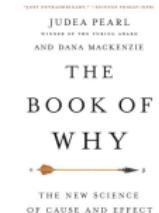
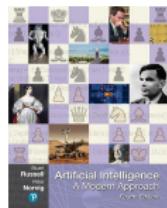


## 8.2: Causal Inference

**Instructor:** Dr. GP Saggese - [gsaggese@umd.edu](mailto:gsaggese@umd.edu)

### References:

- AIMA (Artificial Intelligence: a Modern Approach)
- Pearl et al., The Book of Why, 2017



- *Type of Variables in Causal AI*
- Paths

# Mediator Variable

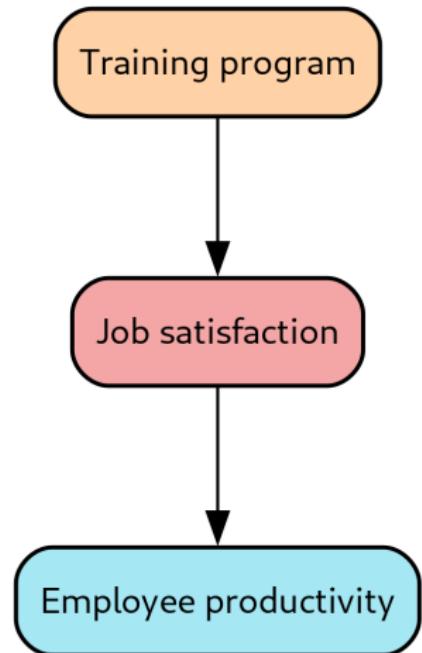
---

- A **mediator variable**  $M$ 
  - Is an intermediate variable that *transmits* the causal effect from  $X$  (treatment) to  $Y$  (outcome)
  - Lies **on the causal path** between  $X$  and  $Y$
  - Captures the **mechanism or process** through which  $X$  influences  $Y$



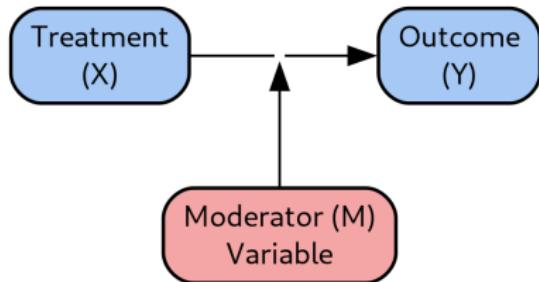
# Mediator Variable: Example

- **Research question:** “Does a training program increase employee productivity?”
- Causal effect may be indirect, through a **mediator**
  - Training might not immediately boost productivity
  - Could enhance job satisfaction, raising productivity
- **Causal interpretation**
  - $X$ : Training Program (cause)
  - $M$ : Job Satisfaction (mediator)
  - $Y$ : Employee Productivity (effect)
  - Path:  $X \rightarrow M \rightarrow Y$
- **Direct vs. Indirect effects**
  - *Indirect effect*  $X$  affects  $Y$  through  $M$
  - *Direct effect*  $X$  affects  $Y$  not through  $M$
  - Controlling for  $M$  separates effects, clarifying training impact



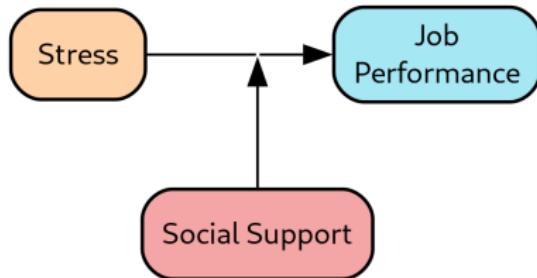
# Moderator Variable

- A **moderator variable**  $M$ 
  - Changes the *strength* or *direction* of the relationship between an independent variable ( $X$ ) and a dependent variable ( $Y$ )
  - Is not part of the causal chain but conditions the relationship



# Moderator Variable: Example

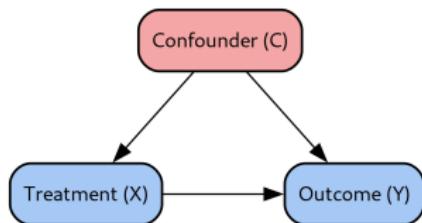
- **Research question:** “*Study relationship between stress X and job performance Y*”
- **Social support M as a moderator**
  - High social support weakens stress's negative effect on performance
  - Low social support strengthens stress's negative effect on performance



