Observational Causal Inference

Lost in Stats: Pitfalls of the Scientific Method

University of Cologne

Mariami Gachechiladze, 24.11.2021

Outline

Outline

Part 1:

Simpson's paradox and its resolution

Outline

Part 1:

Simpson's paradox and its resolution

Part 2:

Observational Causal Inference

Part 1:
Simpson's paradox and its resolution

Recap: Simpson's paradox

	Combined	E	$\neg E$		Recovery Rate
(a)	Drug (C)	20	20	40	50%
	No drug ($\neg C$)	16	24	40	40%
		36	44	80	

	Males	E	$\neg E$		Recovery Rate
(b)	Drug (C)	18	12	30	60%
	No drug ($\neg C$)	7	3	10	70%
		25	15	40	

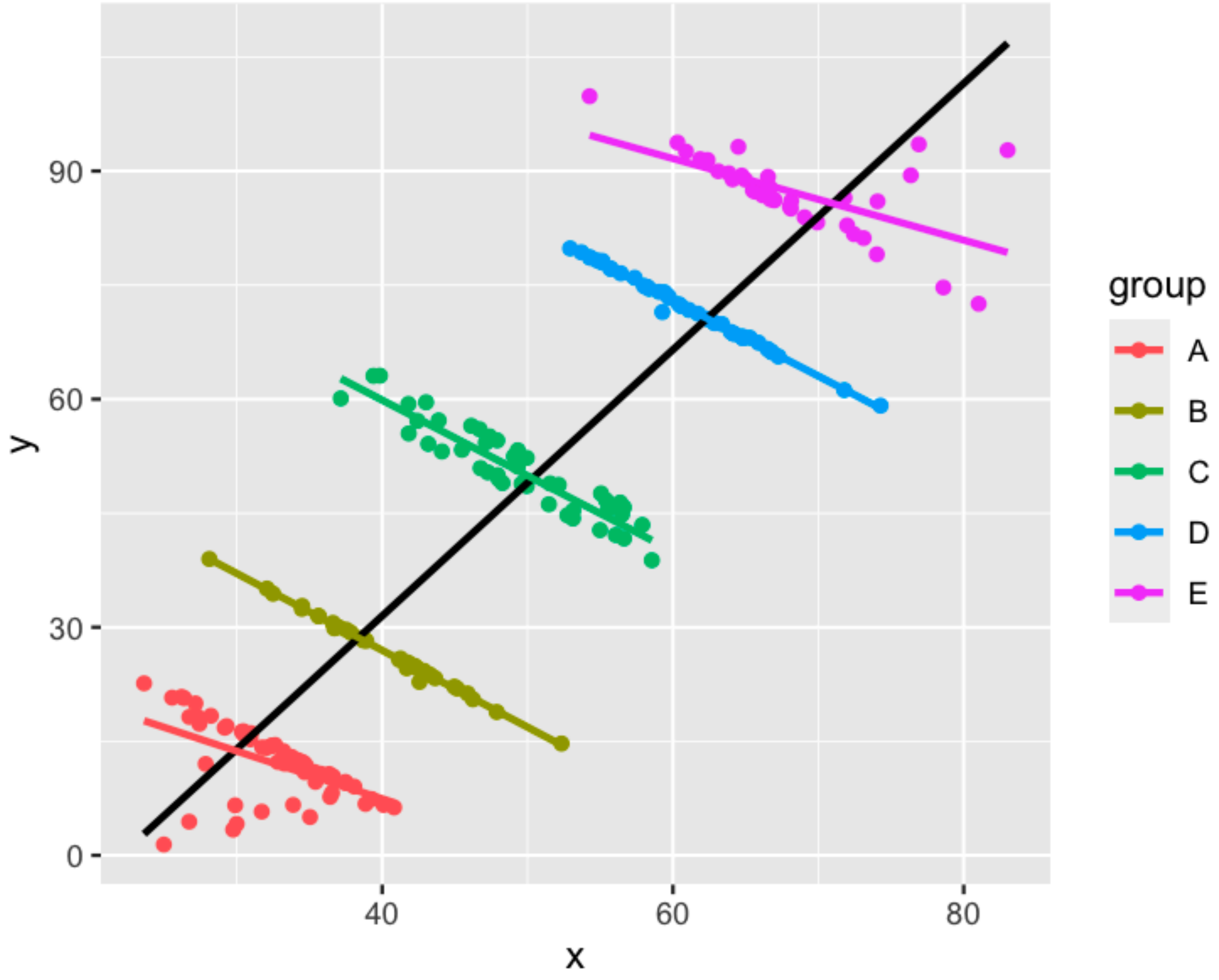
	Females	E	$\neg E$		Recovery Rate
(c)	Drug (C)	2	8	10	20%
	No drug ($\neg C$)	9	21	30	30%
		11	29	40	

Recap: Simpson's paradox

	Combined	E	$\neg E$		Recovery Rate
(a)	Drug (C)	20	20	40	50%
	No drug ($\neg C$)	16	24	40	40%
		36	44	80	

	Males	E	$\neg E$		Recovery Rate
(b)	Drug (C)	18	12	30	60%
	No drug ($\neg C$)	7	3	10	70%
		25	15	40	

	Females	E	$\neg E$		Recovery Rate
(c)	Drug (C)	2	8	10	20%
	No drug ($\neg C$)	9	21	30	30%
		11	29	40	

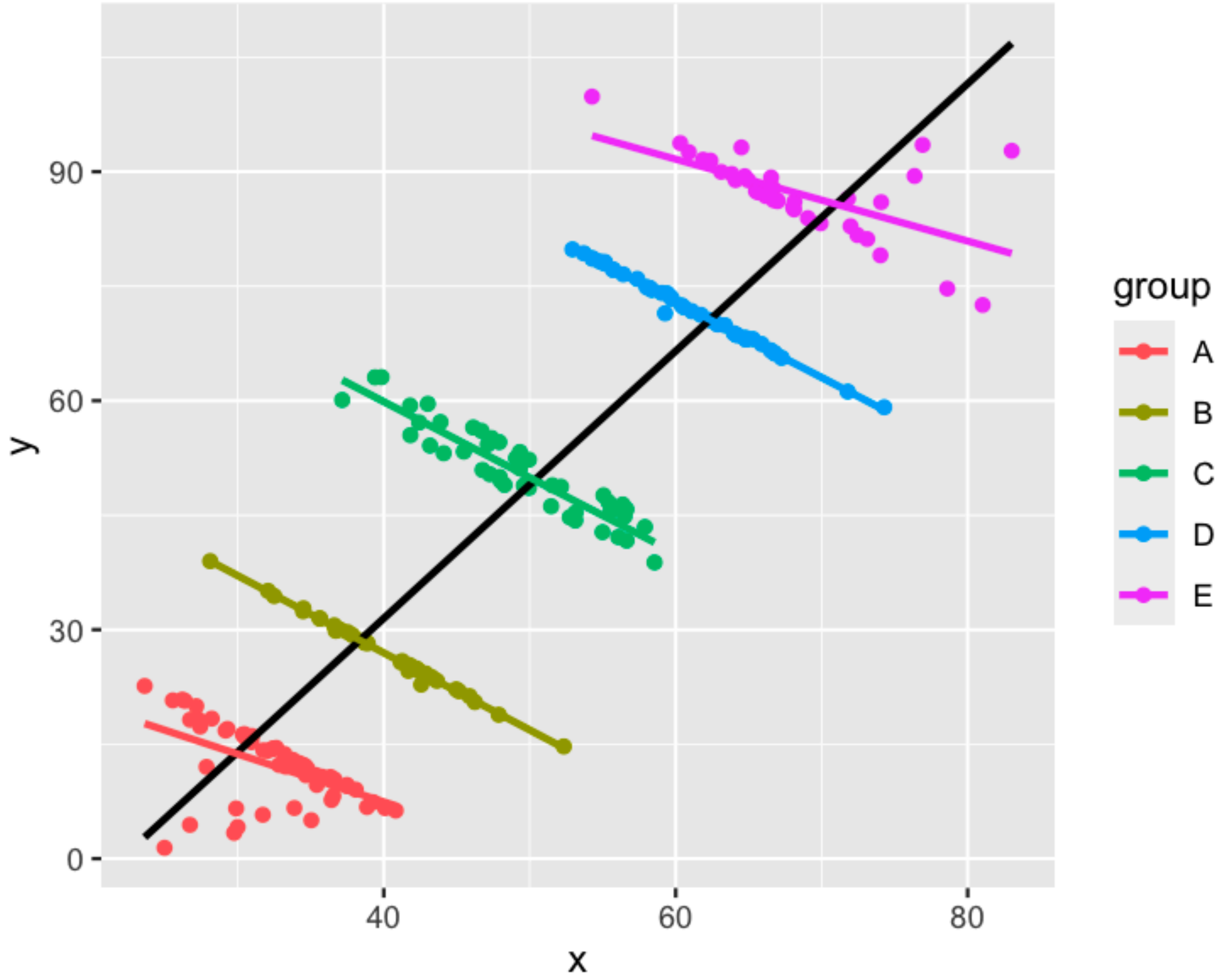


Recap: Simpson's paradox

	Combined	E	$\neg E$		Recovery Rate
(a)	Drug (C)	20	20	40	50%
	No drug ($\neg C$)	16	24	40	40%
		36	44	80	

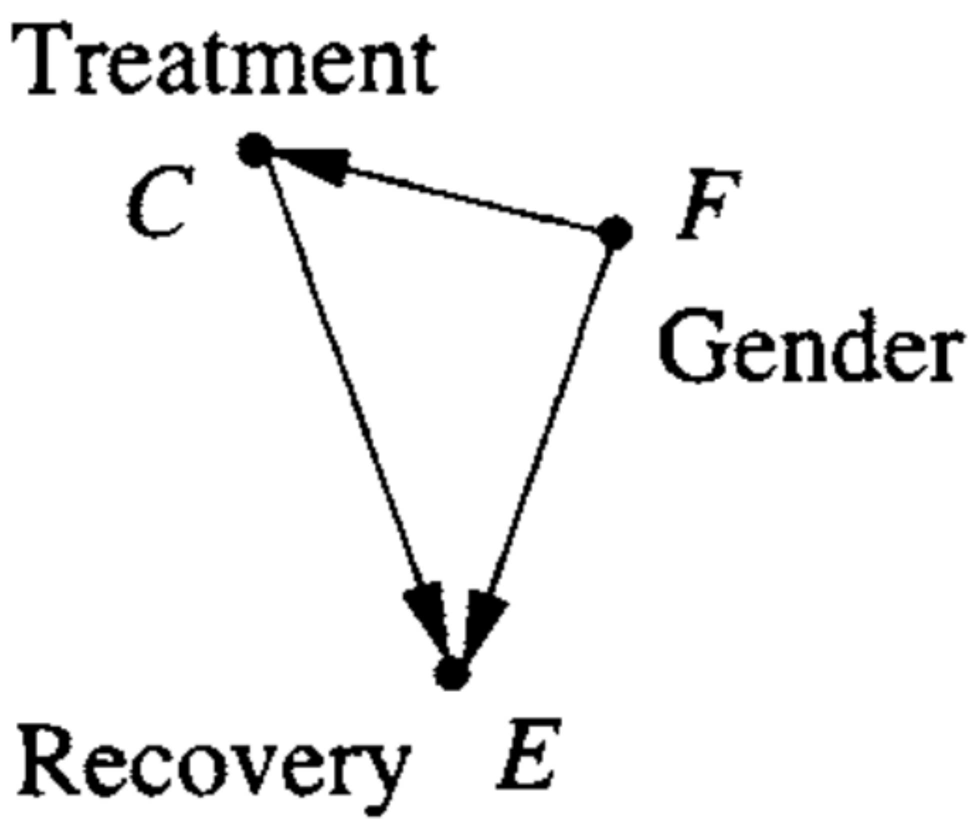
	Males	E	$\neg E$		Recovery Rate
(b)	Drug (C)	18	12	30	60%
	No drug ($\neg C$)	7	3	10	70%
		25	15	40	

	Females	E	$\neg E$		Recovery Rate
(c)	Drug (C)	2	8	10	20%
	No drug ($\neg C$)	9	21	30	30%
		11	29	40	



Question: Shall we give a drug to a next patient or not?

Statistics alone cannot suffice

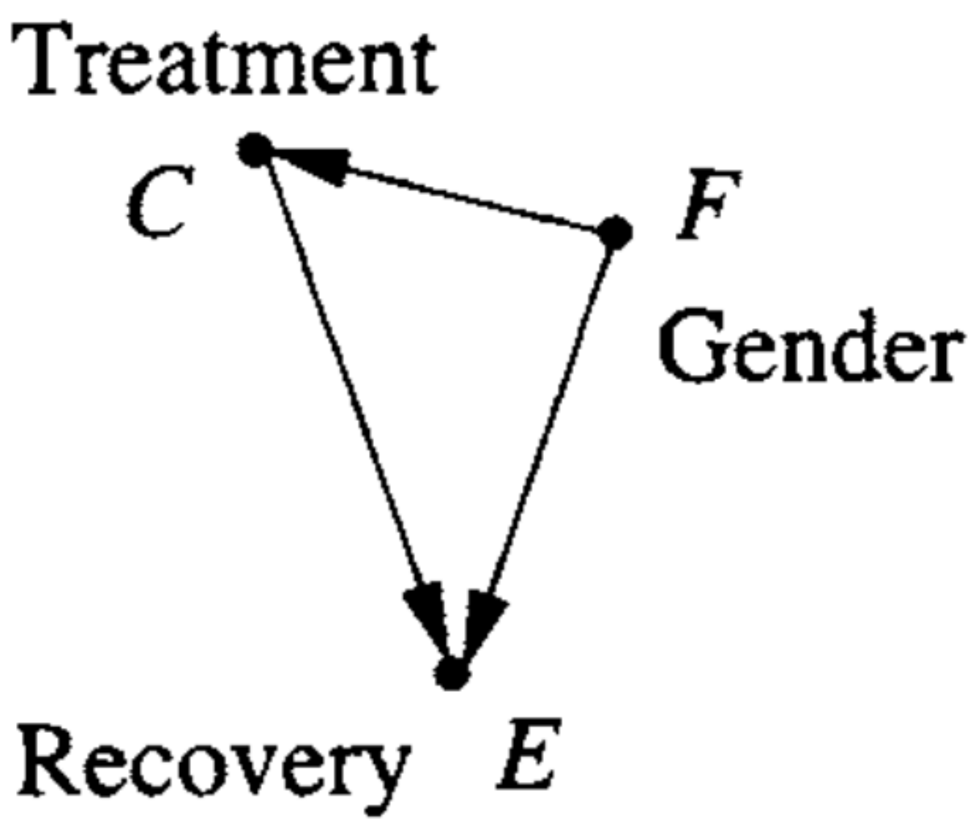


(a)

Statistics alone cannot suffice

Sure-Thing Principle:

Decision makers who would take a certain action if they knew that event F has occurred, and also if they knew that the negation of F has occurred, should also take that same action if they knows nothing about F.



