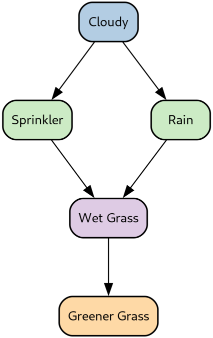
Structural Causal Model: Sprinkler Example



- Structural equations for this model:

$$\begin{cases} C := f_C(\varepsilon_C) \\ R := f_R(C, \varepsilon_R) \\ S := f_S(C, \varepsilon_S) \\ W := f_W(R, S, \varepsilon_W) \\ G := f_G(W, \varepsilon_G) \end{cases}$$

- Unmodeled variables ε_x represent error terms
  - E.g., ε_W is another source of wetness besides *Sprinkler* and *Rain* (e.g., *MorningDew*)
- Assume unmodeled variables are exogenous, independent, with a certain distribution (prior)

- Express joint distribution of five variables as a product of conditional distributions using causal DAG topology:

$$\Pr(C, R, S, W, G) = \Pr(C) \Pr(R|C) \Pr(S|C) \Pr(W|R, S) \Pr(G|W)$$

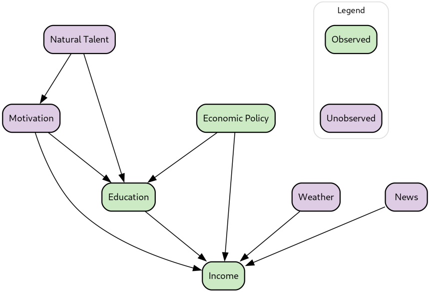
SCIENCE ACADEMY

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

SCIENCE ACADEMY

Observed Vs. Unobserved Variables

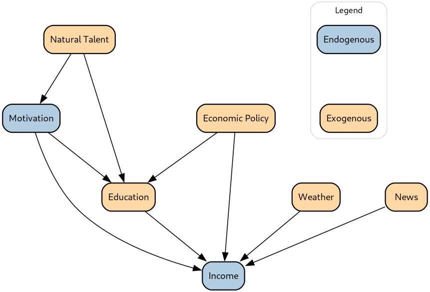
- **Observed variables**
  - Aka "measurable" or "visible"
  - Variables directly measured or collected in a dataset
  - E.g.,
    - Education
    - Income
    - Blood pressure
    - Product price



- **Unobserved variables**
  - Aka "latent" or "hidden"
  - Exist but not measured or included in data
  - E.g.,
    - Natural talent
    - Motivation
    - Company culture
  - Ignoring unobserved variables distorts causal relationships
    - Observed: *IceCreamSales* and *DrowningRates*
    - Unobserved: *Temperature*
    - Misleading conclusion: *IceCream* causes *Drowning*

Endogenous Vs. Exogenous Variables

- **Endogenous variables**
  - Values determined *within* the model
    - Dependent on other variables in the system
  - Represent system's internal behavior and outcomes
  - E.g.,
    - Motivation
    - Income



- **Exogenous variables**
  - Originate *outside* the system being modeled
    - Not caused by other variables in the model
  - Represent background conditions or external shocks
  - E.g.,
    - Natural talent
    - Economic policy
    - Weather
    - News

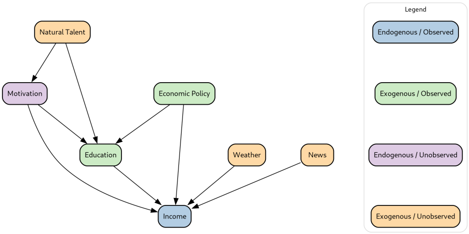
SCIENCE ACADEMY

Endo / Exogenous, Observed / Unobserved Vars

- In **Structural Causal Models**

  $$X_i = f_i(Parents(X_i), \varepsilon_i)$$

  where:
  - X_i: endogenous
  - ε_i: exogenous noise
- **Typically**
  - *Endogenous variables*: focus for prediction and intervention
  - *Exogenous variables*: capture randomness or unknown external factors



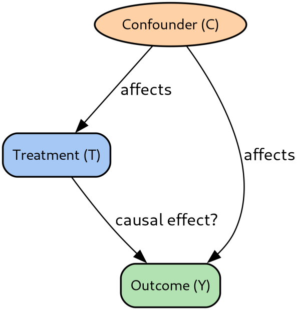
Variable Type	Observability	Example
Endogenous	Observed	Income
Exogenous	Observed	Education
Endogenous	Unobserved	Motivation
Exogenous	Unobserved	Natural Talent

