IDSIA SUPSI

Lecture 4: Structural Causal Models are (solvable by) Credal Nets

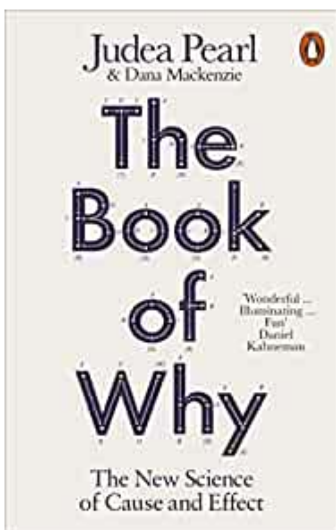
On the Relation between Imprecise Probabilities and Counterfactuals

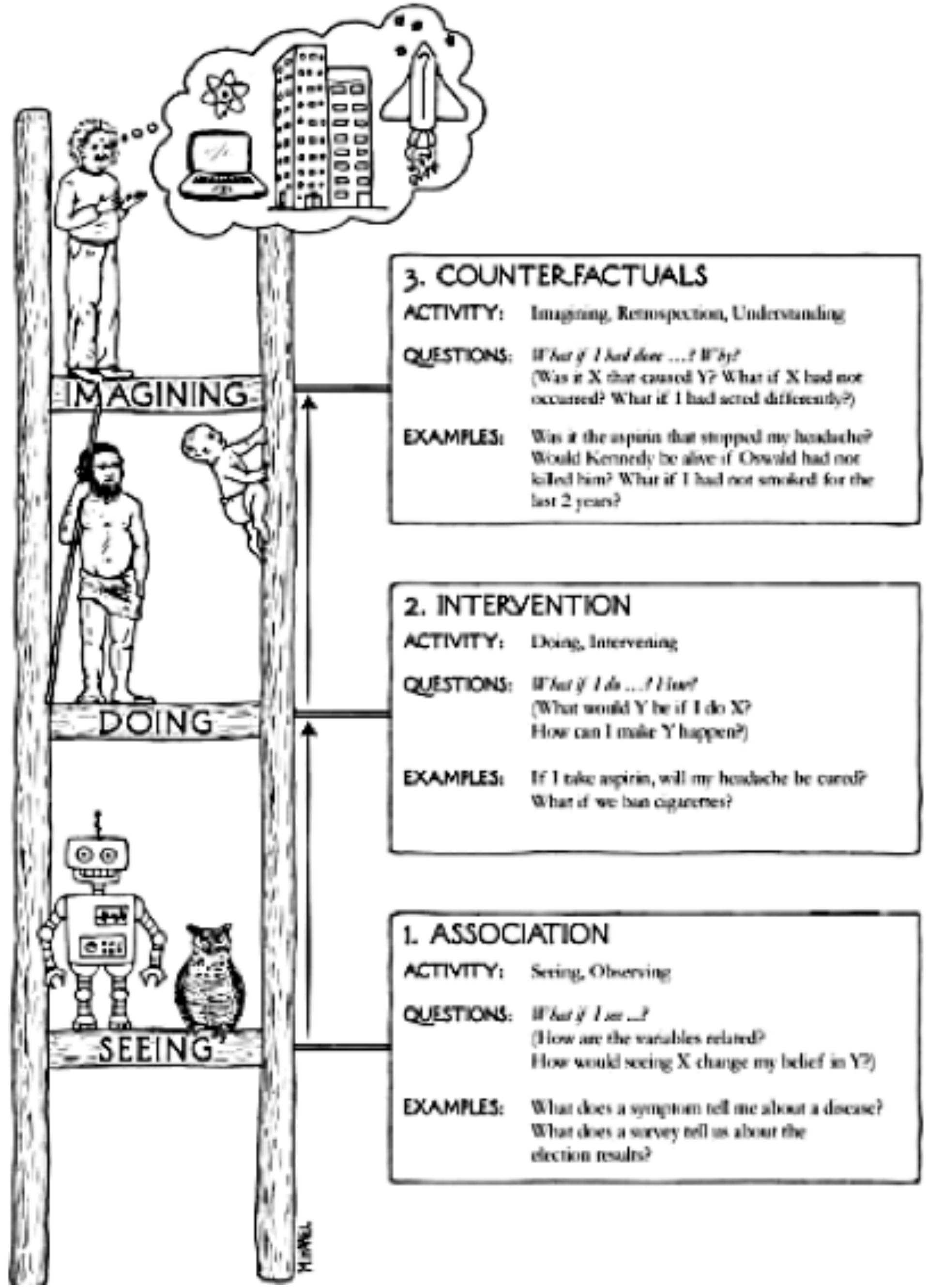
Alessandro Antonucci (alessandro@idsia.ch)
Senior lecturer-researcher IDSIA USI-SUPSI

Sipta Summer School - Bristol (UK) - August 18, 2022

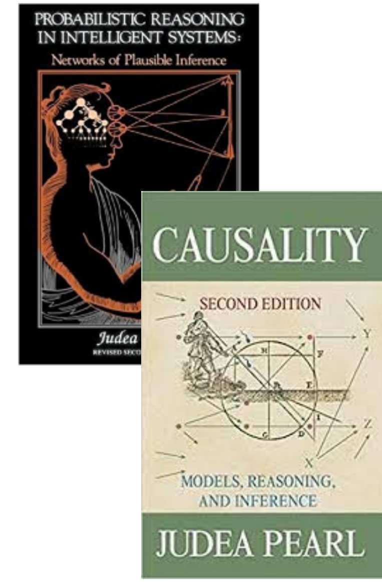
Judea Pearl

image of the ladder from

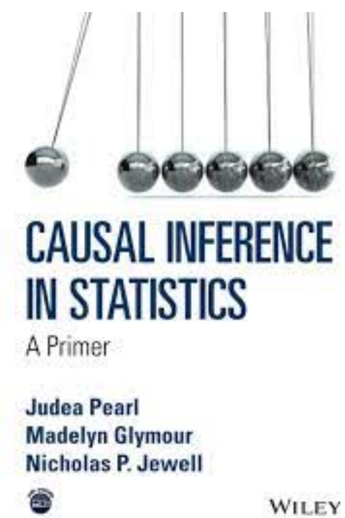




and his Ladder of Causation



earlier books on Bayesian nets and structural causal models

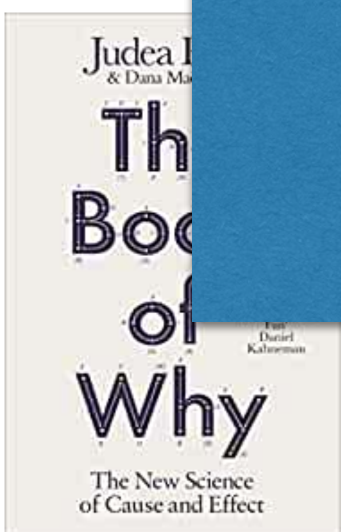


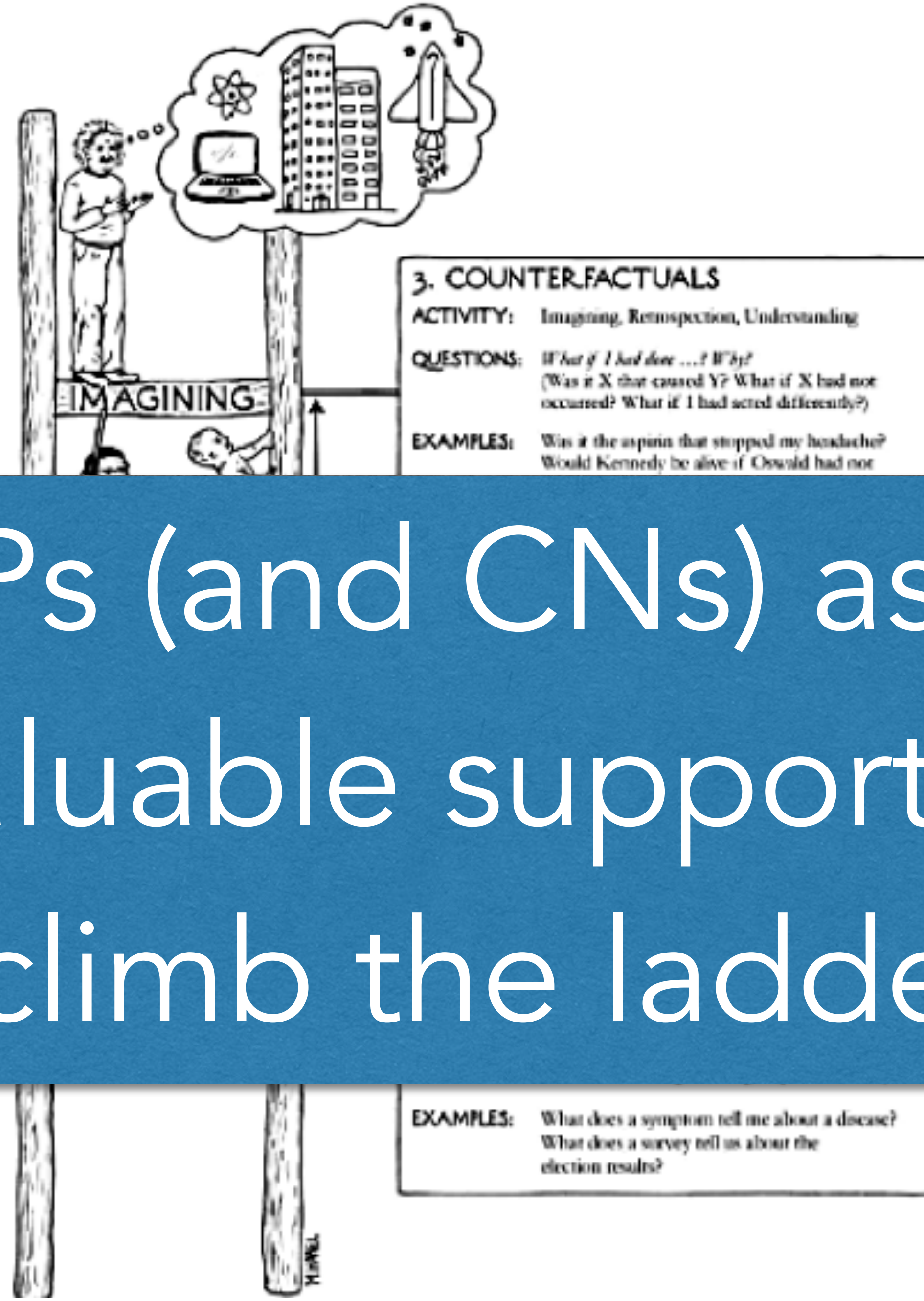
best reference for newcomers interested in causality research

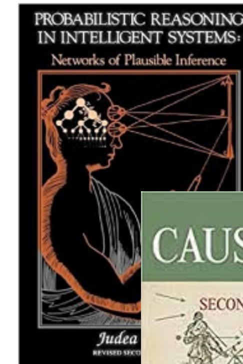
IPs (and CNs) as a valuable support to climb the ladder