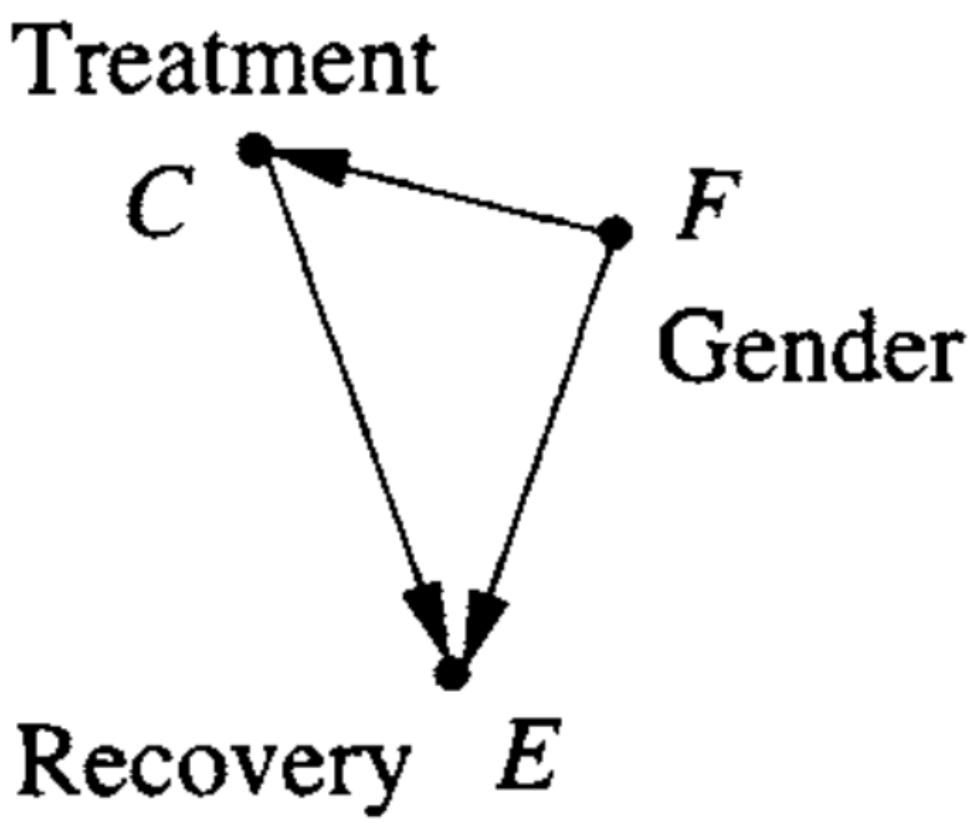
(a)

Statistics alone cannot suffice

Sure-Thing Principle:

Decision makers who would take a certain action if they knew that event F has occurred, and also if they knew that the negation of F has occurred, should also take that same action if they knows nothing about F.

$$P(E \mid C, F) < P(E \mid \neg C, F)$$



(a)

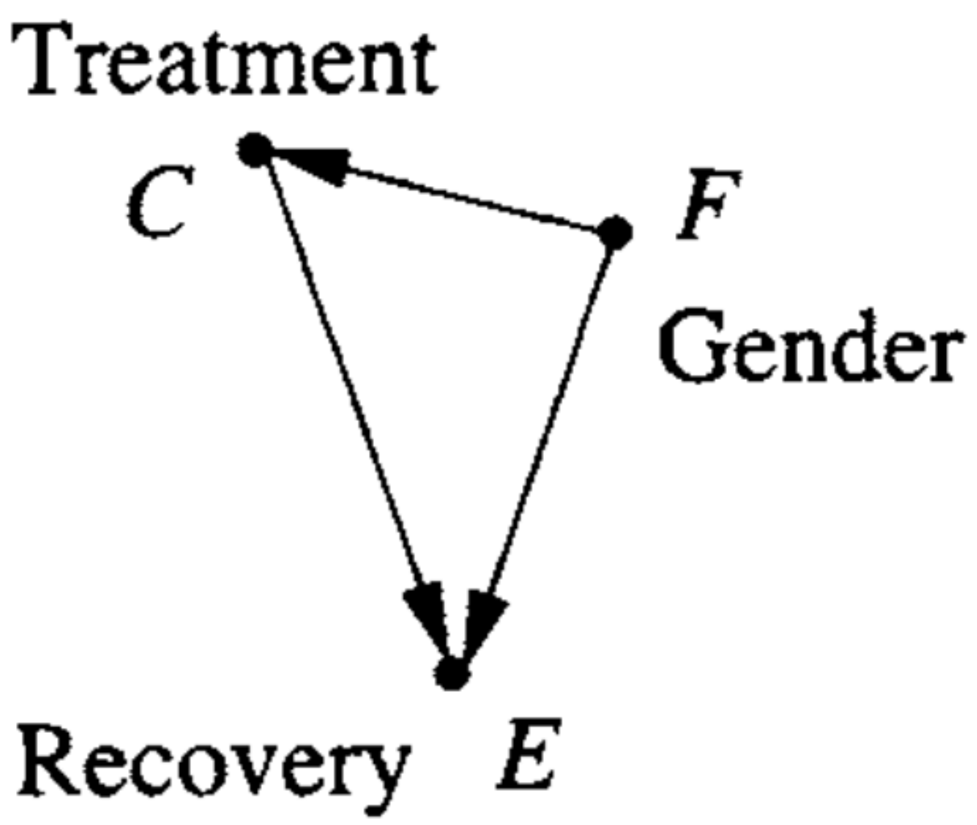
Statistics alone cannot suffice

Sure-Thing Principle:

Decision makers who would take a certain action if they knew that event F has occurred, and also if they knew that the negation of F has occurred, should also take that same action if they knows nothing about F.

$$P(E \mid C, F) < P(E \mid \neg C, F)$$

$$P(E \mid C, \neg F) < P(E \mid \neg C, \neg F)$$



(a)

Statistics alone cannot suffice

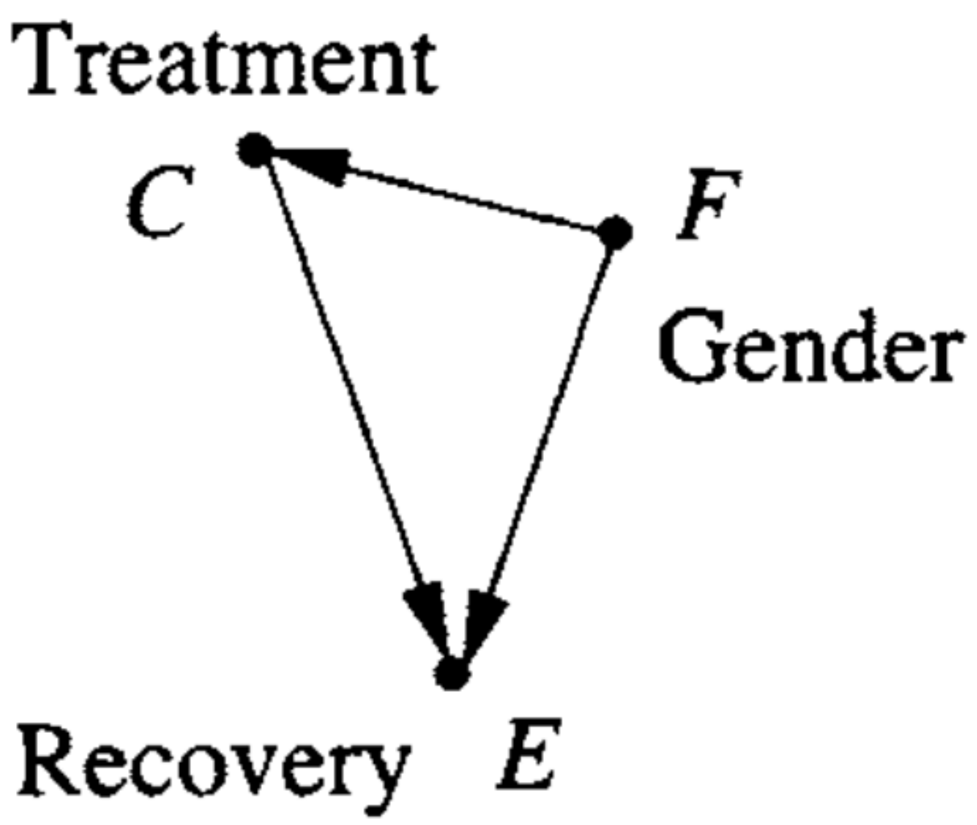
Sure-Thing Principle:

Decision makers who would take a certain action if they knew that event F has occurred, and also if they knew that the negation of F has occurred, should also take that same action if they knows nothing about F.

$$P(E \mid C, F) < P(E \mid \neg C, F)$$

$$P(E \mid C, \neg F) < P(E \mid \neg C, \neg F)$$

$$P(E \mid C) > P(E \mid \neg C)$$



(a)

Statistics alone cannot suffice

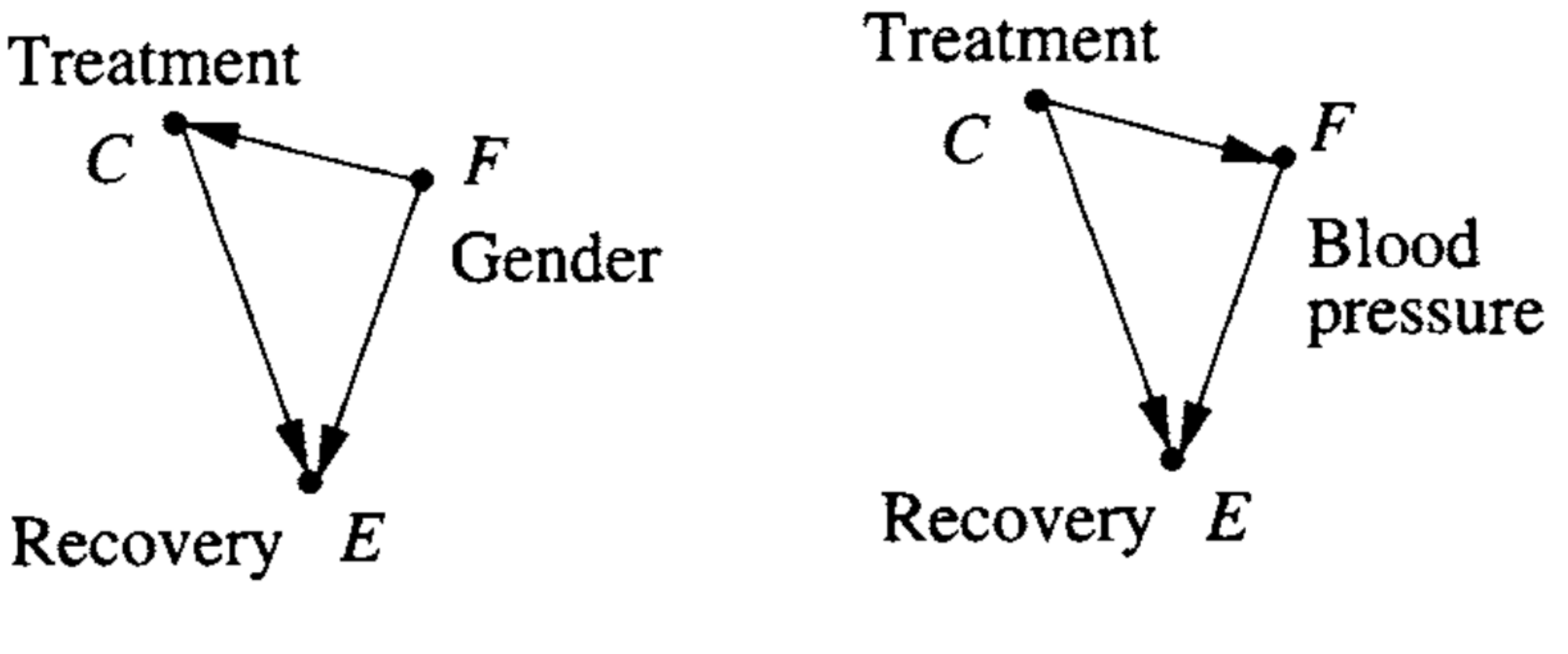
Sure-Thing Principle:

Decision makers who would take a certain action if they knew that event F has occurred, and also if they knew that the negation of F has occurred, should also take that same action if they knows nothing about F.

$$P(E \mid C, F) < P(E \mid \neg C, F)$$

$$P(E \mid C, \neg F) < P(E \mid \neg C, \neg F)$$

$$P(E \mid C) > P(E \mid \neg C)$$



(a) (b)

Statistics alone cannot suffice

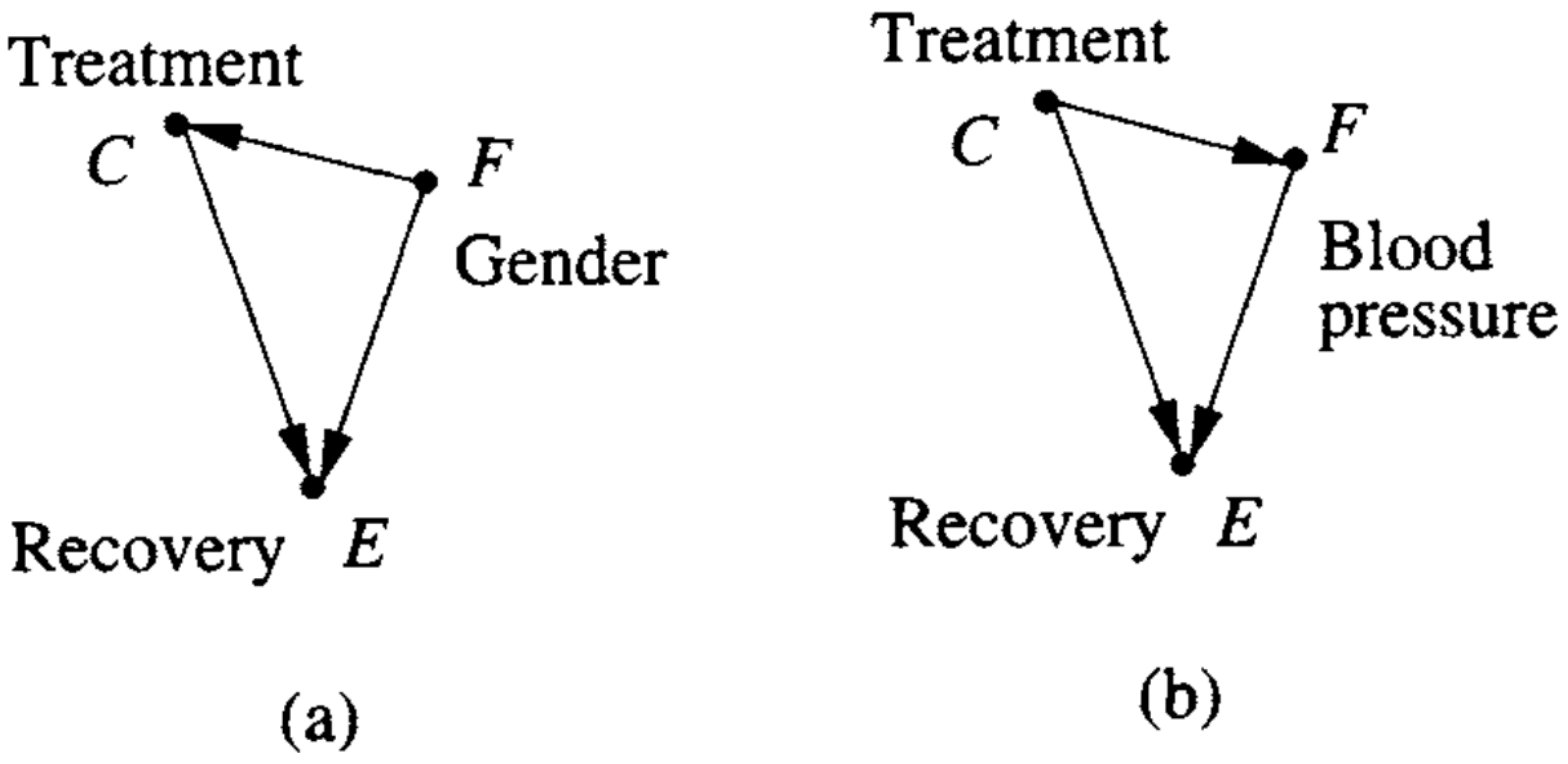
Sure-Thing Principle:

Decision makers who would take a certain action if they knew that event F has occurred, and also if they knew that the negation of F has occurred, should also take that same action if they knows nothing about F.

$$P(E \mid C, F) < P(E \mid \neg C, F)$$

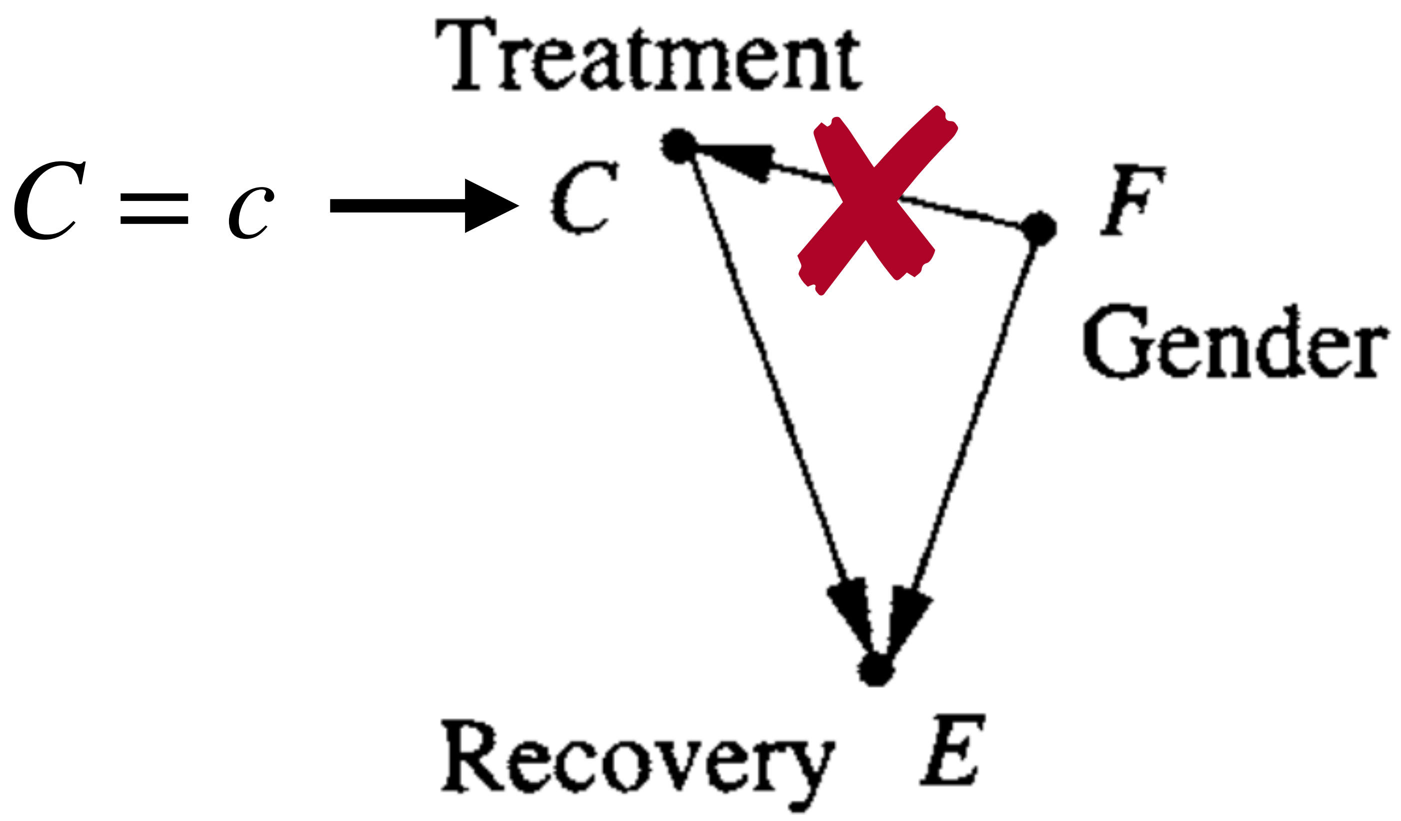
$$P(E \mid C, \neg F) < P(E \mid \neg C, \neg F)$$

