Mental Models of (Causal) Structure in Economics and Psychology

Sandro Ambuehl
University of Zurich

Workshop on Beliefs, Narratives, and Memory

Riederau, September 2025

Based on the review article

Mental Models of (Causal) Structure
in Economics and Psychology

Sandro Ambuehl, Rahul Bhui, Heidi C. Thysen

(in progress, invited by JPE:microeconomics)

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Here

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Here

1. How do people learn the structure of the world? Do they mislearn?

    - E.g. What affects what? What is a symptom, cause, mediator? Which effects are direct, which indirect?

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Here

1. How do people learn the structure of the world? Do they mislearn? (**Structure learning**)
   - E.g. What affects what? What is a symptom, cause, mediator? Which effects are direct, which indirect?

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Here

1. How do people learn the structure of the world? Do they mislearn? (**Structure learning**)
   - E.g. What affects what? What is a symptom, cause, mediator? Which effects are direct, which indirect?
2. How do they learn the magnitudes of causal relations in the world when they have a (possibly wrong) model of the structure of the world?

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Here

1. How do people learn the structure of the world? Do they mislearn? (**Structure learning**)
   - E.g. What affects what? What is a symptom, cause, mediator? Which effects are direct, which indirect?
2. How do they learn the magnitudes of causal relations in the world when they have a (possibly wrong) model of the structure of the world? (**Parameter learning**)

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Here

1. How do people learn the structure of the world? Do they mislearn? (**Structure learning**)
   - E.g. What affects what? What is a symptom, cause, mediator? Which effects are direct, which indirect?
2. How do they learn the magnitudes of causal relations in the world when they have a (possibly wrong) model of the structure of the world? (**Parameter learning**)
3. How do people use their (fitted) structural models of the world to explain events?

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Here

1. How do people learn the structure of the world? Do they mislearn? (**Structure learning**)
   - E.g. What affects what? What is a symptom, cause, mediator? Which effects are direct, which indirect?
2. How do they learn the magnitudes of causal relations in the world when they have a (possibly wrong) model of the structure of the world? (**Parameter learning**)
3. How do people use their (fitted) structural models of the world to explain events? (**Causal reasoning**)

Traditional approach

Much research on belief updating and learning in economics such as balls and urns experiments (see Benjamin, 2019):

Fix a (very simple) structure of the world. Let people update beliefs from information. Do they get the magnitudes right?

Here

1. How do people learn the structure of the world? Do they mislearn? (**Structure learning**)
   - E.g. What affects what? What is a symptom, cause, mediator? Which effects are direct, which indirect?
2. How do they learn the magnitudes of causal relations in the world when they have a (possibly wrong) model of the structure of the world? (**Parameter learning**)
3. How do people use their (fitted) structural models of the world to explain events? (**Causal reasoning**)

Context: Causal Bayes Nets (a.k.a. Directed Acyclic Graphs, DAGs)

Some Examples

The Causal Bayes Nets approach is suitable in all areas of economics in which expectation formation matters

- Inflation expectations matter for monetary policy; if expectation formation is misspecified, central bank must respond (Spiegler, 2022). Laypeople's formation of inflation expectations systematically differs from that of experts (Andre et al., 2022)
- Traders with misspecified models (fewer variables than truth) shape asset prices. Generate well-known asset return patterns (Molavi et al., 2024)
- Morality and attribution of responsibility (Engl, 2022)
- Cycles of populism when people interpret the effect of public policy through wrong model (Levy et al., 2022). Coexistence of conflicting narratives in the public sphere (can predict which ones, can predict comparative statics; Eliaz and Spiegler, 2020).
- Mental models shape founders' perceptions of their firms' competitive advantage (Camuffo et al., 2024)

Plan

1. What is a Causal Bayesian Network?

Plan

1. What is a Causal Bayesian Network?
2. Causal reasoning (structure known, parameters known)

Plan

1. What is a Causal Bayesian Network?
2. Causal reasoning (structure known, parameters known)
3. Parameter learning (structure known, parameters unknown)

Plan

1. What is a Causal Bayesian Network?
2. Causal reasoning (structure known, parameters known)
3. Parameter learning (structure known, parameters unknown)
4. Structure learning (structure unknown, parameters unknown)

Plan

1. What is a Causal Bayesian Network?
2. Causal reasoning (structure known, parameters known)
3. Parameter learning (structure known, parameters unknown)
4. Structure learning (structure unknown, parameters unknown)
5. Measurement

Context in the broader literature

Tools come from a large statistical literature

- To model causality: "The Book of Why" (Pearl and Mackenzie, 2018), **?**
- Techniques about estimation etc. (non-causal): Koller and Friedman (2009)

Context in the broader literature

Tools come from a large statistical literature

- To model causality: "The Book of Why" (Pearl and Mackenzie, 2018), **?**
- Techniques about estimation etc. (non-causal): Koller and Friedman (2009)

Vast literature in cognitive science uses the tools to explain human cognition. Some book-length reviews:

- "Bayesian models of cognition" (Griffiths et al., 2024)
- "Oxford Handbook of Causal Reasoning" (Waldmann, 2017)

Context in the broader literature

Tools come from a large statistical literature

- To model causality: “The Book of Why” (Pearl and Mackenzie, 2018), **?**
- Techniques about estimation etc. (non-causal): Koller and Friedman (2009)

Vast literature in cognitive science uses the tools to explain human cognition. Some book-length reviews:

- “Bayesian models of cognition” (Griffiths et al., 2024)
- “Oxford Handbook of Causal Reasoning” (Waldmann, 2017)

There is a more general economics theory literature on misspecified models (reviewed in Bohren and Hauser, 2024). Has a higher level of abstraction, does not explicitly model structure. Hence, absent additional assumptions, makes far less specific predictions

- Key concept: Berk-Nash equilibrium (Esponda and Pouzo, 2016).

1. What is a Causal Bayesian Network?

Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY}\, X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

X Z

Y

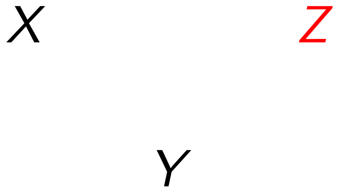
Exogenous:

$X = \beta_X + \varepsilon_X$

Endogenous:

$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$

$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$



Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$$

X Z

Y

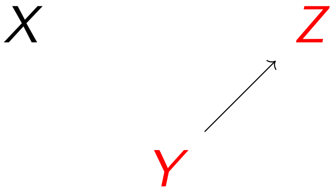
Exogenous:

$X = \beta_X + \varepsilon_X$

Endogenous:

$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$

$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$



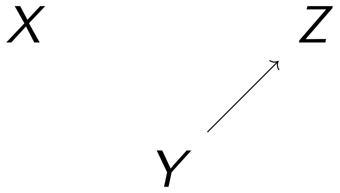
Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY}\, X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$



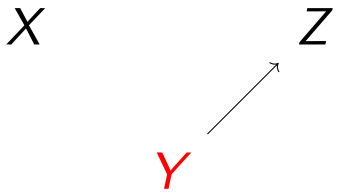
Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$$



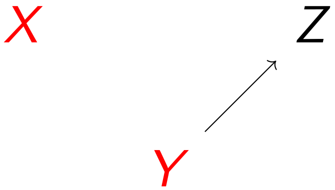
Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$$



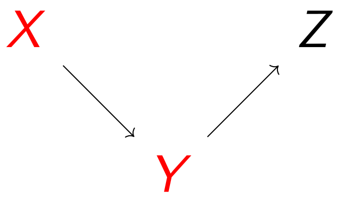
Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$$



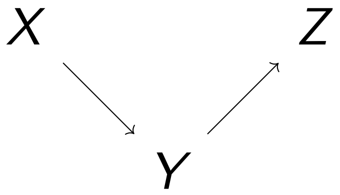
Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$$



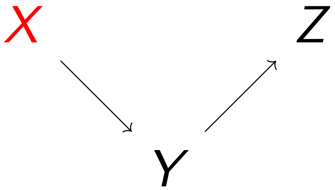
Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$$



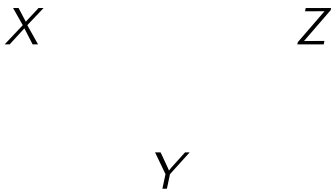
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\,X \\ 0 & \text{with probability } 1 - \left(p_Y + p_{XY}\,X\right) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\,Y \\ 0 & \text{with probability } 1 - \left(p_Z + p_{YZ}\,Y\right) \end{cases}$$



Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - \left(p_Y + p_{XY}\, X\right) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - \left(p_Z + p_{YZ}\, Y\right) \end{cases}$$

X Z

Y

Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - \left(p_Y + p_{XY}\, X\right) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - \left(p_Z + p_{YZ}\, Y\right) \end{cases}$$

X Z Y

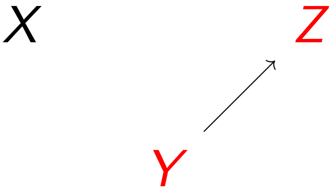
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - (p_Y + p_{XY}\, X) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - (p_Z + p_{YZ}\, Y) \end{cases}$$



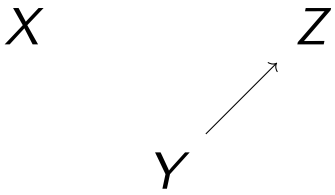
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - \left(p_Y + p_{XY}\, X\right) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - \left(p_Z + p_{YZ}\, Y\right) \end{cases}$$



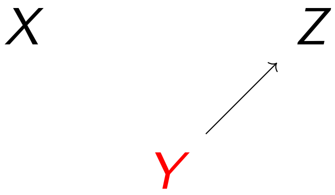
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$\color{red}{Y} = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - \left(p_Y + p_{XY}\, X\right) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - \left(p_Z + p_{YZ}\, Y\right) \end{cases}$$



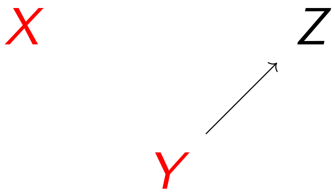
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - \left(p_Y + p_{XY}\, X\right) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - \left(p_Z + p_{YZ}\, Y\right) \end{cases}$$



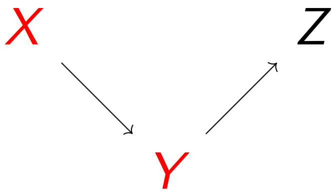
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - (p_Y + p_{XY}\, X) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - (p_Z + p_{YZ}\, Y) \end{cases}$$



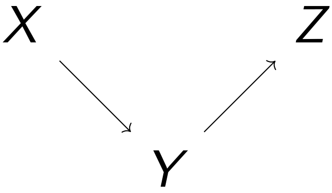
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - \left(p_Y + p_{XY}\, X\right) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - \left(p_Z + p_{YZ}\, Y\right) \end{cases}$$



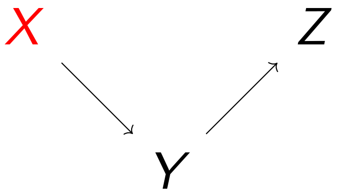
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - \left(p_Y + p_{XY}\, X\right) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - \left(p_Z + p_{YZ}\, Y\right) \end{cases}$$



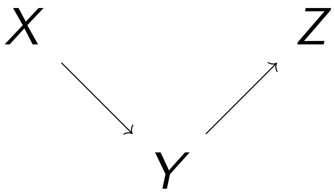
Exogenous:

$$X = \begin{cases} 1 & \text{with probability } p_X \\ 0 & \text{with probability } 1 - p_X \end{cases}$$

Endogenous:

$$Y = \begin{cases} 1 & \text{with probability } p_Y + p_{XY}\, X \\ 0 & \text{with probability } 1 - \big(p_Y + p_{XY}\, X\big) \end{cases}$$

$$Z = \begin{cases} 1 & \text{with probability } p_Z + p_{YZ}\, Y \\ 0 & \text{with probability } 1 - \big(p_Z + p_{YZ}\, Y\big) \end{cases}$$



The above equations specify a complete joint distribution over (X, Y, Z): for each $(x, y, z) \in \{0, 1\}^3$, it defines $P(X = x, Y = y, Z = z)$.

This is a big mess:

	X=0	
	Z=0	Z=1
Y=0	$(1-p_X)(1-p_Y)(1-p_Z)$	$(1-p_X)(1-p_Y)p_Z$
Y=1	$(1-p_X)p_Y(1-p_Z-p_{YZ})$	$(1-p_X)p_Y(p_Z+p_{YZ})$

	X=1	
	Z=0	Z=1
Y=0	$p_X(1-p_Y-p_{XY})(1-p_Z)$	$p_X(1-p_Y-p_{XY})p_Z$
Y=1	$p_X(p_Y+p_{XY})(1-p_Z-p_{YZ})$	$p_X(p_Y+p_{XY})(p_Z+p_{YZ})$

This is a big mess:

X=0		
	Z=0	Z=1
Y=0	$(1-p_X)(1-p_Y)(1-p_Z)$	$(1-p_X)(1-p_Y)p_Z$
Y=1	$(1-p_X)p_Y(1-p_Z-p_{YZ})$	$(1-p_X)p_Y(p_Z+p_{YZ})$

X=1		
	Z=0	Z=1
Y=0	$p_X(1-p_Y-p_{XY})(1-p_Z)$	$p_X(1-p_Y-p_{XY})p_Z$
Y=1	$p_X(p_Y+p_{XY})(1-p_Z-p_{YZ})$	$p_X(p_Y+p_{XY})(p_Z+p_{YZ})$

The key causal information can be represented much more sparsely, intuitively, and insightfully:

$$X \to Y \to Z$$

This is what we gain from using the DAG formalism!

