Causal Structure Search:

Philosophical Foundations and Problems

Richard Scheines & Peter Spirtes

Carnegie Mellon University

Outline

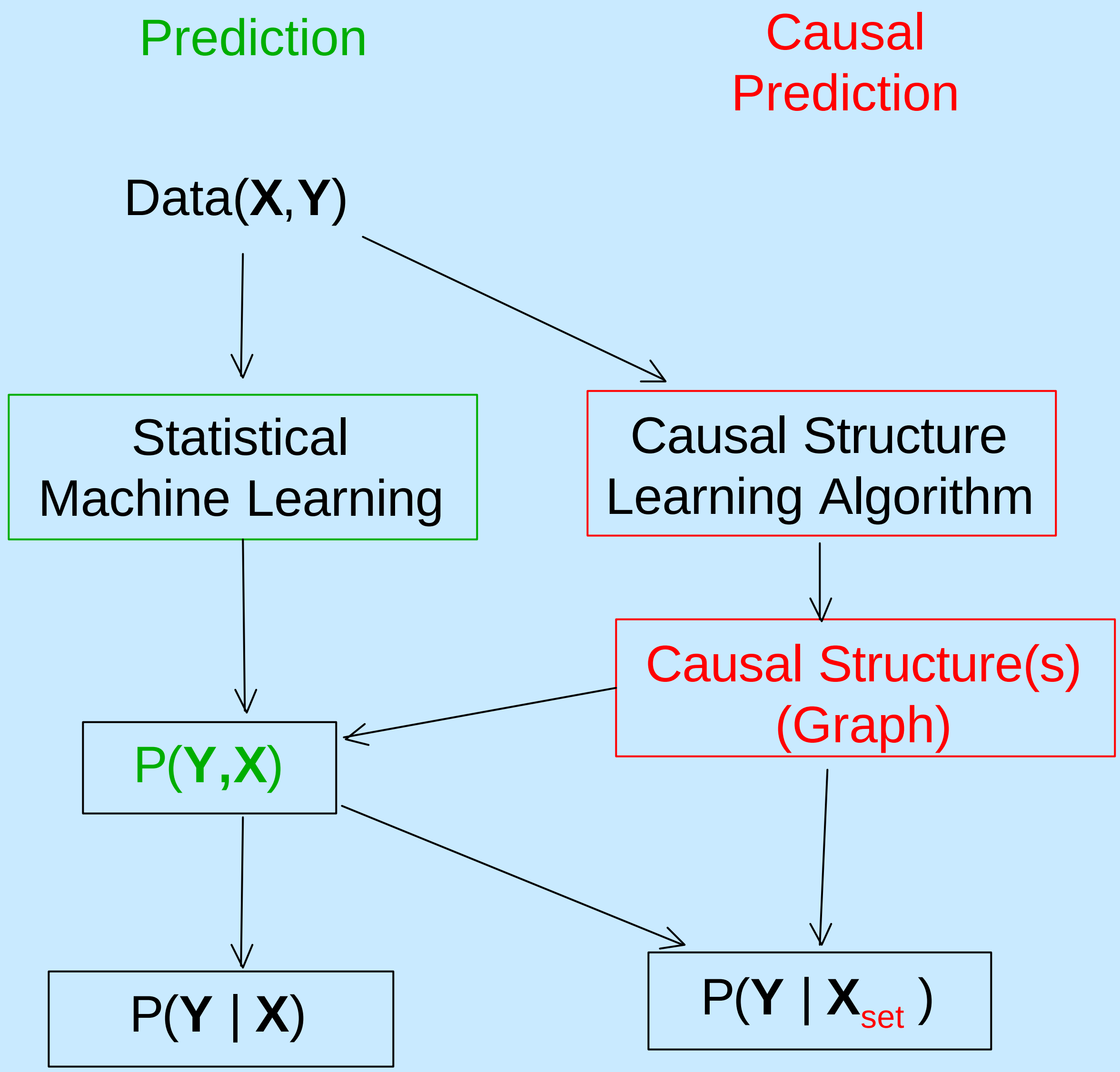
1. Causal Learning (vs. Predictive Learning)
2. Recent Successes
3. Philosophical Foundations of Causal Learning:
   the Standard Set-up
4. Problems with the Standard Set-up

Causal Discovery - Goals

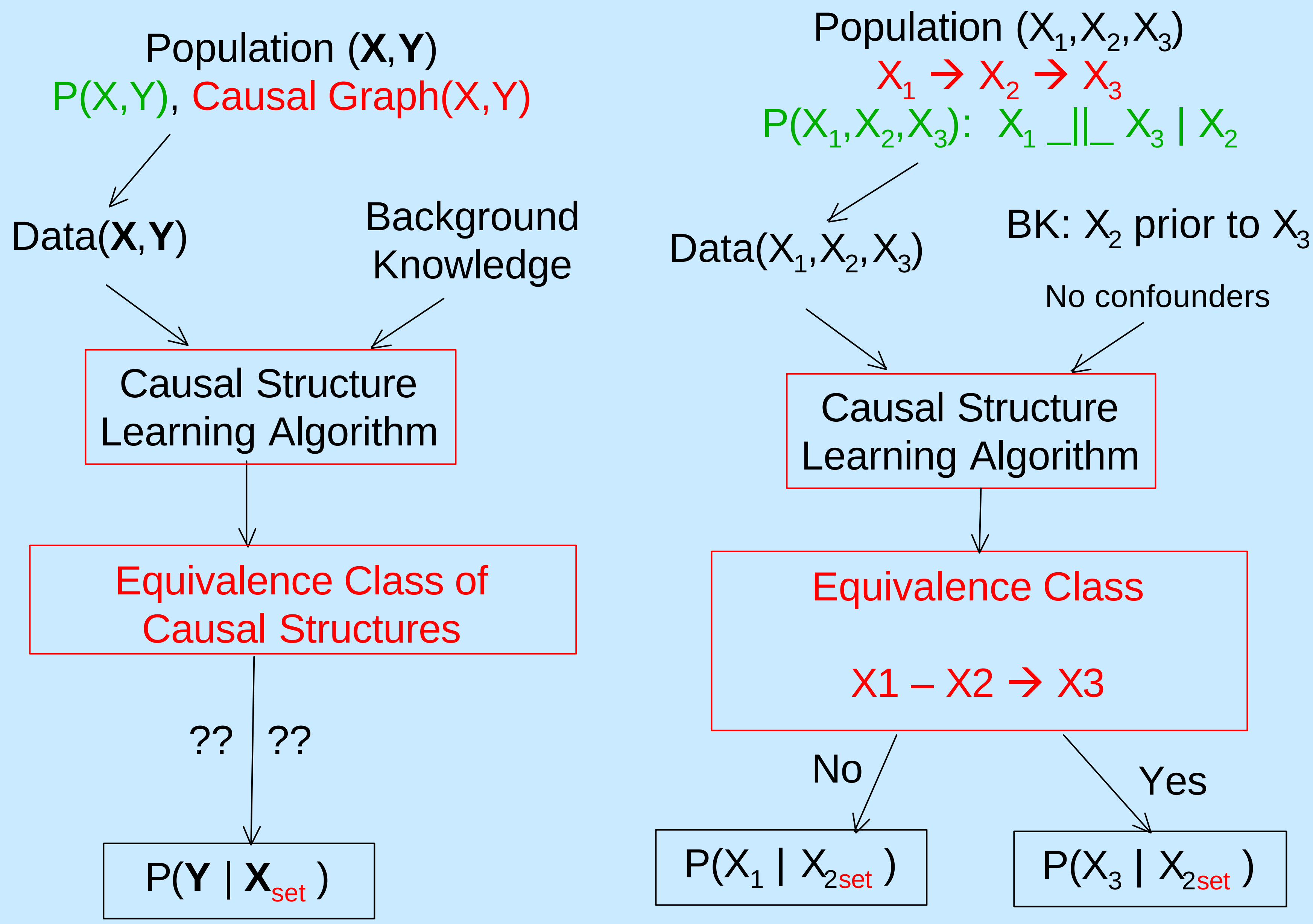
1) Policy, Law, and Science: How can we use data to answer
   a) *subjunctive* questions (effects of future policy interventions), or
   b) *counterfactual* questions (what would have happened had things been done differently (law)?
   c) *scientific* questions (what mechanisms run the world)