SCIENCE ACADEMY

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

Estimating Causal Effects

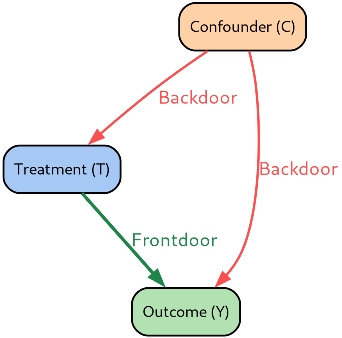
- **Goal**: Determine the causal effect of a treatment variable (aka intervention) T on an outcome Y
- **Example:**
  - T = *"takes drug"*
  - Y = *"recovers"*
  - C = *"overall health"*
- Healthier people may take medicine and recover faster $\implies$ correlation without causation
- In **observational data**
  - Confounding variable C affects both treatment T and outcome Y
  - C creates *spurious correlation* between T and Y



- **Problem**
  - There is a "backdoor path" *Treatment* $\leftarrow$ *Confounder* $\rightarrow$ *Outcome*

Frontdoor and Backdoor Paths: Intuition

- A **backdoor path** is any path from T to Y starting with an arrow into T
  - E.g., $T \leftarrow C \rightarrow Y$
  - Interpretation:
    - C is a common cause of T and Y, confounding their relationship
    - Controlling (conditioning) for C blocks the backdoor path, identifying the causal effect of T on Y



- A **frontdoor path** goes directly or indirectly from T to Y through mediators, following causal flow
  - E.g., $T \rightarrow Y$
  - Interpretation:
    - Direct causal path of interest
    - No mediators, so front-door path is direct causal effect of T on Y

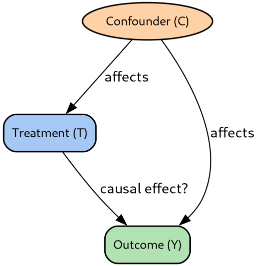
SCIENCE ACADEMY

Randomized Controlled Trials (RCTs)

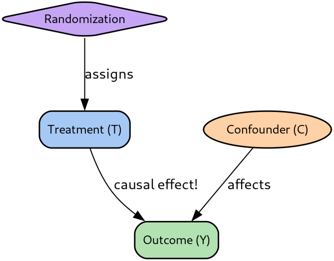
- **Randomized Controlled Trial** is an experimental study to assess causal effect of an intervention or treatment
  - Determine whether an intervention causes an effect, not just associated with it
  - Eliminate selection bias and confounding variables through randomization
- **Key Components**
  - *Randomization*: ensures groups are statistically equivalent at baseline
  - *Control Group*: receives a placebo or standard treatment
  - *Blinding*: participants and/or researchers do not know the assignment to avoid bias
  - *Outcome Measurement*: pre-defined metrics assess the intervention's effect
- **Example**: testing a new drug
  - Treatment group receives the new drug
  - Control group receives a placebo
  - Compare recovery rates after a fixed period
- **Pros**
  - Provides clear causal inference due to randomization
- **Cons**
  - Expensive and time-consuming
  - Ethical or practical constraints may prevent randomization

RCTs Solve the Problem of Confounders

- In **observational data**
  - Confounding variable C affects both treatment T and outcome Y
  - C creates *spurious correlation* between T and Y



- In **experimental settings**
  - Randomization (R) breaks link between C and T
  - Random assignment prevents influence on both treatment and outcome
  - T is independent of C: $T \perp C$
  - Only open path between T and Y is causal path $T \to Y$



SCIENCE ACADEMY

Causal Graphs and Interventions

- **Observing correlations** between variables *does not reveal causality*
  - $\Pr(Y|T)$ confounds direct and indirect influences
- **Randomized Controlled Trials** provide the *gold standard* for causal inference
  - Randomization breaks all back-door (confounding) paths
  - RCTs are expensive, slow, or ethically impossible
- **Alternative solution**
  - Can we estimate the *causal effect* from *observational data alone*?
  - Under *what conditions* and using *which variables*?
- **Idea**: Identify and condition on the right *confounders* to:
  - Block spurious associations between T and Y
  - Recover the true causal effect $\Pr(Y|do(T))$