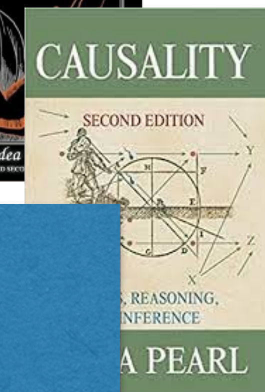
Judea Pearl

image of the l

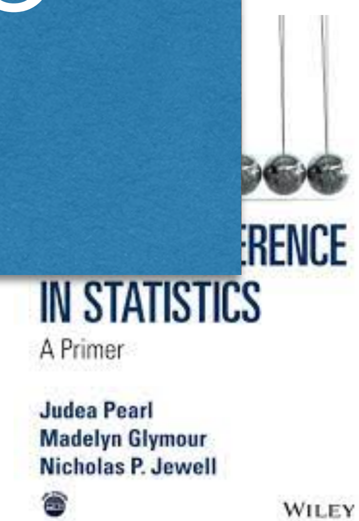








ayesian nets
sal models



best reference for newcomers interested in causality research

and his Ladder of Causation

Recent Research on the Relation between IPs and Causality

- Joint work with IDSIA colleagues (Zaffalon, Cabañas and others)
- Ongoing research (2020 - …)
- Papers and software library available
- So far, credal nets (CNs) mostly used for:
  - decision-support systems
  - robust machine learning
- Lot of research on CN inference/complexity
- Causal ML as a new direction for CNs
- (Causal) EM/sampling for CN inference

Structural Causal Models Are (Solvable by) Credal Networks

Marco Zaffalon ZAFFALON@IDSIA.CH
Alessandro Antonucci ALESSANDRO@IDSIA.CH
Rafael Cabañas RCABANAS@IDSIA.CH
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland

Causal Expectation-Maximisation

Marco Zaffalon, IDSIA (Switzerland), zaffalon@idsia.ch
Alessandro Antonucci, IDSIA (Switzerland), alessandro@idsia.ch
Rafael Cabañas, IDSIA (Switzerland), rcabanas@idsia.ch

Bounding Counterfactuals under Selection Bias

Marco Zaffalon ZAFFALON@IDSIA.CH
IDSIA, Lugano (Switzerland)
Alessandro Antonucci ALESSANDRO@IDSIA.CH
IDSIA, Lugano (Switzerland)
Rafael Cabañas RCABANAS@UAL.ES
Department of Mathematics, University of Almería, Almería (Spain)
David Huber DAVID.HUBER@IDSIA.CH
IDSIA, Lugano (Switzerland)
Dario Azzimonti DARIO.AZZIMONTI@IDSIA.CH
IDSIA, Lugano (Switzerland)

Credici
Credal Inference for Causal Inference

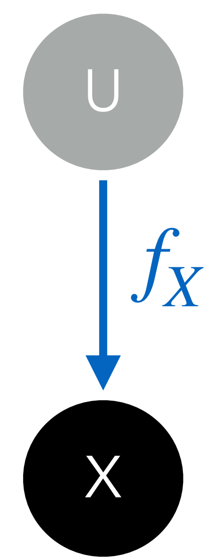
Structural Causal Models

$$K(U) = \{P(U) : P(U=0) + P(U=1) = p\}$$

$$P(U) = \left[\frac{p}{2}, \frac{p}{2}, \frac{1-p}{2}, \frac{1-p}{2}\right]$$

$U \in \{0,1,2,3\}$

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$
- A latent **exogenous** variable U
- States of U determines those of X through a **structural equation** f_X
- f_X surjective but not invertible
- $P(x) = \sum_x P(x|u)P(u) = \sum_u \delta_{f(u),x} P(u)$
- A $P(U)$ giving $P(X)$? More than one!
- Credal set $K(U)$ compatible with $P(X)$



$f_X(U=0)=0$

$f_X(U=1)=0$

$f_X(U=2)=1$

$f_X(U=3)=1$

Boolean X

$P(X=0)=p$

Structural Causal Models

- Manifest **endogenous** variable X
- Observations $\mathscr{D}$ available
- From $\mathscr{D}$ statistical learning of $P(X)$
- A lat
- State
  throu
- f_X su
- $P(x) = \sum_x P(x|u)P(u) = \sum_u \delta_{f(u),x} P(u)$
- A $P(U)$ giving $P(X)$? More than one!
- Credal set $K(U)$ compatible with $P(X)$

$K(U) = \{P(U) : P(U=0) + P(U=1) = p\}$

$P(U) = \left[\frac{p}{2}, \frac{p}{2}, \frac{1-p}{2}, \frac{1-p}{2}\right]$

$U \in \{0,1,2,3\}$

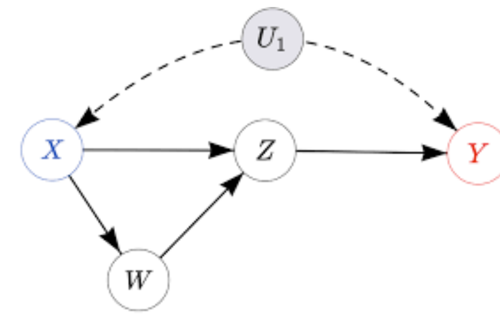
Boolean X

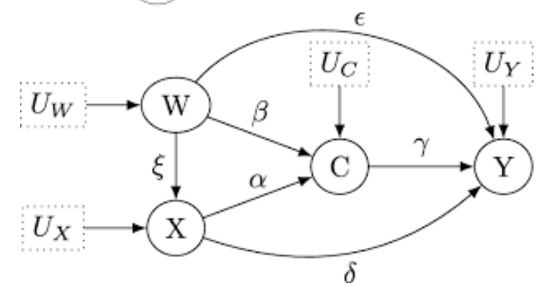
$P(X=0) = p$

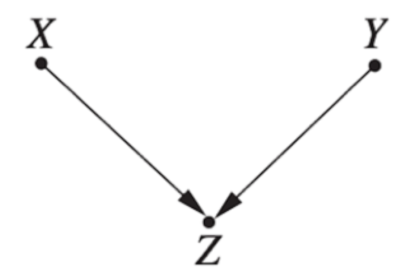
This is a (minimalistic) structural causal model

Structural Causal Models (General Definition)

- $\mathbf{X} := (X_1, \ldots, X_n)$ (endogenous variables)
- $\mathbf{U} := (U_1, \ldots, U_m)$ (exogenous variables)
- Directed graph $\mathscr{G}$ assumed to be semi-Markovian = root in $\mathbf{U}$, non-root in $\mathbf{X}$
- Equation $X = f_X(\mathrm{Pa}_X)$ for each $X \in \mathbf{X}$
- Marginal $P(U)$ for $U \in \mathbf{U}$ (assessed if possible)
- SCM = BN with CPTs $P(X|\mathrm{Pa}_X) = \delta_{X, f_X(\mathrm{Pa}_X)}$
- Joint PMF $P(\mathbf{x}, \mathbf{u}) = \prod_{U \in \mathbf{U}} P(u) \prod_{X \in \mathbf{X}} \delta_{f_X(\mathrm{pa})_X, x}$
- Here discrete vars, continuous case analogous



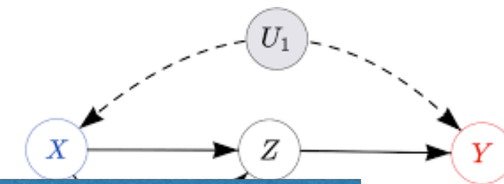




$U = \{X, Y\}, \quad V = \{Z\}, \quad F = \{f_Z\}$

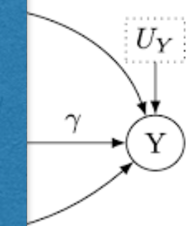
$f_Z : Z = 2X + 3Y$

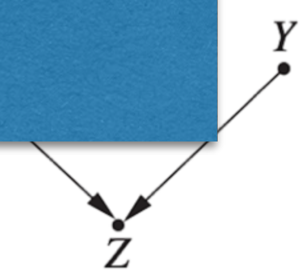
Structural Causal Models (General Definition)



- $\mathbf{X} := (X_1, \ldots, X_n)$ (endogenous variables)
- $\mathbf{U} :=$
- Direc
  semi
- Equa
- Marg
- SCM

SCMs as (one of) the most powerful tools for causal analyses





- Joint PMF $P(\mathbf{x}, \mathbf{u}) = \prod_{U \in \mathbf{U}} P(u) \prod_{X \in \mathbf{X}} \delta_{f_X(\mathrm{pa})_X, x}$
- Here discrete vars, continuous case analogous

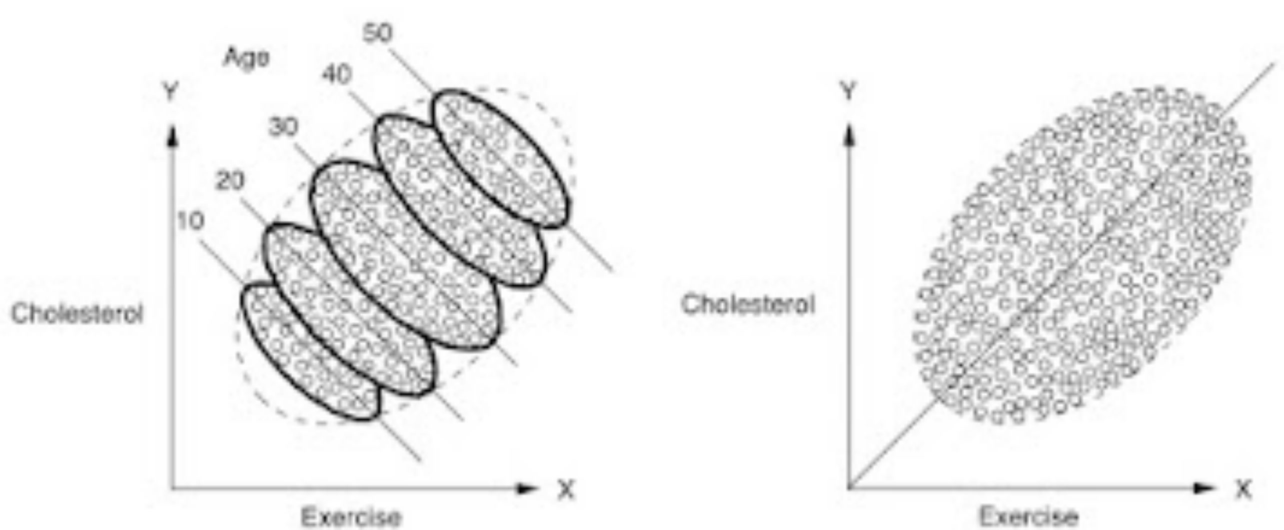
$U = \{X, Y\}, \quad V = \{Z\}, \quad F = \{f_Z\}$

$f_Z : Z = 2X + 3Y$

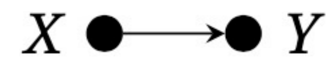
Headache Example (Staying on the First Rung)

- You take aspirin ($X=1$) and headache vanishes ($Y=1$)
- Probability that this has been due to aspirin?
- Observational data $\mathscr{D}$ about the two variables available
- From $\mathscr{D}$, $P(Y=0|X=0)=0.5>P(Y=0|X=1)=0.1$
- Not genuine causal analysis: adding further covariates might give contradictory results (Simpson's paradox)

X	Y	n
0	0	...
0	1	....
1	0	....
1	1	....



