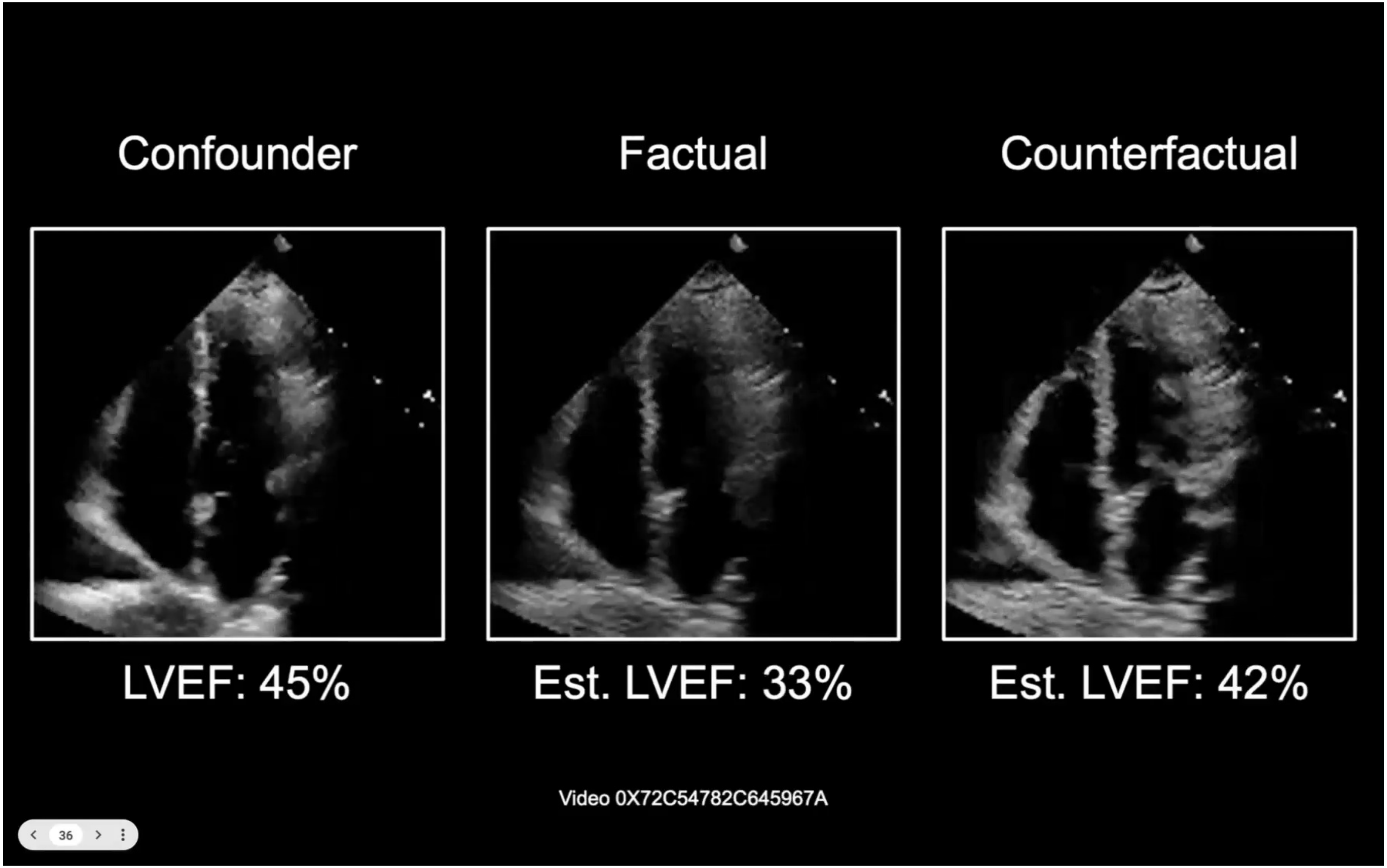


D'ARTAGNAN: Counterfactual Video Generation

Hadrien Reynaud[1,2], Athanasios Vlontzos[2], Mischa Dombrowski[3], Ciarán Gilligan-Lee[4], Arian Beqiri[5,6], Paul Leeson[5,7], and Bernhard Kainz[2,3]