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X \cup Y \cup Z)$$

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z)$$

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z) = P(Z \mid Y, X)\, P(Y, X)$$

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z) = P(Z \mid Y, X)\, P(Y \mid X)\, P(X)$$

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z) = P(Z \mid Y, X)\, P(Y \mid X)\, P(X)$$

IF we know that $X \to Y \to Z$

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z) = P(Z \mid Y, X)\, P(Y \mid X)\, P(X)$$

IF we know that $X \to Y \to Z$

- X affects Z only through Y. Hence, once we know Y, we cannot predict Z any better by also using X.

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z) = P(Z \mid Y, X)\, P(Y \mid X)\, P(X)$$

IF we know that $X \to Y \to Z$

- X affects Z only through Y. Hence, once we know Y, we cannot predict Z any better by also using X.
- Conditional on Y, Z is independent of X. Formally, $P(Z \mid Y, X) = P(Z \mid Y)$.

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z) = P(Z \mid Y, X)\, P(Y \mid X)\, P(X)$$

IF we know that $X \to Y \to Z$

- X affects Z only through Y. Hence, once we know Y, we cannot predict Z any better by also using X.
- Conditional on Y, Z is independent of X. Formally, $P(Z \mid Y, X) = P(Z \mid Y)$.
- Hence, we get the *factorization formula* that describes the DAG

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z) = P(Z \mid Y, X)\, P(Y \mid X)\, P(X)$$

IF we know that $X \to Y \to Z$

- X affects Z only through Y. Hence, once we know Y, we cannot predict Z any better by also using X.
- Conditional on Y, Z is independent of X. Formally, $P(Z \mid Y, X) = P(Z \mid Y)$.
- Hence, we get the *factorization formula* that describes the DAG

$$P(X, Y, Z) = P(Z \mid Y)\, P(Y \mid X)\, P(X)$$

The Factorization Formula for the DAG $X \to Y \to Z$

The definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ implies the *chain rule*

$$P(A \cap B) = P(A \mid B)\, P(B)$$

In the case of three variables

$$P(X, Y, Z) = P(Z \mid Y, X)\, P(Y \mid X)\, P(X)$$

IF we know that $X \to Y \to Z$

- X affects Z only through Y. Hence, once we know Y, we cannot predict Z any better by also using X.
- Conditional on Y, Z is independent of X. Formally, $P(Z \mid Y, X) = P(Z \mid Y)$.
- Hence, we get the *factorization formula* that describes the DAG

$$P(X, Y, Z) = P(Z \mid Y)\, P(Y \mid X)\, P(X)$$

- Which links are absent matters much more than which links are present!

Causal interventions

Consider $X \rightarrow Y$.

E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Causal interventions

Consider $X \rightarrow Y$.
E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Correlation is not causation

Causal interventions

Consider $X \to Y$.
E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Correlation is not causation

- Correlation: If Y was higher, X must have been higher

Causal interventions

Consider $X \to Y$.
E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Correlation is not causation

- Correlation: If Y was higher, X must have been higher
- Causation: If we raise Y, we do <u>not</u> change X

Causal interventions

Consider $X \rightarrow Y$.
E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Correlation is not causation

- Correlation: If Y was higher, X must have been higher
- Causation: If we raise Y, we do <u>not</u> change X

Formally, $P(X|Y) \neq P(X|do(Y))$

Causal interventions

Consider $X \to Y$.
E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Correlation is not causation

- Correlation: If Y was higher, X must have been higher
- Causation: If we raise Y, we do <u>not</u> change X

Formally, $P(X|Y) \neq P(X|do(Y))$

do-operator

The *do*-operator changes the causal model

Causal interventions

Consider $X \to Y$.
E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Correlation is not causation

- Correlation: If Y was higher, X must have been higher
- Causation: If we raise Y, we do not change X

Formally, $P(X|Y) \neq P(X|do(Y))$

do-operator

The *do*-operator changes the causal model

- Disconnect intervention variables from their usual causes (even if endogenous) and replace them with constants

Causal interventions

Consider $X \to Y$.
E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Correlation is not causation

- Correlation: If Y was higher, X must have been higher
- Causation: If we raise Y, we do not change X

Formally, $P(X|Y) \neq P(X|do(Y))$

do-operator

The *do*-operator changes the causal model

- Disconnect intervention variables from their usual causes (even if endogenous) and replace them with constants
  - $do(Y)$ disconnects Y from X, hence does not affect X

Causal interventions

Consider $X \to Y$.
E.g. $Y = \alpha + \beta X + \epsilon$, with $\beta > 0$

Correlation is not causation

- Correlation: If Y was higher, X must have been higher
- Causation: If we raise Y, we do not change X

Formally, $P(X|Y) \neq P(X|do(Y))$

do-operator

The *do*-operator changes the causal model

- Disconnect intervention variables from their usual causes (even if endogenous) and replace them with constants
  - $do(Y)$ disconnects Y from X, hence does not affect X
- Difference between original and resulting distribution is the *causal effect*

Fully specified probabilistic causal models

A. Linear Gaussian

Exogenous:

$$X = \beta_X + \varepsilon_X$$

Endogenous:

$$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z,$$

B. Linear binary variables

Exogenous:

$$X = L(1, 0; p_X)$$

Endogenous:

$$Y = L(1, 0; p_Y + p_{XY} X)$$

$$Z = L(1, 0; p_Z + p_{YZ} A)$$

Abstract representations

C. DAG

$$\begin{array}{ccccc} X & & & & Z \\ & \searrow & & \nearrow & \\ & & Y & & \end{array}$$

D. Factorization formula

$$P(X, Y, Z) = P(Z|Y)P(Y|X)P(X)$$

Level of abstraction at which DAGs operate

A DAG represents

- correlational structure
- causal structure

Level of abstraction at which DAGs operate

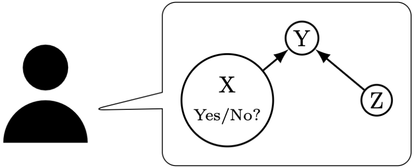
A DAG represents

- correlational structure
- causal structure

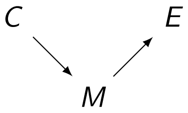
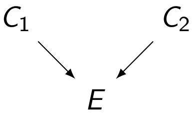
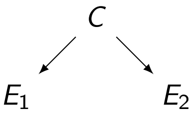
A DAG abstracts from

- Nature of the random variables (discrete, continuous, etc.)
- Whether an effect is positive or negative
- Functional forms

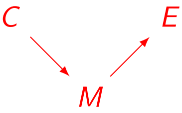
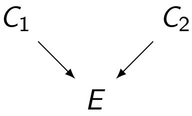
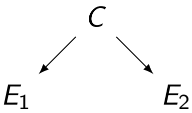
2. Causal reasoning



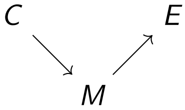
Three archetypical causal structures

	(A)	(B)	(C)
Common labels	Chain, Line	*v*-collider, Common Effect	Fork, Common Cause
			
Example	Salt consumption increases blood pressure which decreases life expectancy	Two paths of admission to Harvard: (i) Smarts, (ii) Wealth	Sickle cell disease decreases life expectancy and protects against malaria

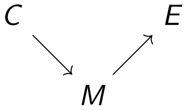
Three archetypical causal structures

	(A)	(B)	(C)
Common labels	Chain, Line	v-collider, Common Effect	Fork, Common Cause
			
Example	Salt consumption increases blood pressure which decreases life expectancy	Two paths of admission to Harvard: (i) Smarts, (ii) Wealth	Sickle cell disease decreases life expectancy and protects against malaria

Chain: Key correlational implications

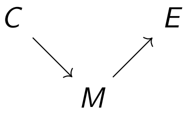


Chain: Key correlational implications



(i) $\text{cov}(C, E) > 0$ (generically), since cause C indirectly affects effect E

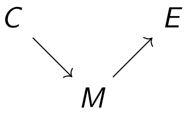
Chain: Key correlational implications



(i) $\text{cov}(C, E) > 0$ (generically), since cause C indirectly affects effect E

(ii) $\text{cov}(C, E|M) = 0$, since holding mediator M fixed blocks the effect that cause C could have on effect E (*blocking*)

Chain: Key correlational implications



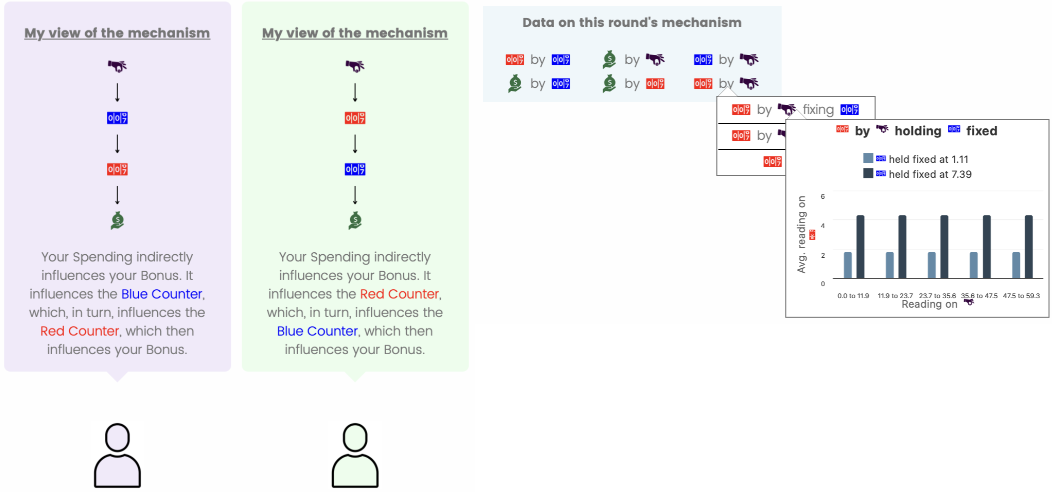
(i) $\text{cov}(C, E) > 0$ (generically), since cause C indirectly affects effect E

(ii) $\text{cov}(C, E|M) = 0$, since holding mediator M fixed blocks the effect that cause C could have on effect E (*blocking*)

(ii) is an example of
The *Causal Markov Condition*: Conditional on immediate predecessors, a variable X is independent of all variables that are not consequences of X

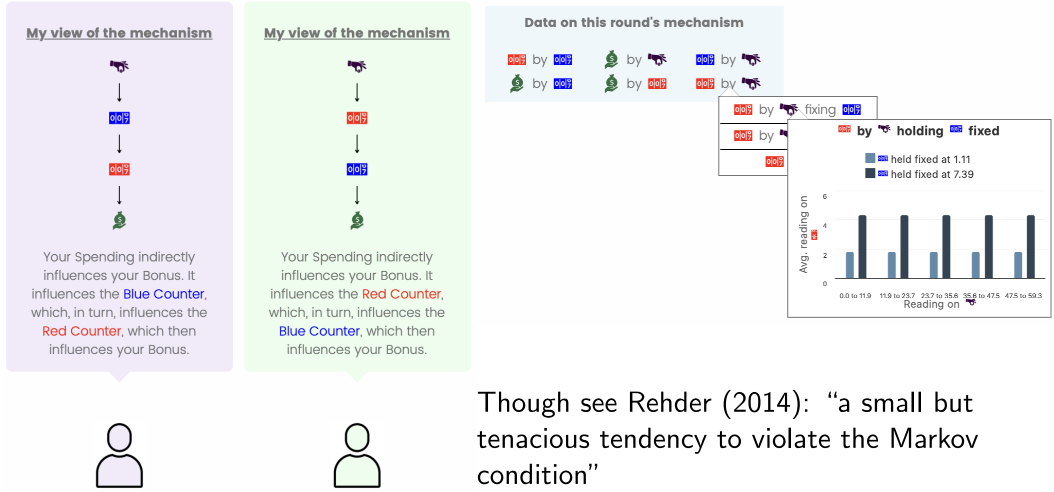
Evidence speaking to humans' understanding of the causal Markov condition

2/3 of subjects in Ambuehl and Thysen (2025) intuitively understand blocking without explanation and connect it to the data to find the correctly specified of two models.



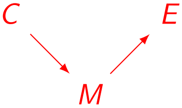
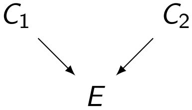
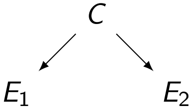
Evidence speaking to humans' understanding of the causal Markov condition

2/3 of subjects in Ambuehl and Thysen (2025) intuitively understand blocking without explanation and connect it to the data to find the correctly specified of two models.