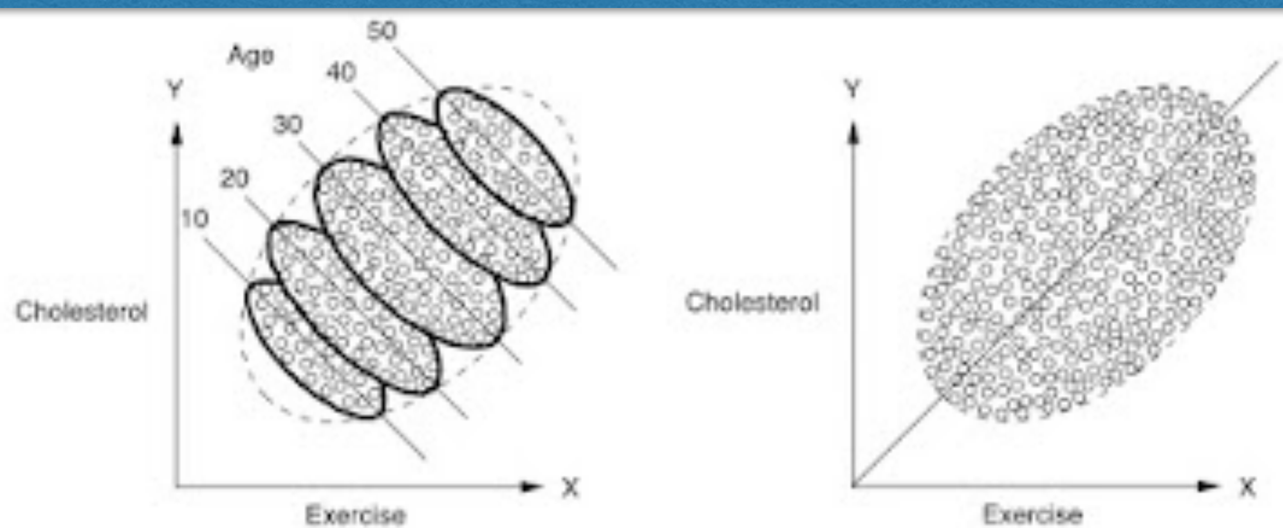
Headache Example (Staying on the First Rung)



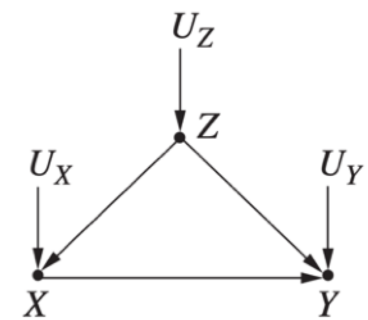
- You take aspirin ($X = 1$) and headache vanishes ($Y = 1$)
- Probability that this has been due to aspirin?
- Obs
- Fron
- Not
  migh

X	Y	n
	0	...
	1	....
	0	....
	1	....

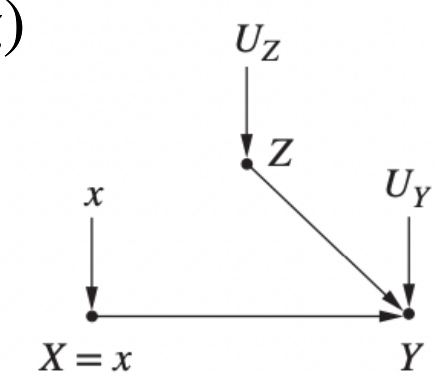
Time to climb up the ladder



Take the Aspirin! (Interventions = Second Rung)

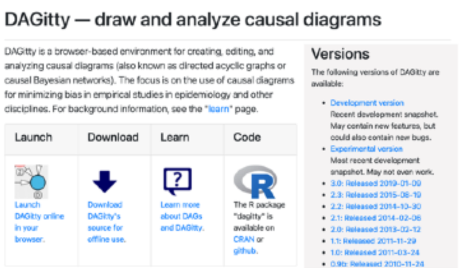


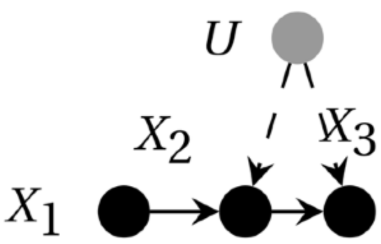
- Gender Z as an additional (endogenous) variable
- Markovian $\mathcal{G}$ (one exo parent for each endo)
- Force people to take aspirin = **intervention** $\mathrm{do}(X = 1)$
- f_X should be modified (constant output), after a **surgery** on $\mathcal{G}$ (incoming arcs removed) intervention = observation
- Pearl's **do calculus** allows to reduce interventional queries to observational ones (solved by BN inference)
- E.g., backdoor $P(y \,|\, \mathrm{do}(X = x)) = \sum_z P(y \,|\, x, z) \cdot P(z)$
- Do calculus only needs $\mathcal{G}$ (and not the SCM)!



Identifiability of Causal Queries

- Do calculus reduces interventional to observational queries by exploiting d-separation in SCMs
- Sound and complete (graph-theoretic) algorithm + inference in the empirical joint PMF
- Alternatively: surgery and inference in the SCM ...
- Not all queries can be computed by do calculus. If not we call the query **unidentifiable**
- Emerging idea: unidentifiable queries are only partially identifiable (bounds can be estimated!) Recent works by Barenboim's and Shipster's groups



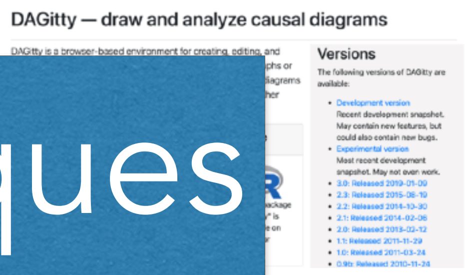


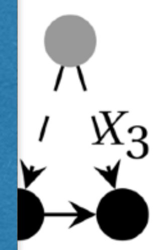
$P(x_3 \,|\, \mathrm{do}(x_2)) \in [l, u]$

Identifiability of Causal Queries

- Do calculus reduces interventional to observational quer
- Sour
  + inf
- Alter
- Not
  If no
- Emerging idea: unidentifiable queries are only partially identifiable (bounds can be estimated!)
  Recent works by Barenboim's and Shipster's groups

Optimisation techniques for IPs to be used for partial identifiability



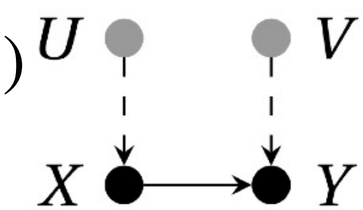


$P(x_3 \mid do(x_2) \in [l, u]$

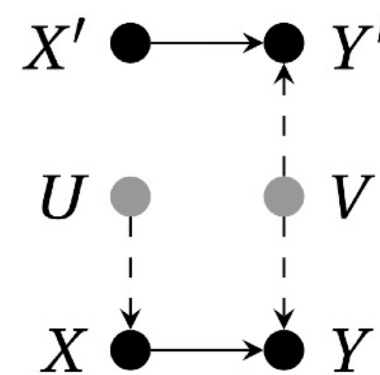
Back to Headache (Moving to the Third Rung)

- **What if** I had not taken the aspirin, would have headache stayed?

- An intervention contrasting the current observation ...
- This is a **counterfactual** query $P(Y_{X=0} = 0 | X = 1, Y = 1)$ (called probability of necessity, PN, sub denote do)



- We need the complete SCM: $\mathscr{G} + \{f_X\}_{X \in \mathbf{X}} + \{P(U)\}_{U \in \mathbf{U}}$
- With complete SCM, an augmented model called **twin network** with duplicated endogenous variables is used for counterfactual analysis after surgery
- (Non-trivial) **counterfactuals are unidentifiable**!

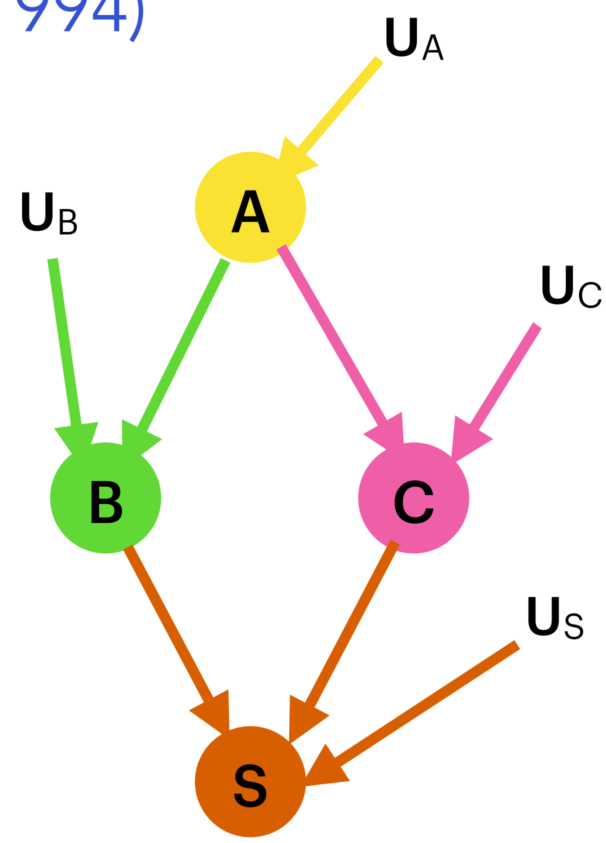


To Compute Counterfactuals ...

- We need a fully specified SCM, i.e.,
  1. Graph $\mathscr{G}$ over $(\mathbf{X}, \mathbf{U})$
     (often available by domain expert or Markovian assumption)
  2. Endogenous equations $\{f_X\}_{X \in \mathbf{X}}$
     (available or obtained by complete enumeration)
  3. Exogenous marginals $\{P(U)\}_{U \in \mathbf{U}}$ (rarely available)
- Latent $P(\mathbf{U}) = \prod P(U)$ unavailable? We have data $\mathscr{D}$ about $\mathbf{X}$
- Compute counterfactual = Compute $\{P(U)\}_{U \in \mathbf{U}}$ from $\mathscr{D}$
- Not a new problem: LP approach for special cases already in Balke and Pearl (1994), but do-calculus reduced attention to CFs

Causal Analysis at the Party (Balke & Pearl 1994)

Ann sometimes goes to parties
Bob is not a party guy,
but he likes Ann
and he might be there
Carl broke up with Ann,
he tries to avoid Ann,
but he likes parties
Carl and Bob hate each other,
they might have a Scuffle
if both at the party



besides such knowledge assume we have observations $\mathscr{D}$ corresponding to a joint mass function $P(A, B, C, S)$ (e.g., in the form of a BN)

Causal Analysis at the Party (Balke & Pearl 1994)

CAUSAL GOSSIP

INTERVENTIONAL

"**A**nn must not be at the party, or **B**ob would be there instead of home"

$$P(B \mid \mathrm{do}(\bar{a})) = ?$$

COUNTERFACTUAL

"If **B**ob were at the party, then Bob and Carl would surely **S**cuffle"

$$P(S_b \mid \bar{b}) = ?$$

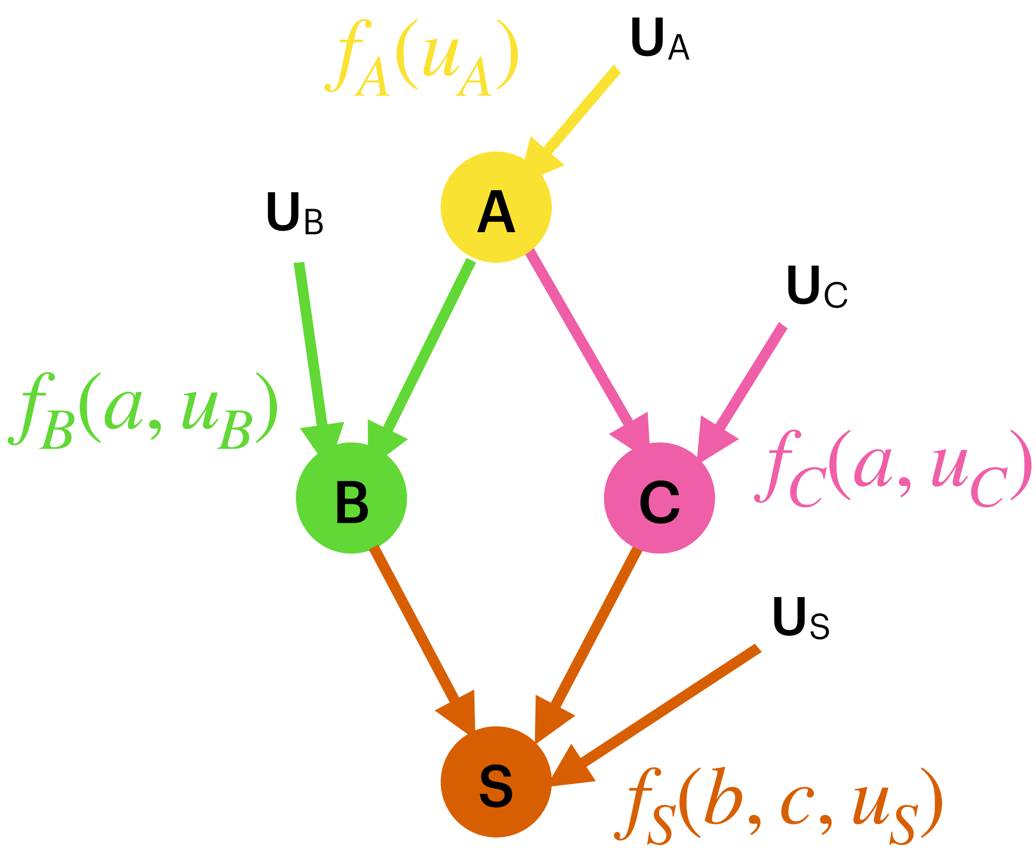
a (fully specified) SCM can answer these questions

Let's (Eventually) Use IPs!