2) Rumsfeld Problem: Do we know what we don’t know: Can we tell when there is or is not enough information in the data to answer causal questions?

Causal Learning is Harder than Prediction



Causal Learning is Limited, but Rumsfeld



Recent Successes

(Partial List!)

- Do-Calculus
- Identification
- Bounding
- Bayesian Search
- Time-varying confounders and conditionally randomized treatment

  (Jamie Robins)
- Dynamic Bayes Nets
- Equivalence Classes

  (patterns, PAGs, Factor Analytic Measurement Models)

Recent Successes
(Partial List!)

- Pointwise Consistent Discovery Algorithms

  (patterns, PAGs, MMs, SEM with pure MM, Linear-Cyclic Models)
- Discovery in Time Series

  (Granger & Swanson, Hoover, Bessler, Moneta)
- Linear, non-Gaussian models (Shimizu, Hoyer, Hyvarinen)
- Active Search

  (Cooper, Eberhardt, Tong, Kohler, Murphy, He & Gong)
- Overlapping Sets of Variables (Tillman & Danks)
- Applications (Ed. Research, Biology, Economics, Sociology, etc.)
- Causality Challenge!!

Philosophical Foundations of Causal Structure Learning

V = {**M, L**} M measured, L = unobserved (latent)

Causal structure over **V** $\Rightarrow$ Constraints in P(**V**)

- *Assumption 1*: Weak Causal Markov Assumption

V_1,V_2 causally disconnected $\Rightarrow$ V_1 _||_ V_2

- *Assumption 2a*:

Causal Markov Axiom

- Assumption 2b: Determinism, e.g.,

Structural Equations

For each $V_i \in \mathbf{V}$, V_i := f(parents(V_i))

Philosophical Foundations of Causal Learning

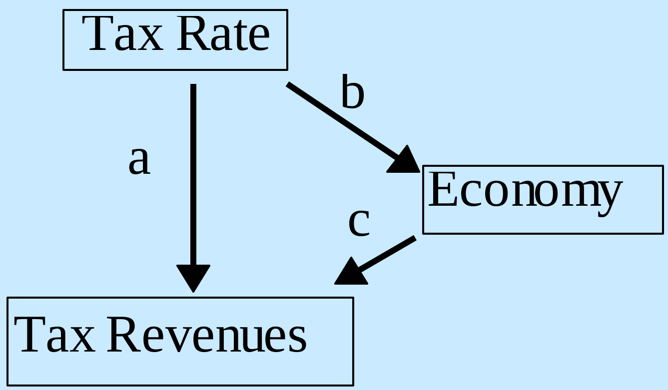
Causal structure over V $\Rightarrow$ Constraints in P(V)

Causal Markov Axiom:

If G is a causal graph, and P a probability distribution over the variables in G, then in P: every variable V is independent of its non-effects, conditional on its immediate causes.

Faithfulness

Constraints on a probability distribution P generated by a causal structure G hold for *all* parameterizations of G.



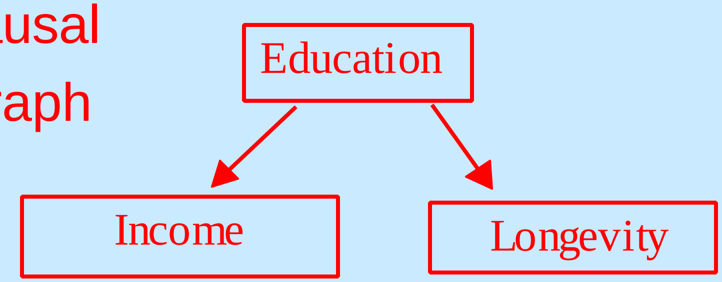
Revenues = aRate + cEconomy + $e_{Rev.}$

Economy = bRate + $e_{Econ.}$

Faithfulness: *a ? -bc*

Modularity of Intervention/Manipulation

Causal Graph



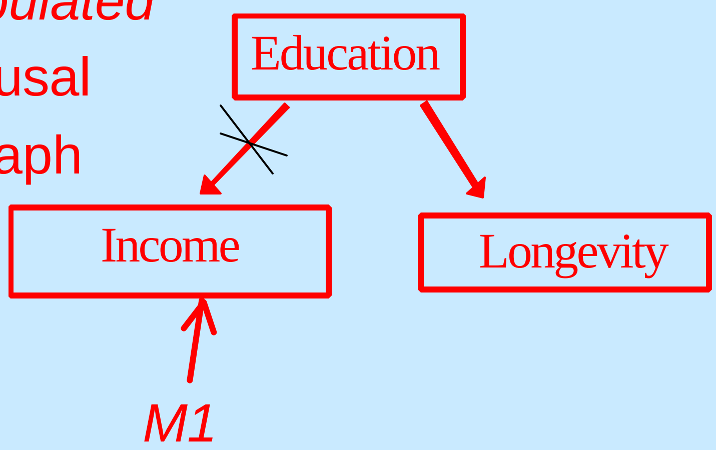
Structural Equations:

$$\text{Education} = \varepsilon_{ed}$$

$$\text{Longevity} = f_1\,(\text{Education}) + \varepsilon_{\text{Longevity}}$$

$$\text{Income} = f_2\,(\text{Education}) + \varepsilon_{\text{income}}$$

Manipulated Causal Graph



Manipulated Structural Equations:

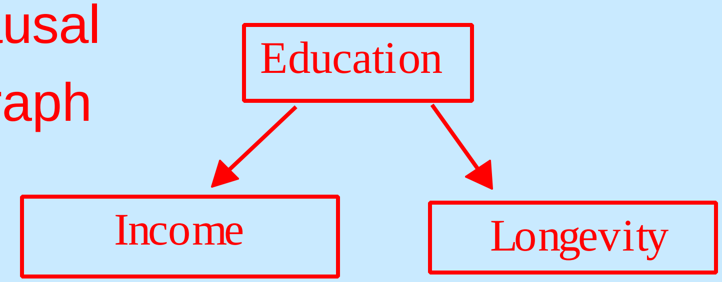
$$\text{Education} = \varepsilon_{ed}$$

$$\text{Longevity} = f_1\,(\text{Education}) + \varepsilon_{\text{Longevity}}$$

Income = f_3 (M1)