Though see Rehder (2014): "a small but tenacious tendency to violate the Markov condition"

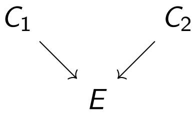
Three archetypical causal structures

	(A)	(B)	(C)
Common labels	Chain, Line	*v*-collider, Common Effect	Fork, Common Cause
			
Example	Salt consumption increases blood pressure which decreases life expectancy	Two paths of admission to Harvard: (i) Smarts, (ii) Wealth	Sickle cell disease decreases life expectancy and protects against malaria

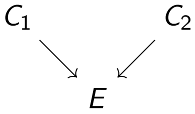
Three archetypical causal structures

	(A)	(B)	(C)
Common labels	Chain, Line	*v*-collider, Common Effect	Fork, Common Cause
	$C \rightarrow M \rightarrow E$	$C_1 \rightarrow E \leftarrow C_2$	$E_1 \leftarrow C \rightarrow E_2$
Example	Salt consumption increases blood pressure which decreases life expectancy	Two paths of admission to Harvard: (i) Smarts, (ii) Wealth	Sickle cell disease decreases life expectancy and protects against malaria

Common cause: Key correlational implications

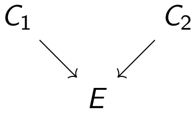


Common cause: Key correlational implications



(i) $\text{cov}(C_1, C_2) = 0$

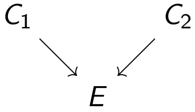
Common cause: Key correlational implications



(i) $\text{cov}(C_1, C_2) = 0$

(ii) $\text{cov}(C_1, C_2|E) \neq 0$ (generically)

Common cause: Key correlational implications



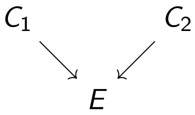
(i) $\text{cov}(C_1, C_2) = 0$

(ii) $\text{cov}(C_1, C_2|E) \neq 0$ (generically)

Collider Bias

(ii) is called *Collider Bias*: conditioning on the effect alters the apparent correlation between the two causes (also see *Berkson's paradox*).

Common cause: Key correlational implications



(i) $\text{cov}(C_1, C_2) = 0$

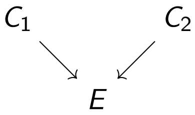
(ii) $\text{cov}(C_1, C_2|E) \neq 0$ (generically)

Collider Bias

(ii) is called *Collider Bias*: conditioning on the effect alters the apparent correlation between the two causes (also see *Berkson's paradox*).

- Example 1: Two ways to get into Harvard: smart or rich. You learn that a Harvard student is rich. How smart do you think they are relative to other Harvard students?

Common cause: Key correlational implications



(i) $\text{cov}(C_1, C_2) = 0$

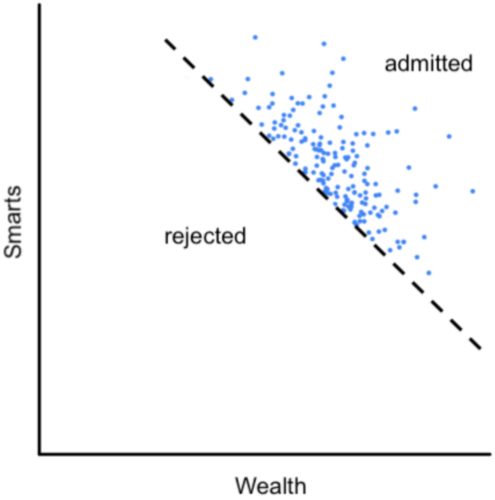
(ii) $\text{cov}(C_1, C_2|E) \neq 0$ (generically)

Collider Bias

(ii) is called *Collider Bias*: conditioning on the effect alters the apparent correlation between the two causes (also see *Berkson's paradox*).

- Example 1: Two ways to get into Harvard: smart or rich. You learn that a Harvard student is rich. How smart do you think they are relative to other Harvard students?
- Example 2: $E = C_1 + C_2$

Prior to conditioning: smarts and wealth uncorrelated



After selection: smarts and wealth correlated

Explaining Away

Under additional assumptions on the DGP, collider bias leads to *Explaining Away*:

Suppose each of two causes can cause an effect. You know the effect happened. Then, learning that cause 1 occurred decreases the posterior that cause 2 occurred:

$$P(C_2 \mid E, C_1) < P(C_2 \mid E).$$

E.g.

$$P(\text{smart} \mid \text{Harvard student}, \text{rich parents}) < P(\text{smart} \mid \text{Harvard student}).$$

Explaining Away

Under additional assumptions on the DGP, collider bias leads to *Explaining Away*:

Suppose each of two causes can cause an effect. You know the effect happened. Then, learning that cause 1 occurred decreases the posterior that cause 2 occurred:

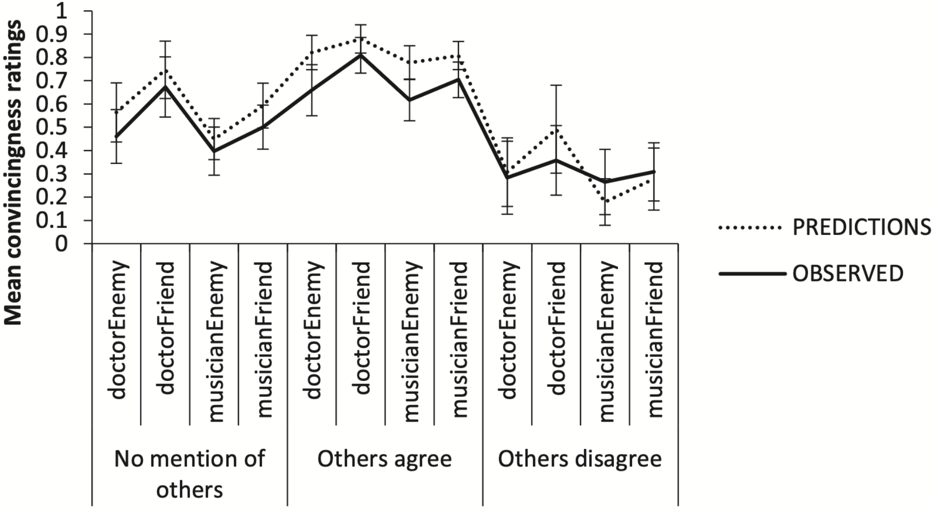
$$P(C_2 \mid E, C_1) < P(C_2 \mid E).$$

E.g.

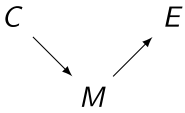
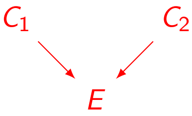
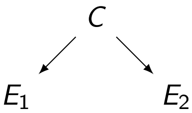
$$P(\text{smart} \mid \text{Harvard student}, \text{rich parents}) < P(\text{smart} \mid \text{Harvard student}).$$

- Experimental subjects generally adhere to the directional predictions (Rottman and Hastie, 2014).

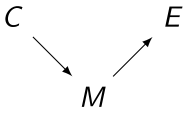
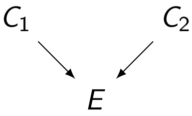
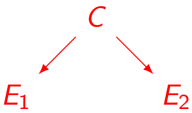
Do people also get the magnitudes right? Harris et al. (2016)



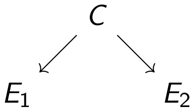
Three archetypical causal structures

	(A)	(B)	(C)
Common labels	Chain, Line	*v*-collider, Common Effect	Fork, Common Cause
			
Example	Salt consumption increases blood pressure which decreases life expectancy	Two paths of admission to Harvard: (i) Smarts, (ii) Wealth	Sickle cell disease decreases life expectancy and protects against malaria

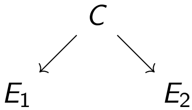
Three archetypical causal structures

	(A)	(B)	(C)
Common labels	Chain, Line	*v*-collider, Common Effect	Fork, Common Cause
			
Example	Salt consumption increases blood pressure which decreases life expectancy	Two paths of admission to Harvard: (i) Smarts, (ii) Wealth	Sickle cell disease decreases life expectancy and protects against malaria

Fork: Key correlational implications

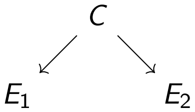


Fork: Key correlational implications



(i) $\text{cov}(E_1, E_2) > 0$ (generically), since cause C affects both effects E_1 and E_2

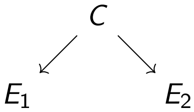
Fork: Key correlational implications



(i) $\text{cov}(E_1, E_2) > 0$ (generically), since cause C affects both effects E_1 and E_2

(ii) $\text{cov}(E_1, E_2 | C) = 0$, since holding common cause C eliminates the sole reason that created the correlation between C_1 and C_2

Fork: Key correlational implications

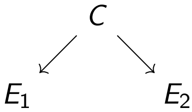


(i) $\text{cov}(E_1, E_2) > 0$ (generically), since cause C affects both effects E_1 and E_2

(ii) $\text{cov}(E_1, E_2 | C) = 0$, since holding common cause C eliminates the sole reason that created the correlation between C_1 and C_2

These correlational implications are the same as for the chain $(E_1 \to C \to E_2)$.

Fork: Key correlational implications



(i) $\text{cov}(E_1, E_2) > 0$ (generically), since cause C affects both effects E_1 and E_2

(ii) $\text{cov}(E_1, E_2|C) = 0$, since holding common cause C eliminates the sole reason that created the correlation between C_1 and C_2

These correlational implications are the same as for the chain ($E_1 \rightarrow C \rightarrow E_2$).

Two DAGs with identical correlational implications are called *Markov-equivalent*

Theorem: "Markov-equivalence = skeleton + *v*-colliders" (Verma and Pearl, 1991)

Two DAGs are Markov-equivalent (have the same set of conditional independence relationships) if and only if they have the same skeleton (i.e. once we drop arrowheads, the DAGs are identical) and the same set of v-colliders).

Theorem: "Markov-equivalence = skeleton + v-colliders" (Verma and Pearl, 1991)

Two DAGs are Markov-equivalent (have the same set of conditional independence relationships) if and only if they have the same skeleton (i.e. once we drop arrowheads, the DAGs are identical) and the same set of v-colliders).

Examples

- Chain ($X \to Y \to Z$) and common cause ($X \leftarrow Y \to Z$) .

Theorem: "Markov-equivalence = skeleton + v-colliders" (Verma and Pearl, 1991)

Two DAGs are Markov-equivalent (have the same set of conditional independence relationships) if and only if they have the same skeleton (i.e. once we drop arrowheads, the DAGs are identical) and the same set of v-colliders).

Examples

- Chain ($X \to Y \to Z$) and common cause ($X \leftarrow Y \to Z$) are Markov-equivalent.

Theorem: "Markov-equivalence = skeleton + v-colliders" (Verma and Pearl, 1991)

Two DAGs are Markov-equivalent (have the same set of conditional independence relationships) if and only if they have the same skeleton (i.e. once we drop arrowheads, the DAGs are identical) and the same set of v-colliders).

Examples

- Chain ($X \to Y \to Z$) and common cause ($X \leftarrow Y \to Z$) are Markov-equivalent.
- Common cause ($X \leftarrow Y \to Z$) and common effect ($X \to Y \leftarrow Z$)

.

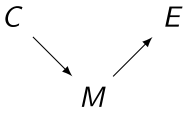
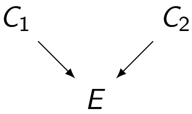
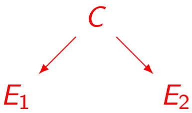
Theorem: "Markov-equivalence = skeleton + v-colliders" (Verma and Pearl, 1991)

Two DAGs are Markov-equivalent (have the same set of conditional independence relationships) if and only if they have the same skeleton (i.e. once we drop arrowheads, the DAGs are identical) and the same set of v-colliders).

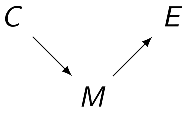
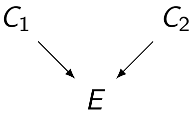
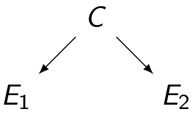
Examples

- Chain ($X \to Y \to Z$) and common cause ($X \leftarrow Y \to Z$) are Markov-equivalent.
- Common cause ($X \leftarrow Y \to Z$) and common effect ($X \to Y \leftarrow Z$) are not Markov-equivalent.

Three archetypical causal structures

	(A)	(B)	(C)
Common labels	Chain, Line	*v*-collider, Common Effect	Fork, Common Cause
			
Example	Salt consumption increases blood pressure which decreases life expectancy	Two paths of admission to Harvard: (i) Smarts, (ii) Wealth	Sickle cell disease decreases life expectancy and protects against malaria

Three archetypical causal structures

	(A)	(B)	(C)
Common labels	Chain, Line	*v*-collider, Common Effect	Fork, Common Cause
			
Example	Salt consumption increases blood pressure which decreases life expectancy	Two paths of admission to Harvard: (i) Smarts, (ii) Wealth	Sickle cell disease decreases life expectancy and protects against malaria

“Correlation does not imply causation”

- True, but correlation carries some information about causation.
- Because different causal structures have different correlational implications.