- Find the exogenous marginals?

  $P(U_A)P(U_B)P(U_C)P(U_S)$

- **Endogenous** (= with $\mathscr{D}$) **consistency**
- This induces global non-linear (so-called Verma) constraints
- Constraints became local and linear ones by marginalisation and conditioning (Zaffalon et al., 2020)



Unknown, Unknown, Unknown, Unknown, Empirical, known

$$\sum_{u_A,u_B,u_C,u_D} \left[p(u_A)\cdot\delta_{a,f_A(u_A)}\cdot p(u_B)\cdot\delta_{b,f_B(a,u_B)}\cdot p(u_C)\cdot\delta_{c,f_C(a,u_c)}\cdot p(u_S)\cdot\delta_{s,f_S(b,c,u_S)}\right] = \tilde{p}(a,b,c,s)$$

Constraining Exogenous Marginals

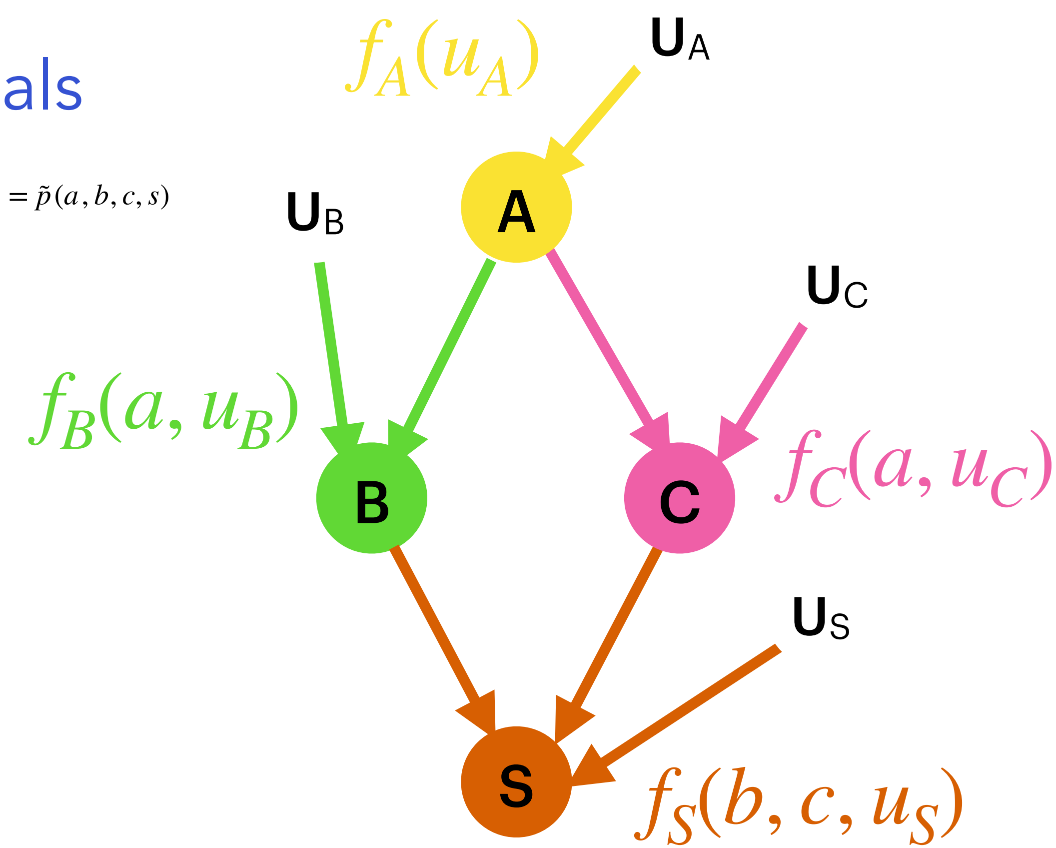
$$\sum_{u_A,u_B,u_C,u_D}\left[p(u_A)\cdot\delta_{a,f_A(u_A)}\cdot p(u_B)\cdot\delta_{b,f_B(a,u_B)}\cdot p(u_C)\cdot\delta_{c,f_C(a,u_C)}\cdot p(u_S)\cdot\delta_{s,f_S(b,c,u_S)}\right]=\tilde{p}(a,b,c,s)$$

$$P(a)=\sum_{u_A}P(a\,|\,u_A)\cdot P(u_A)$$

$$P(b\,|\,a)=\sum_{u_B}P(b\,|\,a,u_B)\cdot P(u_B)$$

$$P(c\,|\,a)=\sum_{u_C}P(c\,|\,a,u_C)\cdot P(u_C)$$

$$P(s\,|\,b,c)=\sum_{u_S}P(s\,|\,b,c,u_S)\cdot P(u_S)$$



- Linear constraints on marginal exogenous probabilities leading to the credal sets specification $K(U_A)$, $K(U_B)$, $K(U_C)$, $K(U_S)$
- Structural equations (= endogenous CPTS) remain unaffected

Constraining Exogenous Marginals

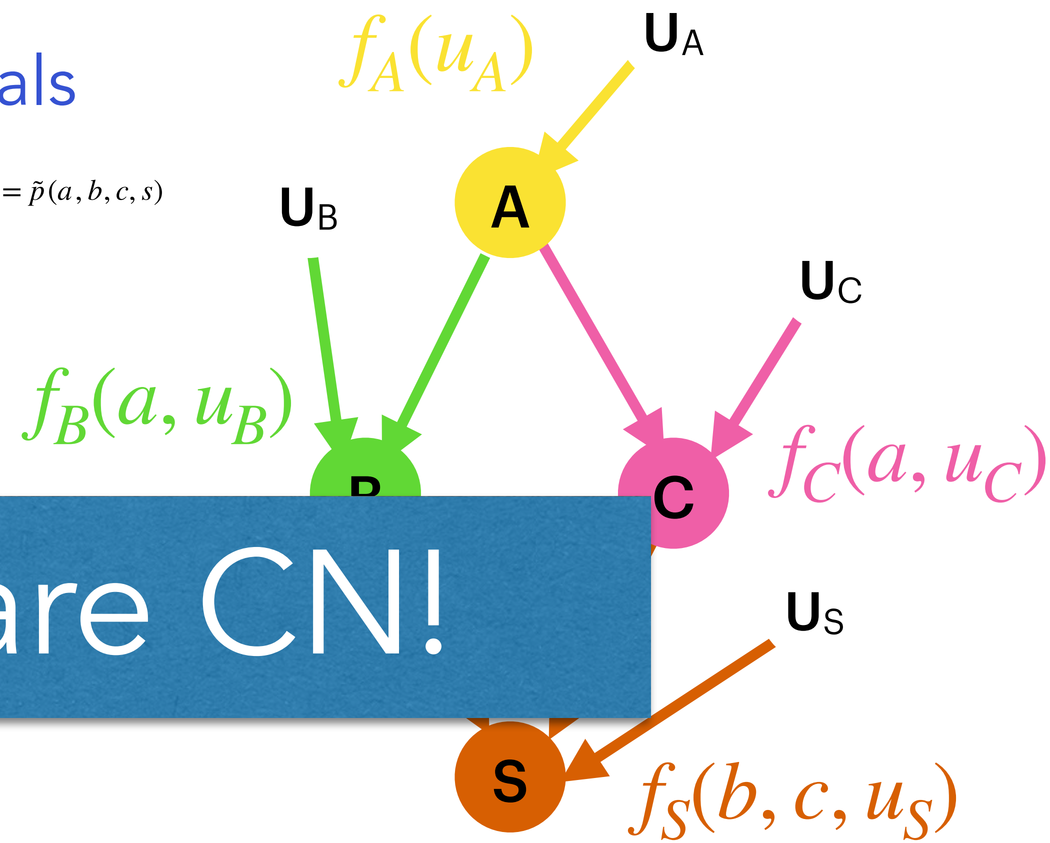
$$\sum_{u_A,u_B,u_C,u_D}\left[p(u_A)\cdot\delta_{a,f_A(u_A)}\cdot p(u_B)\cdot\delta_{b,f_B(a,u_B)}\cdot p(u_C)\cdot\delta_{c,f_C(a,u_C)}\cdot p(u_S)\cdot\delta_{s,f_S(b,c,u_S)}\right]=\tilde{p}(a,b,c,s)$$

$$P(a)=\sum_{u_A}P(a\,|\,u_A)\cdot P(u_A)$$

$$P(b\,|\,a)=\sum_{u_B}\ldots$$

$$P(c\,|\,a)=\sum_{u_C}\ldots$$

$$P(s\,|\,b,c)=\sum_{u_S}P(s\,|\,b,c,u_S)\cdot P(u_S)$$

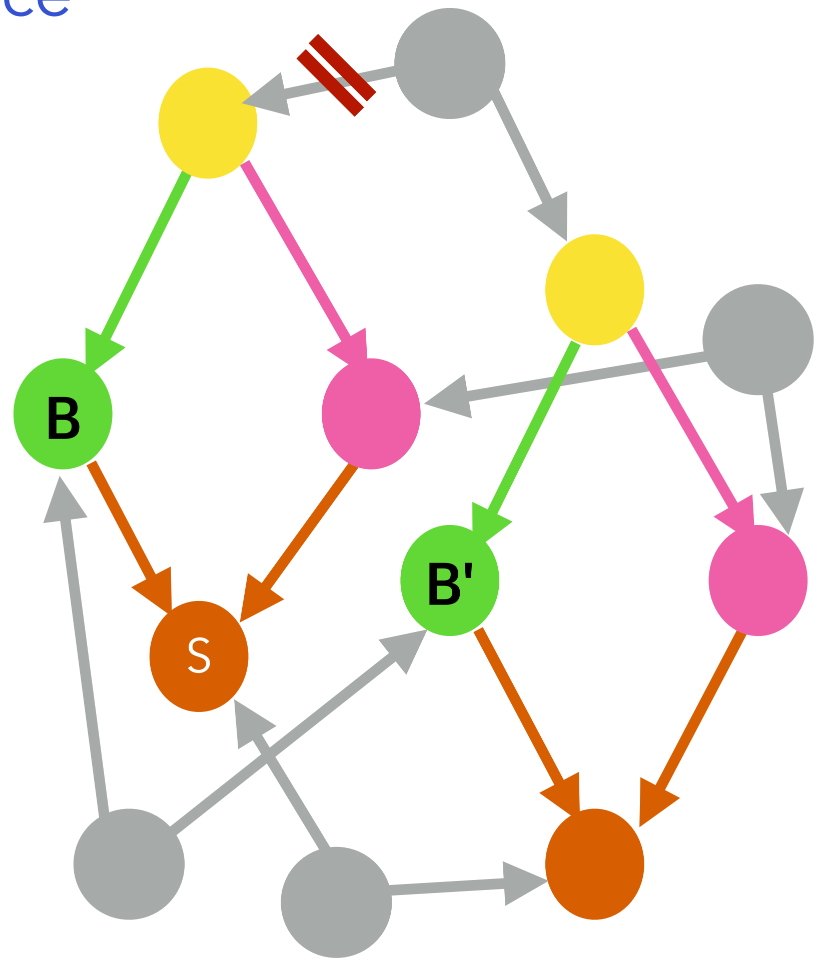


SCMs are CN!