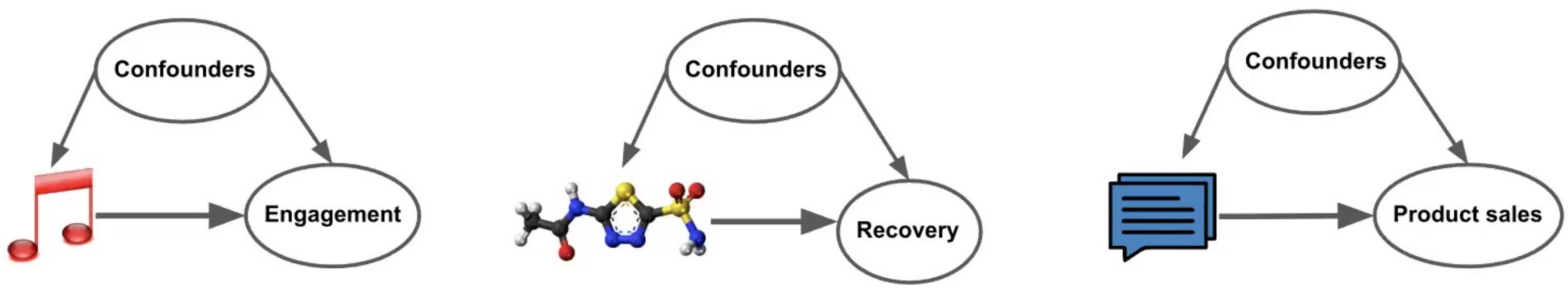
[1] UKRI CDT in AI for Healthcare, Imperial College London, London, UK
hadrien.reynaud19@imperial.ac.uk
[2] Department of Computing, Imperial College London, London, UK
[3] Friedrich–Alexander University Erlangen–Nürnberg, DE
[4] Spotify & University College London, London, UK
[5] Ultromics Ltd, Oxford, UK
[6] King's College London, School of BioEng & Imaging Sciences, London, UK
[7] John Radcliffe Hospital, Cardiovascular Clinical Research Facility, Oxford, UK

Abstract. Causally-enabled machine learning frameworks could help clinicians to identify the best course of treatments by answering counterfactual questions. We explore this path for the case of echocardiograms by looking into the variation of the Left Ventricle Ejection Fraction, the most essential clinical metric gained from these examinations. We combine deep neural networks, twin causal networks and generative adversarial methods for the first time to build D'ARTAGNAN (Deep ARtificial Twin-Architecture GeNerAtive Networks), a novel *causal* generative model. We demonstrate the soundness of our approach on a synthetic dataset before applying it to cardiac ultrasound videos to answer the question: "What would this echocardiogram look like if the patient had a different ejection fraction?". To do so, we generate new ultrasound videos, retaining the video style and anatomy of the original patient, while modifying the Ejection Fraction conditioned on a given input. We achieve an SSIM score of 0.79 and an R2 score of 0.51 on the counterfactual videos. Code and models are available at: https://github.com/HReynaud/dartagnan.


Confounder
Factual
Counterfactual
LVEF: 45%
Est. LVEF: 33%
Est. LVEF: 42%
Video 0X72C54782C645967A
36

