THE LOGIC OF CAUSATION

COUNTERFACTUAL THINKING

LEVI JOHN WOLF

levi.john.wolf@bristol.ac.uk
calendly.com/ljwolf
@levijohnwolf
ljwolf.org

WHAT IS A COUNTERFACTUAL?

WHO CARES?

STRATEGIES

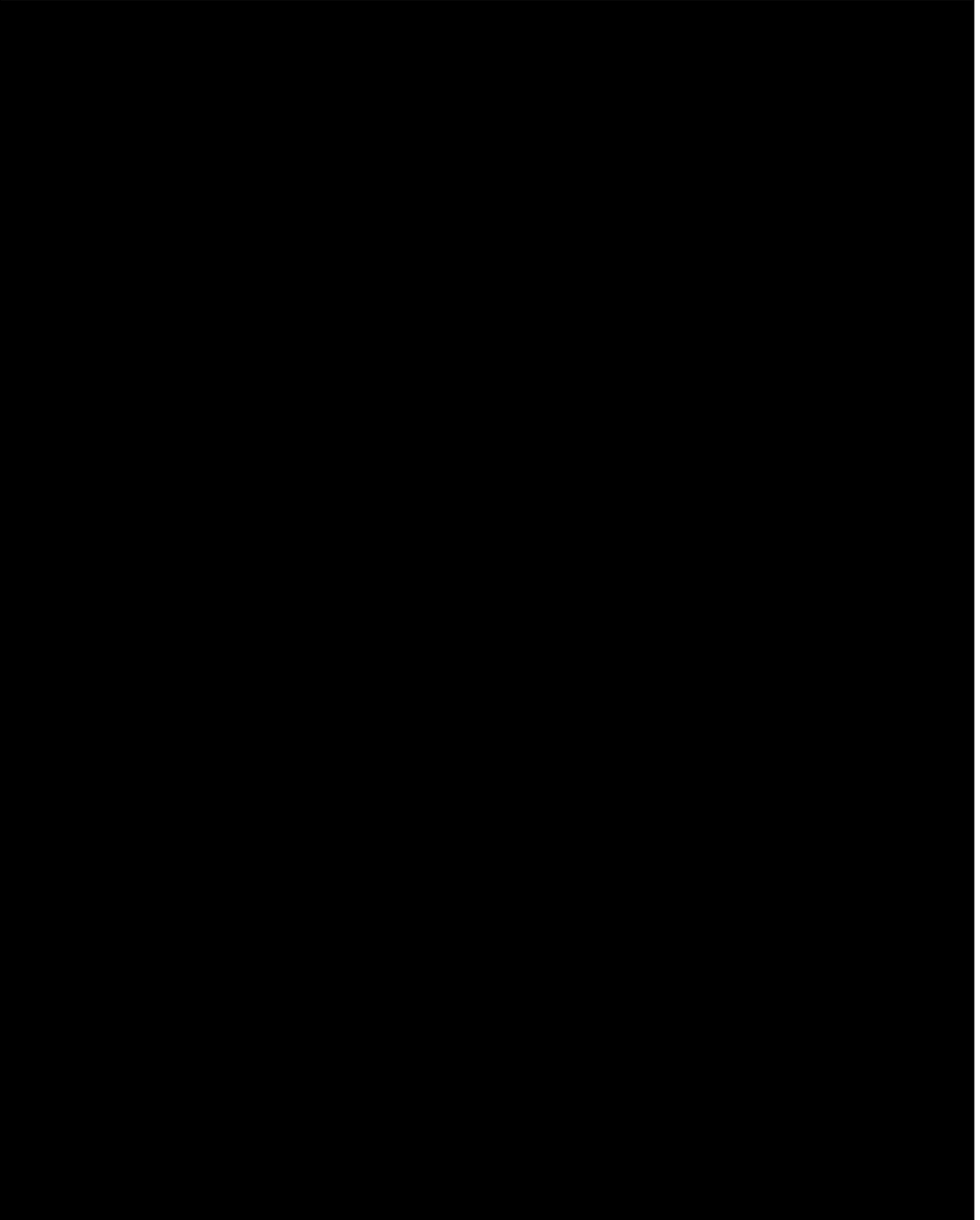
WHAT IS A COUNTERFACTUAL?