# Confounder Variable

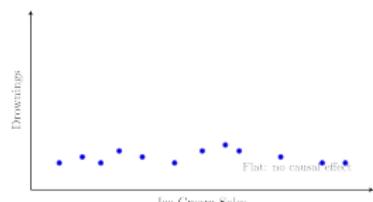
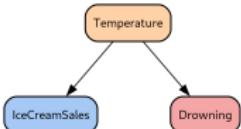
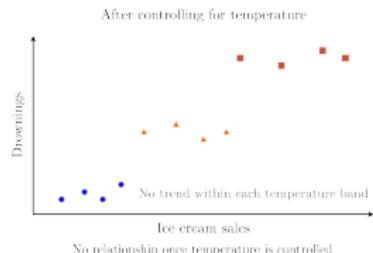
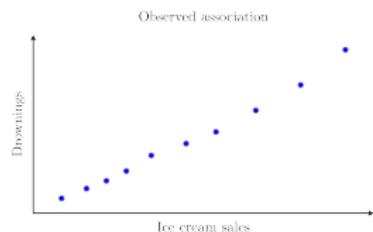
---

- A **confounder**  $C$ 
  - Affects both treatment (cause) and outcome
  - Creates misleading association if not controlled



# Confounder Variable: Example

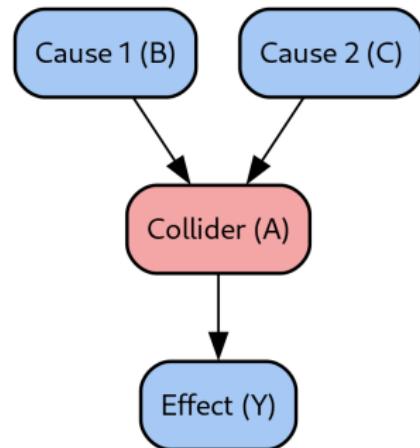
- *IceCreamSales* and *Drowning* move together
  - Correlation-based model claims association, but how to use this relationship?
  - Ban ice cream to prevent drowning?
  - Ice cream maker increase drowning to boost sales?
- No cause-effect between *IceCreamSales* and *Drowning*
  - *Temperature* is a confounder
  - Control for season in regression or intervention, association disappears



# Collider

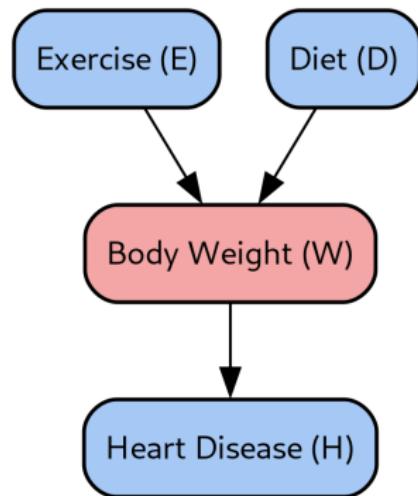
---

- A **collider**  $A$ 
  - Is a variable influenced by multiple variables  $B, C$
  - Complicates understanding relationships between variables  $B, C$  and those it influences,  $Y$



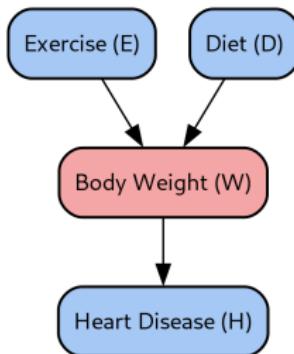
# Collider: Examples

- Study the relationship between *Exercise* and *HeartDisease*
  - *Diet* and *Exercise* influence *BodyWeight*
  - *BodyWeight* influences *HeartDisease*
  - *BodyWeight* is a collider