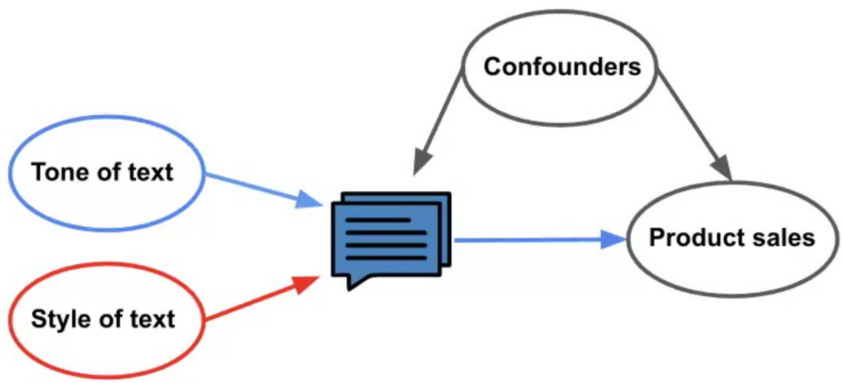
BEYOND BINARY TREATMENTS

- We can now do answer counterfactual questions when confounders and covariates are high dimensional, complex objects—like video.
- But to answer counterfactual questions in real-world settings, like drug discovery or recommender systems, we need to move beyond binary or continuous treatments.
- Treatments in many real-world settings are complex, structured, high-dimensional objects, such as text, audio, graphs, or products in an online marketplace, etc.



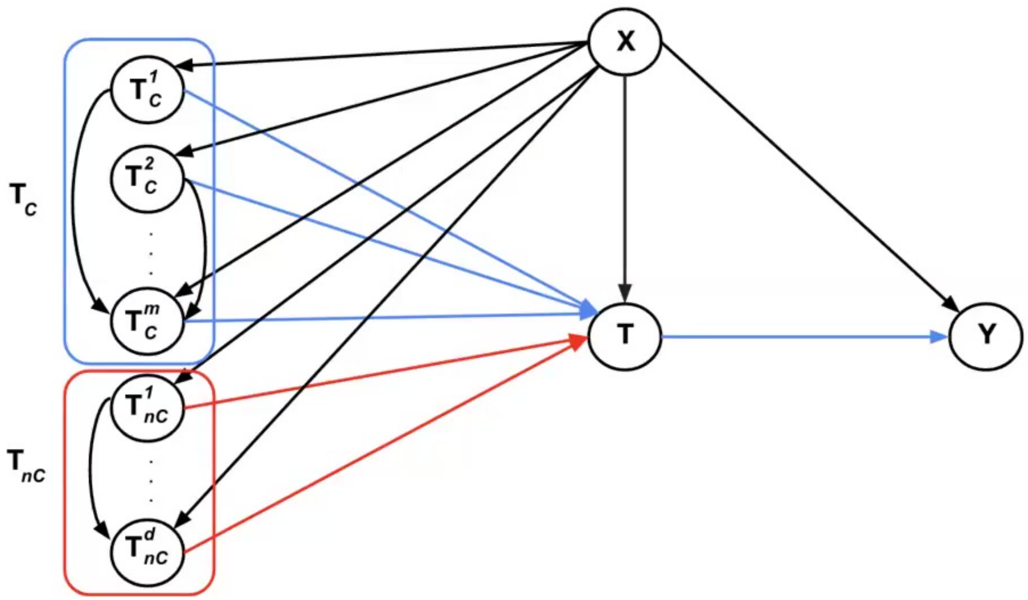
CAUSAL INFERENCE WITH HIGH-DIMENSIONAL, STRUCTURED TREATMENTS

- What's the impact of a product review on sales of that product? Here, the positive or negative tone of the review will likely be the main driver of impact to sales.
- To estimate the effect, however, all we have access to is the full text of the review.
- Other latent aspects of the text—such as style—may not impact sales, yet are mixed together with the tone of the message in the text itself. How do we estimate the effect of the true causal components, given they are latent?



38

THE PROBLEM SET-UP



Causally relevant latents: T_C,
Non-causally relevant latents: T_{nC}.