3. Parameter learning

3. Parameter learning

- So far: Causal reasoning

  “*I know what generally affects what and by how much.*
  What happened in this specific instance?”
  (e.g. explaining away)

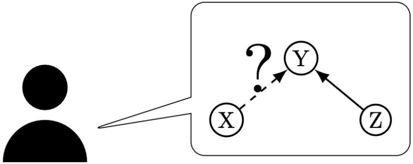
3. Parameter learning

- So far: Causal reasoning

  "I know what generally affects what and by how much. What happened in this specific instance?"
  (e.g. explaining away)

- Now: Parameter learning

  "I know what affects what, but not by how much"



3. Parameter learning

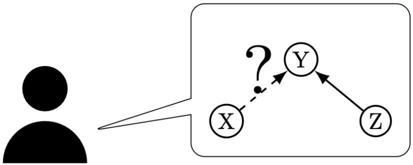
- So far: Causal reasoning

  “*I know what generally affects what and by how much.*
  What happened in this specific instance?”
  (e.g. explaining away)

- Now: Parameter learning

  “*I know what affects what, but not by how much*”

Viewing the world through the lens of a causal model



Parameter learning

When the DAG represents a system of regression equations, then parameter learning is estimating the system (e.g. with OLS)

Parameter learning

When the DAG represents a system of regression equations, then parameter learning is estimating the system (e.g. with OLS)

Questions

Parameter learning

When the DAG represents a system of regression equations, then parameter learning is estimating the system (e.g. with OLS)

Questions

1. When will estimating misspecified DAGs cause errors, and when won't it? How bad can misperceptions get?
   - Mostly not today, see our review paper or Spiegler (2020a)

Parameter learning

When the DAG represents a system of regression equations, then parameter learning is estimating the system (e.g. with OLS)

Questions

1. When will estimating misspecified DAGs cause errors, and when won't it? How bad can misperceptions get?
   - Mostly not today, see our review paper or Spiegler (2020a)
2. When individuals have the wrong DAG in mind but view the world through it (fit it to the data), what are economic implications?

If you are estimating a misspecified model (your model's DAG $\neq$ DGP's DAG), will the misspecification cause wrong choices (interventions on variables)?

If you are estimating a misspecified model (your model's DAG $\neq$ DGP's DAG), will the misspecification cause wrong <u>choices</u> (interventions on variables)?

- If you want to intervene with any node: You need the correct DAG.

If you are estimating a misspecified model (your model's DAG $\neq$ DGP's DAG), will the misspecification cause wrong <u>choices</u> (interventions on variables)?

- If you want to intervene with any node: You need the correct DAG.
- If there's only a single node you can affect, and some 'outcome' nodes: See Spiegler (2016)

If you are estimating a misspecified model (your model's DAG $\neq$ DGP's DAG), will the misspecification cause wrong choices (interventions on variables)?

- If you want to intervene with any node: You need the correct DAG.
- If there's only a single node you can affect, and some 'outcome' nodes: See Spiegler (2016)
  - Choice will be correct if the subjective DAG is Markov-equivalent to some DAG in which the 'outcome' nodes form an ancestral clique

If you are estimating a misspecified model, will the misspecification cause wrong predictions?

If you are estimating a misspecified model, will the misspecification cause wrong predictions?

Answer: The misspecification will not matter as long as your DAG is *Markov-equivalent* to the true DGP (same correlational implications).

If you are estimating a misspecified model, will the misspecification cause wrong predictions?

Answer: The misspecification will not matter as long as your DAG is *Markov-equivalent* to the true DGP (same correlational implications).

Recall: Verma and Pearl, 1991

Two DAGs are Markov-equivalent if and only if they have the same skeleton and the same set of v-colliders.

If you are estimating a misspecified model, will the misspecification cause wrong <u>predictions</u>?

Answer: The misspecification will not matter as long as your DAG is *Markov-equivalent* to the true DGP (same correlational implications).

Recall: Verma and Pearl, 1991

Two DAGs are Markov-equivalent if and only if they have the same skeleton and the same set of v-colliders.

Examples

- If you fit a Chain ($X \to Y \to Z$) even though the actual data-generating process is a Fork ($X \leftarrow Y \to Z$),

If you are estimating a misspecified model, will the misspecification cause wrong predictions?

Answer: The misspecification will not matter as long as your DAG is *Markov-equivalent* to the true DGP (same correlational implications).

Recall: Verma and Pearl, 1991

Two DAGs are Markov-equivalent if and only if they have the same skeleton and the same set of v-colliders.

Examples

- If you fit a Chain ($X \rightarrow Y \rightarrow Z$) even though the actual data-generating process is a Fork ($X \leftarrow Y \rightarrow Z$), you will still make correct predictions.

If you are estimating a misspecified model, will the misspecification cause wrong predictions?

Answer: The misspecification will not matter as long as your DAG is *Markov-equivalent* to the true DGP (same correlational implications).

Recall: Verma and Pearl, 1991

Two DAGs are Markov-equivalent if and only if they have the same skeleton and the same set of v-colliders.

Examples

- If you fit a Chain ($X \to Y \to Z$) even though the actual data-generating process is a Fork ($X \leftarrow Y \to Z$), you will still make correct predictions.
- If you fit a Chain ($X \to Y \to Z$) even though the actual data-generating process is a Collider ($X \to Y \leftarrow Z$),

If you are estimating a misspecified model, will the misspecification cause wrong <u>predictions</u>?

Answer: The misspecification will not matter as long as your DAG is *Markov-equivalent* to the true DGP (same correlational implications).

Recall: Verma and Pearl, 1991

Two DAGs are Markov-equivalent if and only if they have the same skeleton and the same set of v-colliders.

Examples

- If you fit a Chain ($X \rightarrow Y \rightarrow Z$) even though the actual data-generating process is a Fork ($X \leftarrow Y \rightarrow Z$), you will still make correct predictions.
- If you fit a Chain ($X \rightarrow Y \rightarrow Z$) even though the actual data-generating process is a Collider ($X \rightarrow Y \leftarrow Z$), you will make wrong predictions.

If you are estimating a misspecified model, will the misspecification cause wrong predictions?

Answer: The misspecification will not matter as long as your DAG is *Markov-equivalent* to the true DGP (same correlational implications).

Recall: Verma and Pearl, 1991

Two DAGs are Markov-equivalent if and only if they have the same skeleton and the same set of v-colliders.

Examples

- If you fit a Chain ($X \to Y \to Z$) even though the actual data-generating process is a Fork ($X \leftarrow Y \to Z$), you will still make correct predictions.
- If you fit a Chain ($X \to Y \to Z$) even though the actual data-generating process is a Collider ($X \to Y \leftarrow Z$), you will make wrong predictions.

Literature

- Additional, more econ-specific characterizations in Spiegler (2016, 2017, 2020b)
- How bad can the predictions from misspecified models get? Eliaz et al. (2020)

Data-generating process

$$X = \beta_X + \varepsilon_X$$

$$Y = \beta_Y + \beta_{XY}\, X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

Which model (if any) produces correct predictions when fit to the data?

Data-generating process

$$X = \beta_X + \varepsilon_X$$

$$Y = \beta_Y + \beta_{XY}\, X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

Which model (if any) produces correct predictions when fit to the data?

Data-generating process

$$X = \beta_X + \varepsilon_X$$

$$Y = \beta_Y + \beta_{XY} X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ} Y + \varepsilon_Z$$

Which model (if any) produces correct predictions when fit to the data?

Candidate 1

$$X = \beta_X + \varepsilon_X$$

$$Y = \beta_Y + \beta_{XY} X + \beta_{ZY} Z + \varepsilon_Y$$

$$Z = \beta_Z + \varepsilon_Z$$

Data-generating process

$$X = \beta_X + \varepsilon_X$$
$$Y = \beta_Y + \beta_{XY}\, X + \varepsilon_Y$$
$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

Which model (if any) produces correct predictions when fit to the data?

Candidate 1

$$X = \beta_X + \varepsilon_X$$
$$Y = \beta_Y + \beta_{XY}\, X + \beta_{ZY}\, Z + \varepsilon_Y$$
$$Z = \beta_Z + \varepsilon_Z$$

Candidate 2

$$X = \beta_X + \beta_{YX}\, Y + \varepsilon_X$$
$$Y = \beta_Y + \varepsilon_Y$$
$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

Data-generating process

$$X = \beta_X + \varepsilon_X$$

$$Y = \beta_Y + \beta_{XY}\, X + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

$$X \longrightarrow Y \longrightarrow Z$$

Which model (if any) produces correct predictions when fit to the data?

Candidate 1

$$X = \beta_X + \varepsilon_X$$

$$Y = \beta_Y + \beta_{XY}\, X + \beta_{ZY}\, Z + \varepsilon_Y$$

$$Z = \beta_Z + \varepsilon_Z$$

Candidate 2

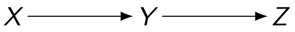
$$X = \beta_X + \beta_{YX}\, Y + \varepsilon_X$$

$$Y = \beta_Y + \varepsilon_Y$$

$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

Data-generating process

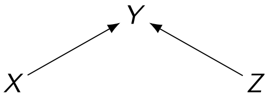
$$X = \beta_X + \varepsilon_X$$
$$Y = \beta_Y + \beta_{XY}\, X + \varepsilon_Y$$
$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$



Which model (if any) produces correct predictions when fit to the data?

Candidate 1

$$X = \beta_X + \varepsilon_X$$
$$Y = \beta_Y + \beta_{XY}\, X + \beta_{ZY}\, Z + \varepsilon_Y$$
$$Z = \beta_Z + \varepsilon_Z$$



Candidate 2

$$X = \beta_X + \beta_{YX}\, Y + \varepsilon_X$$
$$Y = \beta_Y + \varepsilon_Y$$
$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

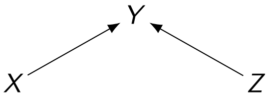
Data-generating process

$$X = \beta_X + \varepsilon_X$$
$$Y = \beta_Y + \beta_{XY}\, X + \varepsilon_Y$$
$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$

Which model (if any) produces correct predictions when fit to the data?

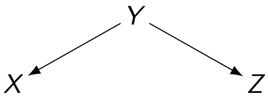
Candidate 1

$$X = \beta_X + \varepsilon_X$$
$$Y = \beta_Y + \beta_{XY}\, X + \beta_{ZY}\, Z + \varepsilon_Y$$
$$Z = \beta_Z + \varepsilon_Z$$



Candidate 2

$$X = \beta_X + \beta_{YX}\, Y + \varepsilon_X$$
$$Y = \beta_Y + \varepsilon_Y$$
$$Z = \beta_Z + \beta_{YZ}\, Y + \varepsilon_Z$$



Economic theory literature

- What systematic mistakes will an agent make whose choices depend on a misspecified model? See before.

Economic theory literature

- What systematic mistakes will an agent make whose choices depend on a misspecified model? See before.
- What if her actions affect the data from which she (mis)learns? Now.

Economic theory literature

- What systematic mistakes will an agent make whose choices depend on a misspecified model? See before.
- What if her actions affect the data from which she (mis)learns? Now.

The assumption that economic agents fit misspecified DAGs to data is a powerful, general tool for modeling 'behavioral' distortions

"Graphical causal models are entirely nonparametric, they are applicable to any static decision problem... the framework thus provides a "general recipe" for transforming a standard rational expectations model into an equilibrium model with nonrational expectations."

Spiegler (2016)

Economic theory literature

- What systematic mistakes will an agent make whose choices depend on a misspecified model? See before.
- What if her actions affect the data from which she (mis)learns? Now.

The assumption that economic agents fit misspecified DAGs to data is a powerful, general tool for modeling 'behavioral' distortions

"Graphical causal models are entirely nonparametric, they are applicable to any static decision problem... the framework thus provides a "general recipe" for transforming a standard rational expectations model into an equilibrium model with nonrational expectations."

Spiegler (2016)

Key example: Dieter's Dilemma

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness
- Blood chemical is a symptom that is mistaken for a mediator

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness
- Blood chemical is a symptom that is mistaken for a mediator
- Assume: (i) Medication masks this symptom (ii) Medication is costly.

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness
- Blood chemical is a symptom that is mistaken for a mediator
- Assume: (i) Medication masks this symptom (ii) Medication is costly.
- Hence, DM with correct model would not take the medication

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness
- Blood chemical is a symptom that is mistaken for a mediator
- Assume: (i) Medication masks this symptom (ii) Medication is costly.
- Hence, DM with correct model would not take the medication
- What will the misspecified DM do?

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness
- Blood chemical is a symptom that is mistaken for a mediator
- Assume: (i) Medication masks this symptom (ii) Medication is costly.
- Hence, DM with correct model would not take the medication
- What will the misspecified DM do?