They are “what-if X” scenarios & imply “but-for X” causation.

WHO CARES?

STRATEGIES

Judea Pearl
& Dana Mackenzie

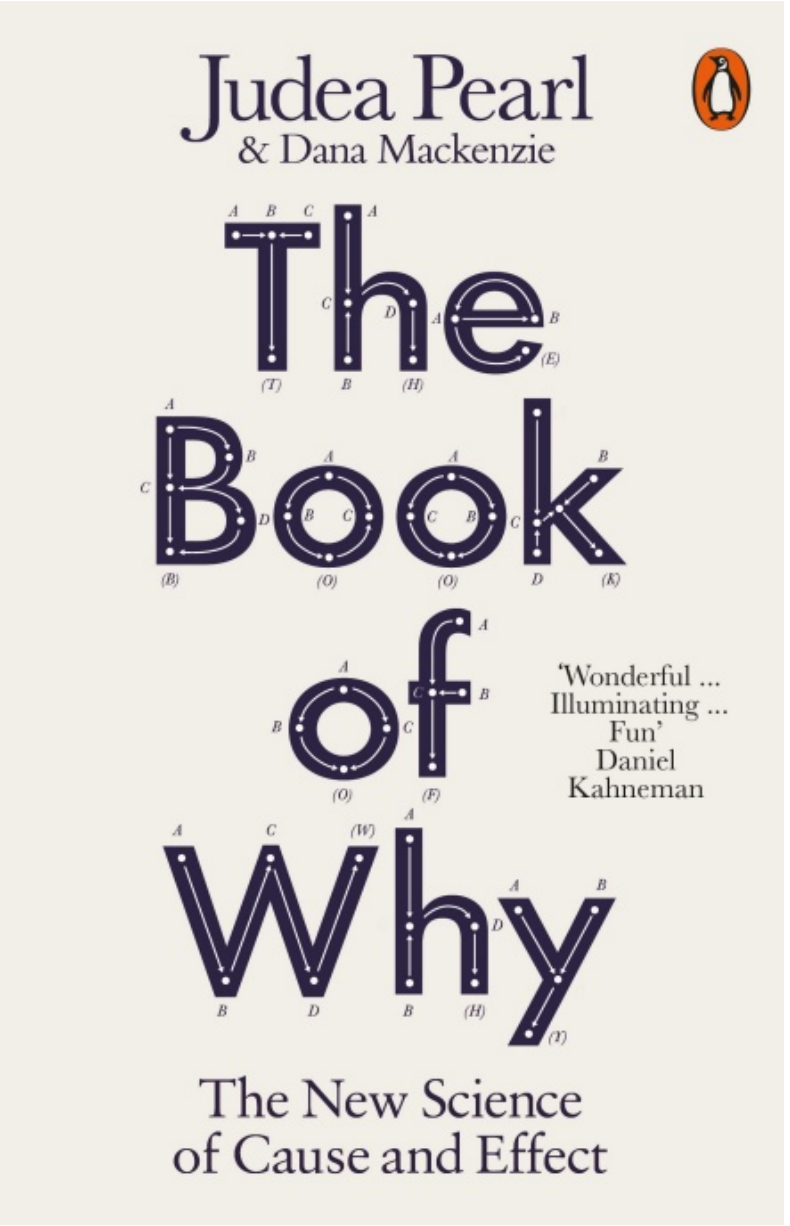


The Book of Why

'Wonderful ... Illuminating ... Fun'
Daniel Kahneman

The New Science of Cause and Effect

"In [Geog methods 1], every student learns to chant: "correlation is not causation." With good reason! [Unfortunately,] this tells us that correlation is not causation, but it does not tell us what causation *is*.



COUNTERFACTUAL

/ˌkaʊntəˈfaktʃʊəl/

characterizing what happens in a scenario that is not observed directly.

COUNTERFACTUAL (STATEMENT)

/ˌkaʊntəˈfaktʃʊəl/

characterizing what happens in a scenario that is not observed directly.