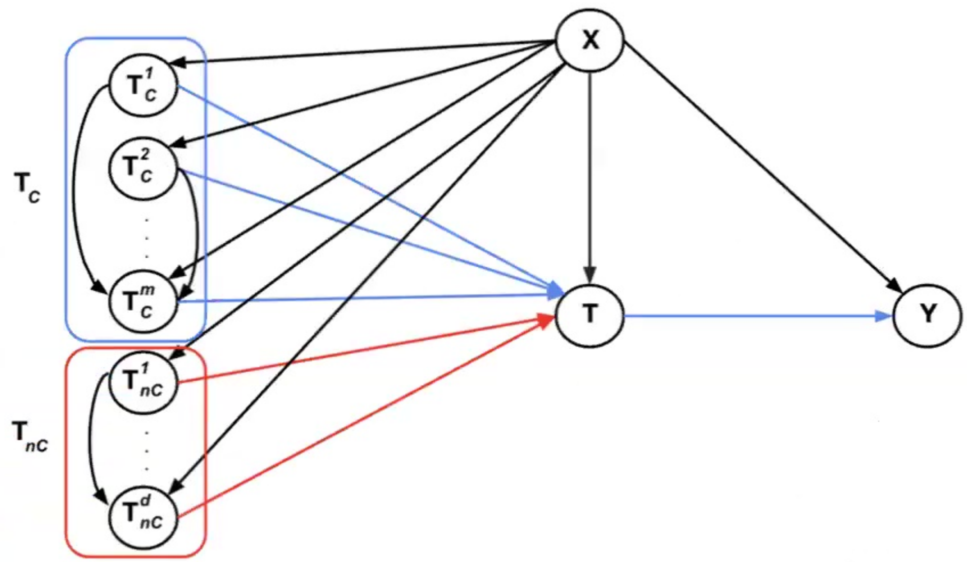
The treatment we're given for a specific problem corresponds to a (potentially non-linear) mixture of causal and non-causal latents, $T=m(T_C, T_{nC})$.

The full causal model is given by:

$X = l(\varepsilon_X)$,
$T_C = g(X, \varepsilon_{T_C})$,
$T_{nC} = h(X, \varepsilon_{T_{nC}})$,
$T = m(T_C, T_{nC})$, and
$Y = f(T_C, X, \varepsilon_Y)$,

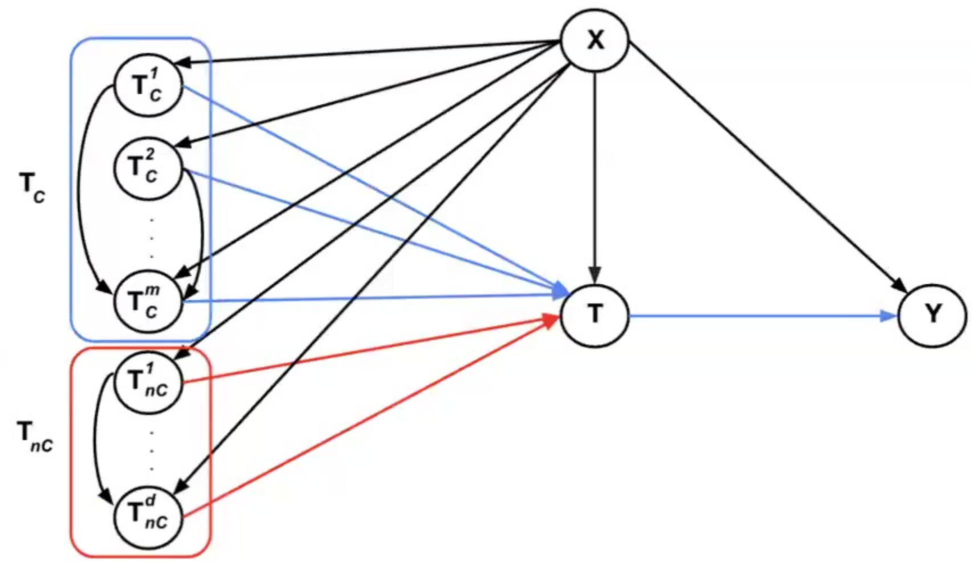
latent noise terms are drawn i.i.d. $\varepsilon_i \sim P(\varepsilon_i)$.