Analysis

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness
- Blood chemical is a symptom that is mistaken for a mediator
- Assume: (i) Medication masks this symptom (ii) Medication is costly.
- Hence, DM with correct model would not take the medication
- What will the misspecified DM do?

Analysis

- Never take medication $\rightarrow$ symptom never masked $\rightarrow$ observe strong correlation between symptom and illness, intepreted causally $\rightarrow$ medication seems effective, start taking it

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness
- Blood chemical is a symptom that is mistaken for a mediator
- Assume: (i) Medication masks this symptom (ii) Medication is costly.
- Hence, DM with correct model would not take the medication
- What will the misspecified DM do?

Analysis

- Never take medication $\rightarrow$ symptom never masked $\rightarrow$ observe strong correlation between symptom and illness, intepreted causally $\rightarrow$ medication seems effective, start taking it
- Always take medication $\rightarrow$ symptom always masked $\rightarrow$ observe no correlation between symptom and illness $\rightarrow$ medication seems ineffective, stop taking it

Dieter's Dilemma (Spiegler, 2016)

Setting

- You believe: Medication $\rightarrow$ Blood chemical level $\rightarrow$ Illness
- True DGP: Medication $\rightarrow$ Blood chemical level $\leftarrow$ Illness
- Blood chemical is a symptom that is mistaken for a mediator
- Assume: (i) Medication masks this symptom (ii) Medication is costly.
- Hence, DM with correct model would not take the medication
- What will the misspecified DM do?

Analysis

- Never take medication $\rightarrow$ symptom never masked $\rightarrow$ observe strong correlation between symptom and illness, intepreted causally $\rightarrow$ medication seems effective, start taking it
- Always take medication $\rightarrow$ symptom always masked $\rightarrow$ observe no correlation between symptom and illness $\rightarrow$ medication seems ineffective, stop taking it
- Interior equilibrium: take the medication sometimes, so that perceived correlation just strong enough that DM indifferent between taking and not taking it

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

A	X	Y
1	0	0
0	1	0
$\vdots$	$\vdots$	$\vdots$

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

A	X	Y
1	0	0
0	1	0
$\vdots$	$\vdots$	$\vdots$

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

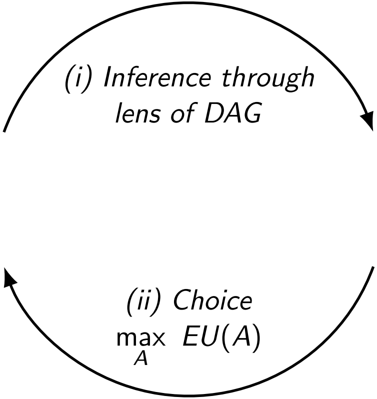
A	X	Y
1	0	0
0	1	0
$\vdots$	$\vdots$	$\vdots$

(i) Inference through lens of DAG

$$EU(A) = \mathbb{E}[\, u(A, Y) \mid do(A)]$$

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

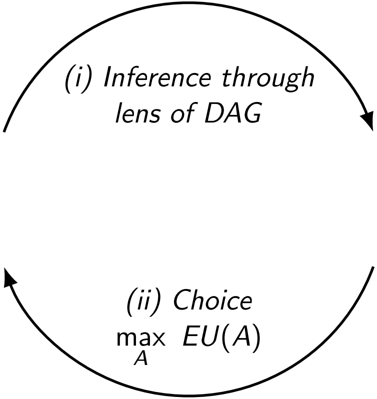
A	X	Y
1	0	0
0	1	0
$\vdots$	$\vdots$	$\vdots$



$$EU(A) = \mathbb{E}[\, u(A, Y) \mid do(A)]$$

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

A	X	Y
1	0	0
0	1	0
⋮	⋮	⋮

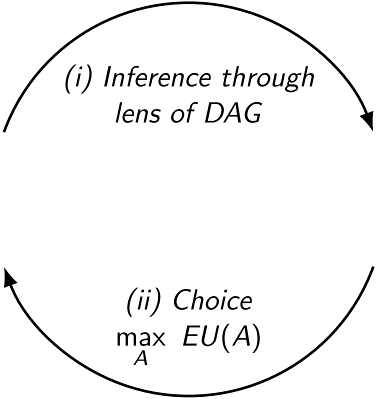


$$EU(A) = \mathbb{E}[\, u(A, Y) \mid do(A)]$$

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

(i) Data (viewed through DAG) justify choices in (ii)

A	X	Y
1	0	0
0	1	0
⋮	⋮	⋮



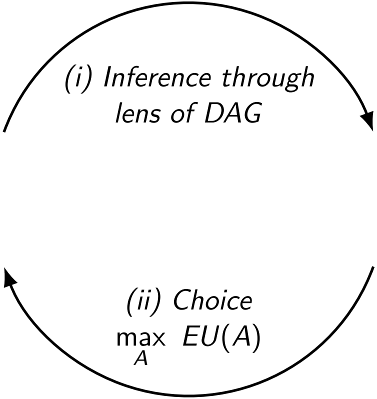
$$EU(A) = \mathbb{E}[\, u(A, Y) \mid do(A)]$$

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

(i) Data (viewed through DAG) justify choices in (ii)

(ii) Choices produce the data in (i)

A	X	Y
1	0	0
0	1	0
$\vdots$	$\vdots$	$\vdots$



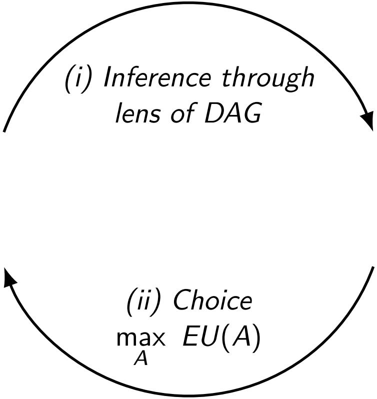
$$EU(A) = \mathbb{E}[\, u(A, Y) \mid do(A)]$$

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

(i) Data (viewed through DAG) justify choices in (ii)

(ii) Choices produce the data in (i)

A	X	Y
1	0	0
0	1	0
$\vdots$	$\vdots$	$\vdots$



$$EU(A) = \mathbb{E}[\, u(A, Y) \mid do(A)]$$

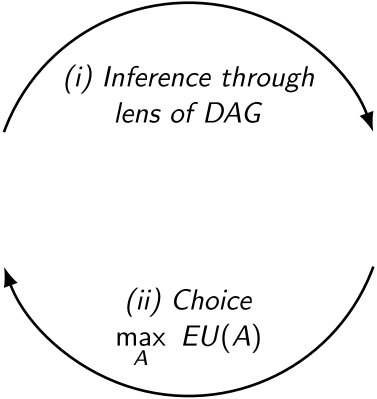
- In many (not all) Personal Equilibria, beliefs about actions' effects are biased.

Dieter's Dilemma illustrates *personal equilibrium* (Spiegler, 2016)

(i) Data (viewed through DAG) justify choices in (ii)

(ii) Choices produce the data in (i)

A	X	Y
1	0	0
0	1	0
⋮	⋮	⋮



$$EU(A) = \mathbb{E}[\, u(A, Y) \mid do(A)]$$

- In many (not all) Personal Equilibria, beliefs about actions' effects are biased.
- Personal Equilibrium often necessary for 'closing' a model

Empirics on misspecified DAGs

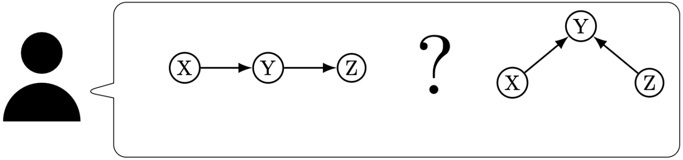
Empirics on misspecified DAGs

- Largely lacking

Empirics on misspecified DAGs

- Largely lacking
- Ambuehl, Huang (in progress): Sequential dieters' dilemma (100 trials)
  - People form the misspecified DAG
  - But do not choose in accordance with those believs, possibly because they misparametrize the DAG

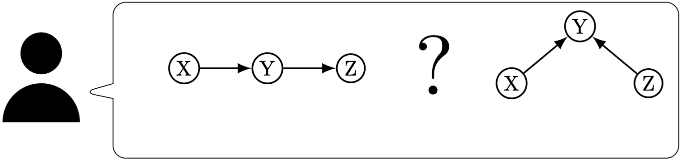
4. Structure learning



4. Structure learning

- So far: Causal reasoning

"I know what generally affects what and by how much.
What happened in this specific instance?"
(e.g. explaining away)



4. Structure learning

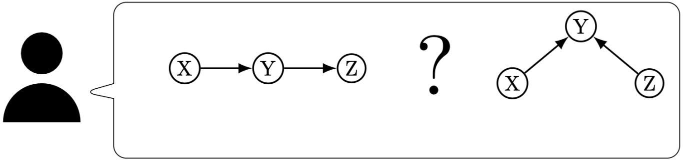
- So far: Causal reasoning

  "I know what generally affects what and by how much. What happened in this specific instance?"
  (e.g. explaining away)

- So far: Parameter learning

  "I know what affects what, but not by how much"

Viewing the world through the lens of a causal model



4. Structure learning

- So far: Causal reasoning

  "I know what generally affects what and by how much. What happened in this specific instance?"
  (e.g. explaining away)

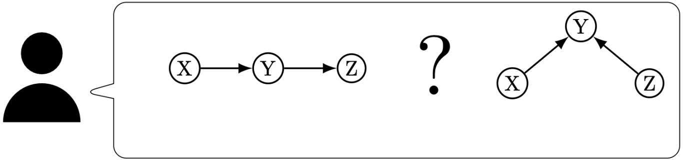
- So far: Parameter learning

  "I know what affects what, but not by how much"

  Viewing the world through the lens of a causal model

- Now: Structure learning

  "What can influence what?"



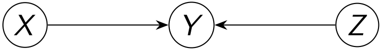
Two approaches to structure learning

Two approaches to structure learning

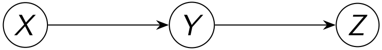
1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it

Constraint-based learning: Example

Collider

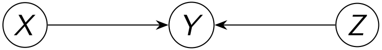


Chain

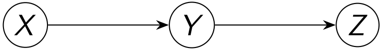


Constraint-based learning: Example

Collider

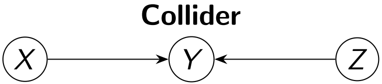


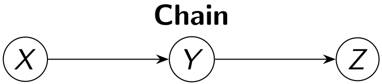
Chain



Data: Suppose we observe the following

Constraint-based learning: Example





Data: Suppose we observe the following

- Conditional on Y, Z is independent of X. But unconditionally, X and Z are correlated.

Constraint-based learning: Example

Data: Suppose we observe the following

- Conditional on Y, Z is independent of X. But unconditionally, X and Z are correlated.
- E.g. if variables are approximately normal, then regress Z on X and Y. Conclude conditional independence if the coefficient on Y is insignificantly different from zero.

Constraint-based learning: Example

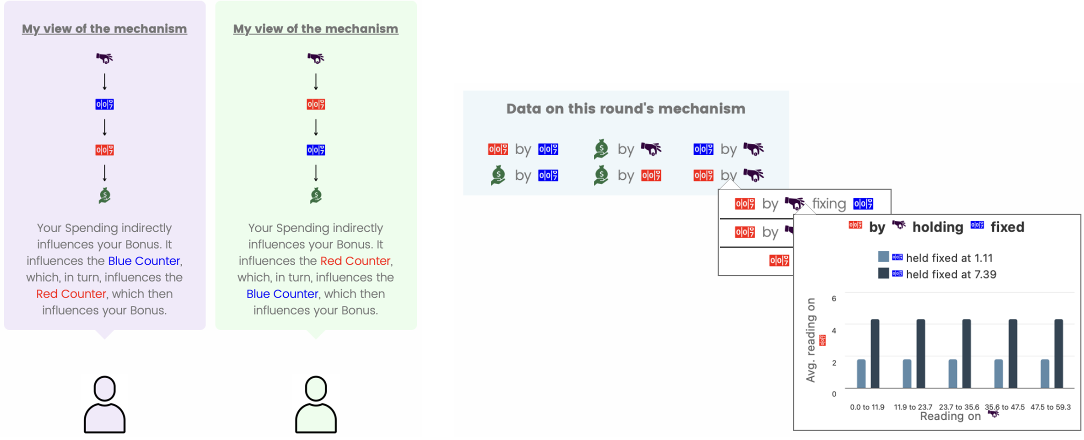
Data: Suppose we observe the following

- Conditional on Y, Z is independent of X. But unconditionally, X and Z are correlated.
- E.g. if variables are approximately normal, then regress Z on X and Y. Conclude conditional independence if the coefficient on Y is insignificantly different from zero.

These data are consistent with the chain and inconsistent with the collider. Infer that the structure is the chain.