Modularity of Intervention/Manipulation

Causal Graph



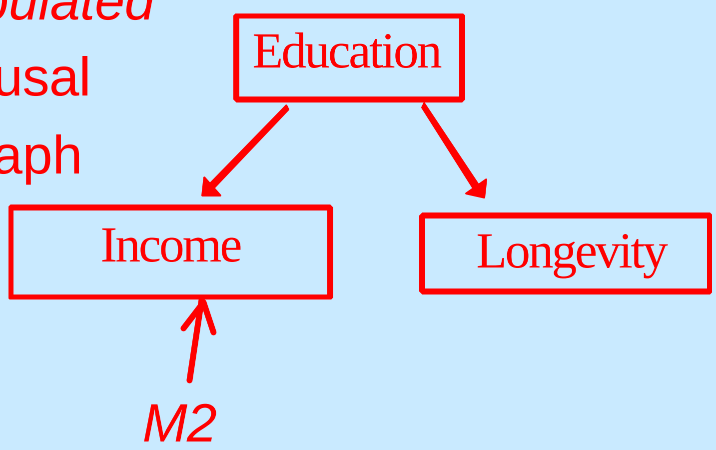
Structural Equations:

Education = ε_{ed}

Longevity = f_1 (Education) + $\varepsilon_{Longevity}$

Income = f_2 (Education) + ε_{income}

Manipulated Causal Graph



Manipulated Structural Equations:

Education = ε_{ed}

Longevity = f_1 (Education) + $\varepsilon_{Longevity}$

Income = f_3 (M2,Education) + e_{income}

The Standard Set-up

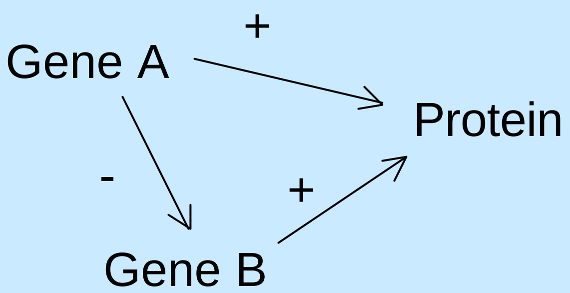
- Meausured Vars **M** given
- **V** = {**M, L**} satisfy Markov, Faithfulness, Modularity
- Tasks:
  - Discover structure (e.g., causal relations) among M
  - Estimate causal parameters
  - Less often:
    - Discover existence of L
    - Discover and estimate causal relations among L

Problems with the Standard Set-up

- Faithfulness in Redundant or Thermostatic Mechanisms
- Measurement
  - Classical Measurement Error
  - Coarsening
  - Aggregation
- Ambiguous Manipulations
- Modularity in Constraint Based, Reversible Systems
- Variable Construction / Decision Theory