BACKDOOR ADJUSTMENT DOESN'T SUFFICE



Theorem: Backdoor adjustment directly using T leads to biased causal effect estimation

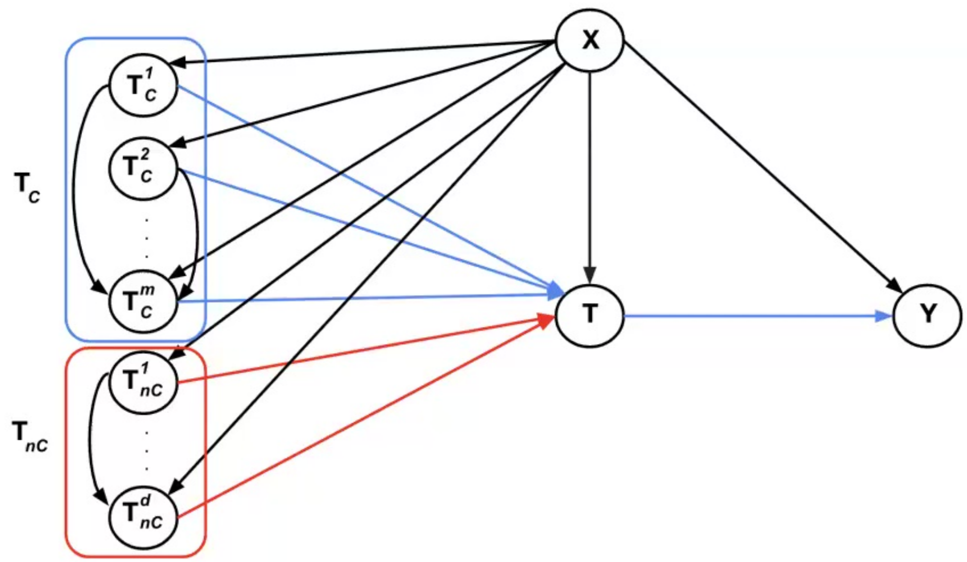
BACKDOOR ADJUSTMENT DOESN'T SUFFICE



Theorem: Backdoor adjustment directly using T leads to biased causal effect estimation

41

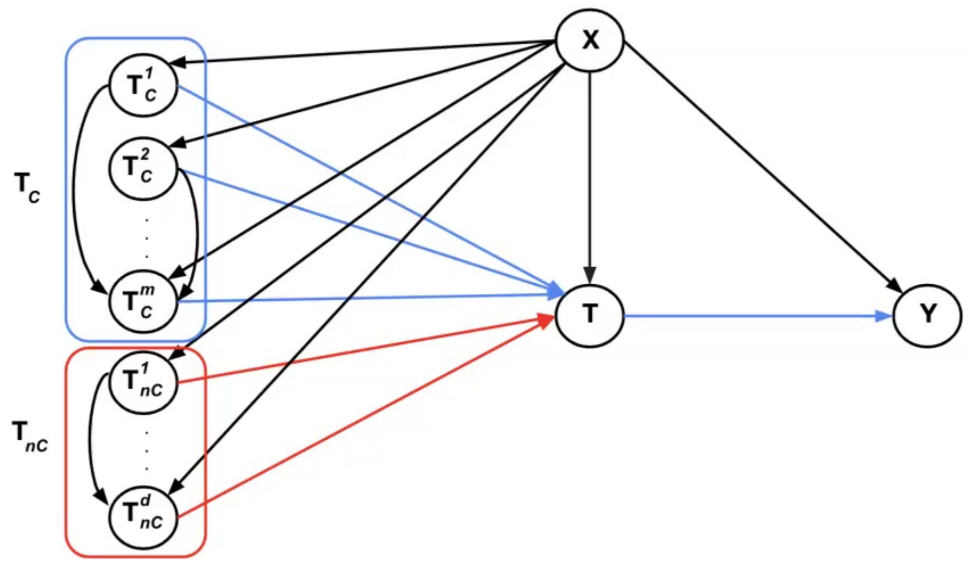
BACKDOOR ADJUSTMENT DOESN'T SUFFICE



Intuition: The non-causal latents T_{nC} are proxies for the confounders, X. Using them in our estimation can make it look like confounders have been controlled for, but when we intervene on the T_{nC} we break the link between T_{nC} and X and reveal that we have not controlled for the true confounders

42

SOLUTION: LEARN REPRESENTATION INVARIANT TO T_{NC}



Proposition: Causal effect estimation is unbiased if and only if a representation of T is used that contains no information about the non-causal latents, T_{nC}

43

LEARNING A REPRESENTATION INVARIANT TO T_{NC}

- Suppose we have two [T, X, Y] data points where the X and Y values are the same, but the T 's are different: [T, x, y], [T′, x, y].