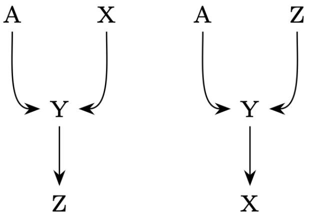
‘Constraint-based’ learning

Subjects are quite well able to derive correlational implications of causal models and, if inconsistent with data, rule out the corresponding model (Ambuehl and Thysen, 2025)

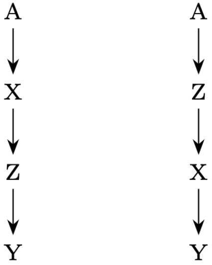


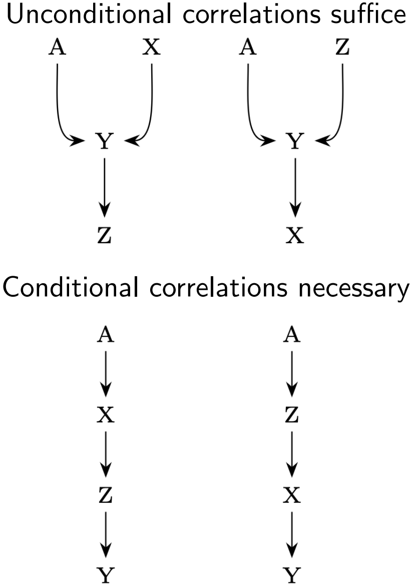
Conditional correlations necessary

Unconditional correlations suffice

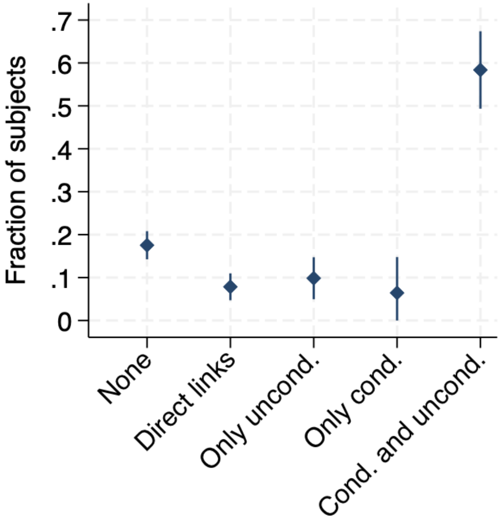


Conditional correlations necessary



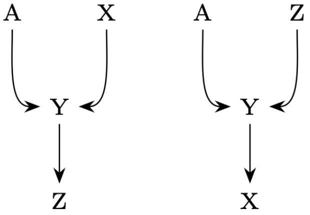


Unconditional correlations suffice
A
X
Y
Z
A
Z
Y
X
Conditional correlations necessary
A
X
Z
Y
A
Z
X
Y

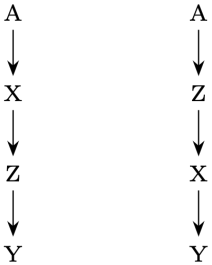


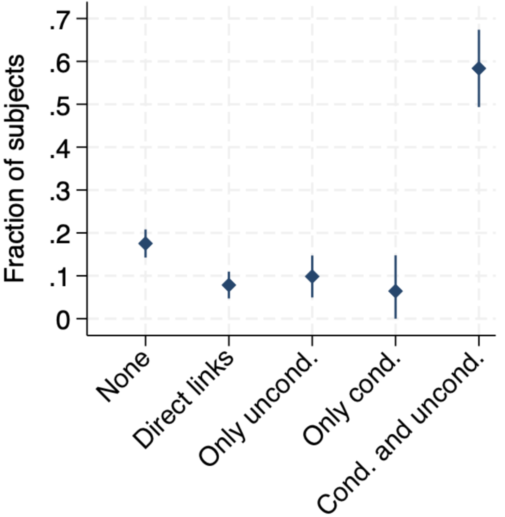
Fraction of subjects
0
.1
.2
.3
.4
.5
.6
.7
None
Direct links
Only uncond.
Only cond.
Cond. and uncond.

Unconditional correlations suffice



Conditional correlations necessary





No effect of threefold increase in stakes
(up to 90EUR)

Two approaches to structure learning

1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it

Two approaches to structure learning

1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it
   - Relies on Null-Hypothesis Significance Testing, and thus on arbitrary statistical significance cutoffs

Two approaches to structure learning

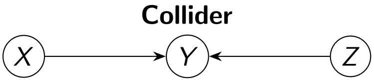
1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it
   - Relies on Null-Hypothesis Significance Testing, and thus on arbitrary statistical significance cutoffs
   - Computationally intensive with unconstrained space of DAGs (because there are so many)

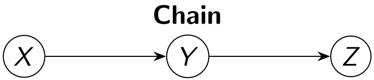
Two approaches to structure learning

1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it
   - Relies on Null-Hypothesis Significance Testing, and thus on arbitrary statistical significance cutoffs
   - Computationally intensive with unconstrained space of DAGs (because there are so many)
2. (Hierarchical) Bayesian structure learning

Bayesian structure learning: Example

Prior: $P(\text{DAG}= \text{collider}) = P(\text{DAG}= \text{chain}) = 0.5$

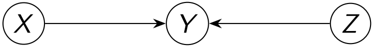




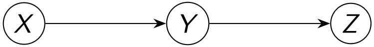
Bayesian structure learning: Example

Prior: $P(\text{DAG} = \text{collider}) = P(\text{DAG} = \text{chain}) = 0.5$

Collider



Chain



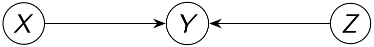
Data Suppose you observed: $(X, Y, Z) = (0, 1, 1)$
Likelihoods:

$$P\big((X, Y, Z) = (0, 1, 1)\big) =$$

Bayesian structure learning: Example

Prior: $P(\text{DAG}=\text{collider}) = P(\text{DAG}=\text{chain}) = 0.5$

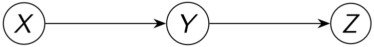
Collider



$$P(X{=}1) = P(Z{=}1) = 0.5$$

$$P(Y{=}1 \mid X, Z) = \begin{cases} 0.25 & X{=}0,\ Z{=}0 \\ 0.75 & \text{otherwise} \end{cases}$$

Chain



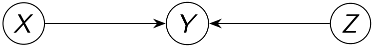
Data Suppose you observed: $(X, Y, Z) = (0, 1, 1)$
Likelihoods:

$$P\big((X, Y, Z) = (0, 1, 1)\big) =$$

Bayesian structure learning: Example

Prior: $P(\text{DAG}= \text{collider}) = P(\text{DAG}= \text{chain}) = 0.5$

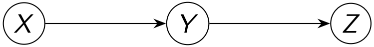
Collider



$$P(X{=}1) = P(Z{=}1) = 0.5$$

$$P(Y{=}1 \mid X, Z) = \begin{cases} 0.25 & X{=}0,\ Z{=}0 \\ 0.75 & \text{otherwise} \end{cases}$$

Chain



$$P(X{=}1) = 0.5$$

$$P(Y{=}1 \mid X) = \begin{cases} 0.75 & X{=}1 \\ 0.25 & X{=}0 \end{cases}$$

$$P(Z{=}1 \mid Y) = \begin{cases} 0.75 & Y{=}1 \\ 0.25 & Y{=}0 \end{cases}$$

Data Suppose you observed: $(X, Y, Z) = (0, 1, 1)$

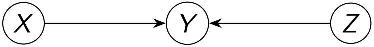
Likelihoods:

$$P\big((X, Y, Z) = (0, 1, 1)\big) =$$

Bayesian structure learning: Example

Prior: $P(\text{DAG} = \text{collider}) = P(\text{DAG} = \text{chain}) = 0.5$

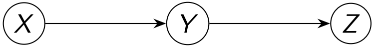
Collider



$$P(X{=}1) = P(Z{=}1) = 0.5$$

$$P(Y{=}1 \mid X, Z) = \begin{cases} 0.25 & X{=}0,\ Z{=}0 \\ 0.75 & \text{otherwise} \end{cases}$$

Chain



$$P(X{=}1) = 0.5$$

$$P(Y{=}1 \mid X) = \begin{cases} 0.75 & X{=}1 \\ 0.25 & X{=}0 \end{cases}$$

$$P(Z{=}1 \mid Y) = \begin{cases} 0.75 & Y{=}1 \\ 0.25 & Y{=}0 \end{cases}$$

Data Suppose you observed: $(X, Y, Z) = (0, 1, 1)$

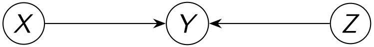
Likelihoods:

$$P\big((X, Y, Z) = (0, 1, 1)\big) = \begin{cases} 0.19 & \text{if DAG} = \text{chain} \\ 0.09 & \text{if DAG} = \text{collider} \end{cases}$$

Bayesian structure learning: Example

Prior: $P(\text{DAG}=\text{collider}) = P(\text{DAG}=\text{chain}) = 0.5$

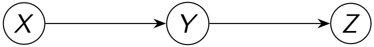
Collider



$$P(X{=}1) = P(Z{=}1) = 0.5$$

$$P(Y{=}1 \mid X, Z) = \begin{cases} 0.25 & X{=}0,\ Z{=}0 \\ 0.75 & \text{otherwise} \end{cases}$$

Chain



$$P(X{=}1) = 0.5$$

$$P(Y{=}1 \mid X) = \begin{cases} 0.75 & X{=}1 \\ 0.25 & X{=}0 \end{cases}$$

$$P(Z{=}1 \mid Y) = \begin{cases} 0.75 & Y{=}1 \\ 0.25 & Y{=}0 \end{cases}$$

Data Suppose you observed: $(X, Y, Z) = (0, 1, 1)$

Likelihoods:

$$P\big((X, Y, Z) = (0, 1, 1)\big) = \begin{cases} 0.19 & \text{if DAG} = \text{chain} \\ 0.09 & \text{if DAG} = \text{collider} \end{cases}$$

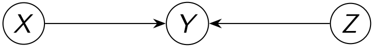
Posterior:

$$P(\text{DAG} = \text{chain} \mid \text{data})$$

Bayesian structure learning: Example

Prior: $P(\text{DAG}=\text{ collider}) = P(\text{DAG}=\text{ chain}) = 0.5$

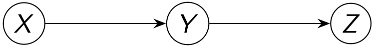
Collider



$$P(X=1) = P(Z=1) = 0.5$$

$$P(Y=1 \mid X, Z) = \begin{cases} 0.25 & X=0,\ Z=0 \\ 0.75 & \text{otherwise} \end{cases}$$

Chain



$$P(X=1) = 0.5$$

$$P(Y=1 \mid X) = \begin{cases} 0.75 & X=1 \\ 0.25 & X=0 \end{cases}$$

$$P(Z=1 \mid Y) = \begin{cases} 0.75 & Y=1 \\ 0.25 & Y=0 \end{cases}$$

Data Suppose you observed: $(X, Y, Z) = (0, 1, 1)$

Likelihoods:

$$P\big((X, Y, Z) = (0, 1, 1)\big) = \begin{cases} 0.19 & \text{if DAG} = \text{chain} \\ 0.09 & \text{if DAG} = \text{collider} \end{cases}$$

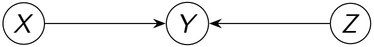
Posterior:

$$P(\text{DAG} = \text{chain} \mid \text{data}) = \frac{P\big((X, Y, Z) = (0, 1, 1) \mid \text{chain}\big) P(\text{DAG} = \text{chain})}{P(\text{data})}$$

Bayesian structure learning: Example

Prior: $P(\text{DAG}=\text{collider}) = P(\text{DAG}=\text{chain}) = 0.5$

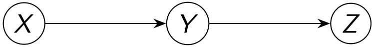
Collider



$$P(X{=}1) = P(Z{=}1) = 0.5$$

$$P(Y{=}1 \mid X, Z) = \begin{cases} 0.25 & X{=}0,\ Z{=}0 \\ 0.75 & \text{otherwise} \end{cases}$$

Chain



$$P(X{=}1) = 0.5$$

$$P(Y{=}1 \mid X) = \begin{cases} 0.75 & X{=}1 \\ 0.25 & X{=}0 \end{cases}$$

$$P(Z{=}1 \mid Y) = \begin{cases} 0.75 & Y{=}1 \\ 0.25 & Y{=}0 \end{cases}$$

Data Suppose you observed: $(X, Y, Z) = (0, 1, 1)$

Likelihoods:

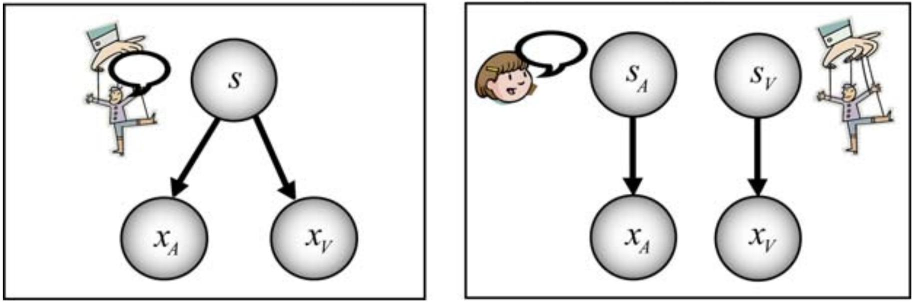
$$P\big((X, Y, Z) = (0, 1, 1)\big) = \begin{cases} 0.19 & \text{if DAG} = \text{chain} \\ 0.09 & \text{if DAG} = \text{collider} \end{cases}$$

Posterior:

$$\begin{aligned} P(\text{DAG} = \text{chain} \mid \text{data}) &= \frac{P\big((X, Y, Z) = (0, 1, 1) \mid \text{chain}\big) P(\text{DAG} = \text{chain})}{P(\text{data})} \\ &= \frac{0.19 \times 0.5}{0.19 \times 0.5 + 0.09 \times 0.5} = 0.69 \end{aligned}$$

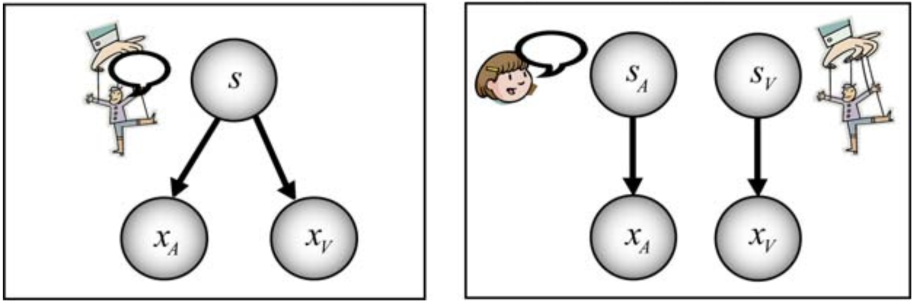
Bayesian structure learning in our perceptual system

Where is the speaker?