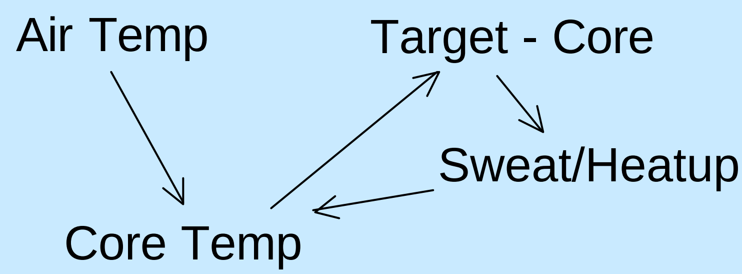
Faithfulness

- **Redundant Mechanisms**



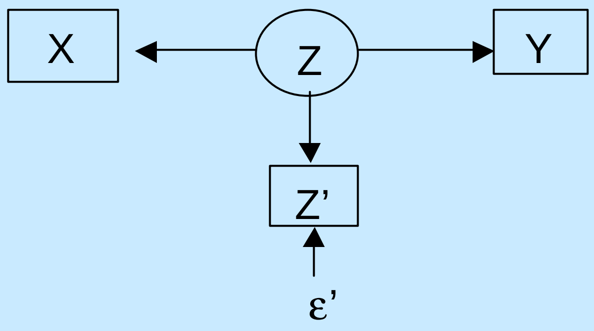
Gene A _||_ Protein

- **Thermostatic Equilibrium**



Air Temp _||_ Core Temp

Classical Measurement Error

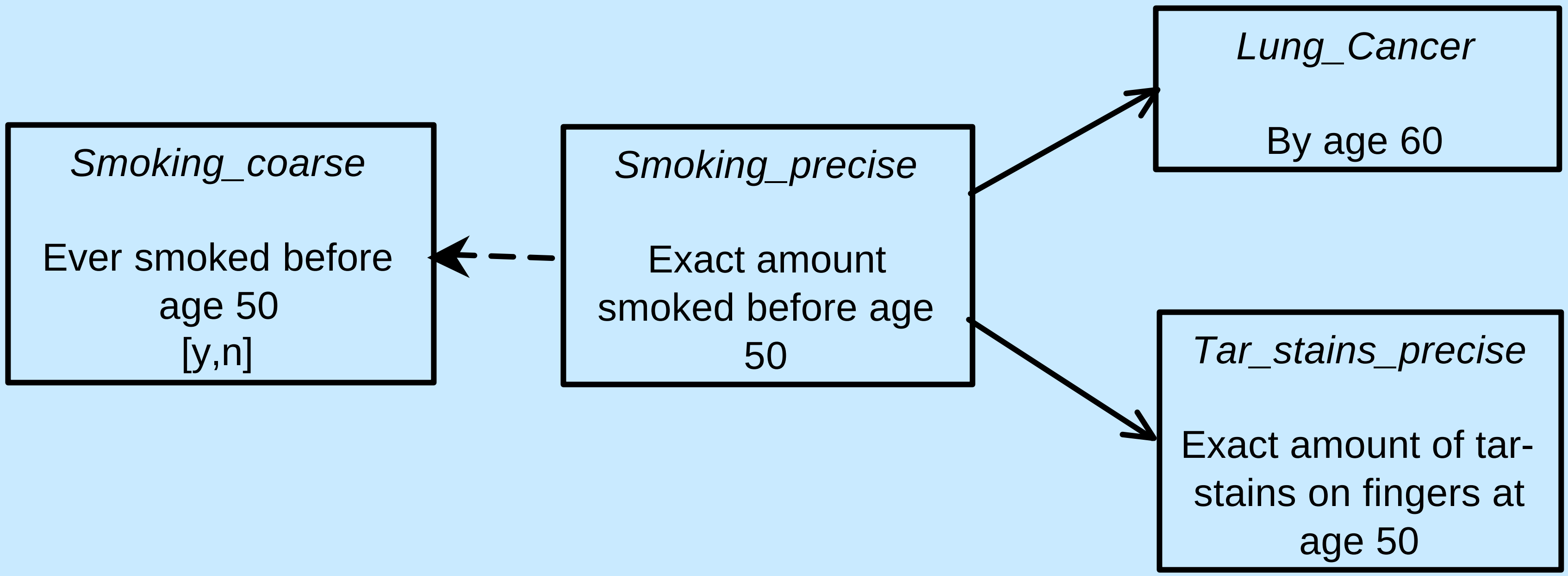


Measurement Error: $Z' = Z + \varepsilon$

X _||_ Y | Z

X ~~_||_~~ Y | Z' unless Var(ε') = 0

Coarsening

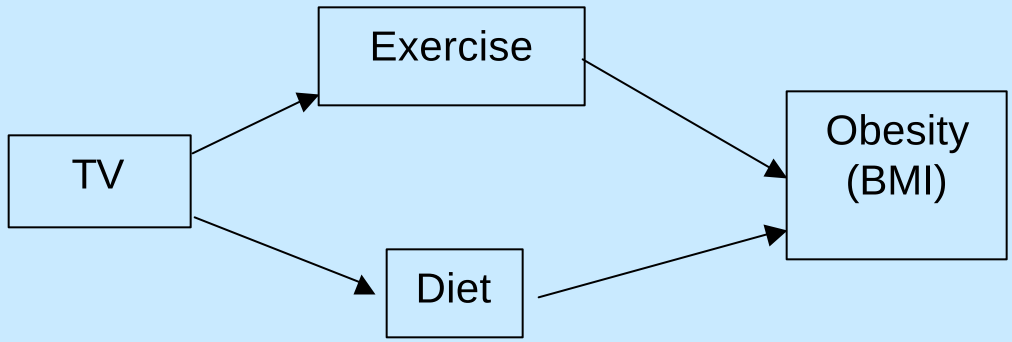


Lung_Cancer _||_ Tar_stains_precise | Smoking_precise

Lung_Cancer _||_ Tar_stains_precise | Smoking_coarse

TV → Obesity

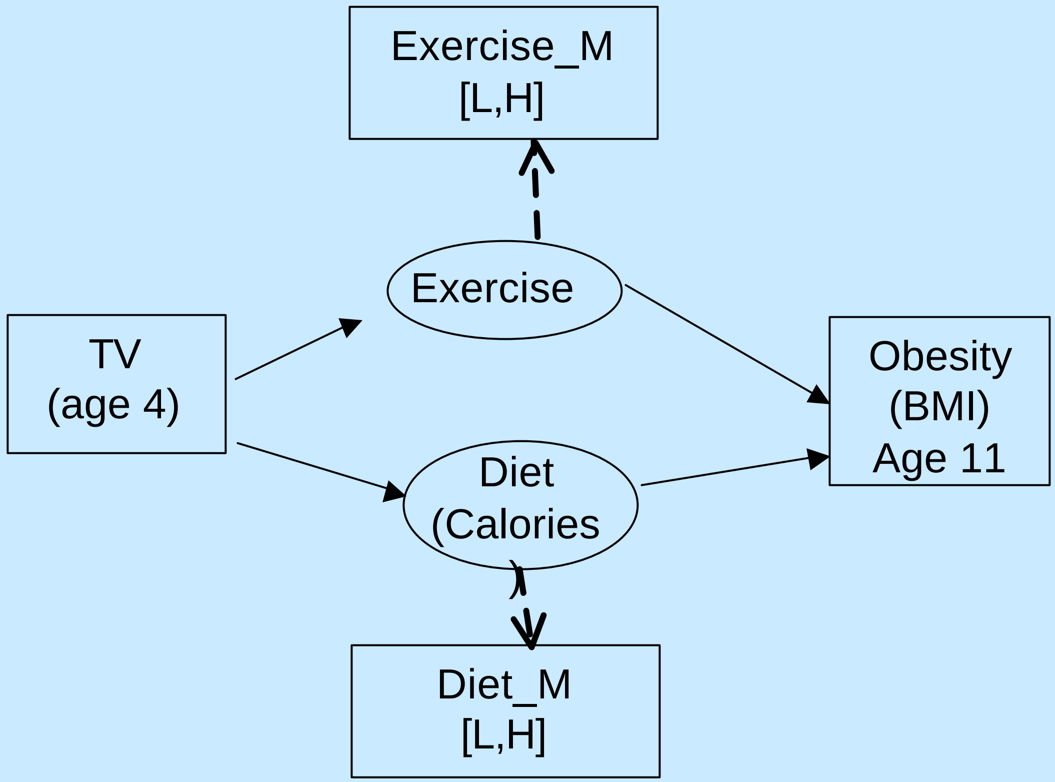
Proctor, et al. (2003). Television viewing and change in body fat from preschool to early adolescence: The Framingham Children's Study *International Journal of Obesity*, 27, 827-833.



Goals:

- Estimate the influence of TV on BMI
- Tease apart the mechanisms (diet, exercise)

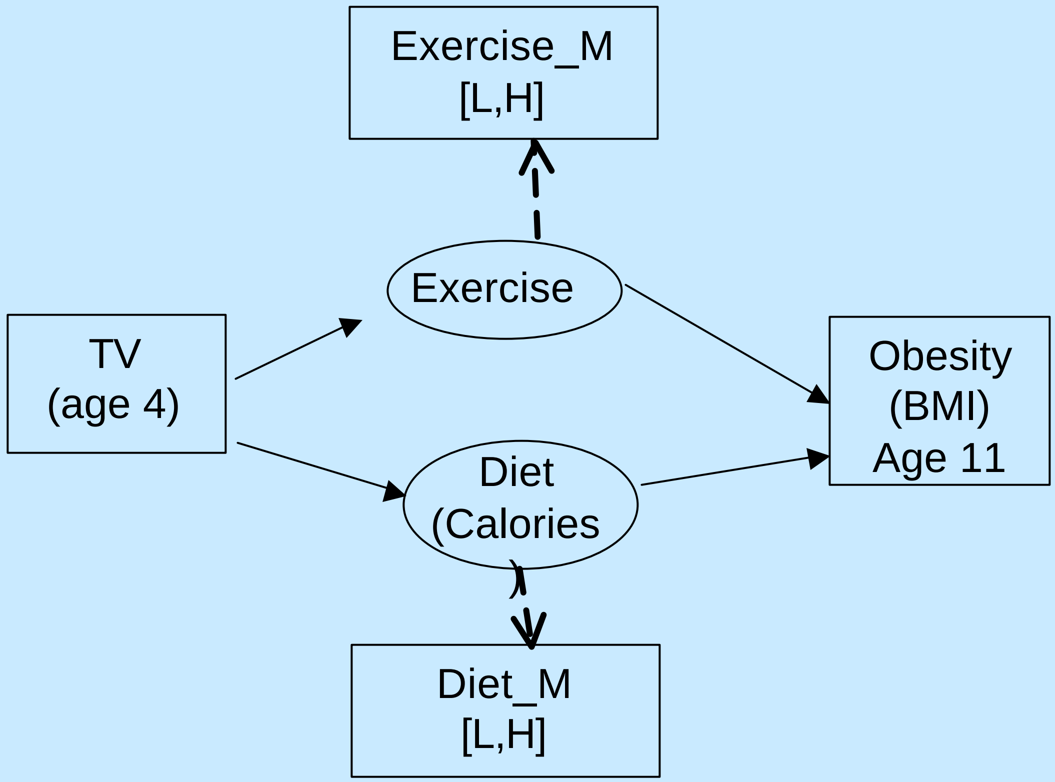
Measures of Exercise, Diet



Exercise_M: L ← Calories expended in exercise in bottom two tertiles
Exercise_M: H ← Calories expended in exercise in top tertile