$$P(E \mid C) > P(E \mid \neg C)$$



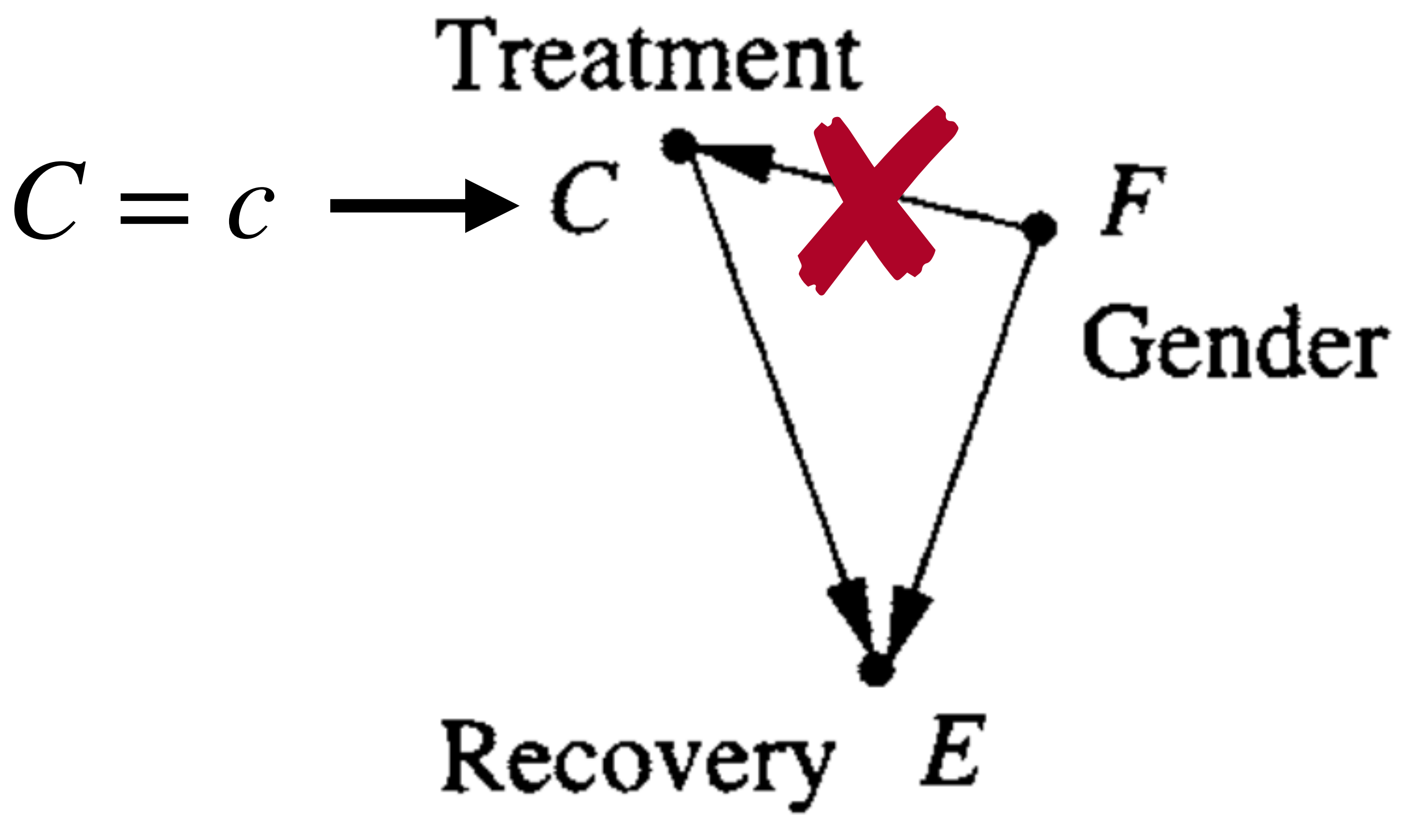
Conclusion: Our actions shall not be based on statistical consideration!

Role of interventions



(a)

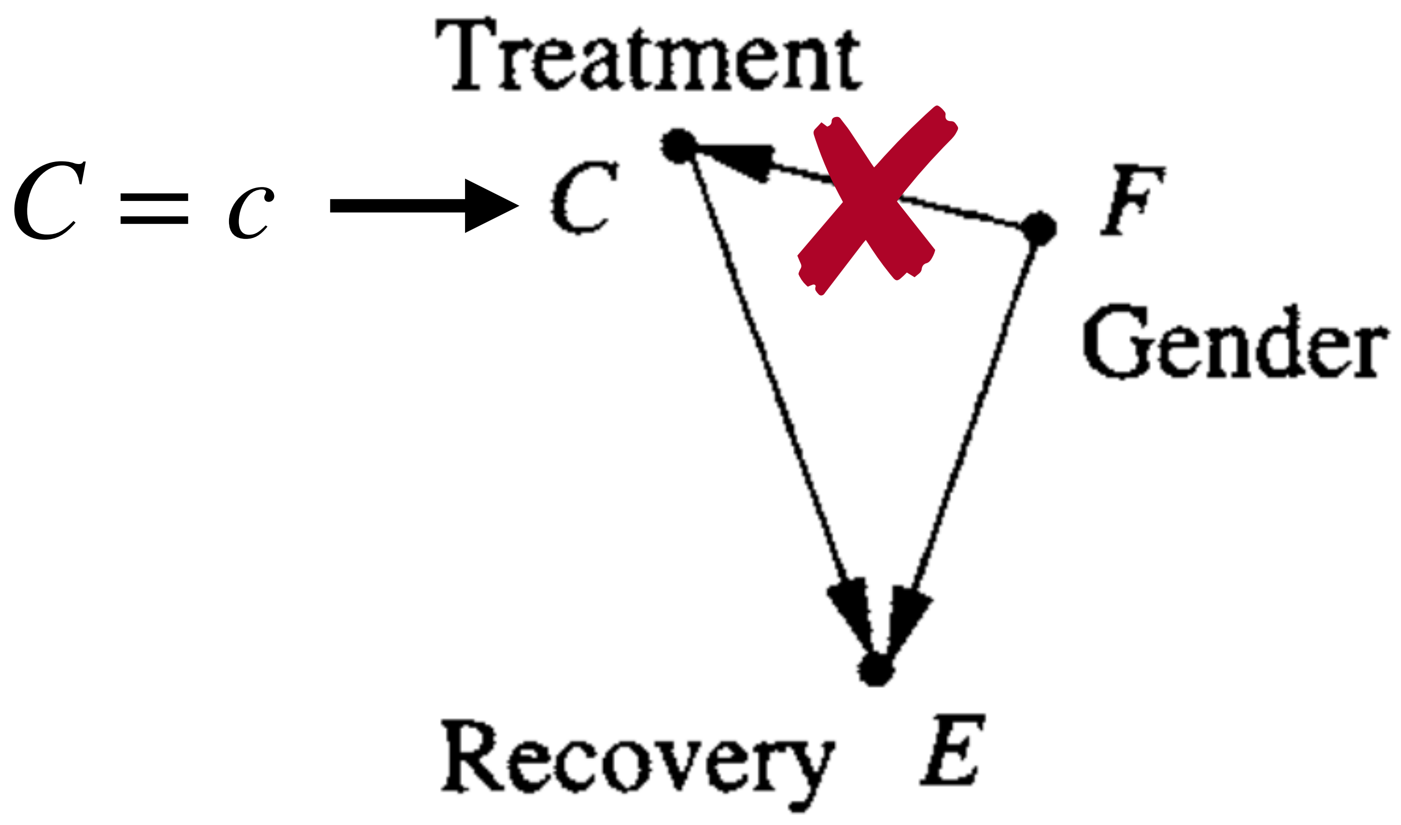
Role of interventions



(a)

$$P(E \mid do(C), F) < P(E \mid do(\neg C), F)$$

Role of interventions

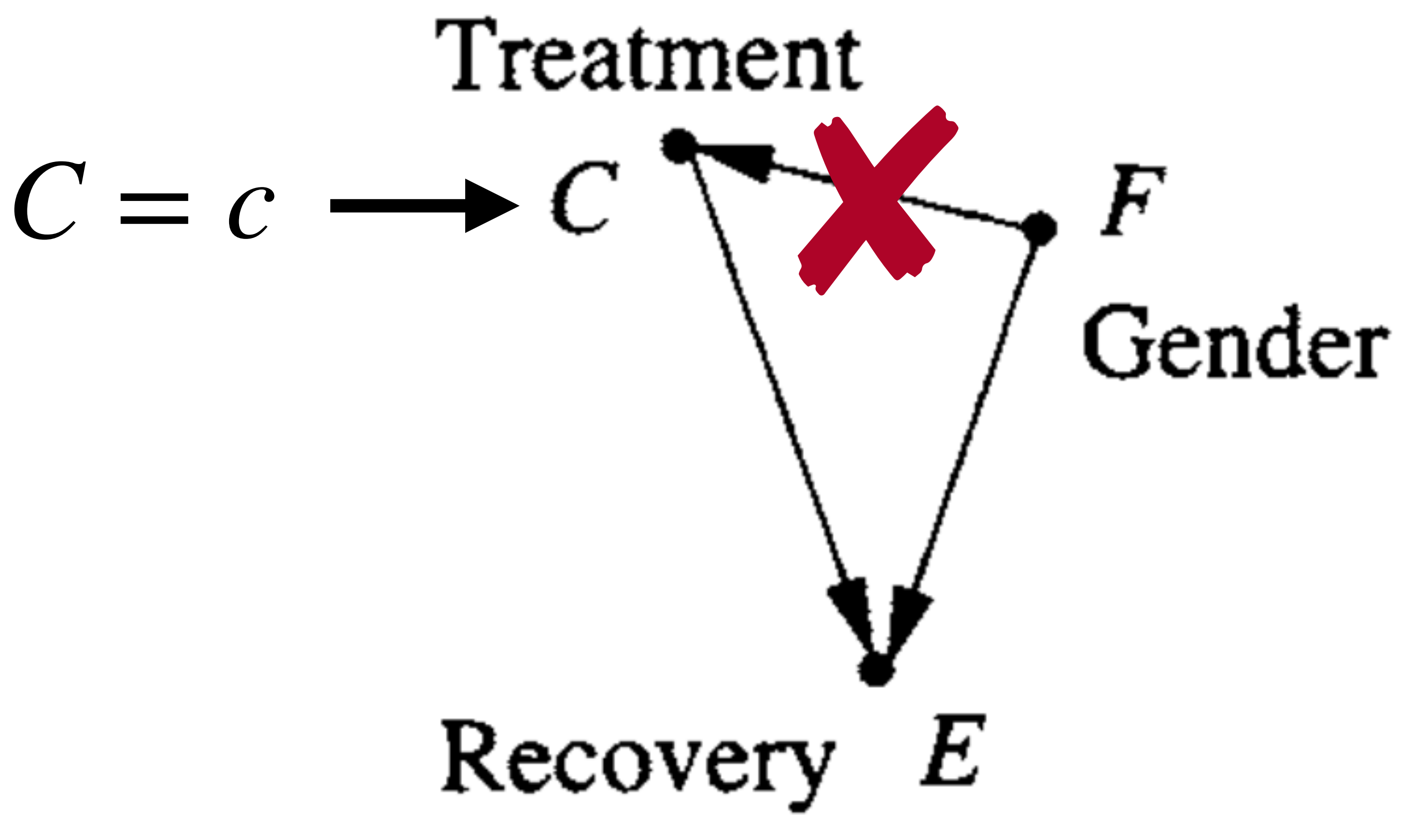


(a)

$$P(E \mid \text{do}(C), F) < P(E \mid \text{do}(\neg C), F)$$

$$P(E \mid \text{do}(\neg C), \neg F) < P(E \mid \text{do}(\neg C), \neg F)$$

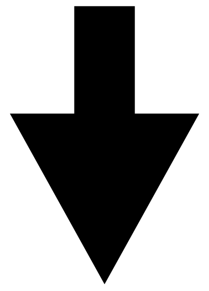
Role of interventions



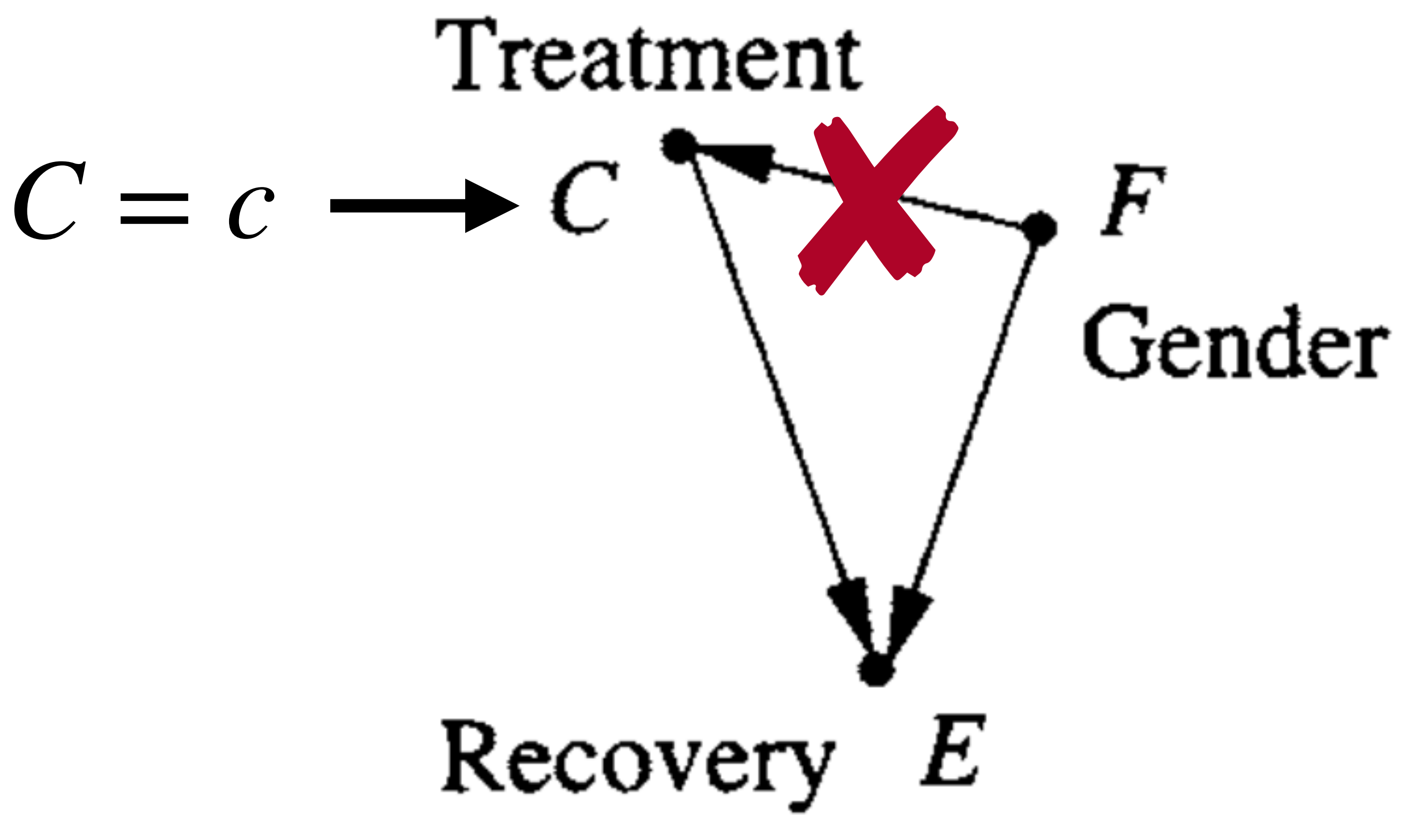
(a)

$$P(E \mid do(C), F) < P(E \mid do(\neg C), F)$$

$$P(E \mid do(\neg C), \neg F) < P(E \mid do(\neg C), \neg F)$$



Role of interventions



(a)

$$P(E \mid \mathrm{do}(C), F) < P(E \mid \mathrm{do}(\neg C), F)$$

$$P(E \mid \mathrm{do}(\neg C), \neg F) < P(E \mid \mathrm{do}(\neg C), \neg F)$$

$$P(E \mid \mathrm{do}(C)) < P(E \mid \mathrm{do}(\neg C))$$

Role of interventions

	Combined	E	$\neg E$		Recovery Rate
(a)	Drug (C)	20	20	40	50%
	No drug ($\neg C$)	16	24	40	40%
		36	44	80	