Bayesian structure learning in our perceptual system

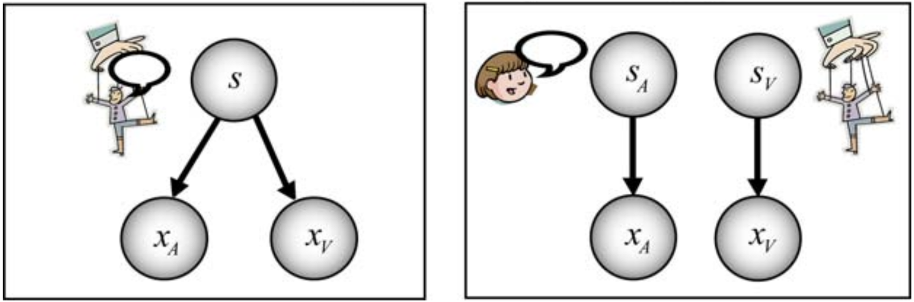
Where is the speaker?



- If Left correct: puppet's location (seeing) provides information about location of the speaker (hearing)

Bayesian structure learning in our perceptual system

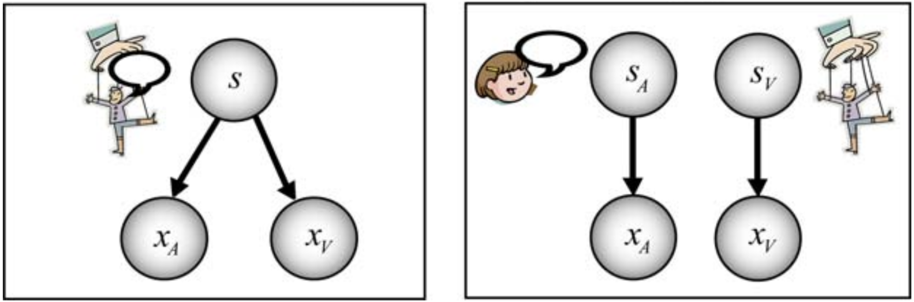
Where is the speaker?



- If Left correct: puppet's location (seeing) provides information about location of the speaker (hearing)
- If Right correct: puppet's location (seeing) provides no information about location of the speaker (hearing)

Bayesian structure learning in our perceptual system

Where is the speaker?



- If Left correct: puppet's location (seeing) provides information about location of the speaker (hearing)
- If Right correct: puppet's location (seeing) provides no information about location of the speaker (hearing)
- Empirically: Our perception selects model based on distance between the cues' perceived locations.

Two approaches to structure learning

1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it
   - Relies on Null-Hypothesis Significance Testing, and thus on arbitrary statistical significance cutoffs
   - Computationally intensive with unconstrained space of DAGs (because there are so many)
2. (Hierarchical) Bayesian structure learning

Two approaches to structure learning

1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it
   - Relies on Null-Hypothesis Significance Testing, and thus on arbitrary statistical significance cutoffs
   - Computationally intensive with unconstrained space of DAGs (because there are so many)
2. (Hierarchical) Bayesian structure learning
   - Maintains uncertainty over DAGs

Two approaches to structure learning

1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it
   - Relies on Null-Hypothesis Significance Testing, and thus on arbitrary statistical significance cutoffs
   - Computationally intensive with unconstrained space of DAGs (because there are so many)
2. (Hierarchical) Bayesian structure learning
   - Maintains uncertainty over DAGs
   - "Dumb" procedures can approximate it (e.g. Gibbs-sampling on local links, see Bramley et al., 2017)

Two approaches to structure learning

1. "Constraint-based algorithms": Identify all conditional independence relationships in the data, find which DAG is consistent with it
   - Relies on Null-Hypothesis Significance Testing, and thus on arbitrary statistical significance cutoffs
   - Computationally intensive with unconstrained space of DAGs (because there are so many)
2. (Hierarchical) Bayesian structure learning
   - Maintains uncertainty over DAGs
   - "Dumb" procedures can approximate it (e.g. Gibbs-sampling on local links, see Bramley et al., 2017)
   - Perceptual system appears to do some of it

6. Measuring and identifying beliefs about structure

Theory

What beliefs about structure can be identified from what data, in principle?

Theory

What beliefs about structure can be identified from what data, in principle?

- Halpern and Piermont (2024); Schenone (2018): N variables, preferences over causal interventions (can choose without restrictions). When are these preferences consistent with beliefs that follow a single DAG?

Theory

What beliefs about structure can be identified from what data, in principle?

- Halpern and Piermont (2024); Schenone (2018): N variables, preferences over causal interventions (can choose without restrictions). When are these preferences consistent with beliefs that follow a single DAG?
- Ellis and Thysen (2025): N variables, one is the action, one is the outcome. Can only observe, not intervene.
  - What can be identified? DAGs that share the same set of most direct causal paths

Theory

What beliefs about structure can be identified from what data, in principle?

- Halpern and Piermont (2024); Schenone (2018): N variables, preferences over causal interventions (can choose without restrictions). When are these preferences consistent with beliefs that follow a single DAG?
- Ellis and Thysen (2025): N variables, one is the action, one is the outcome. Can only observe, not intervene.
  - What can be identified? DAGs that share the same set of most direct causal paths
- These are valuable conceptual first steps, but not ready for practical use to test whether people hold DAGs as causal models, or to elicit them.

Theory

What beliefs about structure can be identified from what data, in principle?

- Halpern and Piermont (2024); Schenone (2018): N variables, preferences over causal interventions (can choose without restrictions). When are these preferences consistent with beliefs that follow a single DAG?
- Ellis and Thysen (2025): N variables, one is the action, one is the outcome. Can only observe, not intervene.
  - What can be identified? DAGs that share the same set of most direct causal paths
- These are valuable conceptual first steps, but not ready for practical use to test whether people hold DAGs as causal models, or to elicit them.
- Most of the body of empirical support that people think in terms of DAGs: Test isolated directional predictions in hypothetical, qualitative environments

Empirical measurement approaches

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to 21%

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to 21%
- Hand-draw the DAGs (e.g. Liefgreen and Lagnado, 2023)

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to 21%
- Hand-draw the DAGs (e.g. Liefgreen and Lagnado, 2023)
  - What do people mean when they draw an arrow? Same interpretation as the theory?

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to 21%
- Hand-draw the DAGs (e.g. Liefgreen and Lagnado, 2023)
  - What do people mean when they draw an arrow? Same interpretation as the theory?
  - Do they realize that the key information is not in the arrows they draw but in those they leave out?

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to 21%
- Hand-draw the DAGs (e.g. Liefgreen and Lagnado, 2023)
  - What do people mean when they draw an arrow? Same interpretation as the theory?
  - Do they realize that the key information is not in the arrows they draw but in those they leave out?
- Verbal judgments about counterfactuals from interventions concerning pairs of variables (Tatlidil et al., 2025)

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to 21%
- Hand-draw the DAGs (e.g. Liefgreen and Lagnado, 2023)
  - What do people mean when they draw an arrow? Same interpretation as the theory?
  - Do they realize that the key information is not in the arrows they draw but in those they leave out?
- Verbal judgments about counterfactuals from interventions concerning pairs of variables (Tatlidil et al., 2025)
  - Large number, grows fast with number of nodes

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to 21%
- Hand-draw the DAGs (e.g. Liefgreen and Lagnado, 2023)
  - What do people mean when they draw an arrow? Same interpretation as the theory?
  - Do they realize that the key information is not in the arrows they draw but in those they leave out?
- Verbal judgments about counterfactuals from interventions concerning pairs of variables (Tatlidil et al., 2025)
  - Large number, grows fast with number of nodes
- Elicit conditional probability distributions (interventional and predictive)

Empirical measurement approaches

- Ask people to describe in words, then have RAs map the words into DAGs (Andre et al., 2024)
  - Verbal descriptions might be incomplete and ambiguous
    - If one coder refers to a factor, the other does so with 88% chance
    - If the coders define the same factors, they have the same DAGs with 84% chance
    - Overall, they have the same DAGs in 51% of cases
    - If there's an indirect effect in their final DAG, this drops to 21%
- Hand-draw the DAGs (e.g. Liefgreen and Lagnado, 2023)
  - What do people mean when they draw an arrow? Same interpretation as the theory?
  - Do they realize that the key information is not in the arrows they draw but in those they leave out?
- Verbal judgments about counterfactuals from interventions concerning pairs of variables (Tatlidil et al., 2025)
  - Large number, grows fast with number of nodes
- Elicit conditional probability distributions (interventional and predictive)
  - Large number, grows fast with number of nodes

Which elicitation method is best?

Which elicitation method is best?

- We don't know

Which elicitation method is best?

- We don't know
- The only (extremely limited) formal comparison of methods is Tatlidil et al. (2025)

Which elicitation method is best?

- We don't know
- The only (extremely limited) formal comparison of methods is Tatlidil et al. (2025)
- Some research suggesting that (i) the elicitation format can affect the type of network that is elicited (Laukkanen and Wang, 2016), (ii) elicitation per se can affect what subjects think (Liefgreen and Lagnado, 2023)

Which elicitation method is best?

- We don't know
- The only (extremely limited) formal comparison of methods is Tatlidil et al. (2025)
- Some research suggesting that (i) the elicitation format can affect the type of network that is elicited (Laukkanen and Wang, 2016), (ii) elicitation per se can affect what subjects think (Liefgreen and Lagnado, 2023)

There are more fundamental things we don't know

Which elicitation method is best?

- We don't know
- The only (extremely limited) formal comparison of methods is Tatlidil et al. (2025)
- Some research suggesting that (i) the elicitation format can affect the type of network that is elicited (Laukkanen and Wang, 2016), (ii) elicitation per se can affect what subjects think (Liefgreen and Lagnado, 2023)

There are more fundamental things we don't know

- Are mental models acyclical, or do they involve feedback loops, etc.

Which elicitation method is best?

- We don't know
- The only (extremely limited) formal comparison of methods is Tatlidil et al. (2025)
- Some research suggesting that (i) the elicitation format can affect the type of network that is elicited (Laukkanen and Wang, 2016), (ii) elicitation per se can affect what subjects think (Liefgreen and Lagnado, 2023)

There are more fundamental things we don't know

- Are mental models acyclical, or do they involve feedback loops, etc.
- Do individuals entertain a single mental model (as in the misspecified models literature) or do they hold multiple models at once (as a Bayesian would)

Which elicitation method is best?

- We don't know
- The only (extremely limited) formal comparison of methods is Tatlidil et al. (2025)
- Some research suggesting that (i) the elicitation format can affect the type of network that is elicited (Laukkanen and Wang, 2016), (ii) elicitation per se can affect what subjects think (Liefgreen and Lagnado, 2023)

There are more fundamental things we don't know

- Are mental models acyclical, or do they involve feedback loops, etc.
- Do individuals entertain a single mental model (as in the misspecified models literature) or do they hold multiple models at once (as a Bayesian would)
- Do people think in terms of causal networks, or do they piece together associations between pairs of variables on demand?

Which elicitation method is best?

- We don't know
- The only (extremely limited) formal comparison of methods is Tatlidil et al. (2025)
- Some research suggesting that (i) the elicitation format can affect the type of network that is elicited (Laukkanen and Wang, 2016), (ii) elicitation per se can affect what subjects think (Liefgreen and Lagnado, 2023)

There are more fundamental things we don't know

- Are mental models acyclical, or do they involve feedback loops, etc.
- Do individuals entertain a single mental model (as in the misspecified models literature) or do they hold multiple models at once (as a Bayesian would)
- Do people think in terms of causal networks, or do they piece together associations between pairs of variables on demand?
- Overall: are people's mental models at all consistent with a single, fixed DAG? That is, is there even a subjective DAG to meaningfully elicit?