Intervention in Structural Equations

- **Purpose of Structural Equations**
  - Capture causal mechanisms among variables
  - Predict impact of external interventions
- **Effect of Intervention** $do(X_j = x_j)$
  - Original equation:
    $$X_j = f_j(Parents(X_j), \varepsilon_j)$$
  - Modified by intervention:
    $$X_j = x_j \text{ (fixed value)}$$
  - "Mutilate" causal network by *removing incoming edges* to X_j
  - Recompute joint distribution of all variables using modified structure
- **Intuition**
  - *do*-operator enforces variable's value externally, breaking causal dependencies
  - Enables reasoning about "what would happen if. . . ?" scenarios

Adjustment Formula in Causal Networks

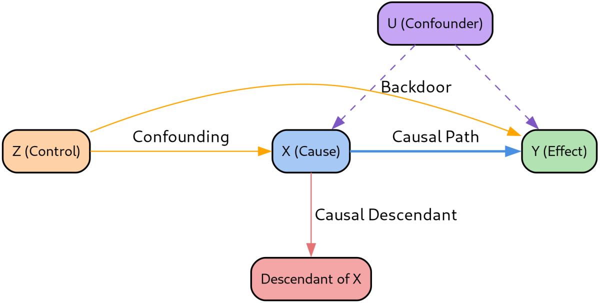
- **Goal**
  - Estimate causal effect of intervention $do(X_j = x_{jk})$ on another variable X_i
- **The Adjustment Formula**
  - Derived from the post-intervention joint distribution:

$$\Pr(X_i = x_i | do(X_j = x_j^*)) = \sum_{Parents(X_j)} \Pr(x_i | x_j^*, Parents(X_j)) \Pr(Parents(X_j))$$

  - The mechanism for X_j is *removed*: it is treated as a fixed cause, not a random variable
- **Interpretation**
  - Computes a *weighted average* of effects of X_j and its parents on X_i
  - Weights come from prior probabilities of the parents' values
- **Back-Door Criterion**
  - A set Z is a valid adjustment set if it blocks *all back-door paths* from X_j to X_i
  - Ensures $X_i \perp \mathrm{Parents}(X_j) | X_j, Z$
- **Why It Matters**
  - Enables causal inference from observational data
  - Estimate treatment and policy effects *without randomized trials*

Backdoor Criterion: Definition

- A set of variables Z satisfies the **backdoor criterion** for variables X (cause) and Y (effect) in a causal graph if:
  1. **No element of Z is a descendant of X**
     - Ensures Z does not "block" part of the causal effect of X on Y
     - Descendants of X may carry information about the causal effect and should not be controlled for
  2. **Z blocks every path between X and Y containing an arrow into X**
     - These paths are *backdoor paths*, representing potential confounding influences
     - Blocking them ensures any remaining association between X and Y is causal, not spurious



Backdoor Criterion: Intuition

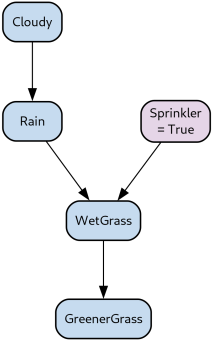
- **Intuition**:
  - The goal is to isolate the causal effect of X on Y by eliminating *confounding bias*
  - Controlling for an appropriate set Z makes the relationship between X and Y as if X were randomly assigned
- **Application**:
  - When Z satisfies the backdoor criterion, we can estimate causal effects from **observational data** (without experiments)
  - The causal effect can be computed using:

$$\Pr(Y|do(X)) = \sum_{z} \Pr(Y|X, Z = z)P(Z = z)$$

Intervention: Sprinkler Example

- "Intervene" by turning the sprinkler on
  - In do-calculus $do(Sprinkler = T)$
  - Sprinkler variable s is independent of cloudy day c
- Structural equations after intervention:

$$\begin{cases} C := f_C(\varepsilon_C) \\ R := f_R(C, \varepsilon_R) \\ S := True \\ W := f_W(R, S, \varepsilon_W) \\ G := f_G(W, \varepsilon_G) \end{cases}$$