	Males	E	$\neg E$		Recovery Rate
(b)	Drug (C)	18	12	30	60%
	No drug ($\neg C$)	7	3	10	70%
		25	15	40	

	Females	E	$\neg E$		Recovery Rate
(c)	Drug (C)	2	8	10	20%
	No drug ($\neg C$)	9	21	30	30%
		11	29	40	

$$P(E \mid \mathrm{do}(C), F) < P(E \mid \mathrm{do}(\neg C), F)$$

$$P(E \mid \mathrm{do}(\neg C), \neg F) < P(E \mid \mathrm{do}(\neg C), \neg F)$$

$$\Downarrow$$

$$P(E \mid \mathrm{do}(C)) < P(E \mid \mathrm{do}(\neg C))$$

Role of interventions

	Combined	E	$\neg E$		Recovery Rate
(a)	Drug (C)	20	20	40	50%
	No drug ($\neg C$)	16	24	40	40%
		36	44	80	

	Males	E	$\neg E$		Recovery Rate
(b)	Drug (C)	18	12	30	60%
	No drug ($\neg C$)	7	3	10	70%
		25	15	40	

	Females	E	$\neg E$		Recovery Rate
(c)	Drug (C)	2	8	10	20%
	No drug ($\neg C$)	9	21	30	30%
		11	29	40	

$$P(E \mid \mathrm{do}(C), F) < P(E \mid \mathrm{do}(\neg C), F)$$

$$P(E \mid \mathrm{do}(\neg C), \neg F) < P(E \mid \mathrm{do}(\neg C), \neg F)$$

$$\Downarrow$$

$$P(E \mid \mathrm{do}(C)) < P(E \mid \mathrm{do}(\neg C))$$

$$P(E \mid \mathrm{do}(C)) = \sum_{F} P(E \mid C, F) P(F)$$

Role of interventions

	Combined	E	$\neg E$		Recovery Rate
(a)	Drug (C)	20	20	40	50%
	No drug ($\neg C$)	16	24	40	40%
		36	44	80	

	Males	E	$\neg E$		Recovery Rate
(b)	Drug (C)	18	12	30	60%
	No drug ($\neg C$)	7	3	10	70%
		25	15	40	

	Females	E	$\neg E$		Recovery Rate
(c)	Drug (C)	2	8	10	20%
	No drug ($\neg C$)	9	21	30	30%
		11	29	40	

$$P(E \mid \mathrm{do}(C), F) < P(E \mid \mathrm{do}(\neg C), F)$$

$$P(E \mid \mathrm{do}(\neg C), \neg F) < P(E \mid \mathrm{do}(\neg C), \neg F)$$

$$\Downarrow$$

$$P(E \mid \mathrm{do}(C)) < P(E \mid \mathrm{do}(\neg C))$$

$$P(E \mid \mathrm{do}(C)) = \sum_F P(E \mid C, F) P(F)$$

Question: Can you still reason, if you are just given the recovery rates $P(E \mid C, F)$?

Role of interventions

	Combined	E	$\neg E$		Recovery Rate
(a)	Drug (C)	20	20	40	50%
	No drug ($\neg C$)	16	24	40	40%
		36	44	80	

	Males	E	$\neg E$		Recovery Rate
(b)	Drug (C)	18	12	30	60%
	No drug ($\neg C$)	7	3	10	70%
		25	15	40	

	Females	E	$\neg E$		Recovery Rate
(c)	Drug (C)	2	8	10	20%
	No drug ($\neg C$)	9	21	30	30%
		11	29	40	

$$P(E \mid do(C), F) < P(E \mid do(\neg C), F)$$

$$P(E \mid do(\neg C), \neg F) < P(E \mid do(\neg C), \neg F)$$

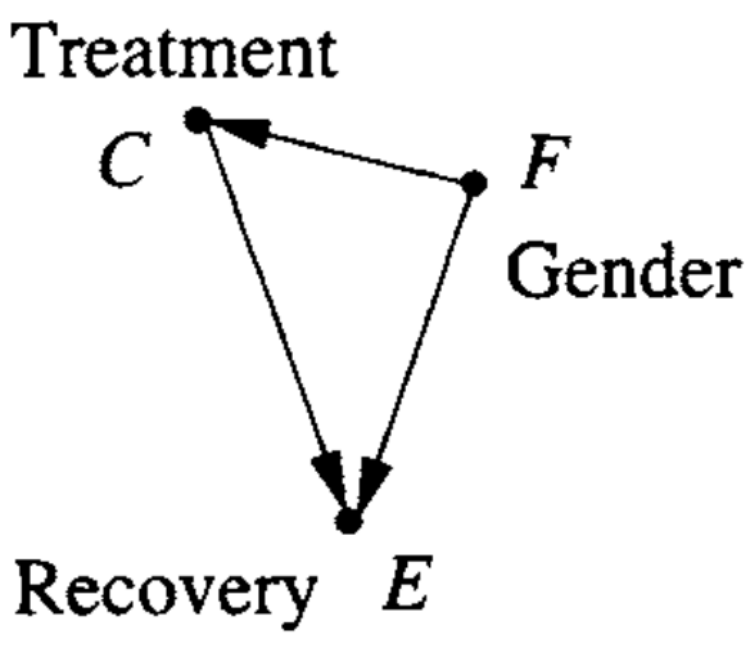
$$P(E \mid do(C)) < P(E \mid do(\neg C))$$

$$P(E \mid do(C)) = \sum_F P(E \mid C, F) P(F)$$

Question: Can you still reason, if you are just given the recovery rates $P(E \mid C, F)$?

Understanding Simpson's paradox

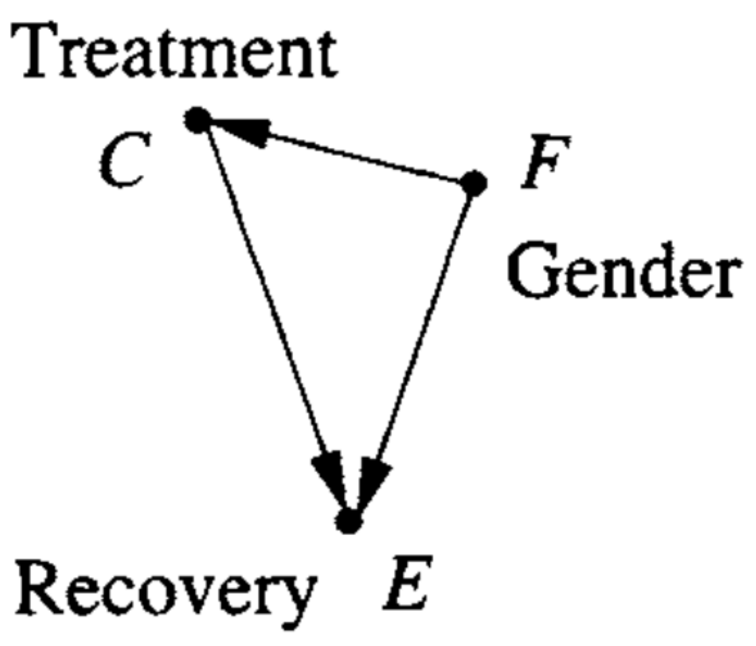
Understanding Simpson's paradox



(a)

Understanding Simpson's paradox

A Path blocking:



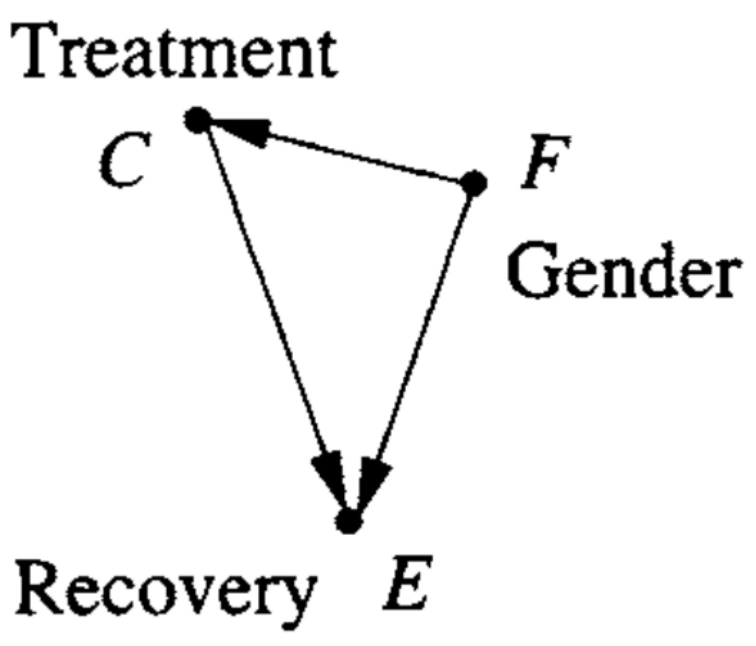
(a)

Understanding Simpson's paradox

A Path blocking:

A path is blocked by F if and only if it contains,

1. a chain $I \rightarrow F \rightarrow J$ or $I \leftarrow F \rightarrow J$, or
2. an inverted fork $I \rightarrow M \leftarrow J$, where $M \neq F$ and M is not a parent of F.



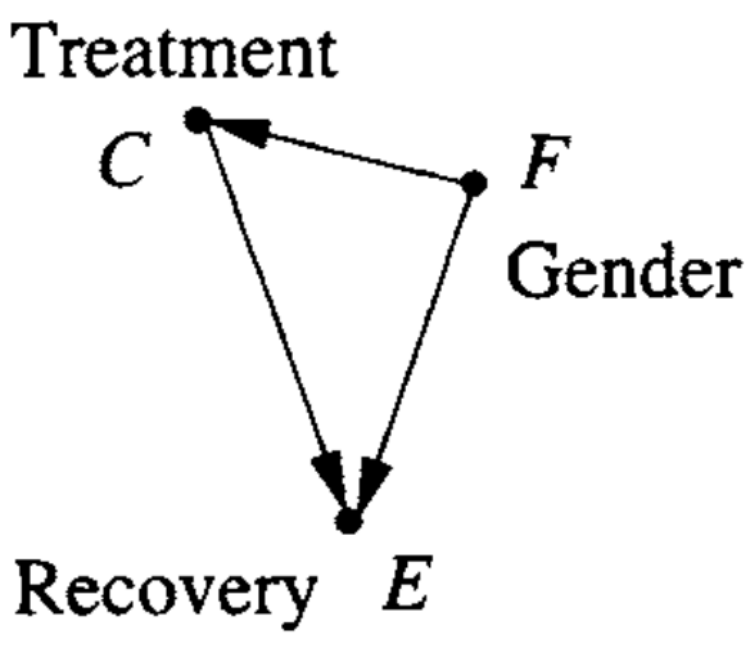
(a)

Understanding Simpson's paradox

A Path blocking:

A path is blocked by F if and only if it contains,

1. a chain $I \to F \to J$ or $I \leftarrow F \to J$, or
2. an inverted fork $I \to M \leftarrow J$, where $M \neq F$ and M is not a parent of F.



(a)

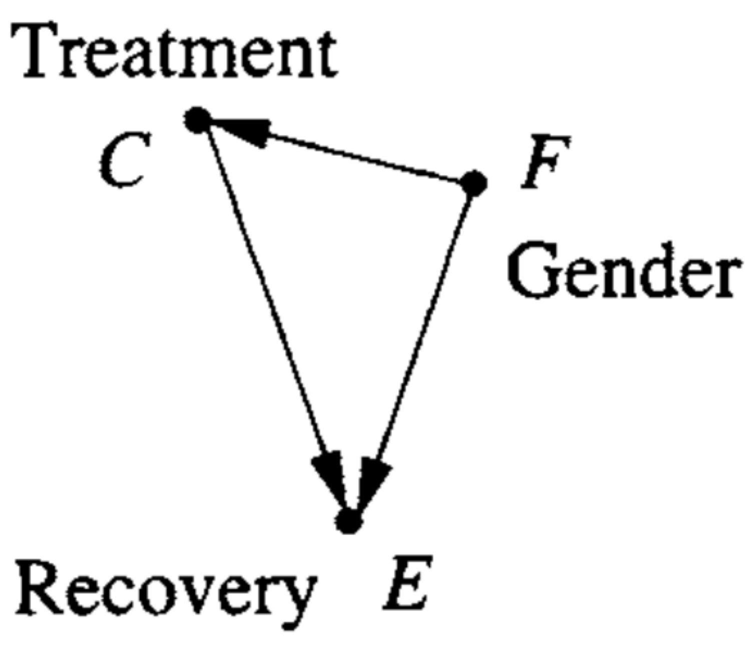
The back-door criteria:

Understanding Simpson's paradox

A Path blocking:

A path is blocked by F if and only if it contains,

1. a chain $I \rightarrow F \rightarrow J$ or $I \leftarrow F \rightarrow J$, or
2. an inverted fork $I \rightarrow M \leftarrow J$, where $M \neq F$ and M is not a parent of F.

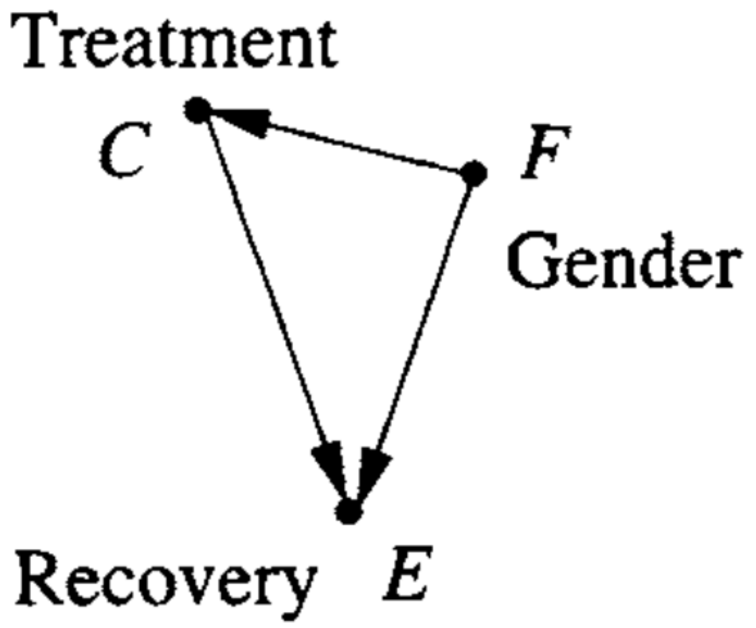


(a)

The back-door criteria:

1. F is not a descendant of C
2. F blocks every path between C and E that contains arrow into C.

Understanding Simpson's paradox



(a)

A Path blocking:

A path is blocked by F if and only if it contains,

1. a chain $I \rightarrow F \rightarrow J$ or $I \leftarrow F \rightarrow J$, or
2. an inverted fork $I \rightarrow M \leftarrow J$, where $M \neq F$ and M is not a parent of F.

The back-door criteria:

1. F is not a descendant of C
2. F blocks every path between C and E that contains arrow into C.

$\Rightarrow$ **We must condition on F.**

Understanding Simpson's paradox

A Path blocking:

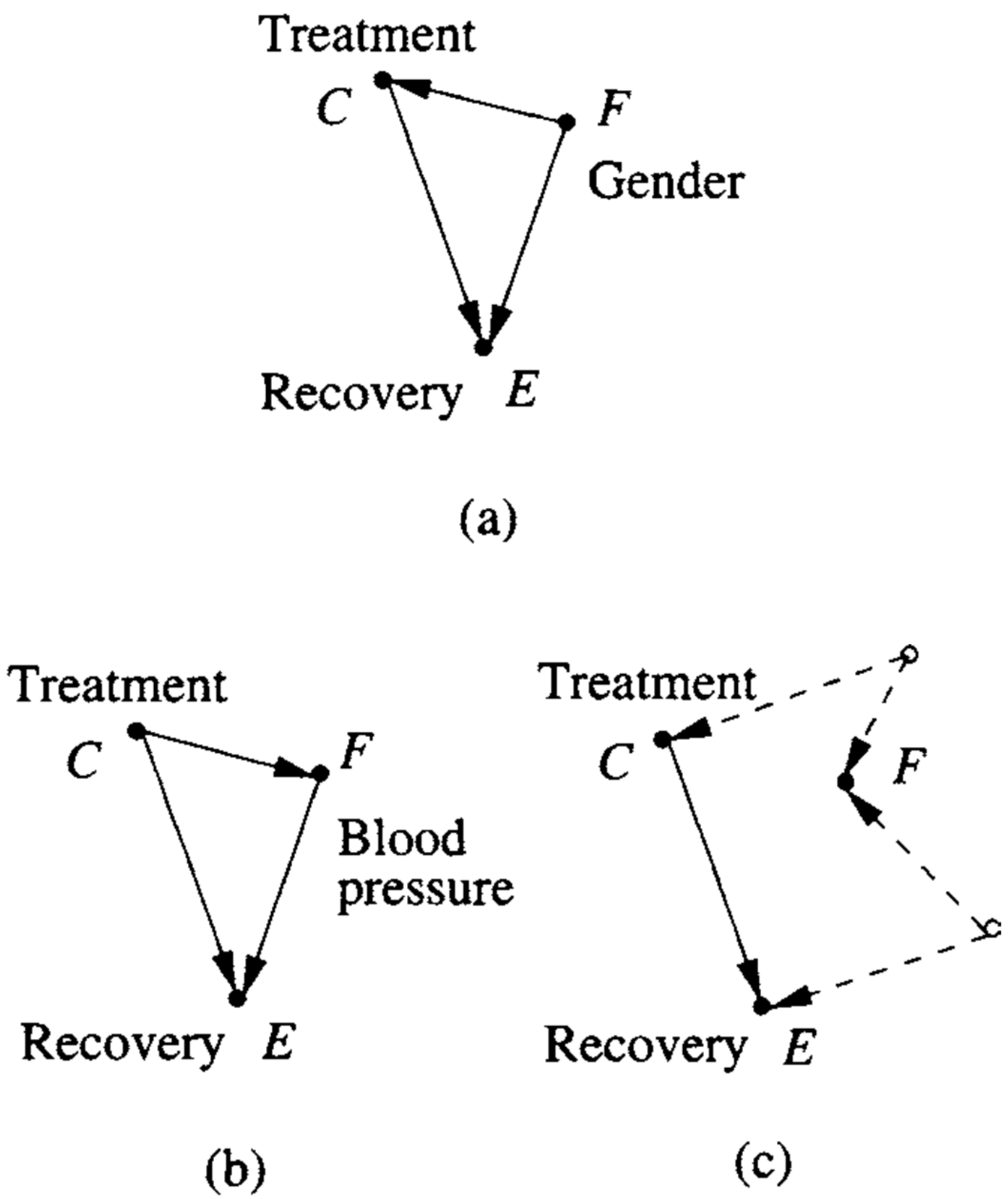
A path is blocked by F if and only if it contains,

1. a chain $I \to F \to J$ or $I \leftarrow F \to J$, or
2. an inverted fork $I \to M \leftarrow J$, where $M \neq F$ and M is not a parent of F.