Diet_M: L ← Calories consumed in bottom two tertiles
Diet_M: H ← Calories consumed in top tertile

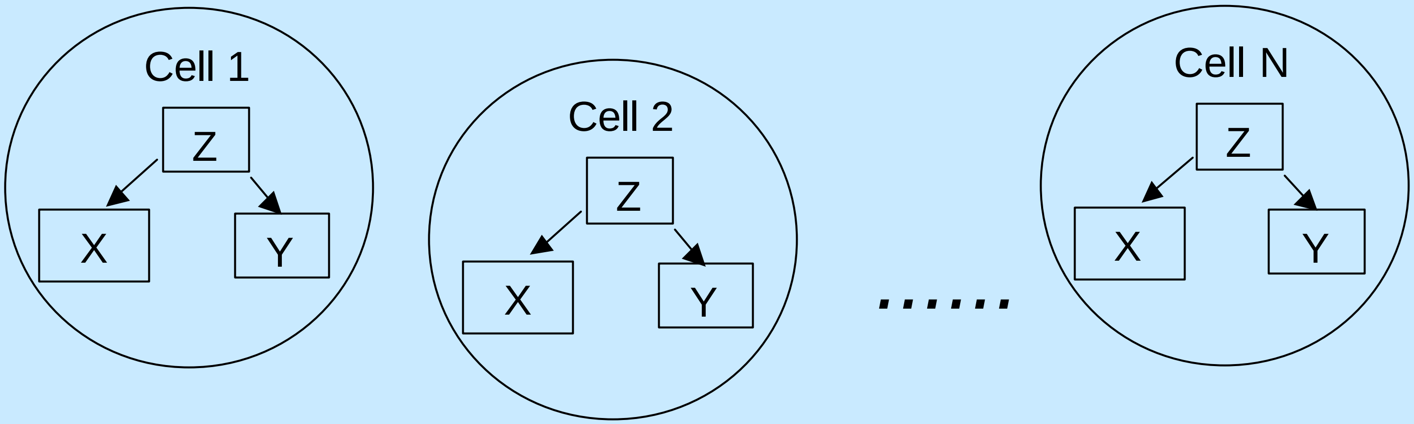
Measures of Exercise, Diet



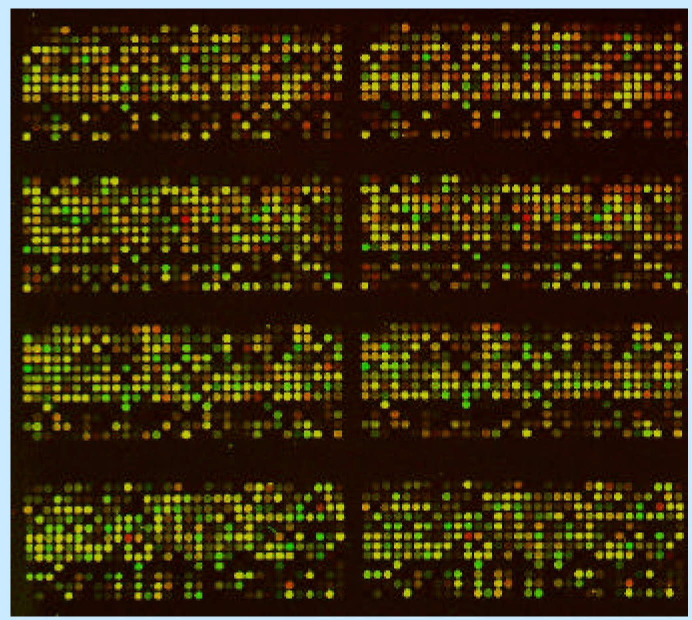
Findings:

- TV and Obesity NOT screened off by Exercise_M & Diet_M
- Bias in mechanism estimation unknown

Screening Off and Aggregation: Genetic Regulatory Network Discovery



Microarrays: measured gene expressions are *sums* of gene expression across all cells in tissue sample