COUNTERFACTUAL (SCENARIO)

/ˌkaʊntəˈfaktʃʊəl/

an alternative present where the past has been changed.

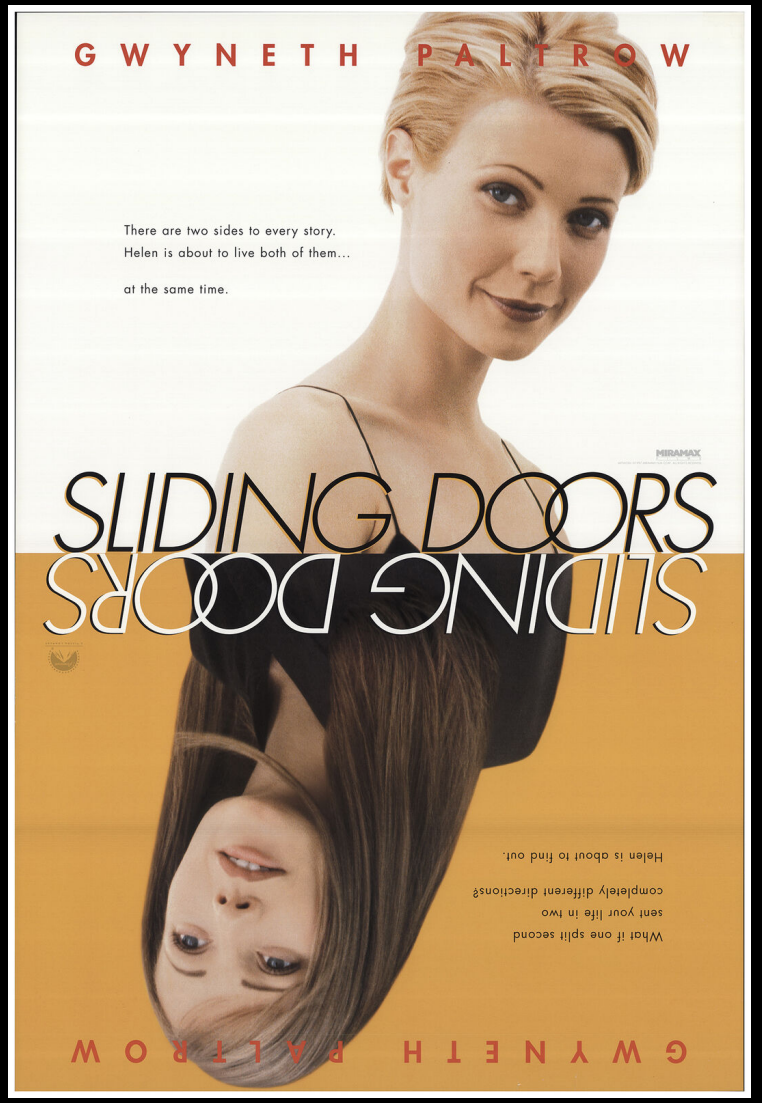
COUNTERFACTUAL (SCENARIO)?

/ˌkaʊntəˈfaktʃʊəl/

IF THINGS HAD GONE DIFFERENTLY, WOULD THE OUTCOME HAVE BEEN DIFFERENT?