Summary

Causal DAGs are an extremely powerful tool

Summary

Causal DAGs are an extremely powerful tool

- Capture fundamental patterns and intuitions regarding causal inference

Summary

Causal DAGs are an extremely powerful tool

- Capture fundamental patterns and intuitions regarding causal inference
- Applications throughout economics

Summary

Causal DAGs are an extremely powerful tool

- Capture fundamental patterns and intuitions regarding causal inference
- Applications throughout economics
- Vast cogSci literature provides evidence that it describes human cognition

Summary

Causal DAGs are an extremely powerful tool

- Capture fundamental patterns and intuitions regarding causal inference
- Applications throughout economics
- Vast cogSci literature provides evidence that it describes human cognition

What we did

Summary

Causal DAGs are an extremely powerful tool

- Capture fundamental patterns and intuitions regarding causal inference
- Applications throughout economics
- Vast cogSci literature provides evidence that it describes human cognition

What we did

1. Causal reasoning (structure known, parameters known)
    - Path blocking / Causal Markov condition
    - Collider bias / explaining away
    - Markov equivalence: same skeleton and same v colliders

Summary

Causal DAGs are an extremely powerful tool

- Capture fundamental patterns and intuitions regarding causal inference
- Applications throughout economics
- Vast cogSci literature provides evidence that it describes human cognition

What we did

1. Causal reasoning (structure known, parameters known)
   - Path blocking / Causal Markov condition
   - Collider bias / explaining away
   - Markov equivalence: same skeleton and same v colliders
2. Parameter learning (structure known, parameters unknown)
   - Fitting misspecified models
   - Personal equilibrium (Dieter's Dilemma)

Summary

Causal DAGs are an extremely powerful tool

- Capture fundamental patterns and intuitions regarding causal inference
- Applications throughout economics
- Vast cogSci literature provides evidence that it describes human cognition

What we did

1. Causal reasoning (structure known, parameters known)
   - Path blocking / Causal Markov condition
   - Collider bias / explaining away
   - Markov equivalence: same skeleton and same v colliders
2. Parameter learning (structure known, parameters unknown)
   - Fitting misspecified models
   - Personal equilibrium (Dieter's Dilemma)
3. Structure learning (structure unknown, parameters unknown)
   - Constraint-based learning
   - Bayesian structure learning

… and there's so much more…

- What do subjects conceive as a node? "The economy" vs. "unemployment, GDP, and stock market valuations" (ontology)
- The predictive/diagnostic reasoning asymmetry
- DAGs explain how people categorize and stereotype
- Illusory causation (Matute et al., 2015)
- The causal frame problem (Icard III and Goodman, 2015)
- Alternative causal approaches, e.g. reduction in Kolmogoroff complexity afforded by a causal explanation (Alexander and Gilboa, 2023)

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes

B. that a blue eyed girls' mother also has blue eyes

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes

B. that a blue eyed girls' mother also has blue eyes

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes

B. that a blue eyed girls' mother also has blue eyes

Eye color distribution stationary $\rightarrow$ both are equal. Yet, subjects believe (i) more likely.

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes (predictive)

B. that a blue eyed girls' mother also has blue eyes

Eye color distribution stationary → both are equal. Yet, subjects believe (i) more likely.

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes (predictive)

B. that a blue eyed girls' mother also has blue eyes (diagnostic)

Eye color distribution stationary $\rightarrow$ both are equal. Yet, subjects believe (i) more likely.

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes (predictive)

B. that a blue eyed girls' mother also has blue eyes (diagnostic)

Eye color distribution stationary $\rightarrow$ both are equal. Yet, subjects believe (i) more likely.

Further research on the asymmetry (e.g. Fernbach et al., 2011, 2010):

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes (predictive)

B. that a blue eyed girls' mother also has blue eyes (diagnostic)

Eye color distribution stationary → both are equal. Yet, subjects believe (i) more likely.

Further research on the asymmetry (e.g. Fernbach et al., 2011, 2010):

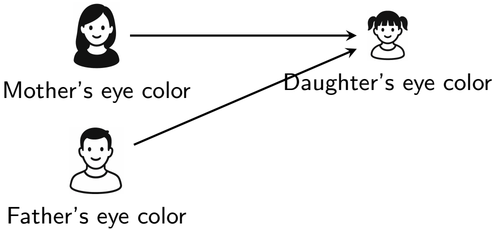
- When reasoning from effect to causes, individuals think of alternative causes

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes (predictive)

B. that a blue eyed girls' mother also has blue eyes (diagnostic)



Eye color distribution stationary $\rightarrow$ both are equal. Yet, subjects believe (i) more likely.

Further research on the asymmetry (e.g. Fernbach et al., 2011, 2010):

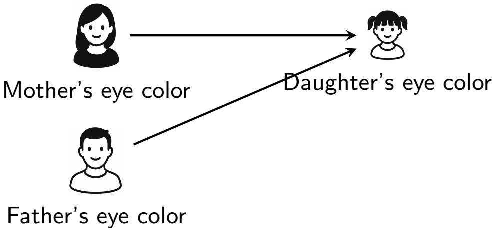
- When reasoning from effect to causes, individuals think of alternative causes

The predictive/diagnostic reasoning asymmetry

Which is more likely (*cf.* Tversky and Kahneman, 1980):

A. that a blue-eyed mother's daughter also has blue eyes (predictive)

B. that a blue eyed girls' mother also has blue eyes (diagnostic)



Eye color distribution stationary $\rightarrow$ both are equal. Yet, subjects believe (i) more likely.

Further research on the asymmetry (e.g. Fernbach et al., 2011, 2010):

- When reasoning from effect to causes, individuals think of alternative causes
- When reasoning from cause to effect, individuals forget alternative causes. Hence overestimate the predictive power of a given cause.

The asymmetry seems fundamental to cognition; it also occurs with LLMs

The asymmetry seems fundamental to cognition; it also occurs with LLMs

ChatGPT 5 Share ···

who is the son of Nina Bruce (née Warren)

Nina Bruce (née Warren) is the mother of **Condoleezza Rice**, the former U.S. Secretary of State. She did not have a son — only one child, Condoleezza.

The asymmetry seems fundamental to cognition; it also occurs with LLMs

ChatGPT 5

Share

who is the son of Nina Bruce (née Warren)

Nina Bruce (née Warren) is the mother of **Condoleezza Rice**, the former U.S. Secretary of State. She did not have a son — only one child, Condoleezza.

who is the mother of george clooney

George Clooney's mother is **Nina Bruce Warren Clooney**, a former city councilwoman and beauty queen.

References I

Yotam Alexander and Itzhak Gilboa. Subjective causality. *Revue économique*, 74(4):619–633, 2023.

Sandro Ambuehl and Heidi C. Thysen. Fact-based versus utility-based choice with unknown causal structures: An experimental study. *Unpublished*, 2025.

Peter Andre, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. Subjective models of the macroeconomy: Evidence from experts and representative samples. *The Review of Economic Studies*, 89(6):2958–2991, 2022.

Peter Andre, Ingar Haaland, Christopher Roth, Mirko Wiederholt, and Johannes Wohlfart. Narratives about the macroeconomy. Technical report, SAFE Working Paper, 2024.

Marco Angrisani, Anya Samek, and Ricardo Serrano-Padial. Competing narratives in action: An empirical analysis of model adoption dynamics. *Unpublished*, 2023.

Daniel J Benjamin. Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186, 2019.

J Aislinn Bohren and Daniel N Hauser. Misspecified models in learning and games. *Annual Review of Economics*, 17, 2024.

Neil R. Bramley, Peter Dayan, Thomas L. Griffiths, and David A. Lagnado. Formalizing neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, 124(3):301–338, 2017. doi: 10.1037/rev0000061.

References II

Arnaldo Camuffo, Alfonso Gambardella, Danilo Messinese, Elena Novelli, Emilio Paolucci, and Chiara Spina. A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. *Strategic Management Journal*, 45(6):1209–1237, 2024.

Kevin Dorst. Bayesians commit the gambler's fallacy. *Available at SSRN 4683064*, 2024.

Kfir Eliaz and Ran Spiegler. A model of competing narratives. *American Economic Review*, 110(12): 3786–3816, 2020.

Kfir Eliaz, Ran Spiegler, and Yair Weiss. Cheating with models. *American Economic Review: Insights*, 2020.

Andrew Ellis and Heidi Christina Thysen. Subjective causality in choice. *arXiv preprint arXiv:2106.05957*, 2025.

Florian Engl. A theory of causal responsibility attribution. *Available at SSRN 2932769*, 2022.

Ignacio Esponda and Demian Pouzo. Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, 84(3):1093–1130, 2016.

Erik Eyster and Matthew Rabin. Cursed equilibrium. *Econometrica*, 73(5):1623–1672, 2005.

Philip M Fernbach, Adam Darlow, and Steven A Sloman. Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science*, 21(3):329–336, 2010.

References III

Philip M Fernbach, Adam Darlow, and Steven A Sloman. Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, 140(2):168, 2011.

Samuel J Gershman and Mina Cikara. Structure learning principles of stereotype change. *Psychonomic Bulletin & Review*, 30(4):1273–1293, 2023.

Thomas L. Griffiths, Nick Chater, and Joshua Tenenbaum. *Bayesian Models of Cognition: Reverse Engineering the Mind*. MIT Press, 2024.

Joseph Y Halpern and Evan Piermont. A representation theorem for causal decision making. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, volume 21, pages 410–419, 2024.

Adam JL Harris and Ulrike Hahn. Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35 (5):1366, 2009.

Adam JL Harris, Ulrike Hahn, Jens K Madsen, and Anne S Hsu. The appeal to expert opinion: Quantitative support for a bayesian network approach. *Cognitive Science*, 40(6):1496–1533, 2016.

Thomas F Icard III and Noah D Goodman. A resource-rational approach to the causal frame problem. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 37, 2015.

Philippe Jehiel. Analogy-based expectation equilibrium. *Journal of Economic theory*, 123(2):81–104, 2005.

References IV

Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

Mauri Laukkanen and Mingde Wang. *Comparative causal mapping: The CMAP3 method*. Routledge, 2016.

Gilat Levy, Ronny Razin, and Alwyn Young. Misspecified politics and the recurrence of populism. *American Economic Review*, 112(3):928–962, 2022.

Alice Liefgreen and David A Lagnado. Drawing conclusions: Representing and evaluating competing explanations. *Cognition*, 234:105382, 2023.

Helena Matute, Fernando Blanco, Ion Yarritu, Marcos Díaz-Lago, Miguel A Vadillo, and Itxaso Barberia. Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in psychology*, 6:146427, 2015.

Pooya Molavi, Alireza Tahbaz-Salehi, and Andrea Vedolin. Model complexity, expectations, and asset prices. *Review of Economic Studies*, 91(4):2462–2507, 2024.

Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic books, 2018.

Bob Rehder. Independence and dependence in human causal reasoning. *Cognitive psychology*, 72: 54–107, 2014.

References V

Benjamin Margolin Rottman and Reid Hastie. Reasoning about causal relationships: Inferences on causal networks. *Psychological bulletin*, 140(1):109, 2014.

Pablo Schenone. Causality: a decision theoretic approach. *arXiv preprint arXiv:1812.07414*, 2018.

Heiner Schumacher and Heidi Christina Thysen. Equilibrium contracts and boundedly rational expectations. *Theoretical Economics*, 17(1):371–414, 2022.

Joshua Schwartzstein and Adi Sunderam. Using models to persuade. *American Economic Review*, 111 (1):276–323, 2021.

Ran Spiegler. Bayesian networks and boundedly rational expectations. *The Quarterly Journal of Economics*, 131(3):1243–1290, 2016.

Ran Spiegler. "data monkeys": a procedural model of extrapolation from partial statistics. *The Review of Economic Studies*, 84(4):1818–1841, 2017.

Ran Spiegler. Behavioral implications of causal misperceptions. *Annual Review of Economics*, 12: 81–106, 2020a.

Ran Spiegler. Can agents with causal misperceptions be systematically fooled? *Journal of the European Economic Association*, 18(2):583–617, 2020b.

Ran Spiegler. On the behavioral consequences of reverse causality. *European Economic Review*, 149: 104258, 2022.

References VI

Semir Tatlidil, Steven A Sloman, Semanti Basu, Tiffany Tran, Serena Saxena, Moon Hwan Kim, and Iris Bahar. A comparison of methods to elicit causal structure. *Frontiers in Cognition*, 4:1544387, 2025.

Amos Tversky and Daniel Kahneman. Causal schemas in judgments under uncertainty. In *Progress in social psychology*, pages 49–72. Erlbaum, Hillsdale, 1980.

Michael Waldmann. *The Oxford handbook of causal reasoning*. Oxford University Press, 2017.