The back-door criteria:

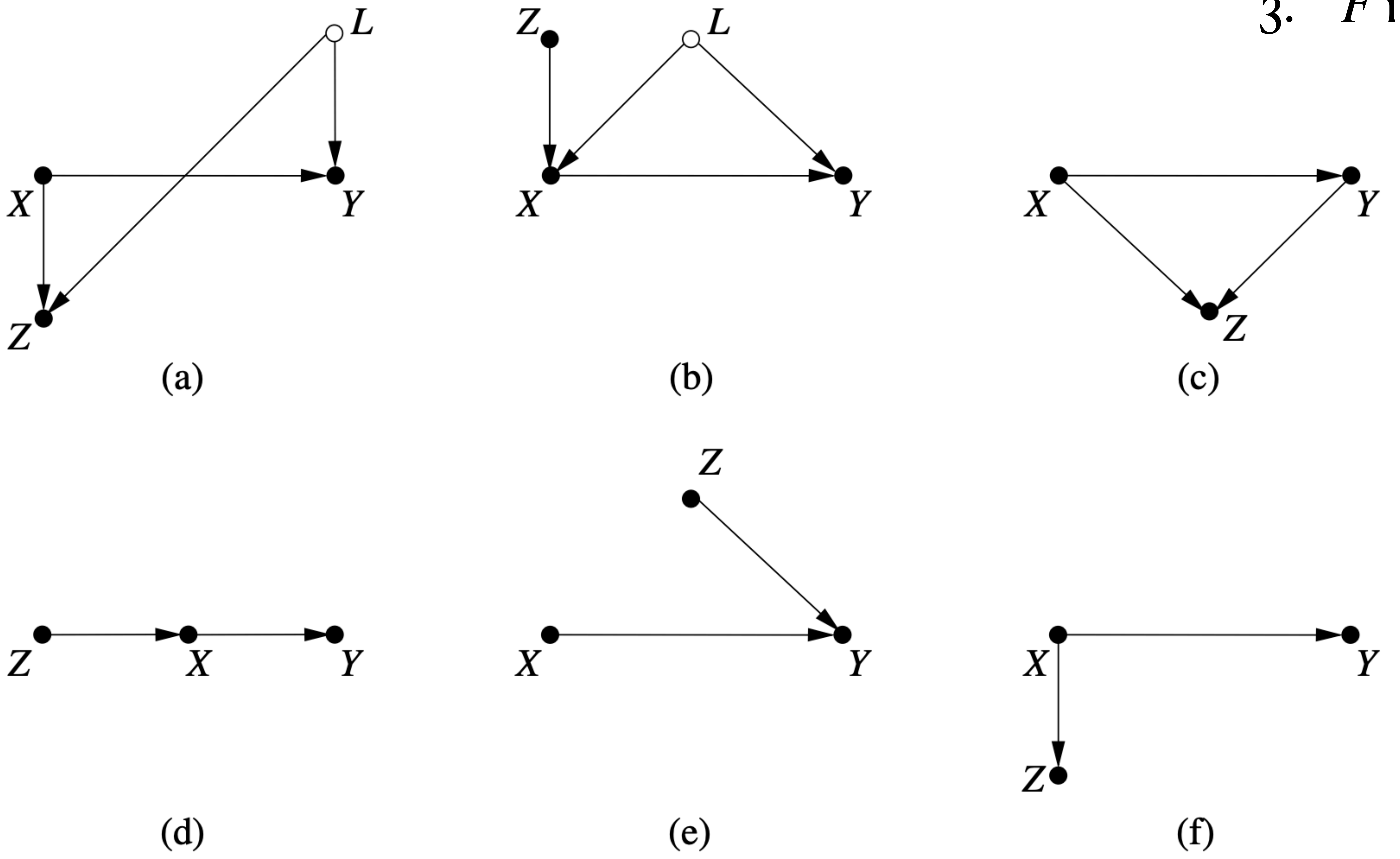
1. F is not a descendant of C
2. F blocks every path between C and E that contains arrow into C.

$\Rightarrow$ **We must condition on F.**



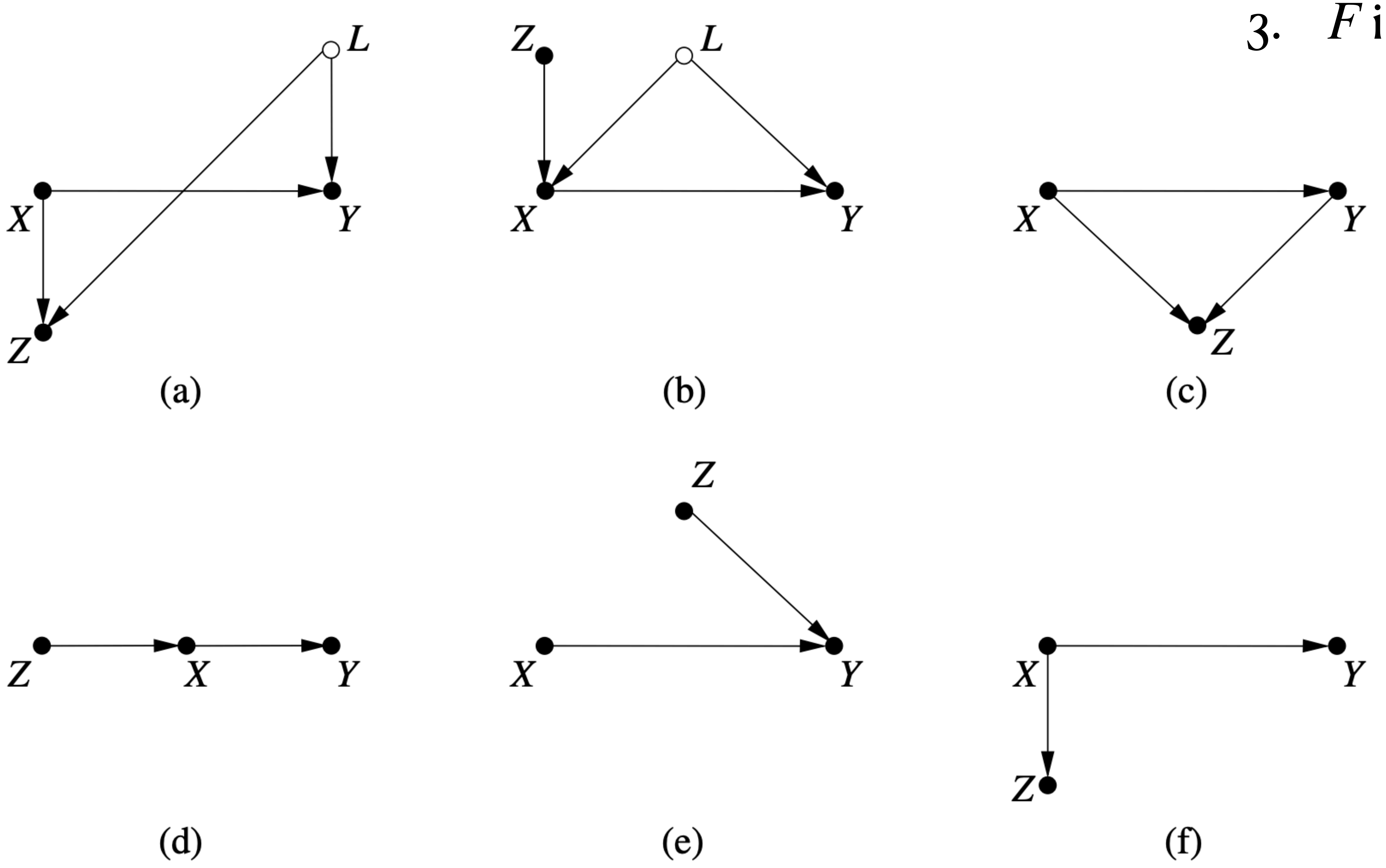
More examples

1. $I \to F \to J$ or $I \leftarrow F \to J$, or
2. $I \to M \leftarrow J$, where $M \neq F$.
3. F is not a descendant of C



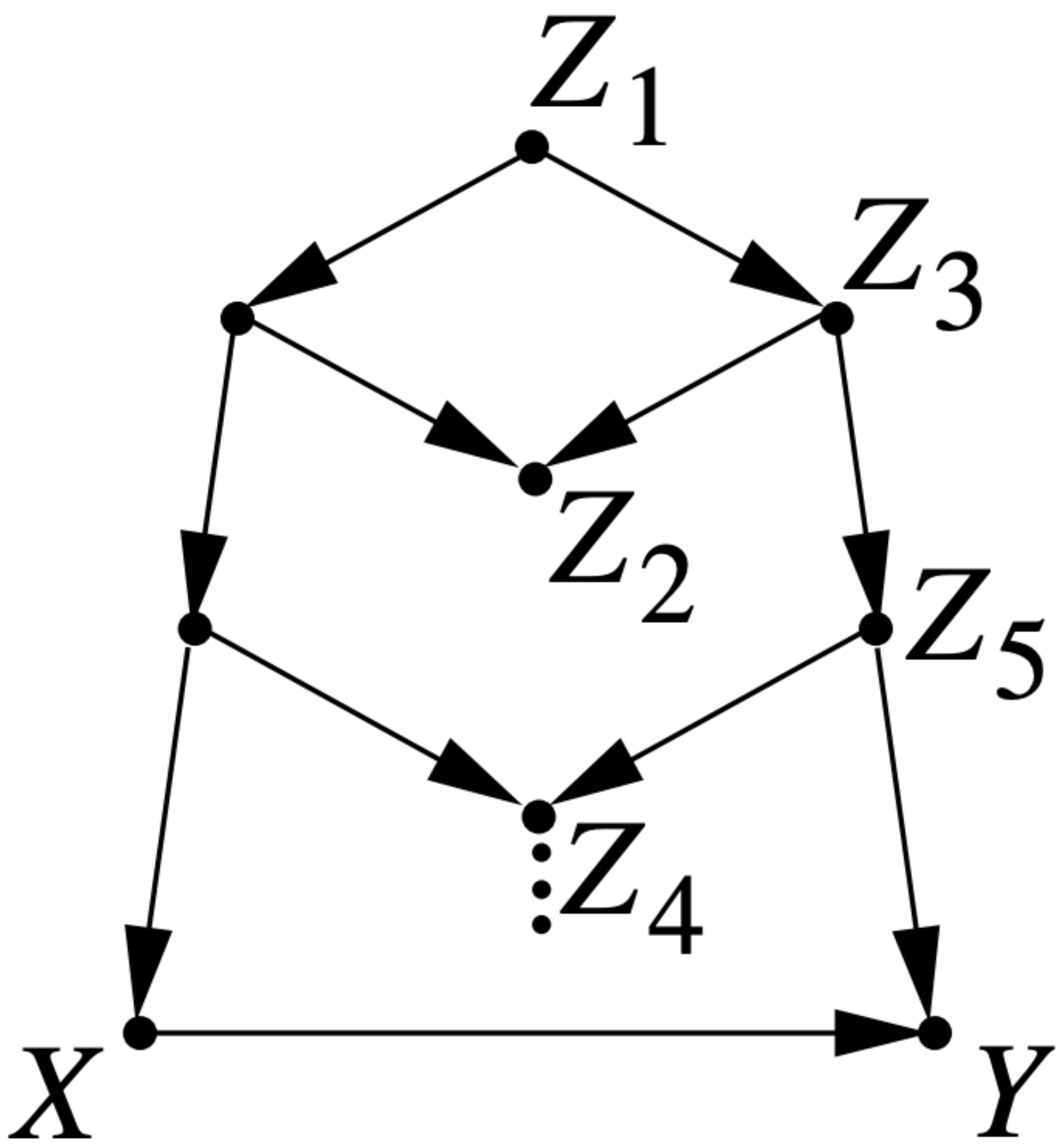
More examples

1. $I \rightarrow F \rightarrow J$ or $I \leftarrow F \rightarrow J$, or
2. $I \rightarrow M \leftarrow J$, where $M \neq F$.
3. F is not a descendant of C



Bonus Question: In which of these causal structures is Simpson's reversal realisable?