- $\Pr(S|C) = 1$ and $\Pr(W|R,S) = \Pr(W|R, S = T)$ and the joint probability becomes:

$$\Pr(C, R, W, G|do(S = True)) = \Pr(C)\Pr(R|C)\Pr(W|R, S = True)\Pr(G|W)$$

- Only descendants of manipulated variable *Sprinkler* are affected

Intervention vs. Observation in Causal Models

- **Intervention** conceptually *breaks* normal causal dependencies
  - Intervening on *Sprinkler* removes causal link from *Weather* to *Sprinkler*
  - After intervention, causal graph excludes arrow $Weather \to Sprinkler$
  - *Weather* and *Sprinkler* become independent under intervention
- **Observation vs. Intervention**
  - **Observation**: seeing $Sprinkler = T$
    - Expressed as $\Pr(\cdot|Sprinkler = T)$
    - Reflects *passive observation* — sprinkler on provides information about weather
    - Since *Weather* influences *Sprinkler*, observing $Sprinkler = T$ makes it *less likely Weather* is cloudy
  - **Intervention**: forcing $Sprinkler = T$
    - Expressed as $\Pr(\cdot|do(Sprinkler = T))$
    - *Active manipulation* — set sprinkler on regardless of weather
    - Causal link from *Weather* to *Sprinkler* is cut, weather distribution remains unchanged
- **Key intuition**
  - Observation $\to$ correlation (information flows along causal links)
  - Intervention $\to$ causation (links into manipulated variable are removed)
  - Thus, $\Pr(Weather|Sprinkler = T) \neq \Pr(Weather|do(Sprinkler = T))$

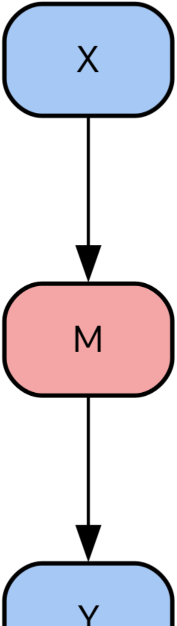
Controlling for a Variable in Causal Analysis

- **Definition**
  - To *control* a variable means to hold it constant (statistically or experimentally) to isolate the causal effect of another variable
- **Example**
  - Does exercise (X) cause weight loss (Y)?
  - Confounder: Diet (Z) affects both exercise and weight
  - By controlling for diet (e.g., comparing people with similar diets), you can estimate the effect of exercise more accurately
- In **regression analysis**
  - Include Z as an additional independent variable
  - E.g., in $Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$
    - β_1 measures the effect of X *controlling for* Z
    - Coefficient $\beta_1 =$ *change in* Y with a one-unit change in X_1, *holding* X_2 *constant*
    - Isolates X_1's unique contribution
    - Compares individuals with the same X_2 but different X_1
- In **experiments**
  - Keep Z constant or randomize it

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

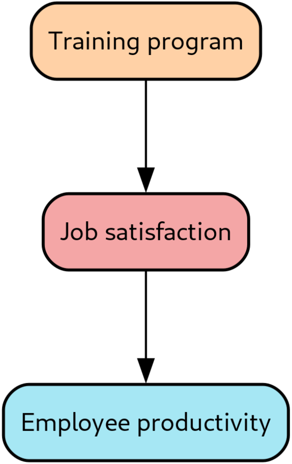
Mediator Variable

- 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



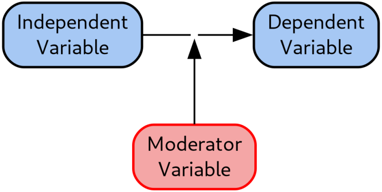
Mediator Variable: Example

- Research question: does a *training program* increase *employee productivity*?
- The causal effect may be **indirect**, operating through a **mediator**
  - The training program might not immediately boost productivity
  - Instead, it could enhance **job satisfaction**, which in turn raises productivity
- **Causal interpretation**
  - X: Training Program (cause)
  - M: Job Satisfaction (mediator)
  - Y: Employee Productivity (effect)
  - Path: $X \to M \to Y$
- **Direct vs. Indirect effects**
  - *Indirect effect* X affects Y through M
  - *Direct effect* X affects Y not through M
  - Controlling for M separates these two effects, clarifying *how* training impacts outcomes



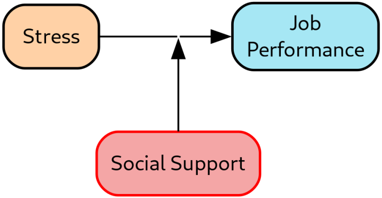
Moderator Variable

- A **moderator variable** changes the *strength* or *direction* of the relationship between an independent variable (X) and a dependent variable (Y)
  - Moderator 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



SCIENCE ACADEMY