# Collider Bias

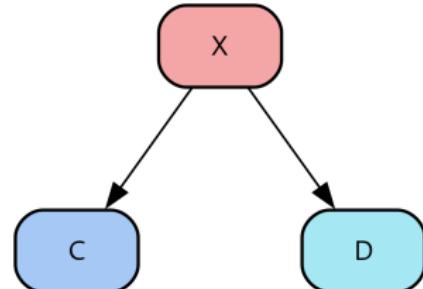
- Aka “Berkson’s paradox”
- **Conditioning on a collider** can introduce a spurious association between its parents by “*opening a path that is blocked*”
- Consider the variables:
  - Diet (D)
  - Exercise (E)
  - BodyWeight (W)
  - HeartDisease (H)
- **Without conditioning on W**
  - E and D are independent
    - E.g., knowing exercise level  $E$  doesn't inform about diet  $D$ , and vice versa
  - Collider  $W$  blocks association between  $E$  and  $D$
- **After conditioning on W**
  - E.g., individuals with specific body weight
  - Introduce dependency between  $E$  and  $D$
  - With  $W$  fixed, changes in  $E$  balanced by changes in  $D$ , inducing spurious correlation between  $E$  and  $D$



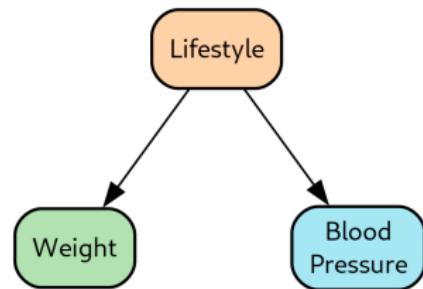
- Type of Variables in Causal AI
- ***Paths***

# Fork Structure

- A **fork**  $D \leftarrow X \rightarrow C$  occurs when a single variable causally influences two or more variables
- $X$  is a **confounder** (common cause) of  $C$  and  $D$
- Forks induce statistical dependence between  $C$  and  $D$  even if  $C$  and  $D$  are not causally linked
- Conditioning on  $X$  blocks the path and removes spurious correlation

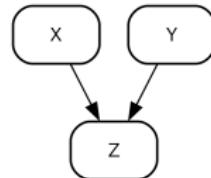


- **Example**
  - *Lifestyle* is a confounder that affects both *Weight* and *BloodPressure*
  - These outcomes may appear correlated due to shared cause



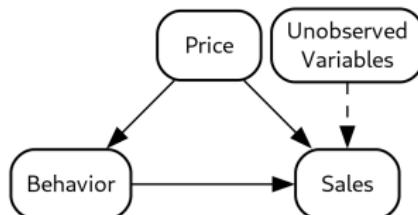
# Inverted Fork

- An **inverted fork** occurs when two or more arrows converge on a common node
  - Also known as a **collider**
- Colliders block associations unless the collider or its descendants are conditioned on
- Conditioning on a collider “opens” a path, inducing spurious correlations
- Example:
  - Sales influenced by multiple independent causes
  - *MarketingSpend* and *ProductQuality* both influence *Sales*
  - Conditioning on *Sales* can induce false dependence between *MarketingSpend* and *ProductQuality*



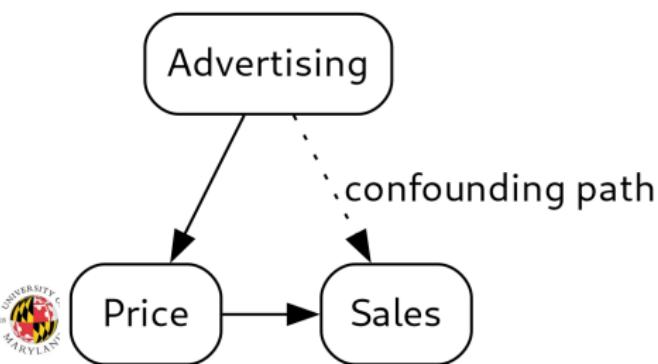
# Path connecting unobserved variables

- **Unobserved variables** affect the model but we don't have a direct measure of it
- E.g., consider the causal DAG
  - A retailer does market research, expecting *Price* to influence *Sales* in a predictable way
  - A retailer sets the *Price* of a new product based on market research
  - The retailer can observe and measure *Behavior*, e.g.,
    - Discounts
    - Promotional campaign
  - There are unobserved vars that influence the model, e.g.,
    - Social media buzz
    - Word-of-mouth recommendation