- As Y is only a function of the casual latents, $Y = f(T_C, X)$, and as the X, Y values in the two data points are the same, we have that $f(T_C, X) = f(T'_C, X)$. Assuming this function is invertible, the causal components of T and T′ are equal.

- Moreover, data points [T, x, y], [T′, x, y’] where the X values are the same, but the Y values are different must have different causal components, T_C.

CONTRASTIVE REPRESENTATIONS OF HIGH-DIMENSIONAL, STRUCTURED TREATMENTS

Theorem: *a contrastive learning approach with these positive pairs yields a representation of T that* T_C

Intuition: Given a datapoint and it's positive pair, T_C is invariant for these two treatments T, T'. We prove that the non-causal components T_{nC} are not invariant between positive pairs, and use the fact that contrastive algorithms identify invariant features.

Contrastive representations of high-dimensional, structured treatments

Oriol Corcoll
Spotify
oriolc@spotify.com

Athanasios Vlontzos
Spotify
Imperial College London
athanasiosv@spotify.com

Michael O'Riordan
Spotify
moriordan@spotify.com

Ciarán M. Gilligan-Lee
Spotify
University College London
ciaranl@spotify.com
ciaran.lee@ucl.ac.uk

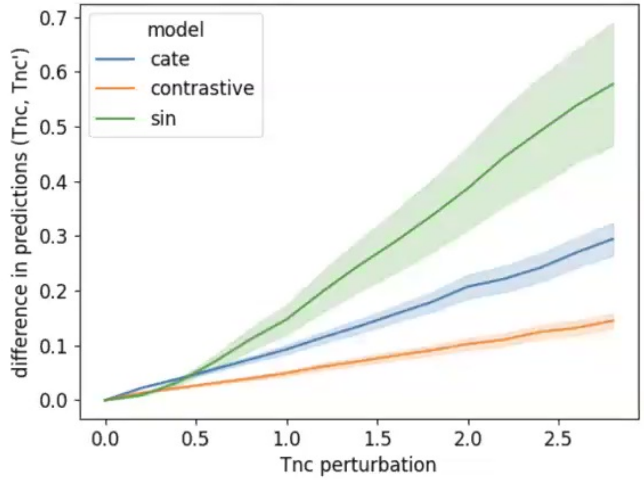
Abstract

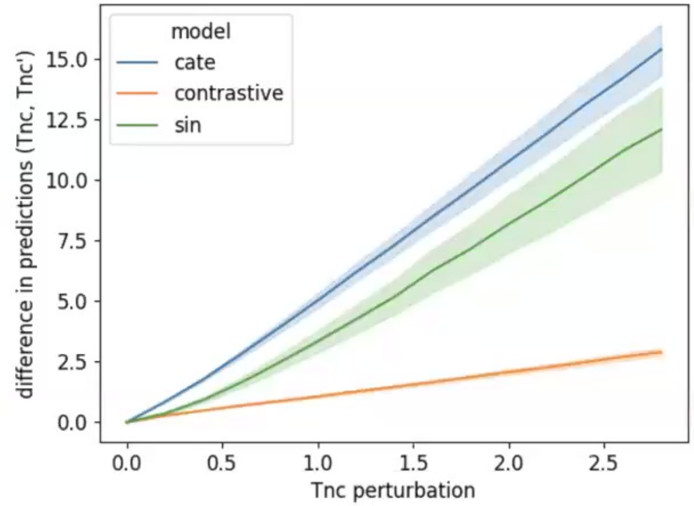
Estimating causal effects is vital for decision making. In standard causal effect estimation, treatments are usually binary- or continuous-valued. However, in many important real-world settings, treatments can be structured, high-dimensional objects, such as text, video, or audio. This provides a challenge to traditional causal effect estimation. While leveraging the shared structure across different treatments can help generalize to unseen treatments at test time, we show in this paper that using such structure blindly can lead to biased causal effect estimation. We address this challenge by devising a novel contrastive approach to learn a representation of the high-dimensional treatments, and prove that it identifies underlying causal factors and discards non-causally relevant factors. We prove that this treatment representation leads to unbiased estimates of the causal effect, and empirically validate and benchmark our results on synthetic and real-world datasets.