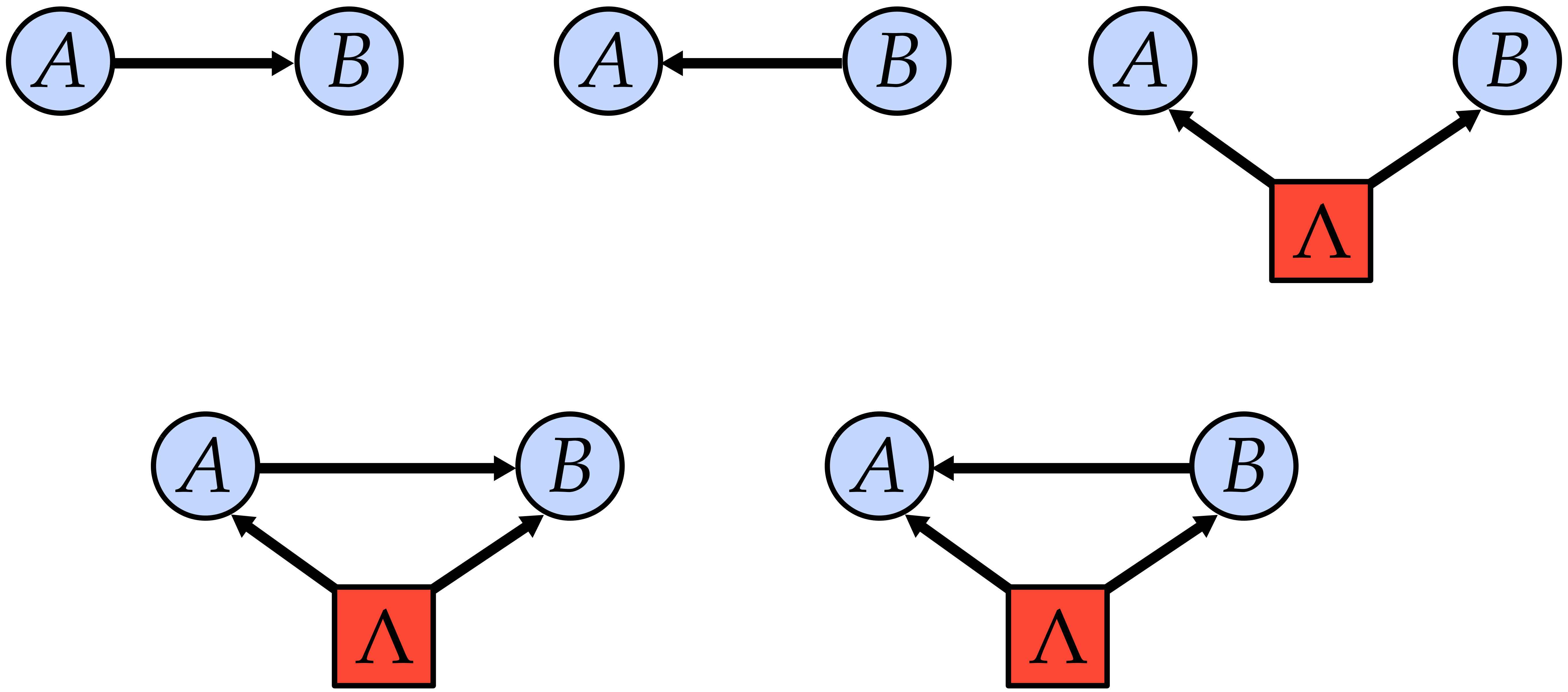
A multi-stage Simpson's paradox



Part 2:
Observational Causal Inference

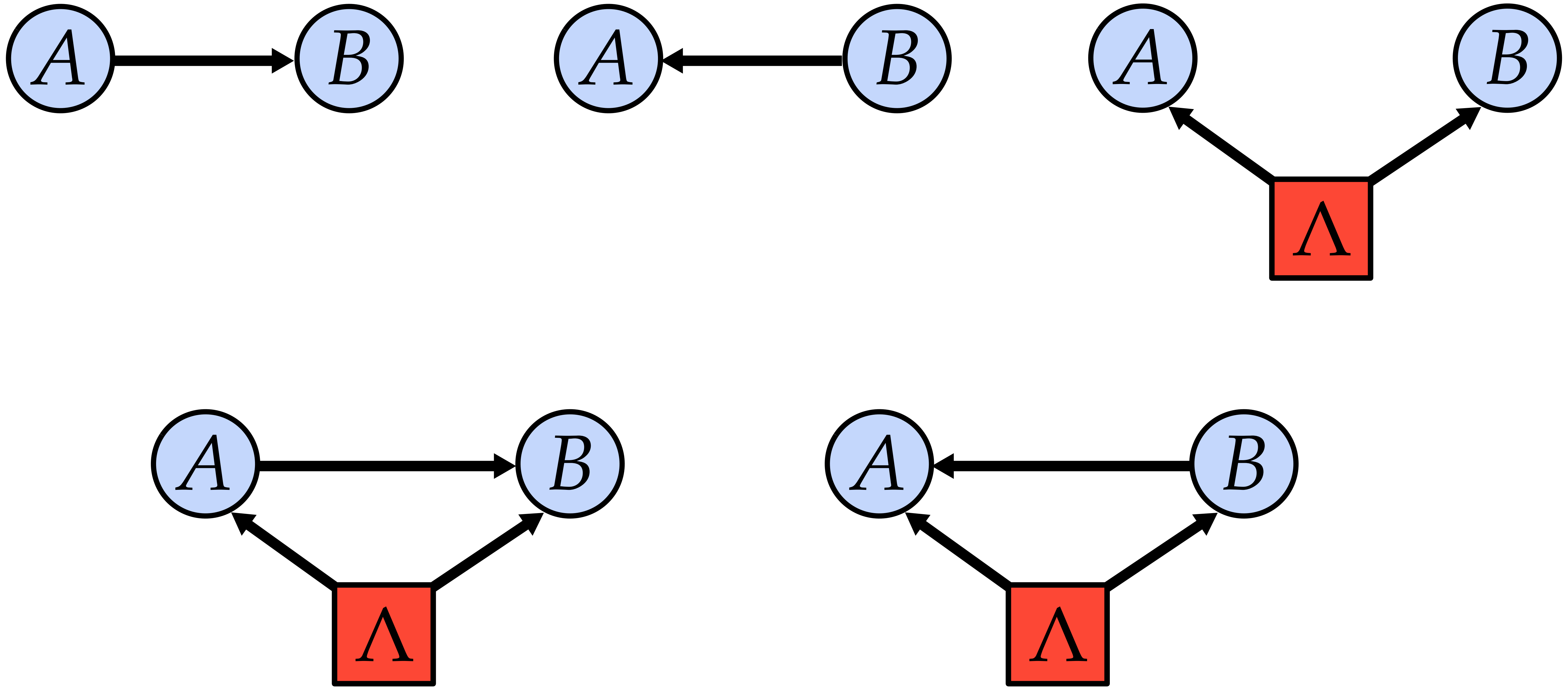
Recap: Causal Reasoning

Causal Reasoning



Example: A – smoking, B – cancer, Λ – genetics

Causal Reasoning



Example: A – smoking, B – cancer, Λ – genetics

Notations: A, B, Λ – random variables, a, b, λ – values, $p(a) = p(A = a)$.

Can we measure how much
A is influencing B?

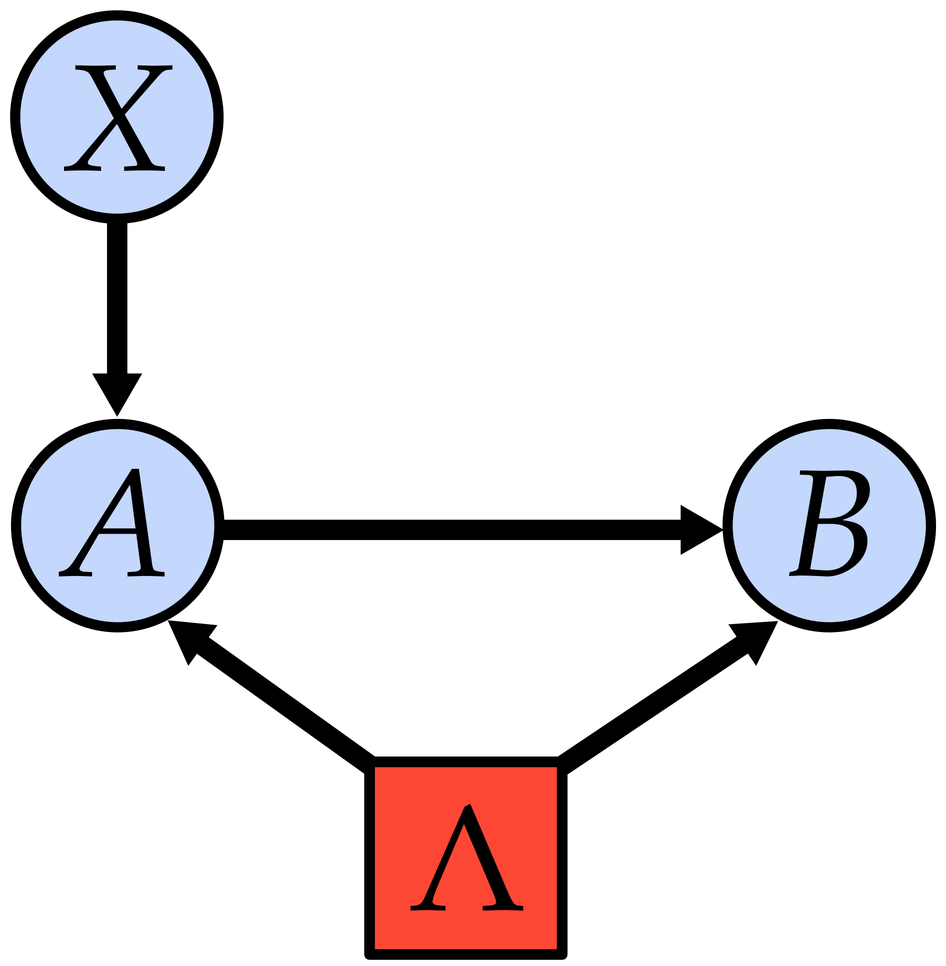
Measuring Causality: Interventions

Measuring Causality: Interventions

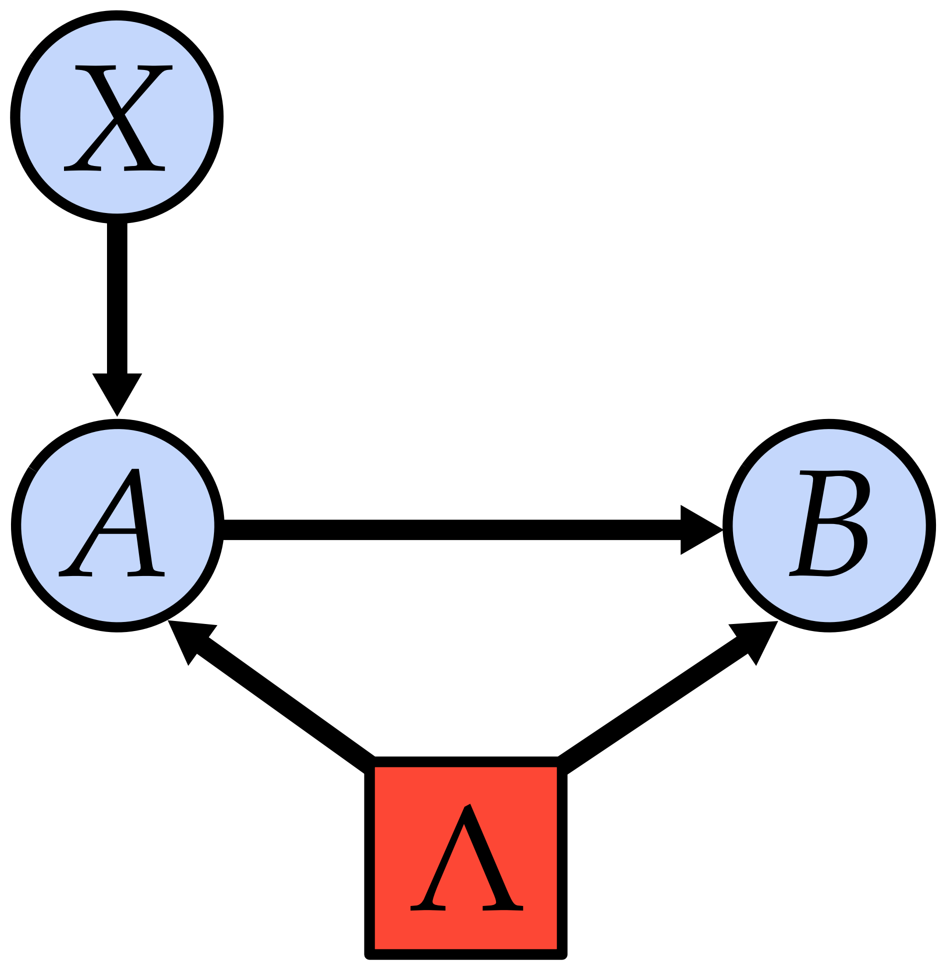
Average causal effect[1]:

$$\mathrm{ACE}_{A\to B} = \max_{a,a',b}\Big(p(b|do(a)) - p(b|do(a'))\Big). \tag{1}$$

Measuring Causality: Instrument

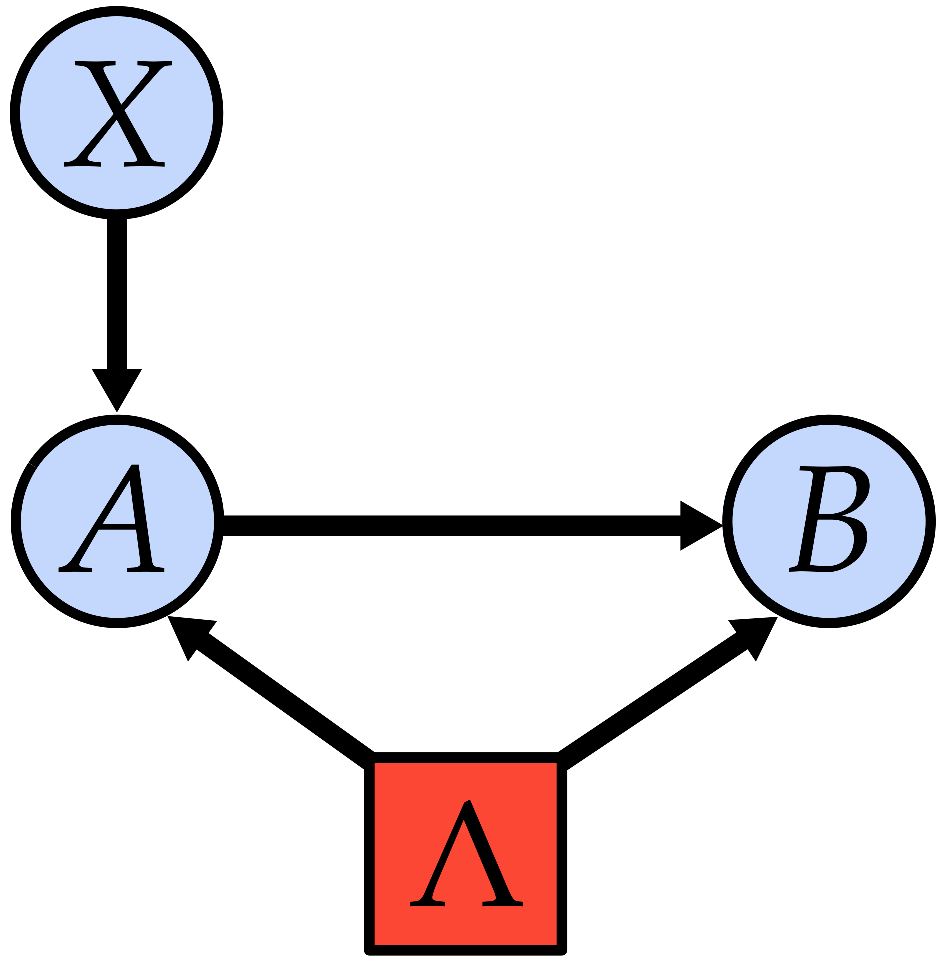


Measuring Causality: Instrument



Example: A – smoking, B – cancer, Λ – genetics, X – taxation of tobacco.

Measuring Causality: Instrument



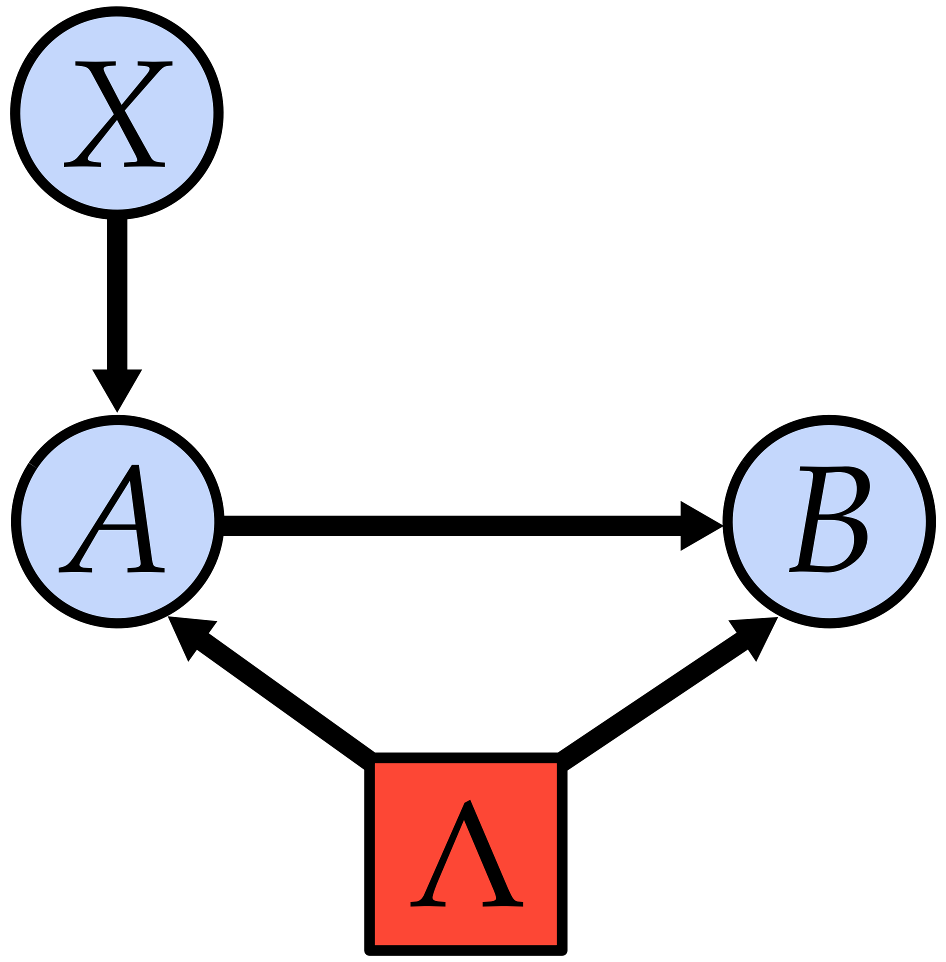
Example: A – smoking, B – cancer, Λ – genetics, X – taxation of tobacco.

X – instrumental variable,

Statistics: $p(a, b|x) = \sum_\lambda p(a|x, \lambda)p(b|a, \lambda)p(\lambda)$.

Assumption: $p(\lambda, x) = p(\lambda)p(x)$

Measuring Causality: Instrument



Example: A – smoking, B – cancer, Λ – genetics, X – taxation of tobacco.