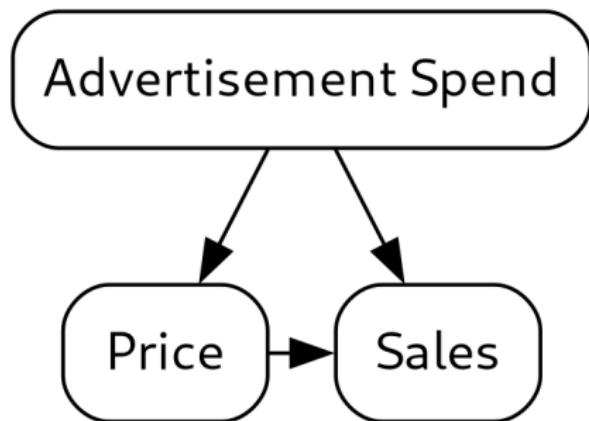
# Front-door Paths in Causal Inference

- A front-door path reveals causal influence through an observable mediator
  - The causal effect flows:  $A \rightarrow P \rightarrow S$
- Requirements for identifiability:
  - All confounders of  $A \rightarrow P$  and  $P \rightarrow S$  are observed and controlled
  - There are no back-door paths from  $A$  to  $S$  through unobserved variables
- Enables causal estimation when back-door adjustment is infeasible
- Example:
  - Advertising impacts sales through customer perception of price
  - $A$ : Advertising,  $P$ : Price perception,  $S$ : Sales
- Pearl's front-door criterion provides a formal method for adjustment
  - Estimate  $P(P|A)$ ,  $P(S|P, A)$ , and  $P(A)$  from data to compute causal effect



# Back-Door Paths

- A company wants to understand the causal effect of price on sales



- Price → Sales is the front-door path
- A confounder is Advertising spend since it can affect both:
  - The price the company can set (e.g., the cost increases to cover advertisement costs and the product is perceived as more valuable)
  - The sales (directly)

The back-door path goes from Sales to Price via Advertising spend



# Frontdoor and Backdoor Paths

- Question: *Will increasing our customer satisfaction increase our sales?*
- Assume that the Causal DAG is



- **Front-door path** (i.e., a direct causal relationship):  
 $CustomerSatisfaction \rightarrow Sales$

- **Backdoor path:**  
*ProductQuality* is a common cause (confounder) of both *CustomerSatisfaction* and *Sales*
- To analyze the relationship between customer satisfaction and sales, we need to:
  - Control for *ProductQuality* to close the backdoor path
  - Eliminate the confounding effect
- In reality there are more confounding effects (e.g., price)