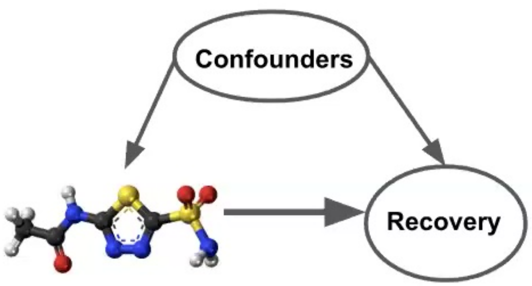
Coming to an arXiv near you very soon

46

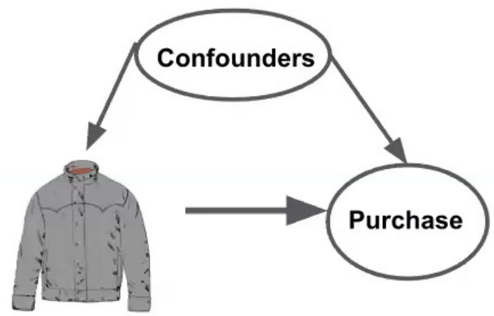
EXPERIMENTS







Impact of drug



Impact of product

IS CAUSAL INFERENCE USEFUL FOR PHYSICS?

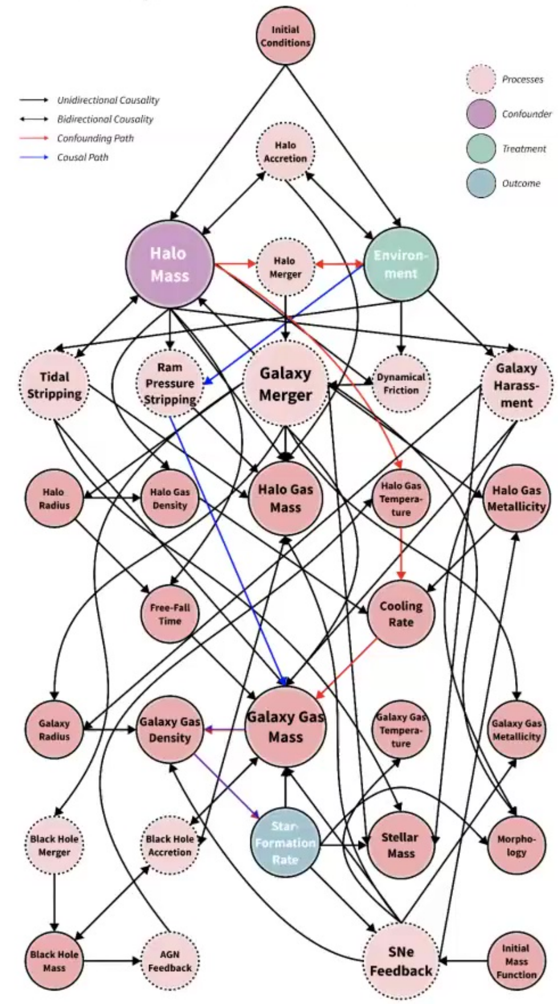
- In the local Universe, there are two distinct populations of galaxies: the *red sequence* of massive, early-type galaxies, and the *blue sequence* of less-massive, late-type, star-forming galaxies.

- These two populations exist in different environments, with the red sequence found in clusters and the blue sequence located in isolation.