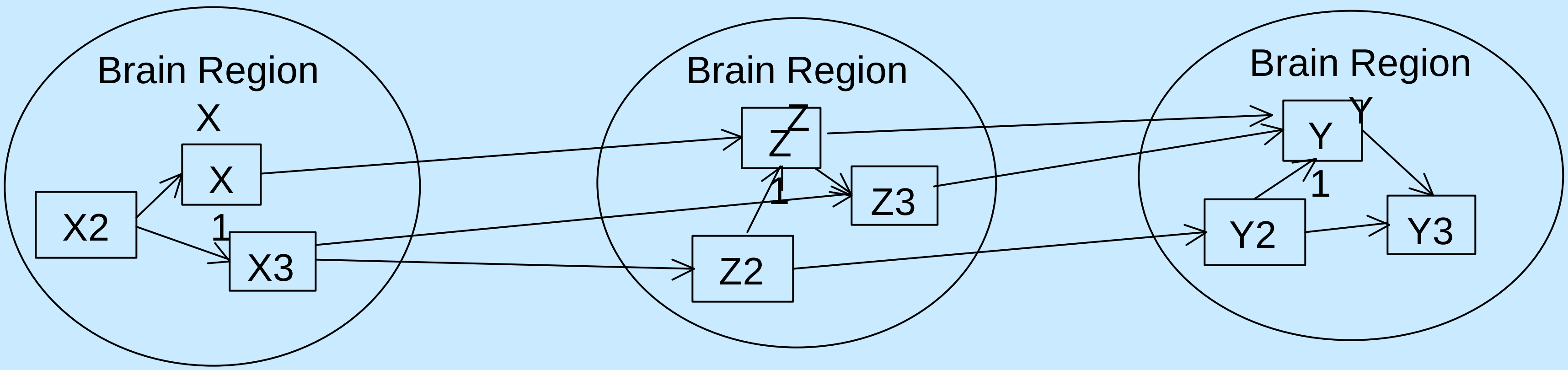


$\forall$ Cells: $X \perp\!\!\!\perp Y \mid Z$

$$\Sigma_n X \not\perp\!\!\!\perp \Sigma_n Y \mid \Sigma_n Z$$

unless P(X,Y,Z) is special,
e.g., Gaussian

Causal Discovery in fMRI

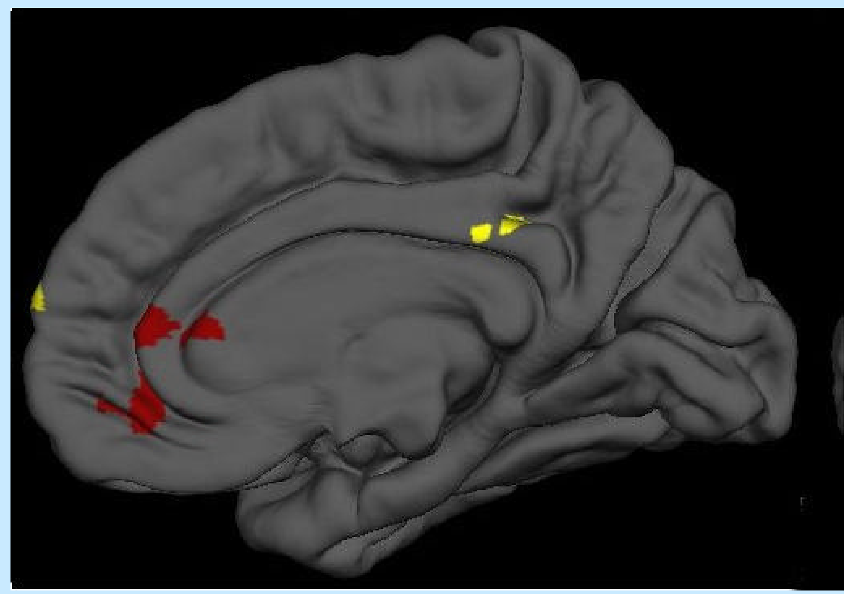


$$\forall i,j : X_i \;\_||\_\; Y_j \mid \{Z\}$$

fMRI measures *aggregate* activity in a voxel

Variables aggregate activity over voxels

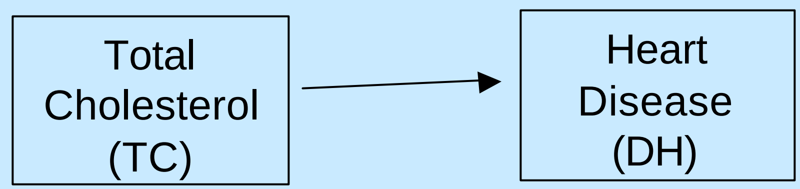
$$\Sigma X \;\not{\_||\_}\; \Sigma Y \mid \Sigma Z$$



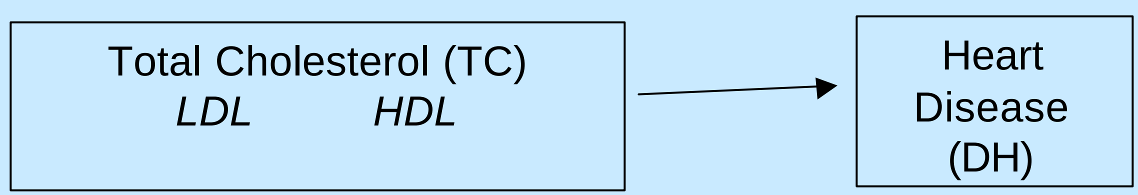
Ambiguous Manipulations

- 1960s : In RCTs, drugs that reduce *TC* (total cholesterol), reduce the risk of *DH* (Heart Disease).

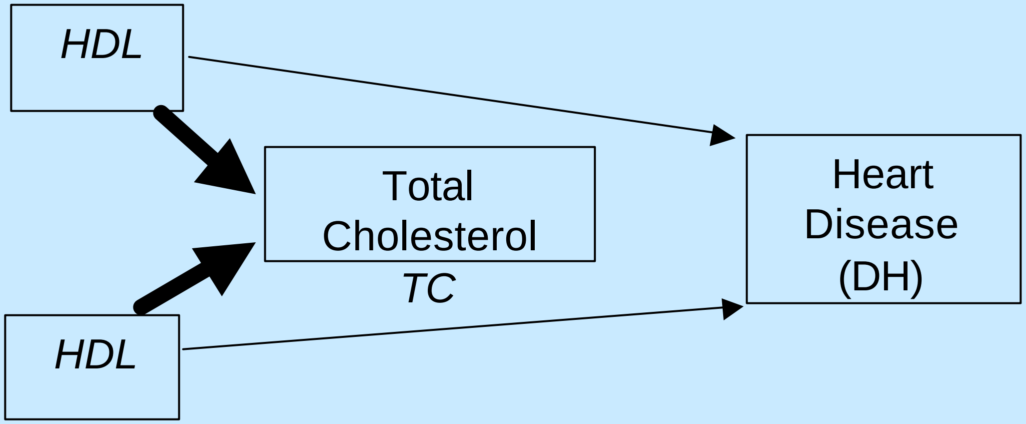
- $P(DH \mid TC_{set})$ identifiable.



- $TC \equiv_{def} f(LDL,HDL)$, high-density & low-density cholesterol



Ambiguous Manipulations



- *TC* [H,M,L], *HDL*[H,L], *LDL*[H,L], DH[Y,N]

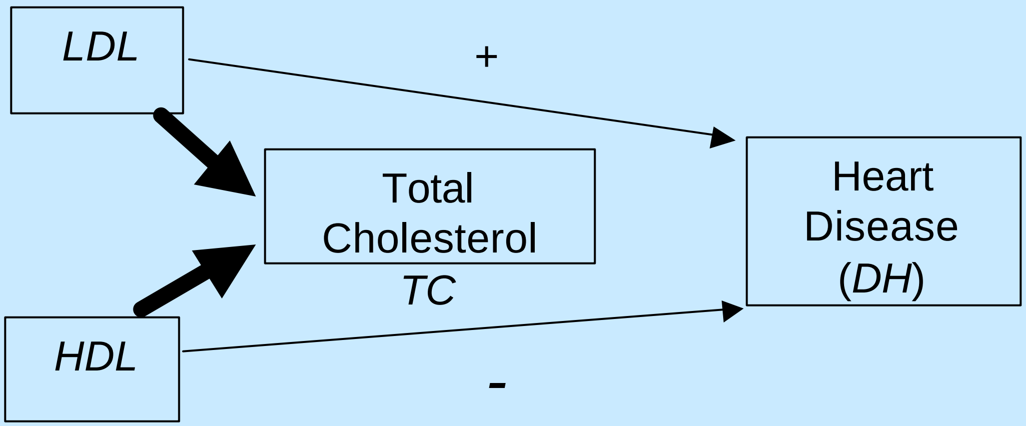
HDL=L, *LDL*=L → *TC*=L

HDL=L, *LDL*=H }
HDL=H, *LDL*=L } ® *TC*=*M*

HDL=H, *LDL*=H → *TC*=H

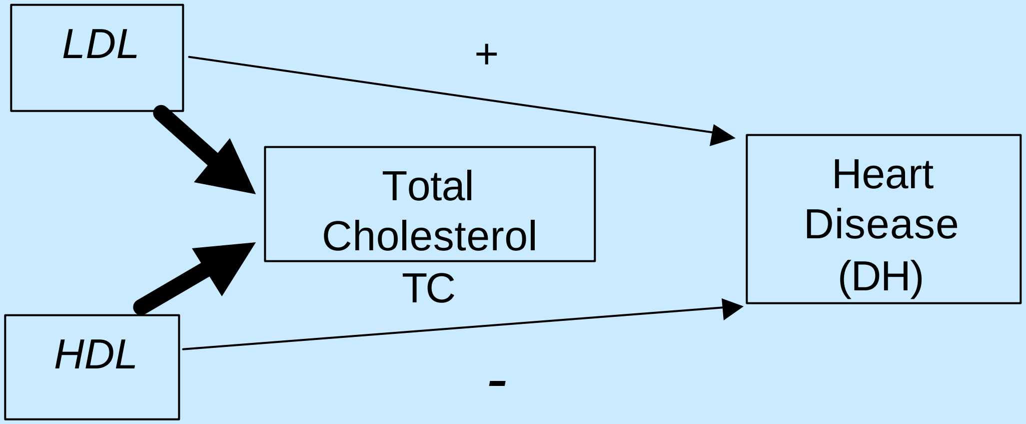
- arrows in boldface are definitional links