- Linear constraints on marginal exogenous probabilities leading to the credal sets specification $K(U_A)$, $K(U_B)$, $K(U_C)$, $K(U_S)$
- Structural equations (= endogenous CPTS) remain unaffected

Reducing Causal Queries to CN Inference

- Consistent SCMs as a single CN
- d-separation holds for CNs, we can do surgery à la Pearl
- CN algs to compute bounds!
- Interventions are straightforward $P(B\,|\,\mathrm{do}(\overline{a})) \in [\underline{P}'(B\,|\,\overline{a}), \overline{P}'(B\,|\,\overline{a})]$
- Counterfactuals require twin nets $P(S_b\,|\,\overline{b}) \in [\underline{P}(S\,|\,b,\overline{b}'), \overline{P}(S\,|\,b,\overline{b}')]$
- Identifiable? $\underline{P} = \overline{P}$

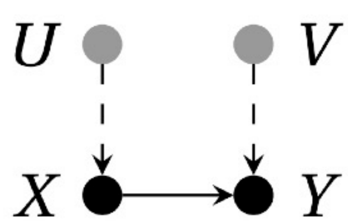


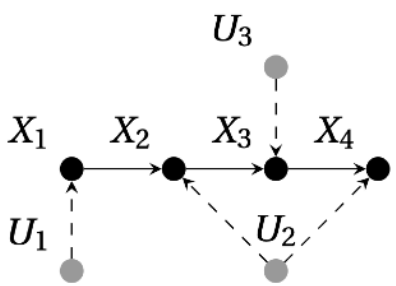
Markovian and Quasi-Markovian SCMs as CNs

Algorithm 1 Given an SCM M and a PMF $\tilde{P}(\boldsymbol{X})$, return CSs $\{K(U)\}_{U\in\boldsymbol{U}}$

for $X \in \boldsymbol{X}$ **do**
  $U \leftarrow \mathrm{Pa}(X) \cap \boldsymbol{U}$ // U as the unique exogenous parent of X
  $\underline{\mathrm{Pa}}(X) \leftarrow \mathrm{Pa}(X) \setminus \{U\}$ // Endogenous parents of X
  if $\underline{\mathrm{Pa}}(X) = \emptyset$ **then**
    $K(U) \leftarrow \{P'(U) : \sum_{u \in f_X^{-1}} P'(u) = \tilde{P}(x), \forall x \in \Omega_X\}$ // Eq. (4)
  else
    $K(U) \leftarrow \{P'(U) : \sum_{u \in f^{-1}_{X|\underline{\mathrm{pa}}(X)}(x)} P'(u) = \tilde{P}(x|\underline{\mathrm{pa}}(X)), \forall x \in \Omega_X, \forall \underline{\mathrm{pa}}(X) \in \Omega_{\underline{\mathrm{Pa}}_X}\}$ // Eq. (6)
  end if
end for

Markovian Models





Quasi-Markovian Models

Algorithm 2 Given an SCM M and a PMF $\tilde{P}(\boldsymbol{X})$, return CSs $\{K(U)\}_{U\in\boldsymbol{U}}$

for $U \in \boldsymbol{U}$ **do**
  $\{X_U^k\}_{k=1}^{n_U} \leftarrow \mathrm{Sort}[X \in \boldsymbol{X} : U \in \mathrm{Pa}(X)]$ // Children of U in topological order
  $\gamma \leftarrow \emptyset$
  for $(x_U^1, \ldots, x_U^{n_U}) \in \times_{k=1}^{n_U} \Omega_{X_U^k}$ **do**
    for $(\underline{\mathrm{pa}}(X_U^1), \ldots, \underline{\mathrm{pa}}(X_U^{n_U})) \in \times_{k=1}^{n_U} \Omega_{\underline{\mathrm{Pa}}(X_U^k)}$ **do**
      $\Omega'_U \leftarrow \bigcap_{k=1}^{n_U} f^{-1}_{X_U^k|\underline{\mathrm{pa}}(X_U^k)}(x_U^k)$
      $\gamma \leftarrow \gamma \cup \left\{\sum_{u \in \Omega'_U} P(u) = \prod_{k=1}^{n_U} \tilde{P}(x_U^k | x_U^1, \ldots, x_U^{k-1}, \underline{\mathrm{pa}}(X_U^1)), \ldots, \underline{\mathrm{pa}}(X_U^k))\right\}$
    end for
  end for
  $K(U) \leftarrow \{P(U) : \gamma\}$ // CS by linear constraints on $P(U)$
end for

GitHub

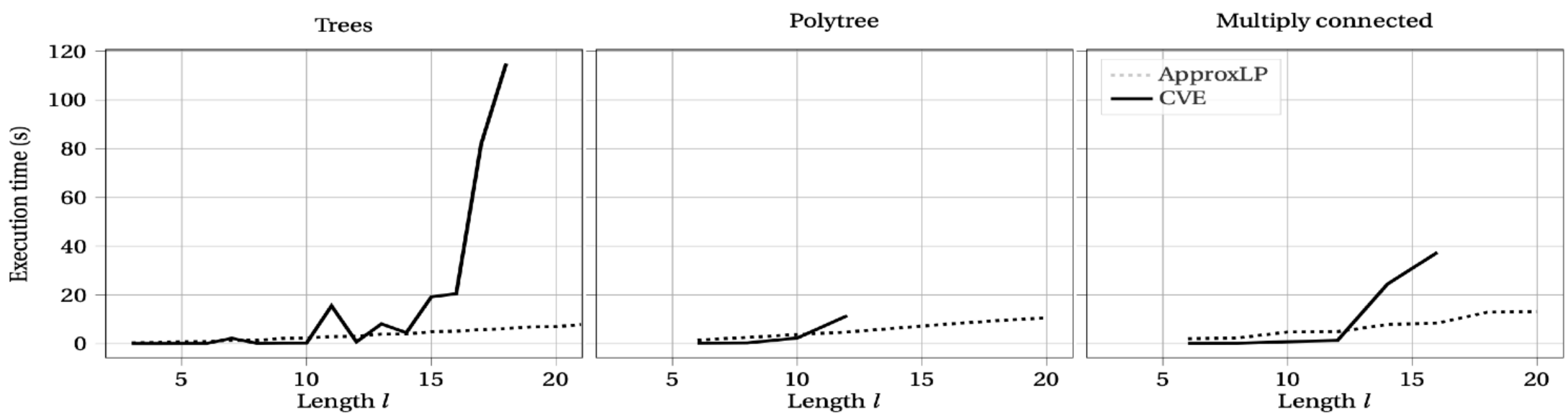
Software and Experiments

Crema
Credal Models Algorithms

Java library for CNs

Credici
Credal Inference for Causal Inference

Java library for Causal Inference built on the top of CREMA



Exact inference by credal variable elimination only for small models
ApproxLP (Antonucci et al., 2014) allows to process larger models
RMSE always <0.7%

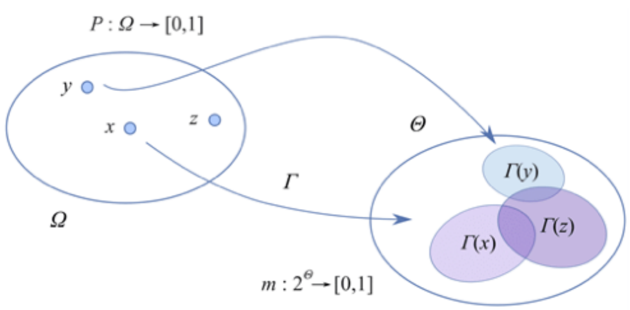
Intermezzo: Belief Functions (as Credal Sets)

- Linear constraints for CN induced by SCM have a peculiar form
- These are CS corresponding to **belief functions** (Dempster '68, Shafer '76)
- Class of generalised probabilistic models
- PMF distributes mass over the singletons, BF over (poss. overlapping) sets
- Dempster's **multi-valued mapping**, in SCMs $\mathbf{U} = f^{-1}(\mathbf{X})$, $\mathrm{BF}(\mathbf{U}) := f^{-1}[P(\mathbf{X})]$
- Dedicated conditioning/combination rules

$$\sum_{u:\text{condition}} P(u) = \text{const}$$

$$\sum_{u \in \Omega'_U} P(u) = \prod_{k=1}^{n_U} \tilde{P}(x_U^k | x_U^1, \ldots, x_U^{k-1}, \underline{\mathrm{pa}}(X_U^1)), \ldots, \underline{\mathrm{pa}}(X_U^k))$$

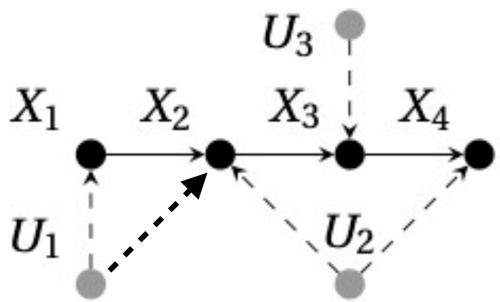
$$\{P'(U) : \sum_{u \in f^{-1}_{X|\underline{\mathrm{pa}}(X)}(x)} P'(u) = \tilde{P}(x|\underline{\mathrm{pa}}(X)), \forall x \in \Omega_X, \forall \underline{\mathrm{pa}}(X) \in \Omega_{\underline{\mathrm{Pa}}_X}\}$$



Credits: Fabio Cuzzolin

Back to SCM2CN: General (Non Quasi-Markovian) Case

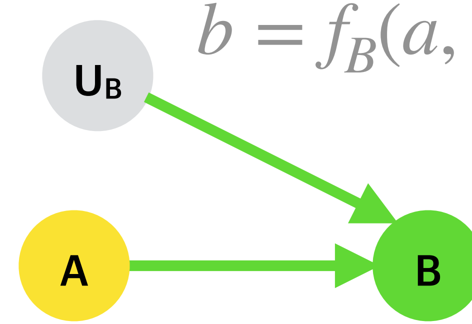
- Non Quasi-Markovian? Non-Linear constraint
- E.g., $\sum P(u_1) \cdot P(u_2) = \ldots$
- Merge exogenous variables $U := (U_1, U_2)$
- Independence constraints can be disregarded (but higher exogenous dimensionality)
- Again CN approximate inference to solve causal queries
- State space dimensionality affects complexity
- We might have very large latent spaces ...