SLIDING DOORS

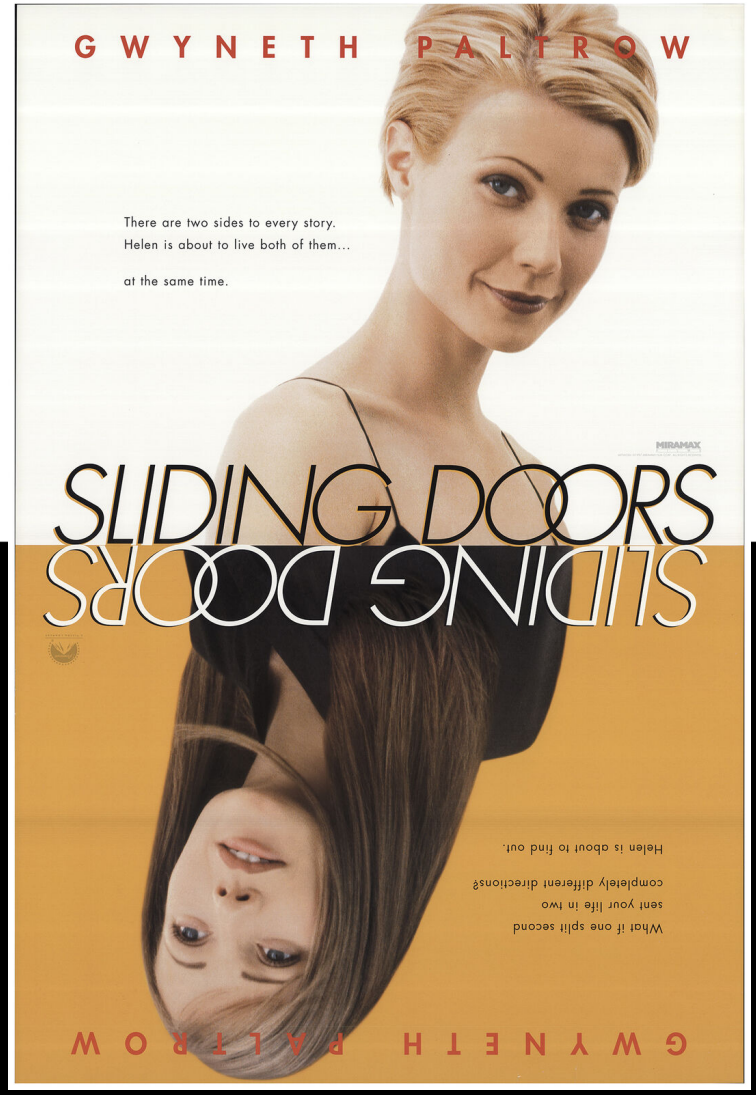
Gwyneth tries to catch the tube at Waterloo. In one timeline she fails, and in the other she succeeds. When she succeeds, she dies. When she fails, she lives.



We observe an outcome y

And some conditions x

We'd like to make statements about how likely it is that y happens, given that we saw x



In one timeline she fails, and in the other she succeeds. When she succeeds, she dies. When she fails, she lives.