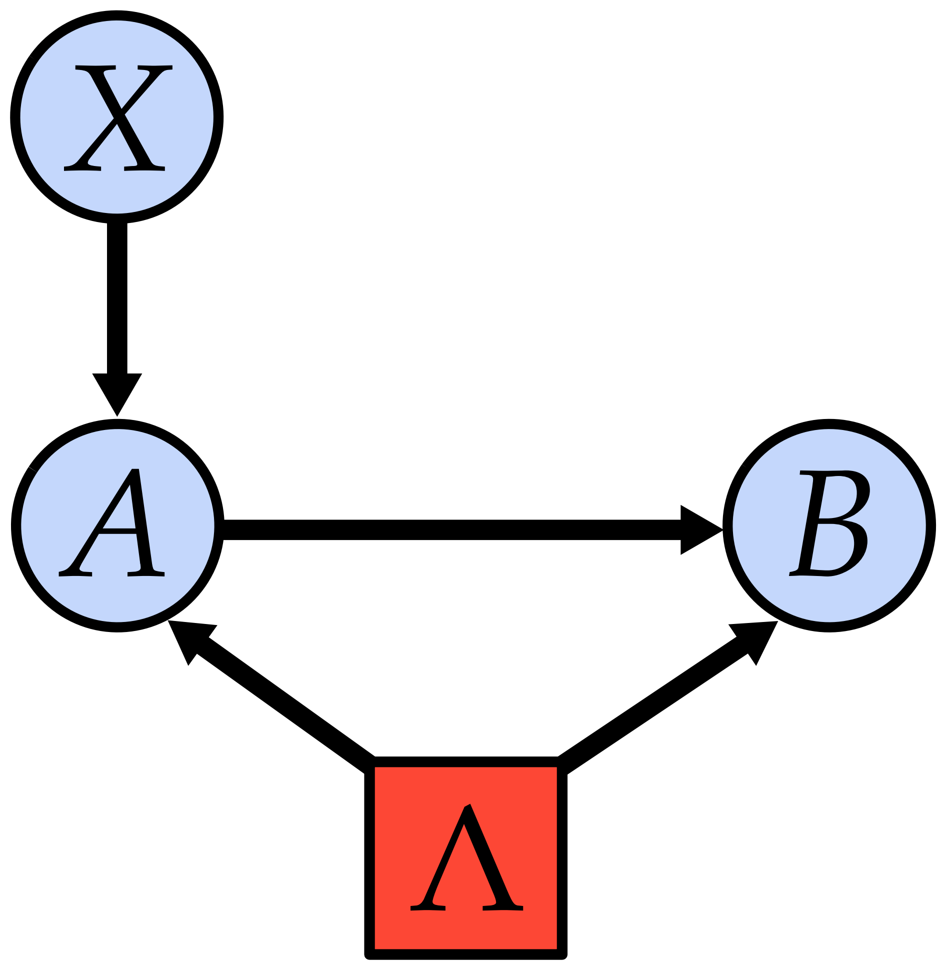
X – instrumental variable,

Statistics: $p(a,b|x) = \sum_\lambda p(a|x,\lambda)p(b|a,\lambda)p(\lambda)$.

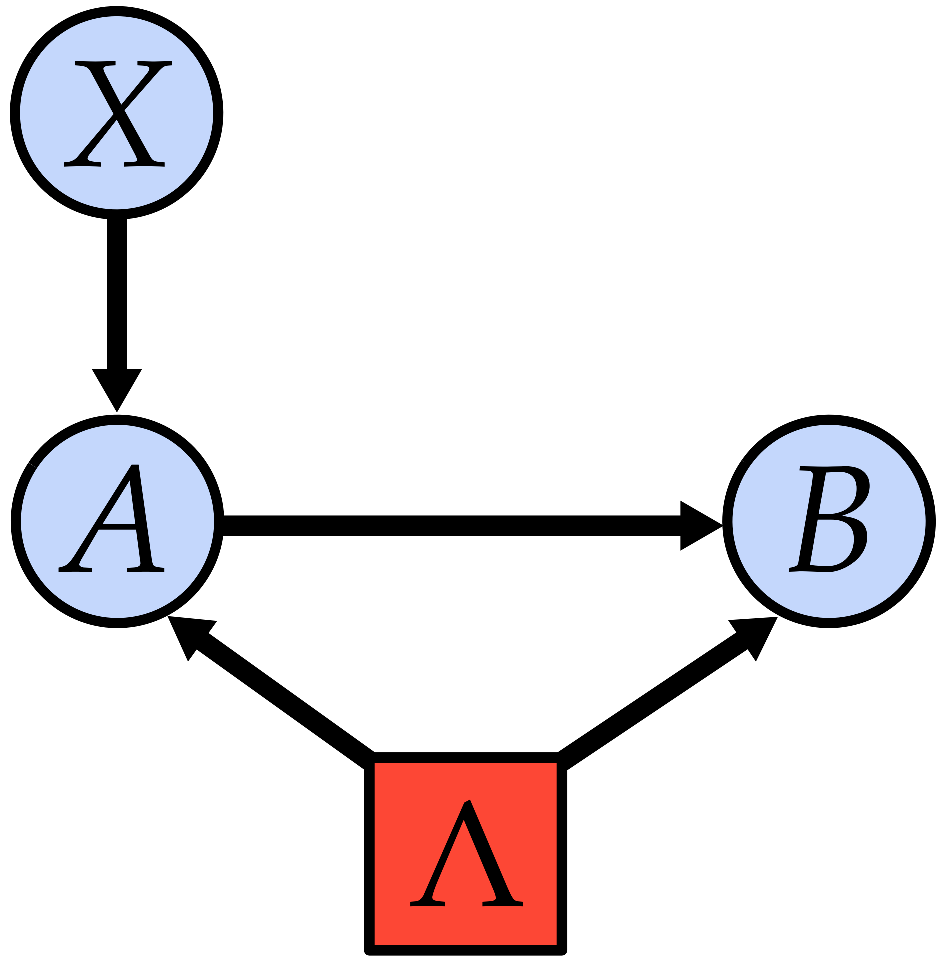
Assumption: $p(\lambda, x) = p(\lambda)p(x)$

If $b = \kappa a + \lambda$, then $\kappa = \dfrac{C(X,B)}{C(X,A)}$, where $C(X,B) = \langle X, B\rangle - \langle X\rangle\langle B\rangle$ – covariance, and $C(X,\Lambda) = 0$.

Simplest Instrumental Scenario

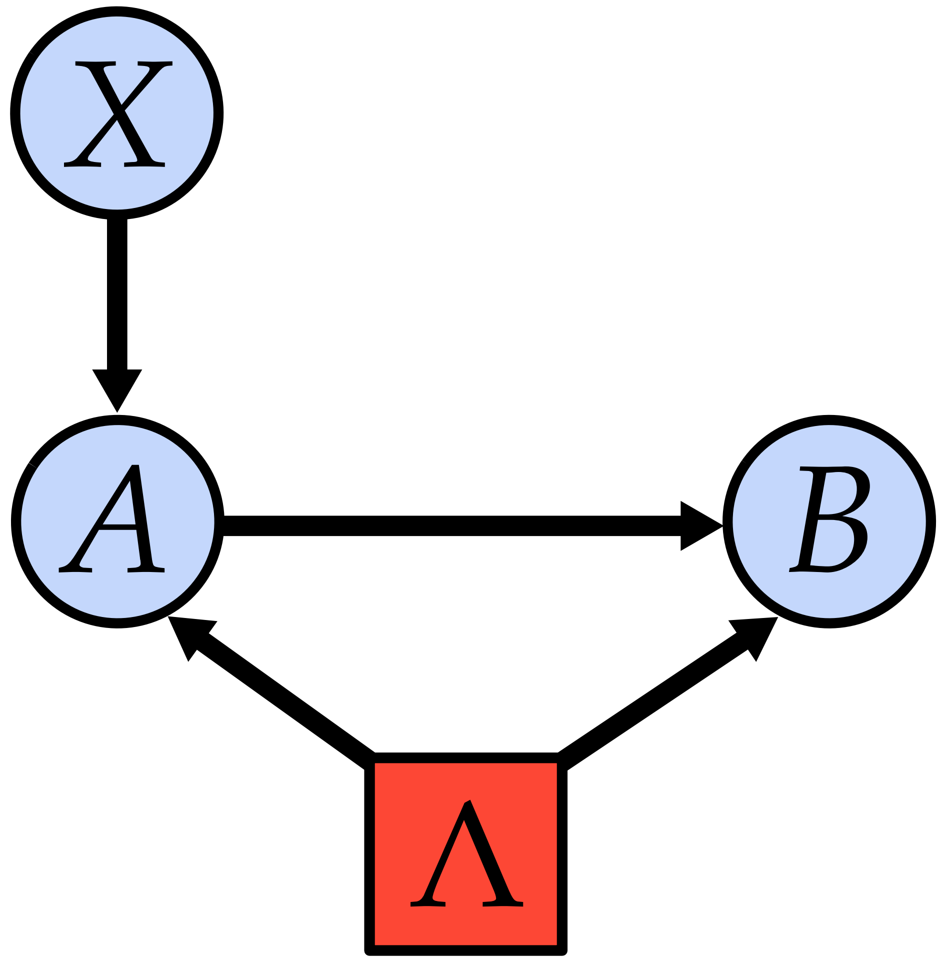


Simplest Instrumental Scenario



Example: A – smoking, B – cancer, Λ – genetics, X – taxation of tobacco.

Simplest Instrumental Scenario

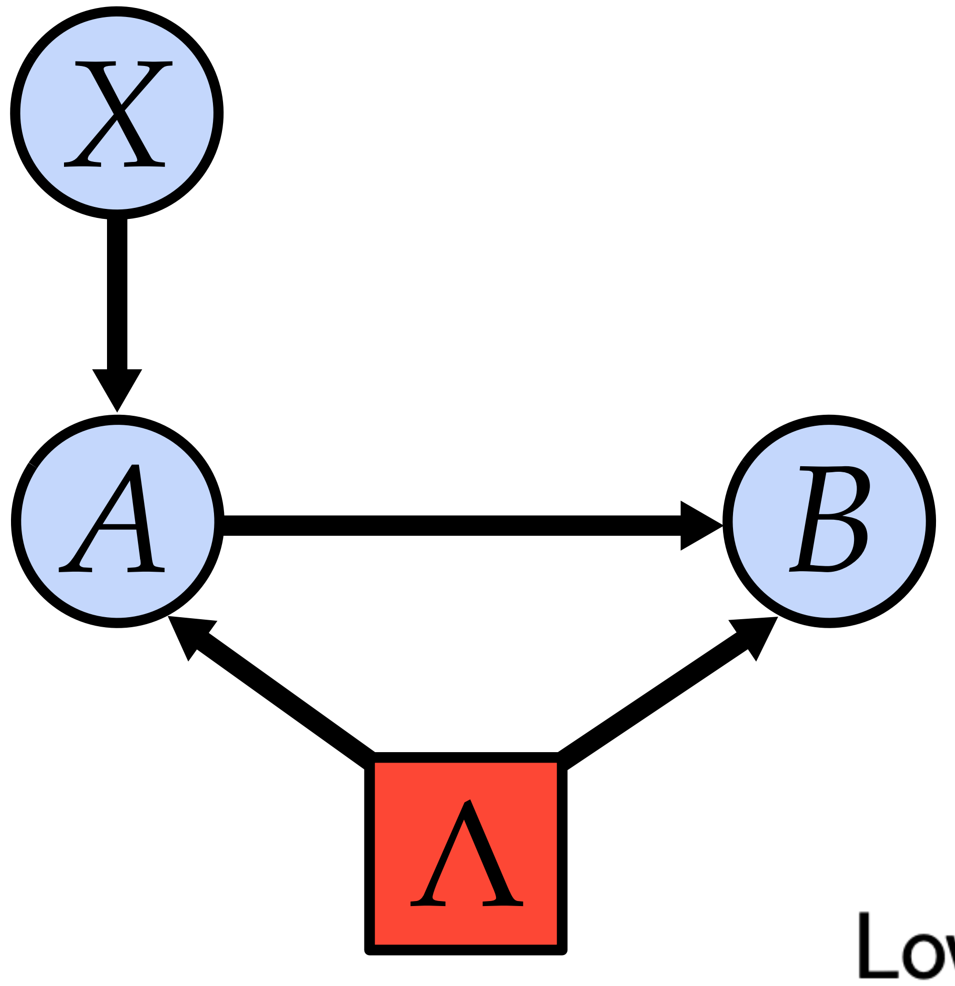


Example: A – smoking, B – cancer, Λ – genetics, X – taxation of tobacco.

$x, a, b \in \{0, 1\}$

$p(a, b|x)$

Simplest Instrumental Scenario



Example: A – smoking, B – cancer, Λ – genetics, X – taxation of tobacco.

$x, a, b \in \{0, 1\}$

$p(a, b|x)$

Lower bound[2]:

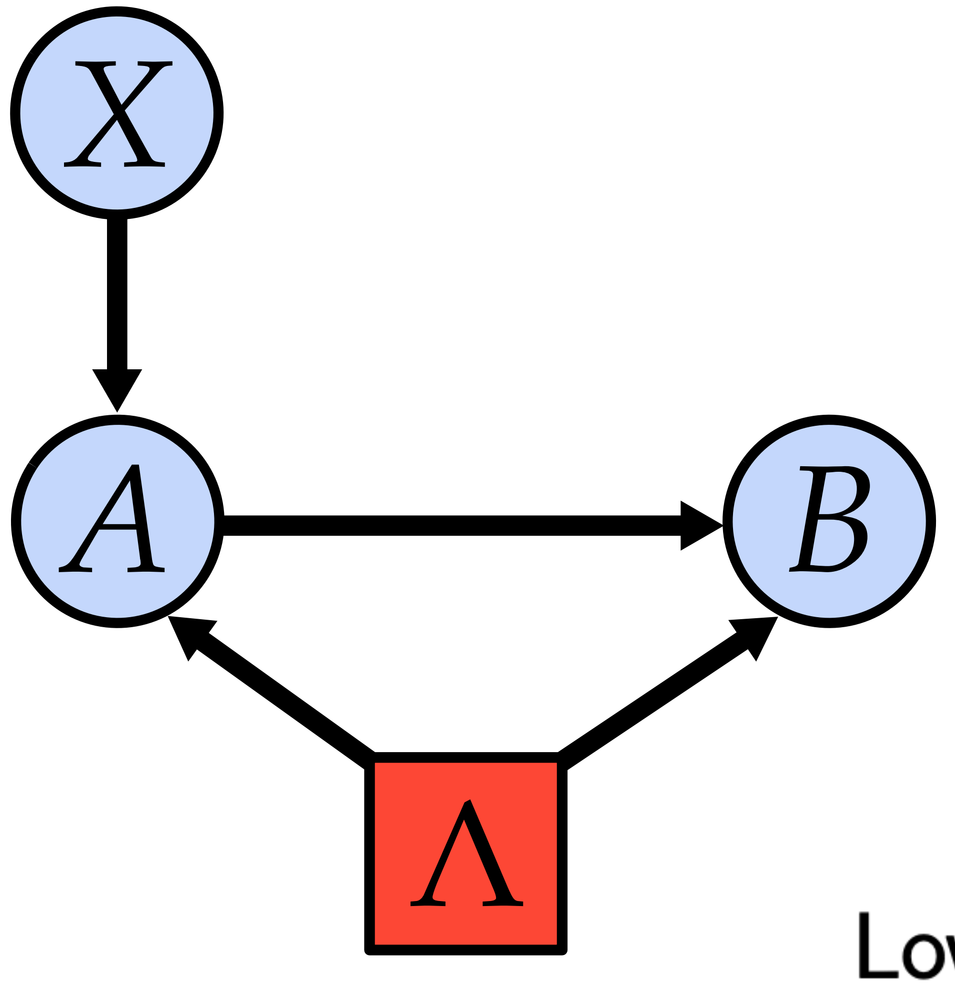
$$\mathrm{ACE}_{A\to B} \geq 2p(0,0|0) + p(1,1|0) + p(0,1|1) + p(1,1|1) - 2, \quad (3)$$

where

$$\mathrm{ACE}_{A\to B} = \max_{a,a',b} \Big(p(b|do(a)) - p(b|do(a')) \Big).$$

[2] A. Balke and J. Pearl, "Bounds on treatment effects from studies with imperfect compliance," Journal of the American Statistical Association 92, 1171–1176 (1997).

Simplest Instrumental Scenario



Example: A – smoking, B – cancer, Λ – genetics, X – taxation of tobacco.

$x, a, b \in \{0, 1\}$

$p(a, b|x)$

Lower bound[2]:

$$\mathrm{ACE}_{A\to B} \geq 2p(0,0|0) + p(1,1|0) + p(0,1|1) + p(1,1|1) - 2, \tag{3}$$

where

$$\mathrm{ACE}_{A\to B} = \max_{a,a',b} \left(p(b|do(a)) - p(b|do(a')) \right).$$

[2] A. Balke and J. Pearl, "Bounds on treatment effects from studies with imperfect compliance," Journal of the American Statistical Association 92, 1171–1176 (1997).

MORE ACM AWARDS

A.M. TURING AWARD

A.M. TURING AWARD LAUREATES BY...