Conservative Specification of Structural Equations

- Finding the equations given $\mathcal{G}$ only
- $P(B\,|\,A)$ should be a deterministic CPT

$$b = f_B(a, u_B)$$



$$P(B\,|\,A)$$

	A=0	A=1
B=0	1	1
B=1	0	0

$B = 0$

	A=0	A=1
B=0	1	0
B=1	0	1

$B = A$

	A=0	A=1
B=0	0	1
B=1	1	0

$B = \neg A$

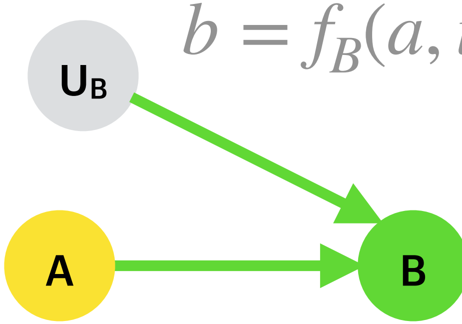
	A=0	A=1
B=0	0	0
B=1	1	1

$B = 1$

Conservative Specification of Structural Equations

- Finding the equations given $\mathscr{G}$ only
- $P(B \mid A)$ should be a deterministic CPT
- U_B indexing all these deterministic CPTs

$$b = f_B(a, u_B)$$

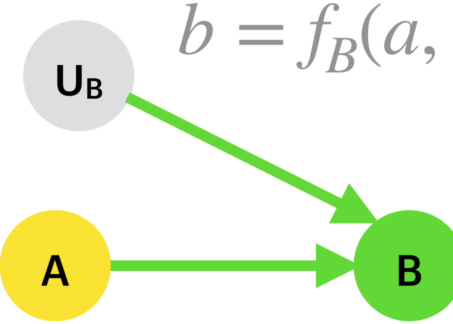


$$P(B \mid A, U)$$

	A=0	A=1	A=0	A=1	A=0	A=1	A=0	A=1
B=0	1	1	1	0	0	1	0	0
B=1	0	0	0	1	1	0	1	1
	U=0		U=1		U=2		U=3	
	$B = 0$		$B = A$		$B = \neg A$		$B = 1$	

Conservative Specification of Structural Equations

$$b = f_B(a, u_B)$$



- Finding the equations given $\mathscr{G}$ only
- $P(B\,|\,A)$ should be a deterministic CPT
- U_B indexing all these deterministic CPTs
- Knowledge might discard some states (ex., Bob goes to the party if Ann does)
- With Boolean parent & child) $|\,U\,| = 4$ in general (exp size) :

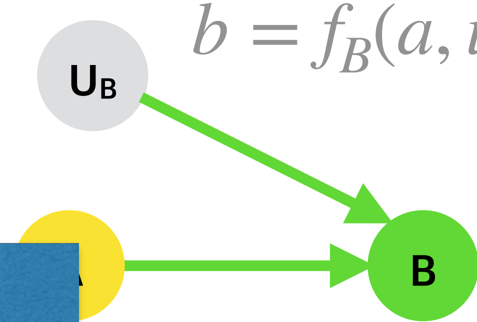
$$|\,U\,| = |\,X\,|^{\prod_{Y \in \mathrm{Pa}_Y} |\mathrm{Y}|}$$

$$P(B\,|\,A, U)$$

	A=0	A=1	A=0	A=1	A=0	A=1	A=0	A=1
B=0			1	0			0	0
B=1			0	1			1	1
	U=0		U=1		U=2		U=3	
	$B=0$		$B=A$		$B=\neg A$		$B=1$	

Conservative Specification of Structural Equations

$b = f_B(a, u_B)$



- Finding the equations given $\mathscr{G}$ only
- $P(B|A)$ should be a deterministic CPT
- U_B indexing
- Knowledge
  (ex., Bob goe
- With Boolea
  in general (exp size) :

$$|U| = |X|^{\prod_{Y \in \mathrm{Pa}_Y} |Y|}$$

CFs based on $\mathscr{G}$ and $\mathscr{D}$ only

$B|A, U)$

	A=1	A=0	A=1	A=0	A=1
B=0	0			0	0
B=1	1			1	1

	U=0	U=1	U=2	U=3
B=0		1, 0		0, 0
B=1		0, 1		1, 1
	$B = 0$	$B = A$	$B = \neg A$	$B = 1$

An Application: Counterfactual Analysis in Palliative Cares

- Study of terminally ill cancer patients' preferences wrt their place of death (home or hospital)
- $\mathscr{G}$ obtained by expert knowledge and data
- Exogenous variables?
- Markovian assumption (= no confounders)

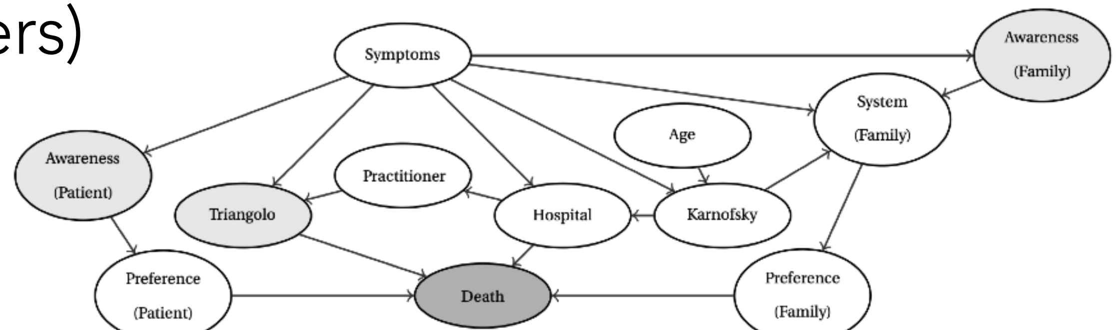
ASSOCIAZIONE TRIANGOLO

qualitépalliative marchio di qualità "in cure palliative"

servizio cure palliative e domiciliari

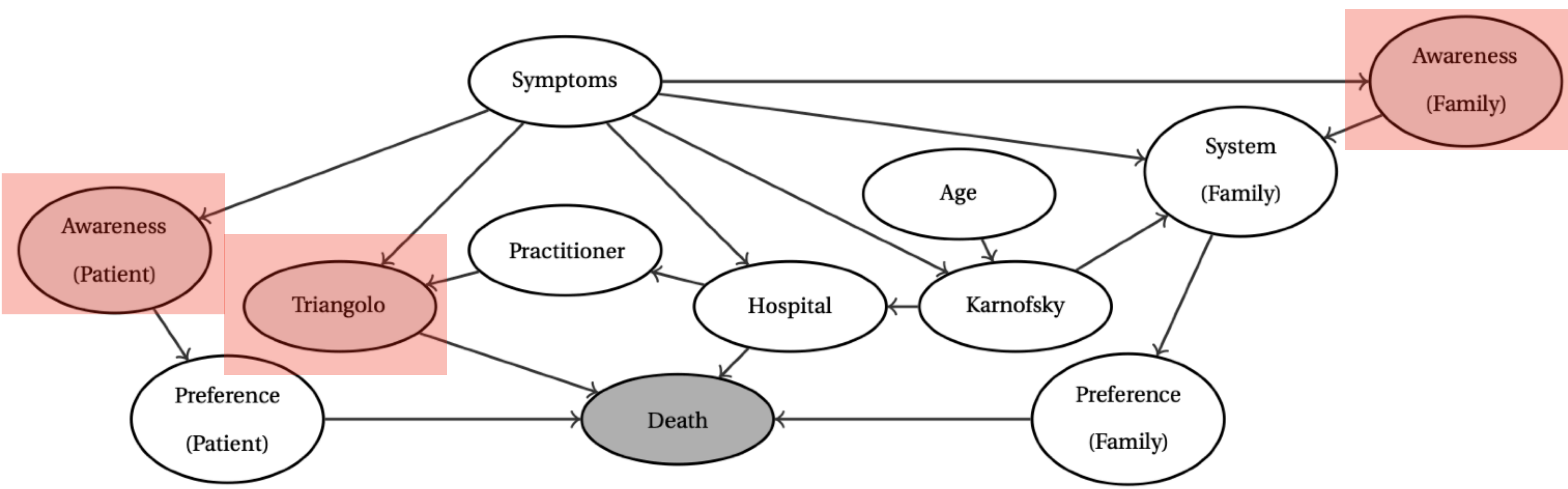

Impact on place of death in cancer patients: a causal exploration in southern Switzerland

Heidi Kern [1], Giorgio Corani [2], David Huber [2], Nicola Vermes [2], Marco Zaffalon [2], Marco Varini [3], Claudia Wenzel [4], André Fringer [5]



An Application: Counterfactual Analysis in Palliative Cares

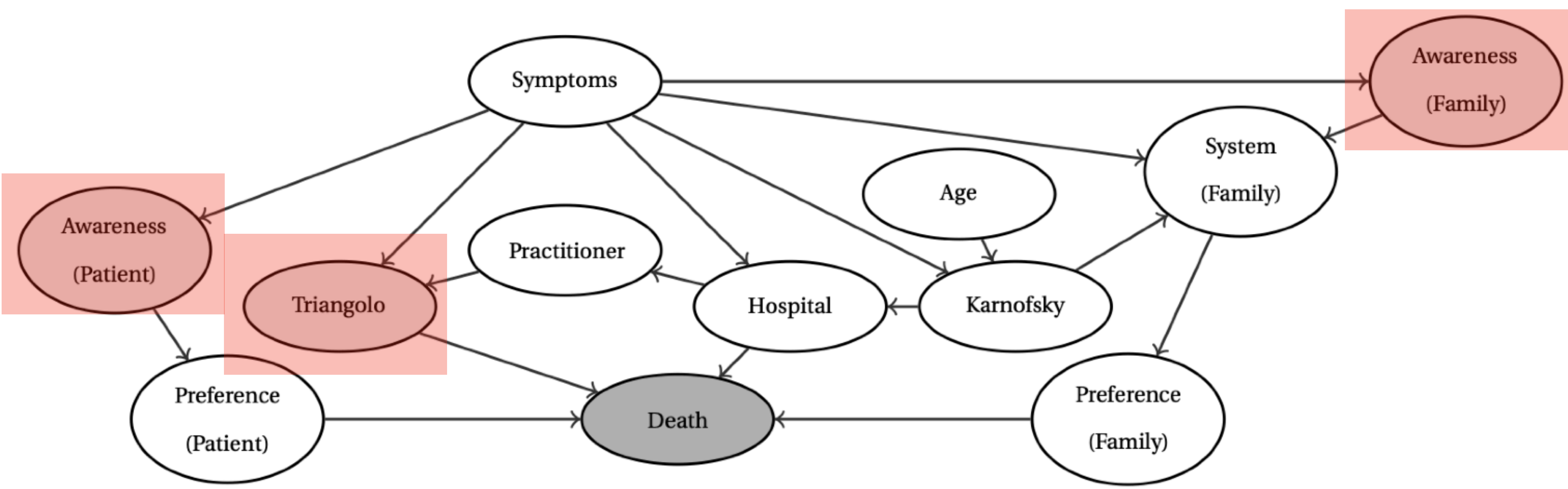
- Most patients prefer to die at home
- But a majority actually die in institutional settings
- Interventions by health care professionals can facilitate dying at home?