Question: is the difference in environment responsible for the difference in star formation rate?

- Previous work has looked at the correlation between environment and star formation rate, or controlled for stellar mass. But does this suffice to estimate the causal effect of environment?

Causal Model of Galaxy Formation and Evolution



48

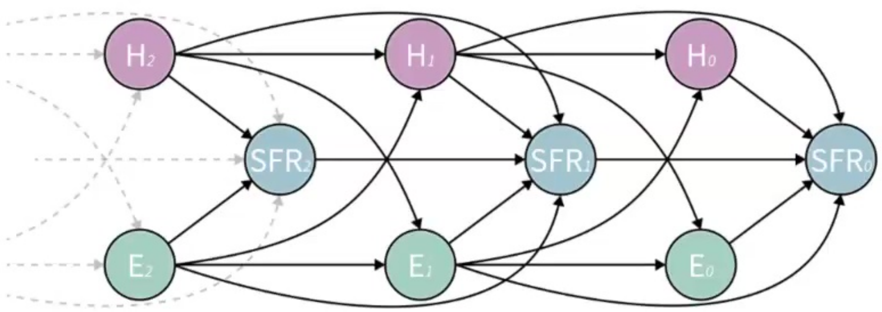
IS CAUSAL INFERENCE USEFUL FOR PHYSICS?

Answer: No!

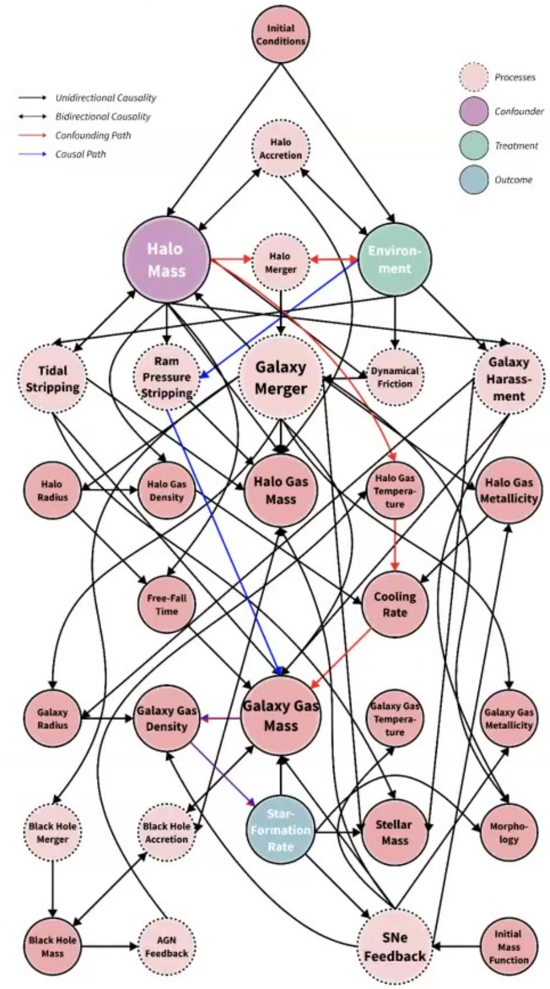
We condensed the last 50 years of astronomy literature into a causal structure and used IllustrisTNG data to estimate the causal effect.

➔ Environment suppresses star formation rate, but in the early universe it had a positive effect—challenging consensus that environment always plays a negative role.

Joint work with S. Mucesh, W. Hartley, O. Lahav from UCL, on arXiv soon



Causal Model of Galaxy Formation and Evolution



49

CONCLUSION

- Being able to answer causal questions enables actionable decision making
- Lots of remaining problems to solve if we want to apply causal inference reliably in the real world
- Many more causal inference applications at Spotify beyond what we've discussed today, reach out of you're interested!

ciaranl@spotify

@quantumciaran

50

TRAINING DEEP TWIN NETWORKS

Enforce during training that functions learned by network satisfy **identification** conditions. These amount to constraints on function class learned by the neural network