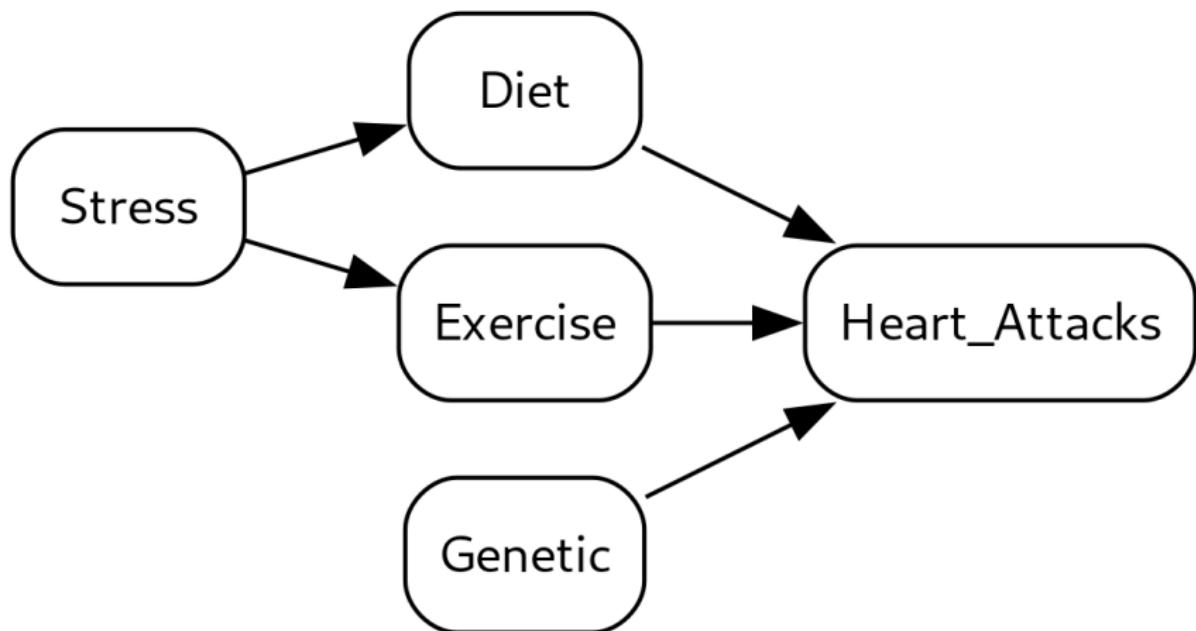
# Building a DAG

---

- **Causal models** visually represent complex environments and relationships
- Nodes are like “nouns” in the model:
  - E.g., “price”, “sales”, “revenue”, “birth weight”, “gestation period”
  - Variables can be endogenous/exogenous and observed/unobserved
  - Complex relationships between variables:
    - Parents, children (direct relationships)
    - Descendants, ancestors (along the path)
    - Neighbors
- **Iterative Refinement:**
  - Models are continuously updated with new variables and insights
- **Modeling as a Communication Tool:**
  - A shared language that bridges gaps between technical and non-technical team members
- **Unobservable Variables:**
  - Supports inclusion of variables not empirically observed but known to exist
  - E.g., trust or competitor activity can be modeled despite lack of direct data

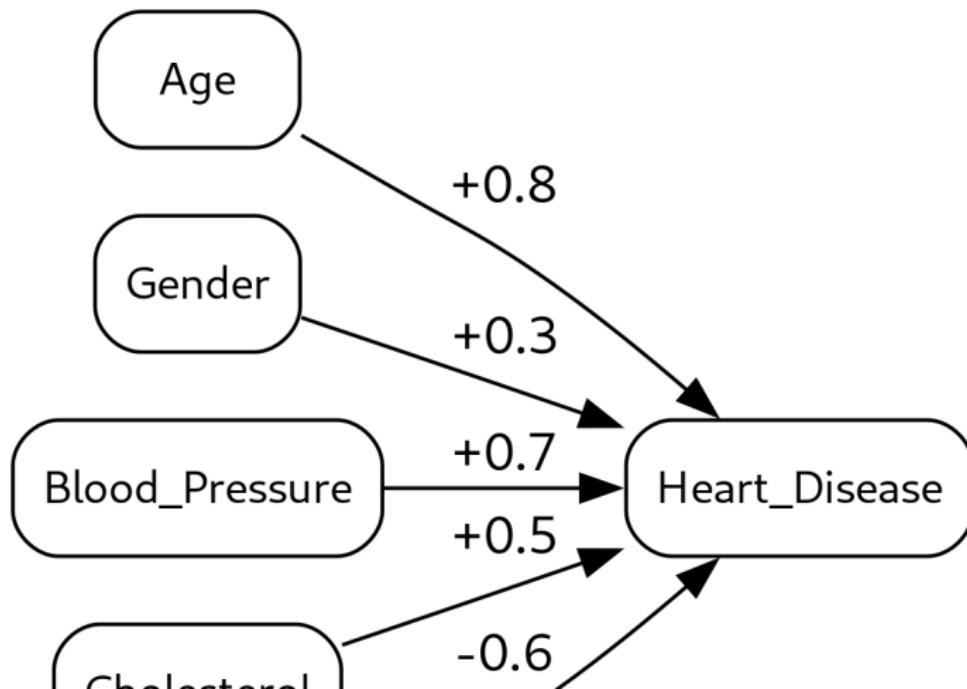
# Heart Attack: Example

- What's the relationship between stress and heart attacks?
  - Stress is the treatment
  - Heart attack is the outcome
  - Stress is not a direct cause of heart attack
    - E.g., a stressed person tends to have poor eating habits



# Weights

- Weights can be assigned to paths to represent the strength of the causal relationship
  - Weights can be estimated using statistical methods
- Sign represents the direction



# Counterfactuals

---

- A **counterfactual** describes what would have happened under a different scenario
  - *"What would the outcome have been if X had been different?"*
  - *"If kangaroos had no tails, they would topple over"*
  - *"What if we had two suppliers of our product, rather than one? Would we have more sales?"*
  - *"Would customers be more satisfied if we could ship products in one week, rather than three weeks?"*
- **Causal reasoning:**
  - Goes beyond correlation and association
  - Requires a causal model (like an SCM) to simulate alternate realities
  - E.g.,
    - Actual: A student received tutoring and scored 85%
    - Counterfactual: What if the student didn't receive tutoring?
    - Causal model estimates the alternative outcome (e.g., 70%)
- **Challenges:**
  - Requires strong assumptions and accurate models
  - Difficult to validate directly since counterfactuals are unobservable