ALPHABETICAL LISTING | YEAR OF THE AWARD | RESEARCH SUBJECT

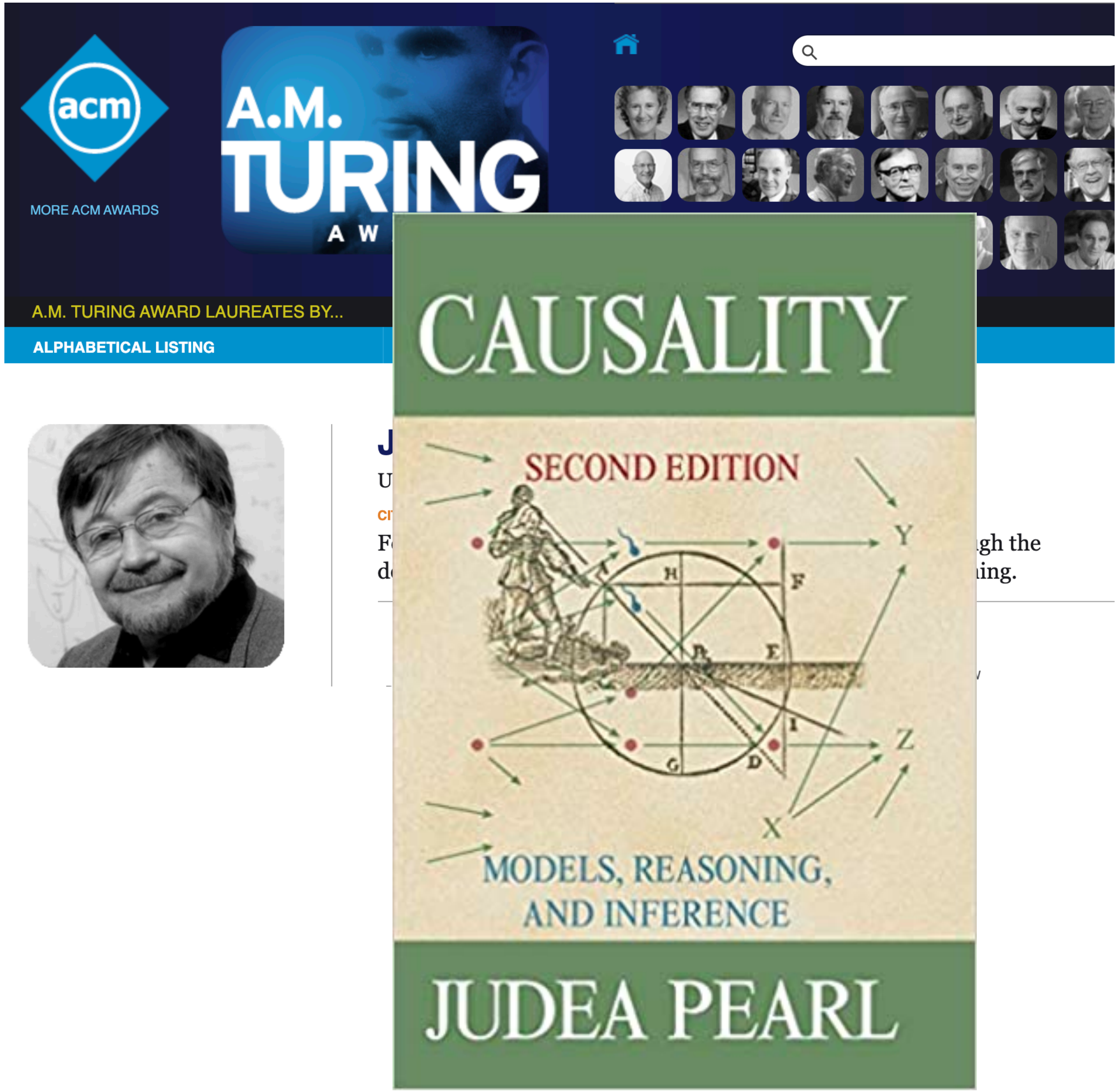
JUDEA PEARL

United States – 2011

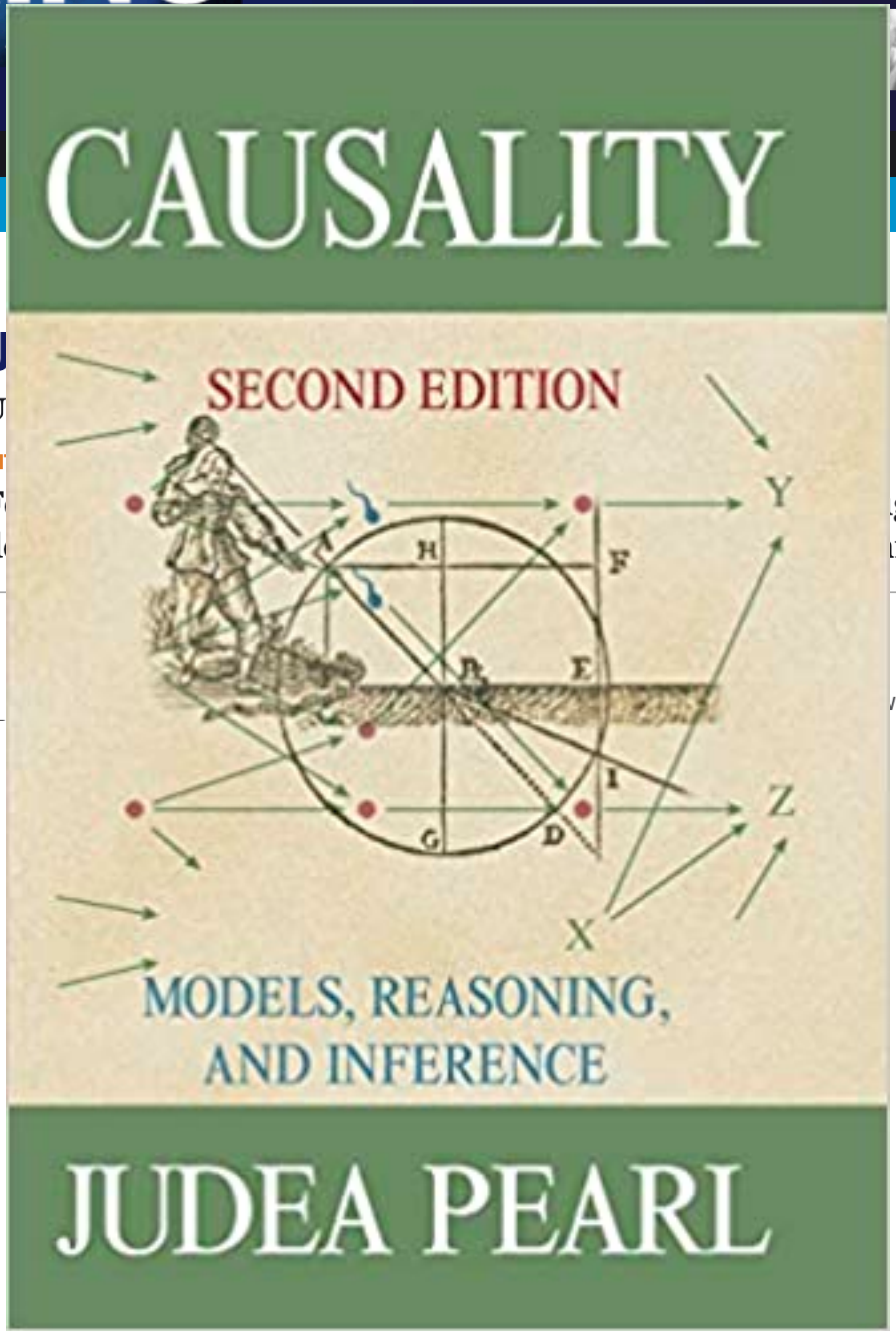
CITATION

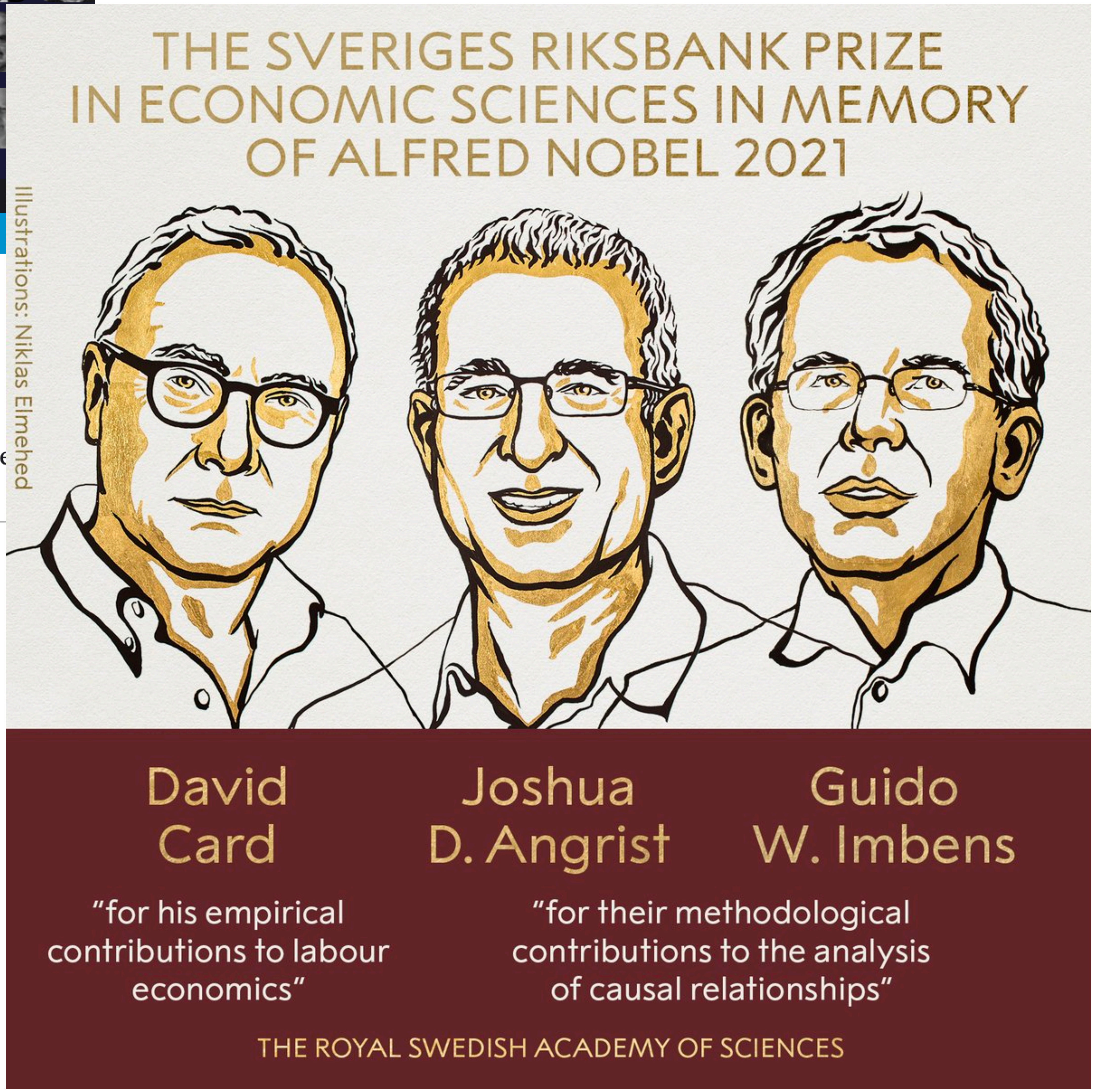
For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.

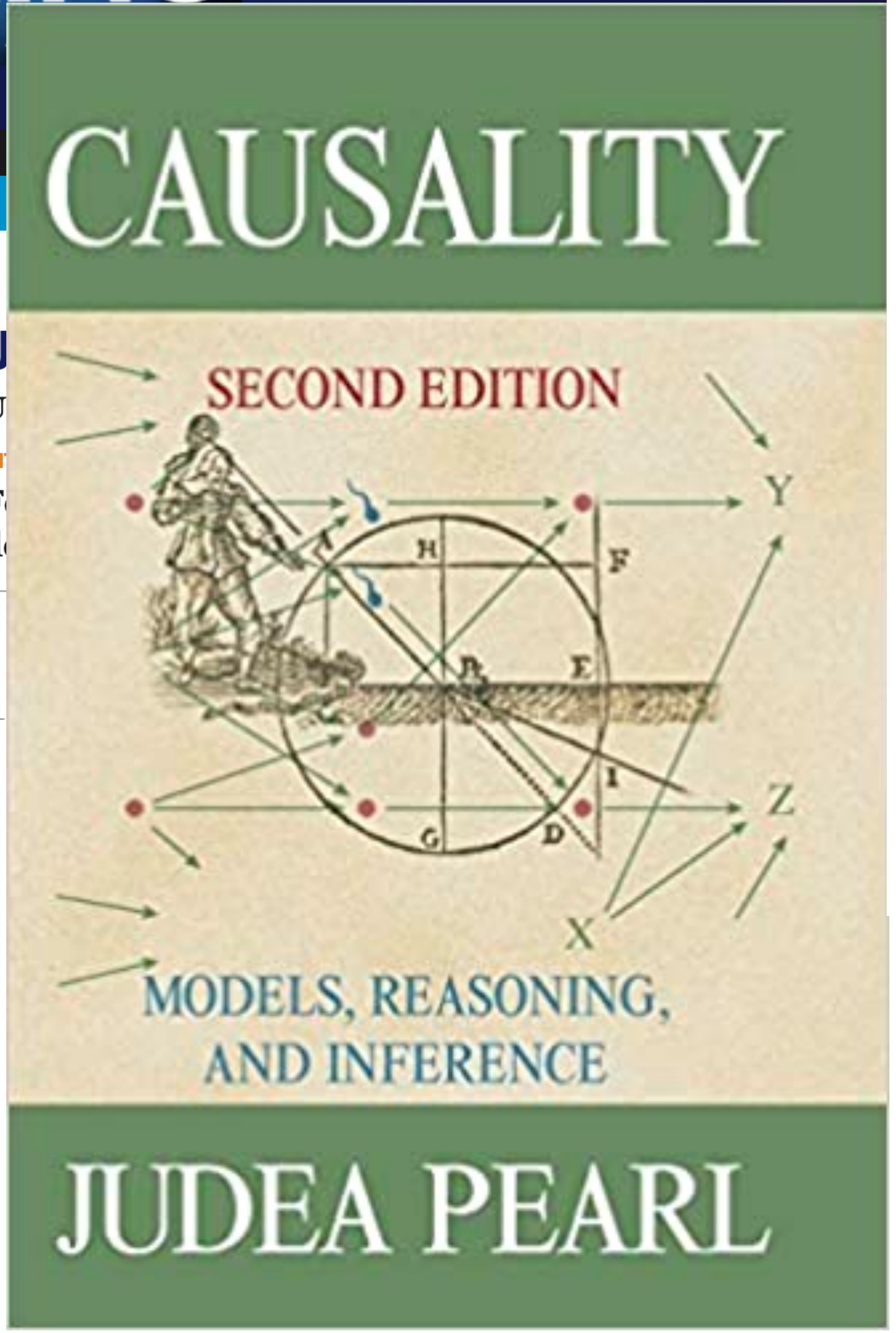
SHORT ANNOTATED | ACM TURING AWARD | RESEARCH SUBJECTS | ADDITIONAL MATERIALS | VIDEO INTERVIEW


acm
MORE ACM AWARDS
A.M. TURING
A.M. TURING AWARD LAUREATES BY...
ALPHABETICAL LISTING
CAUSALITY
SECOND EDITION
MODELS, REASONING, AND INFERENCE
JUDEA PEARL


acm
MORE ACM AWARDS
A.M.
TURING
A.M. TURING AWARD LAUREATES BY...
ALPHABETICAL LISTING


CAUSALITY
SECOND EDITION
MODELS, REASONING,
AND INFERENCE
JUDEA PEARL


THE SVERIGES RIKSBANK PRIZE
IN ECONOMIC SCIENCES IN MEMORY
OF ALFRED NOBEL 2021
Illustrations: Niklas Elmehed
David
Card
Joshua
D. Angrist
Guido
W. Imbens
"for his empirical contributions to labour economics"
"for their methodological contributions to the analysis of causal relationships"
THE ROYAL SWEDISH ACADEMY OF SCIENCES



Josh Angrist

MIT
Verified email at mit.edu
Labor Economics Econometrics

FOLLOW

TITLE	CITED BY	YEAR
Mostly harmless econometrics JD Angrist, JS Pischke Mostly Harmless Econometrics	18941	2008
Identification of causal effects using instrumental variables JD Angrist, GW Imbens, DB Rubin Journal of the American statistical Association 91 (434), 444-455	6450	1996
Identification and estimation of local average treatment effects J Angrist, G Imbens National Bureau of Economic Research	5437	1995
Does compulsory school attendance affect schooling and earnings? JD Angrist, AB Keueger The Quarterly Journal of Economics 106 (4), 979-1014	3321	1991
Instrumental variables and the search for identification: From supply and demand to natural experiments JD Angrist, AB Krueger	2804	2001

David Card — "for his empirical contributions to labour economics"

Joshua D. Angrist, Guido W. Imbens — "for their methodological contributions to the analysis of causal relationships"

THE ROYAL SWEDISH ACADEMY OF SCIENCES

Conclusions

Conclusions

Part 1:

Simpson's paradox and its resolution using the back-door criteria

Conclusions

Part 1:

Simpson's paradox and its resolution using the back-door criteria

Part 2:

Observational Causal Inference

Average causal effect and its estimates

Things that we skipped (maybe next week)

- How to derive the bounds?
- Bottlenecks of the instrumental scenario
- And how to treat them?

Things that we skipped (maybe next week)

- How to derive the bounds?
- Bottlenecks of the instrumental scenario
- And how to treat them?