Ambiguous Manipulations



- Suppose *HDL, LDL* unobserved
- *TC* cannot be manipulated independently of both *HDL* and *LDL*
- "Set *TC* to M" is ambiguous over:
  - *HDL* = H and *LDL* = L
  - *HDL* = L and *HDL* = H

Ambiguous Manipulations



- Suppose

  HDL = H and *LDL* = L prevents *H*, and
  HDL = L and *HDL* = H promotes *H*?

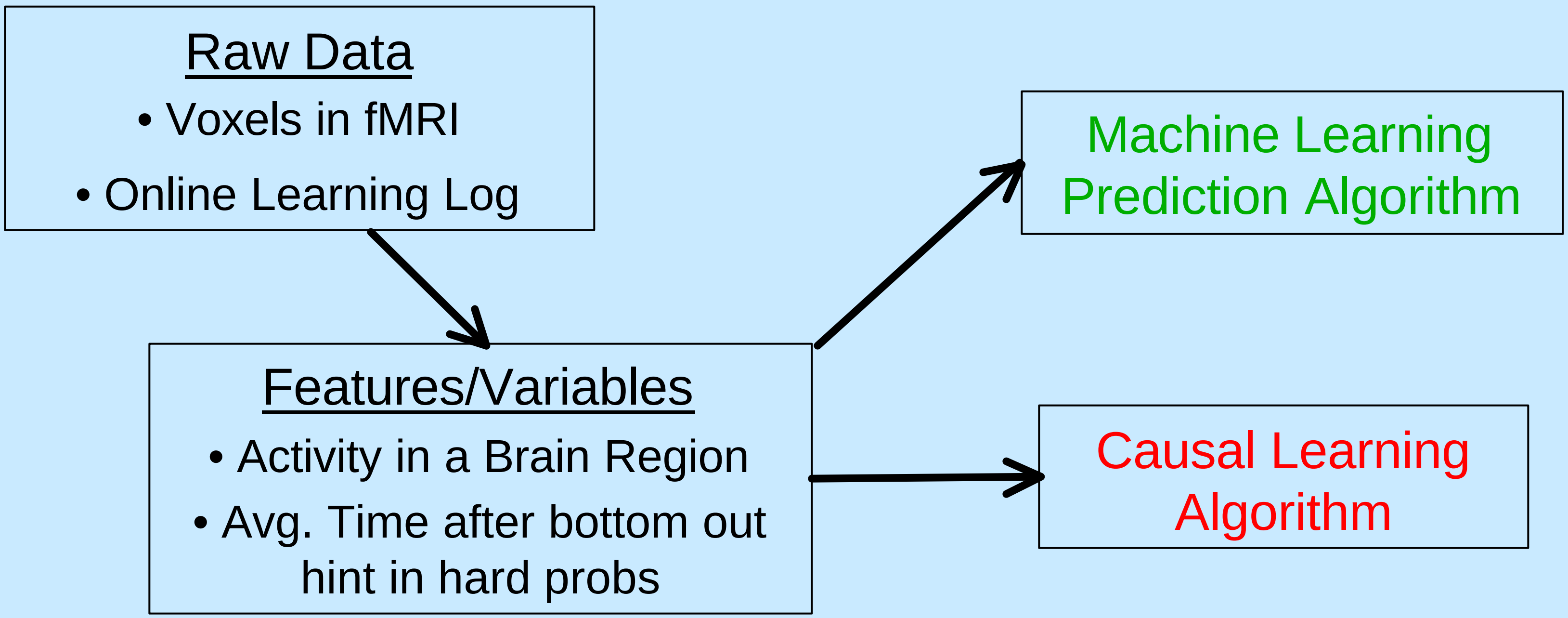
- What is P(*DH* | TC_{set}= M)?

- Can ambiguity be detected?
  - Need additional assumptions? Yes, e.g., variability
  - From observational data? Sometimes
  - Will positive causal hypotheses be inferred involving variables whose effect is ambiguous? Probably not

Reversible/Constraint Systems

- PV = nRT
- Constraint persists, even with surgical interventions
- "joint" part of P(V,T,P) remains unaltered by any intervention.
- Is there a causal graph and parameterization thereof such that the constraint holds for any permissable set of surgically altered equations?
- Can such systems be learned without intervention?

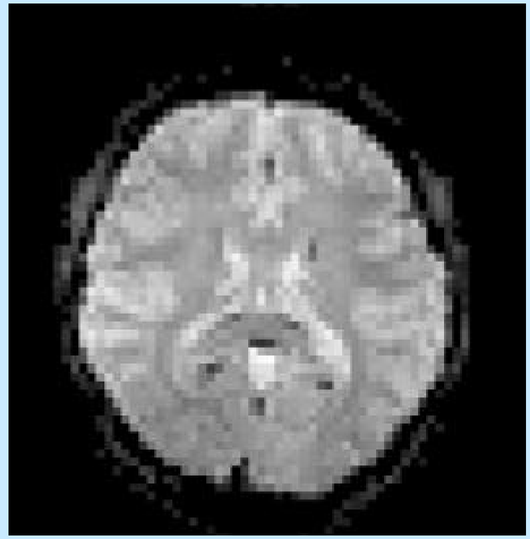
Decision Theory/Variable Construction for Causal Learning