An Application: Counterfactual Analysis in Palliative Cares

- Importance of a variable?
- Probability of necessity and sufficiency

$$PNS := P(Y_{X=1} = 1, Y_{X=0} = 0)$$



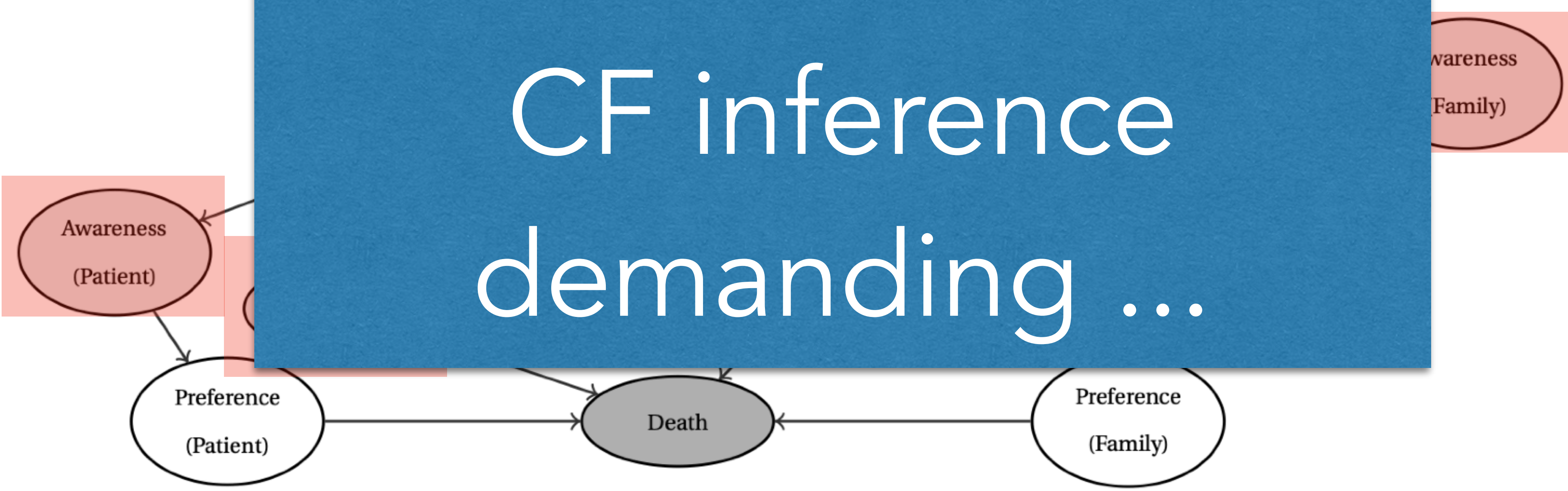
An Application: Counterfactual Analysis in Palliative Cares

- Importance of a variable?
- Probab

P

Small CN but large cardinalities

CF inference demanding ...



Causal Expectation Maximisation (Zaffalon et al., 2021)

- Exogenous variables are always missing (MAR, asystematic, way)
- Expectation Maximisation (Dempster 1977)
  – Random initialisation of P(U)
  – E-step: Missing data completion by expected (fractional) counts
  – M-step: "completed" data to retrain P(U)
  – Iterate until convergence
- EM goes to a (local/global) max of $\log P(\mathscr{D})$

U1	U2	X1	X2	n
*	*	0	0	...
*	*	0	1	...
*	*	1	0	...
*	*	1	1	...

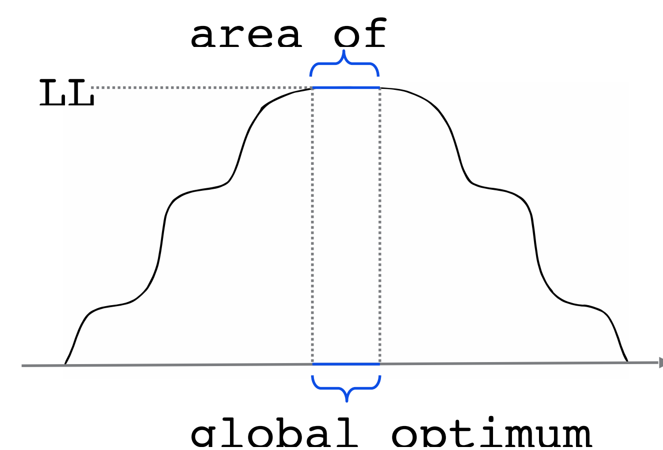
Casual EM: Likelihood Unimodality

- Causal EM reduce should converge to global maxima only the corresponding $P(U)$ belongs to credal set $K(U)$
- Sampling initialisations = sampling of $K(U)$
- For each sample we obtain an inner point

Theorem 1. *Let $\mathcal{K}$ denote the set of quantifications for $\{P(U)\}_{U \in \boldsymbol{U}}$ consistent with the following constraint to be satisfied for each $c \in \mathcal{C}$ and each $\boldsymbol{y}^{(c)}$:*

$$\sum_{\substack{\boldsymbol{u}^{(c)}: f_X(\mathrm{pa}_X)=x \\ \forall X \in \boldsymbol{X}^{(c)}}} \prod_{U \in \boldsymbol{U}^c} P(u) = \prod_{X \in \boldsymbol{X}^{(c)}} \hat{P}(x|\boldsymbol{y}_X^{(c)}), \quad (8)$$

where the values of u, x and $\boldsymbol{y}_X^{(c)}$ are those consistent with $\boldsymbol{u}^{(c)}$ and $\boldsymbol{y}^{(c)}$. If $\mathcal{K} \neq \emptyset$, the log-likelihood in Eq. (7) achieves its global maximum if and only if $\{P(U)\}_{U \in \boldsymbol{U}} \in \mathcal{K}$. If $\mathcal{K} = \emptyset$, the marginal log-likelihood in Eq. (7) can only take values strictly lower than the global maximum.



Casual EM: Guarantees?

- We first reduced causal queries to CN inference
- Cau … e
- Ide…
  (a n…
- Uni…
- Cre…

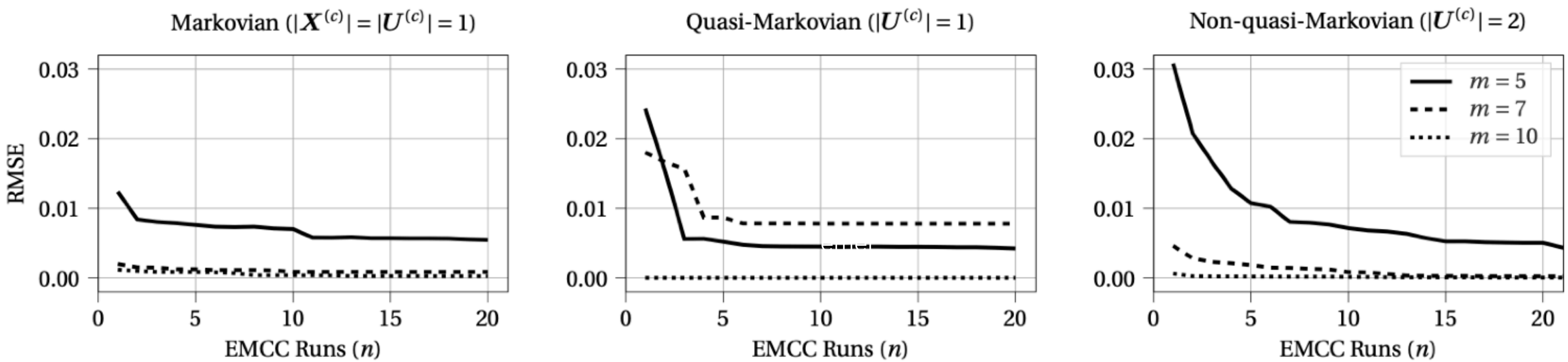
In practice?
20 EM runs to get close to the actual bounds with 95% credibility
For identifiable queries 9 runs to be sure with 99% credibility

Theore… *are the…*
and $b := \max_{i=1} r_i$. *By construction* $a^* \leq a \leq b \leq b^*$. *The following inequality holds:*

$$P\left(a - \varepsilon L \leq a^* \leq b^* \leq b + \varepsilon L \,\middle|\, \rho\right) = \frac{1 + (1+2\varepsilon)^{2-n} - 2(1+\varepsilon)^{2-n}}{(1 - L^{n-2}) - (n-2)(1-L)L^{n-2}}, \tag{13}$$

where $L := (b - a)$ *and* $\varepsilon := \delta/(2L)$ *is the relative error at each extreme of the interval obtained as a function of the absolute allowed error* $\delta \in (0, L)$.

Causal EM: Experiments



PNS for artificial SMCs: quick convergence
(= much faster than direct CN approach)

Credici
Credal Inference for Causal Inference

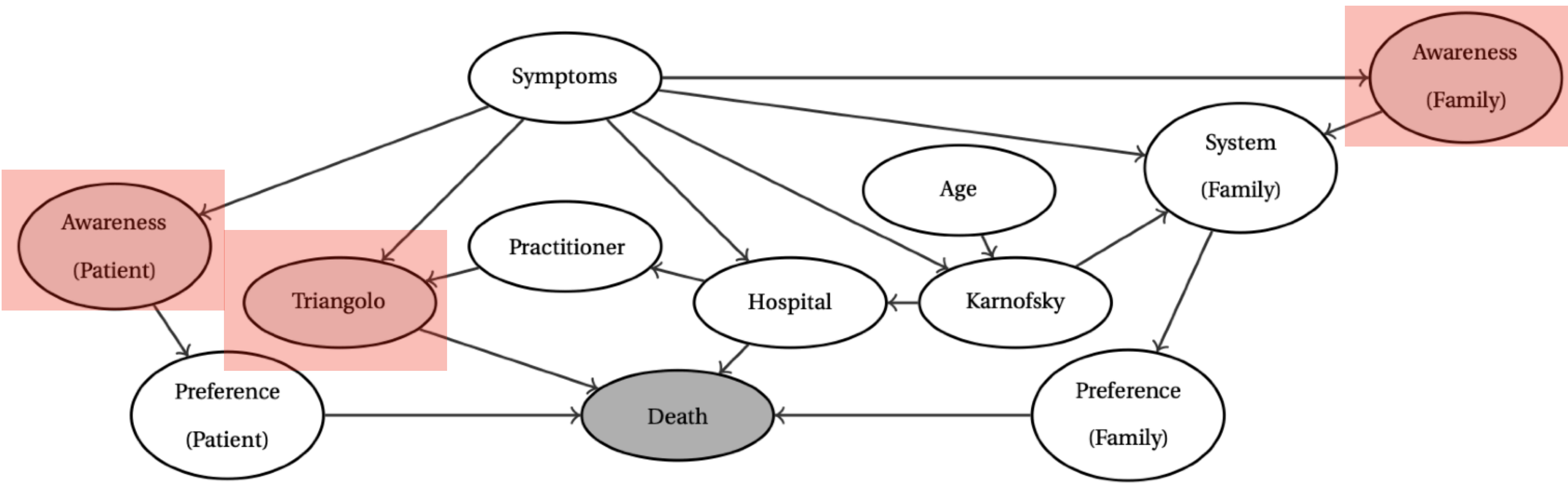
Counterfactual Analysis in Palliative Cares by Causal EM

- Importance of a variable?
- Probability of necessity and sufficiency

$$PNS := P(Y_{X=1} = 1, Y_{X=0} = 0)$$

- 15 EM runs before convergence

PNS(Family_Awareness) ∈ [0.06,0.10]



PNS(Patient_Awareness) ∈ [0.03,0.10]

PNS(Triangolo) ∈ [0.30,0.31]

Coun

- Im
- Pr
- 15

One should act on Triangolo first: for instance, by making Triangolo available to all patients, we should expect a reduction of people at the hospital by 30%

This would save money too, and would allow politicians to do economic considerations as to which amount it is even economically profitable to fund Triangolo, and have patients die at home, rather than spending more to have patients die at the hospital

.06,0.10]

Aware (Patie

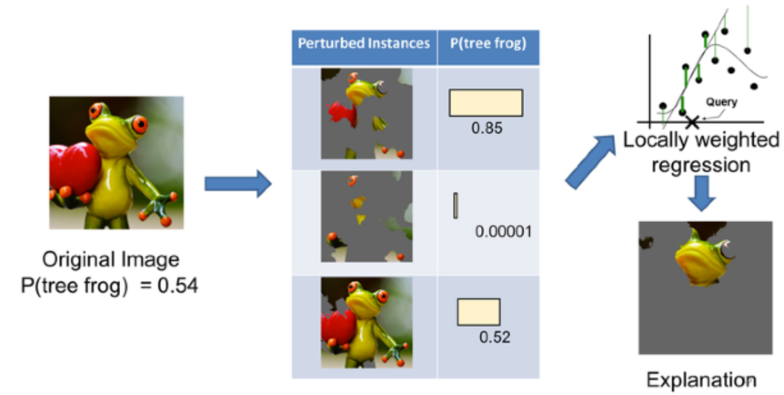
reness mily)