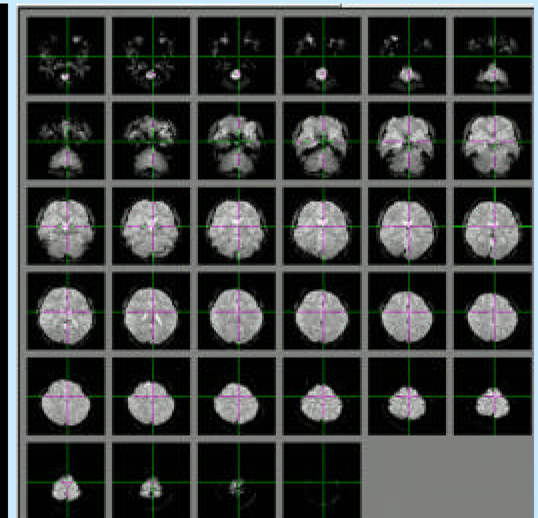


Variable construction can be framed as a search problem, thus a decision problem

Decision problem for prediction ? decision problem for causal learning

Variable Construction for Causal Learning





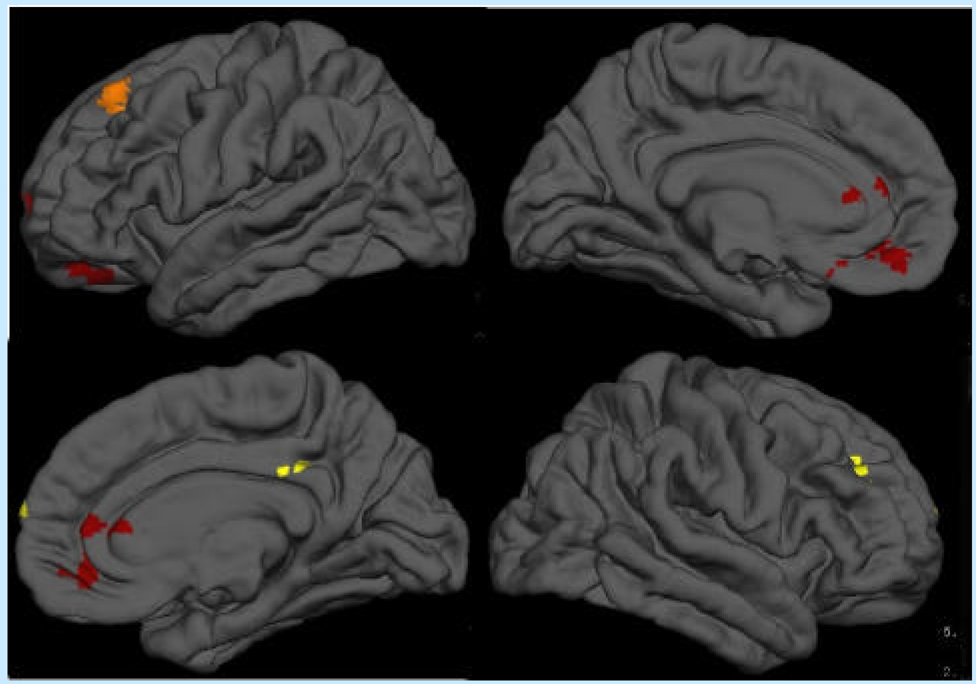
Raw Data

- Voxels in fMRI

↓

Features/Variables

- Activity in a Brain Region



Variable Construction for Causal Learning

Raw Data

- Voxels in fMRI

↓

Features/Variables

- Activity in a Brain Region

- Adjust for interunit – anatomical matching
- Correct for time lag of hemodynamic response & scan time
- Identify voxels with statistically improbable signals
- Cluster, usually by eyeball
- Variables constructed =
  - mean of signal intensity in cluster
  - one of the first 4 principal components
  - average intensity of top X% variance voxels
  - maximum variance voxel
  - non-contiguous regions
  - possibly overlapping

Decision Theory for Causal Learning

- Positive utility on increasing an output from baseline

(e.g., learning in online course, brain activity in region associated with emotional intelligence among autistic children)

- Intervention on 1 variable, leave cost aside.

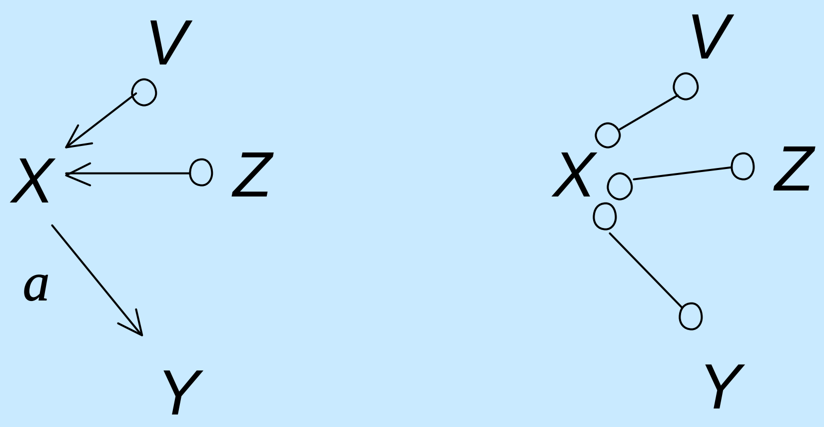
- Raw data → constructed variables → causal search algorithm

- Compute expected utility of intervention

- Uncertainty over :
  - causal structure
  - parameters in a given causal structure

Model Uncertainty for *1 set* of constructed variables

Search Output



Equiv Class 1 **Equiv Class 2**

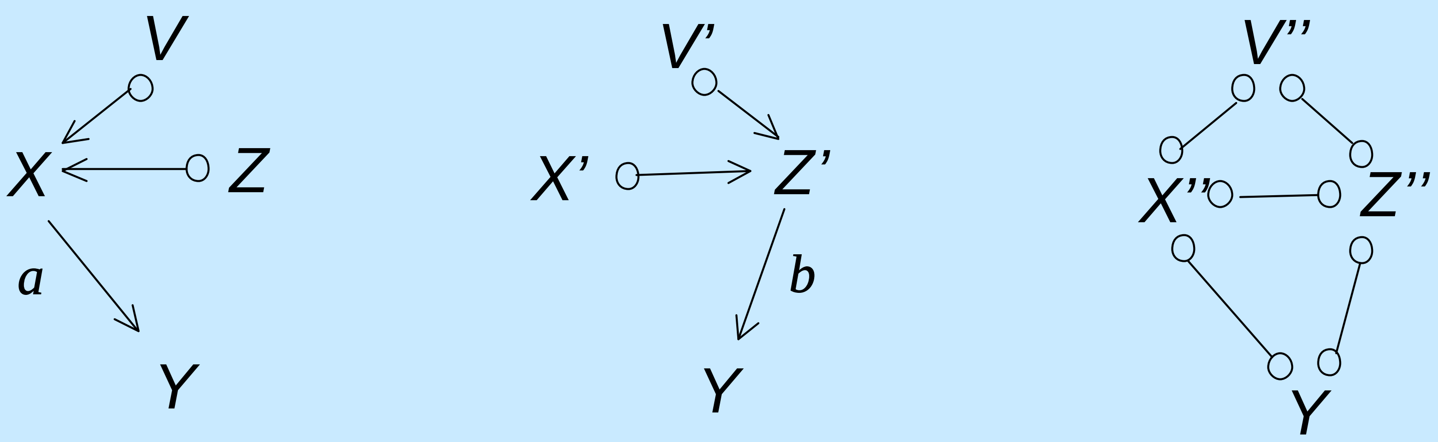
Model Averaging

EU(do(X)) = EU(do(X) in *EC1*) P(EC1) + EU(do(X) in *EC2*)) P(EC2)

EU(do(X) in *EC1*) = EU(do(X) in DAG_i in EC1) P(DAG_i in EC1)) + ...

EU(do(X) in DAG_i in EC1) = ?EU(do(X) in DAG_i,α = x) *dx*

Model Uncertainty for *many sets* of constructed variables



Result :VC 1 **Result :VC 2** **Result :VC 3**

- EU(do(X)) vs. EU(do(X')) vs. EU(do(X'')))?
- Meaningful prior over models in output for each VC regime?

Thanks