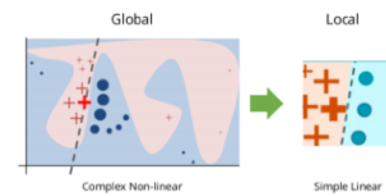
PNS(Patient_Awareness) ∈ [0.03,0.10]

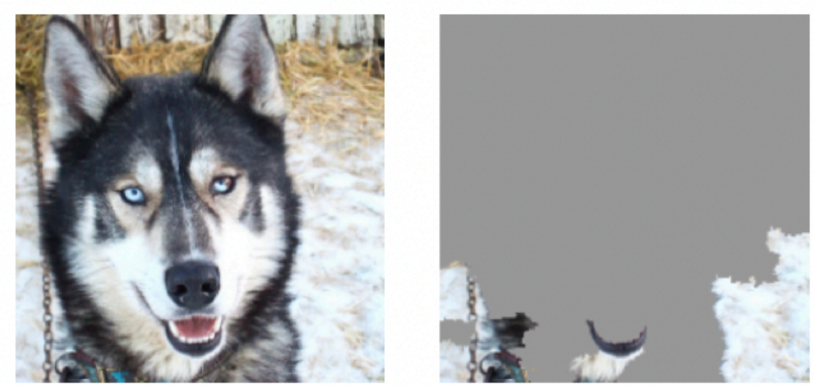
PNS(Triangolo) ∈ [0.30,0.31]

Reasons for Causal AI: XAI

- (Model-agnostic) XAI tools are observational
- Ex. Local Interpretable Model-agnostic Explanations (LIME)
- No genuine CF analysis
- Results prone to attacks/ contractions

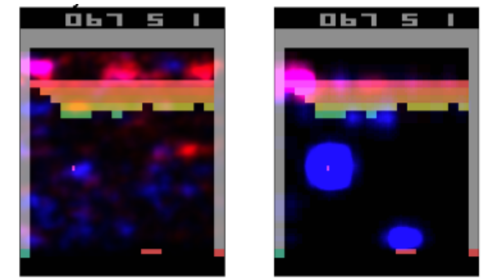




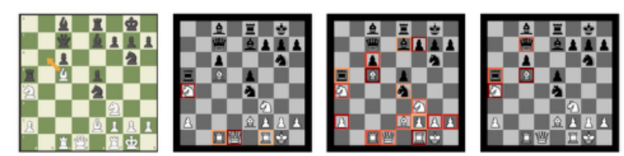

(a) Husky classified as wolf (b) Explanation

Explaining Reinforcement Learning Agents

- Agent operating in state space $\mathcal{S}$
- Set of actions $\mathcal{A}_s$
- Q(uality)-value function $Q(s, a)$ available for each $s \in \mathcal{S}$ and $a \in \mathcal{A}_s$
- Greedy agent $\hat{a} = \arg\max_a Q(a, s)$
- For each feature f compute its saliency $S[f]$
- s' perturbation of s obtained by changing the value of f
- $S[f]$ corresponds to the Q-value change
- E.g., Iyer (2018): $S[f] = Q(s, \hat{a}) - Q(s', \hat{a})$



- Saliency maps can be created by means of the computed saliency levels

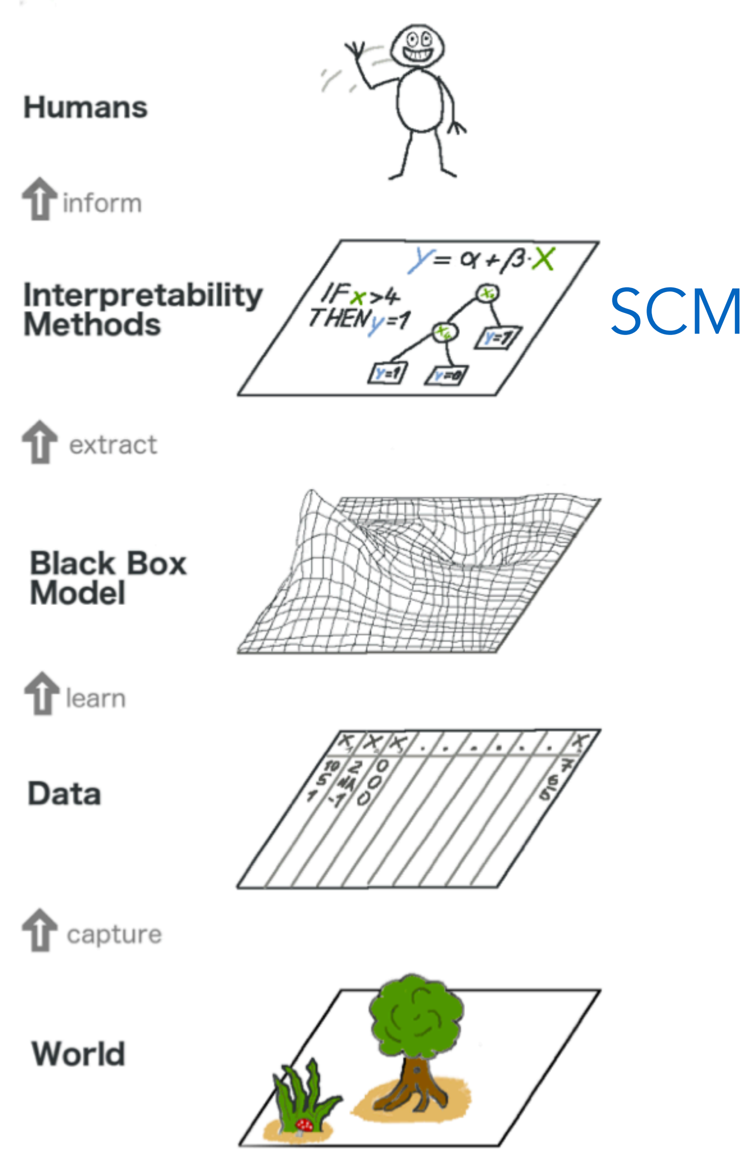


(a) Original Position (b) Iyer et al. (2018) (c) Greydanus et al. (2018) (d) SARFA

same issues as for classifiers/ regressors

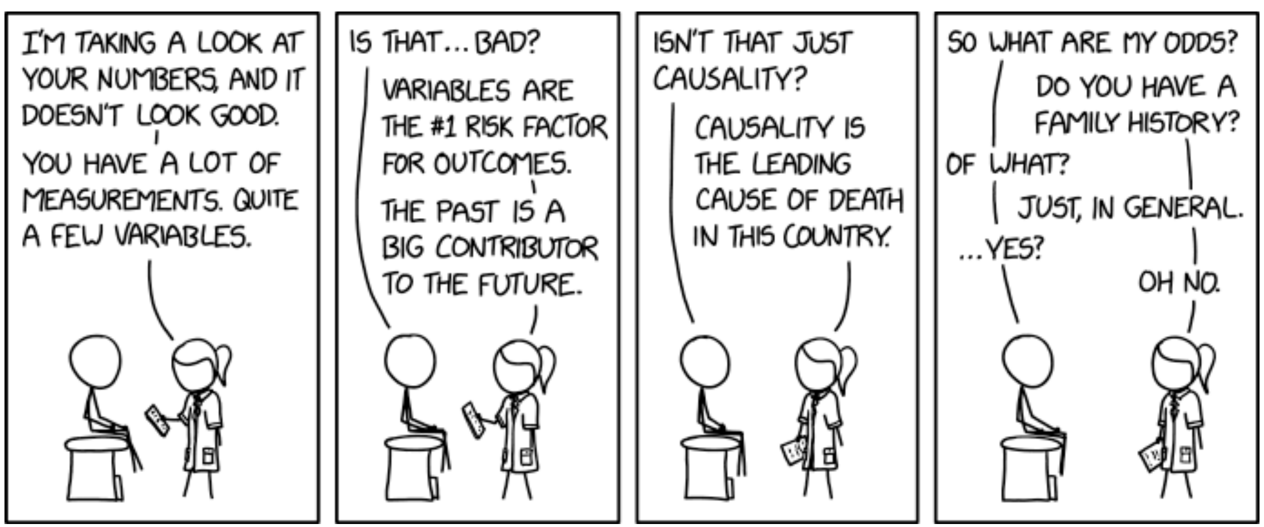
Counterfactual Explanations

- Causal analysis distringuishes between observations and interventions $P(X \mid y) \neq P(X \mid \mathrm{do}(y))$
- This allows for WHAT-IF reasoning Counterfactuals? $P(x' \mid x, y, \mathrm{do}(y'))$
- "if an input datapoint were x′ instead of x, then an ML model's output would be y′ instead of y



Conclusions

- Causality has an intimate connection with IPs
- Past CN research might offer new tools for causal analysis
- But more than that IPs offer formalism for a deeper understanding of those (structural causal) models
- Lot of works has to be done, causal machine (and reinforcement) learning are just at the beginning!



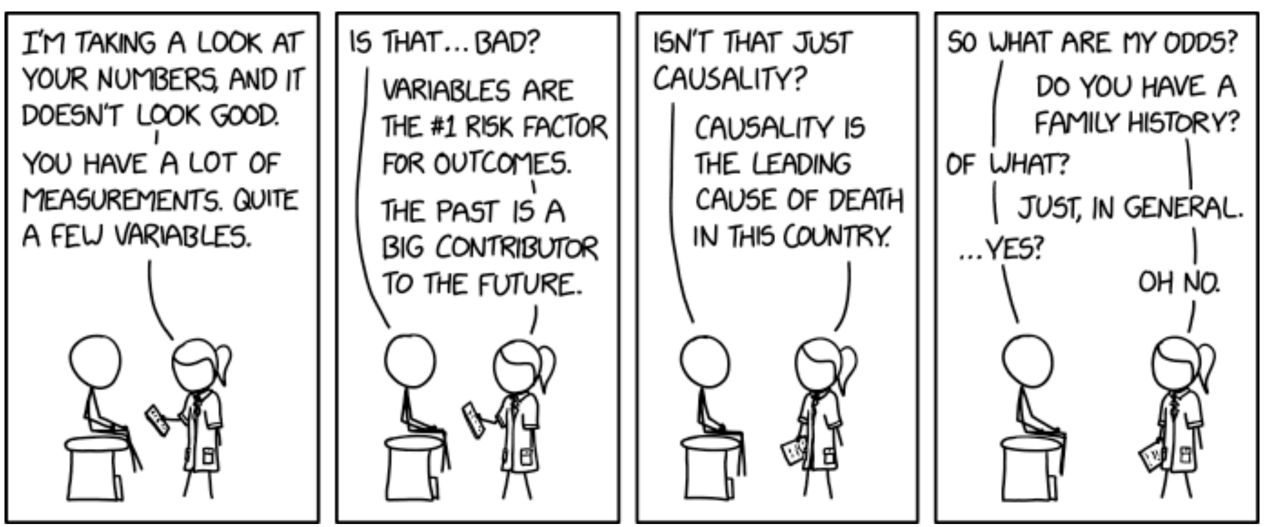
https://xkcd.com/2620/

Conclusions

- Causality has an intimate connection with IPs
- Past CN research might offer new tools for causal analysis
- But more
  understa

alessandro@idsia.ch

- Lot of works has to be done, causal machine (and reinforcement) learning are just at the beginning!



https://xkcd.com/2620/