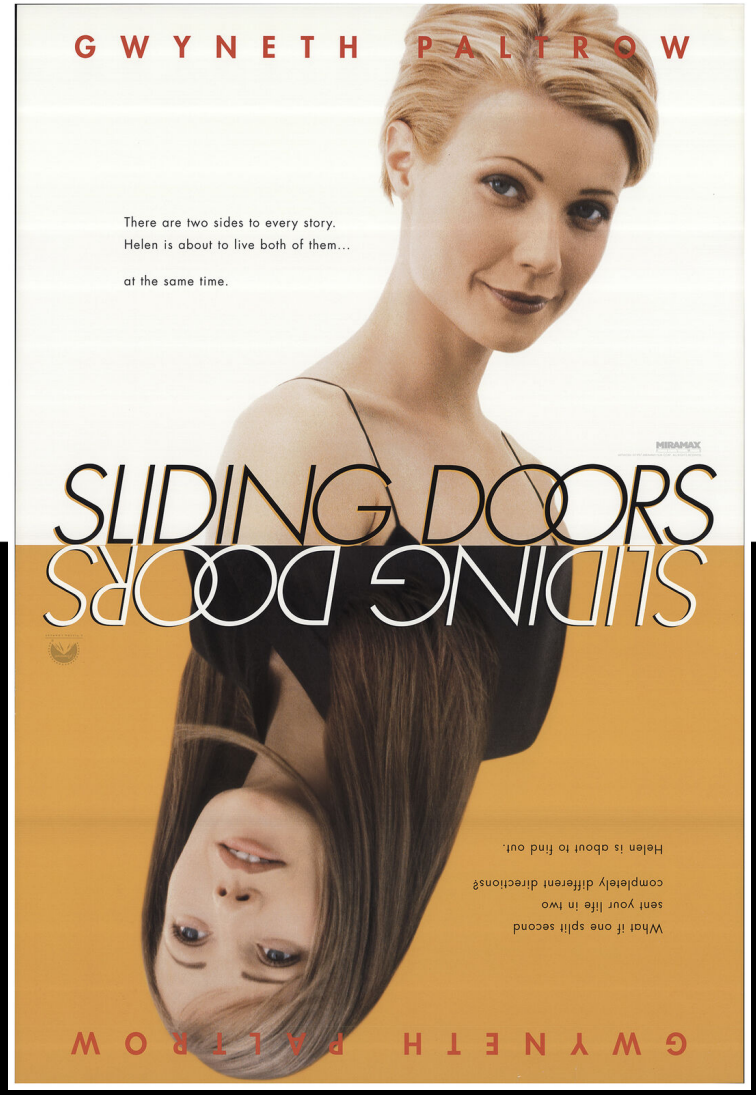
We observe an outcome y

And some conditions x

We'd like to make statements about

$$p(y \mid x)$$

In one timeline she fails, and in the other she succeeds. When she succeeds, she dies. When she fails, she lives.

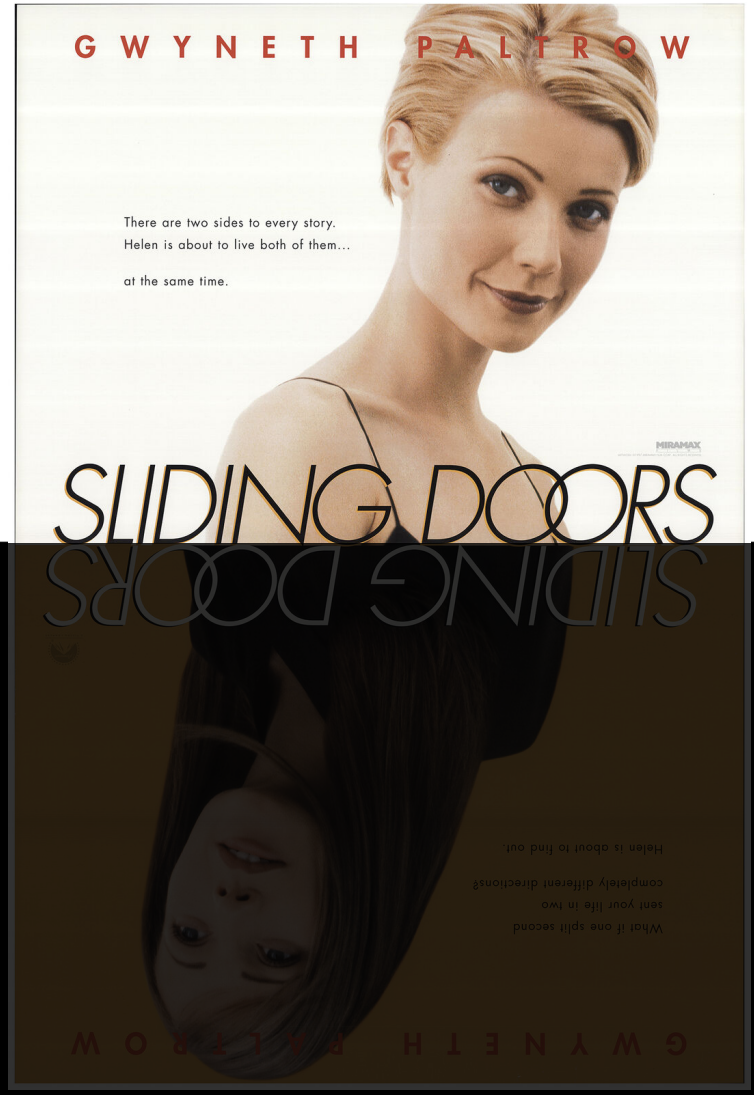


We observe an outcome y

And some conditions x

We'd like to make statements about

$$p(y \mid x)$$



In one timeline she fails, and in the other she succeeds. When she succeeds, she dies. When she fails, she lives.

We observe an outcome y

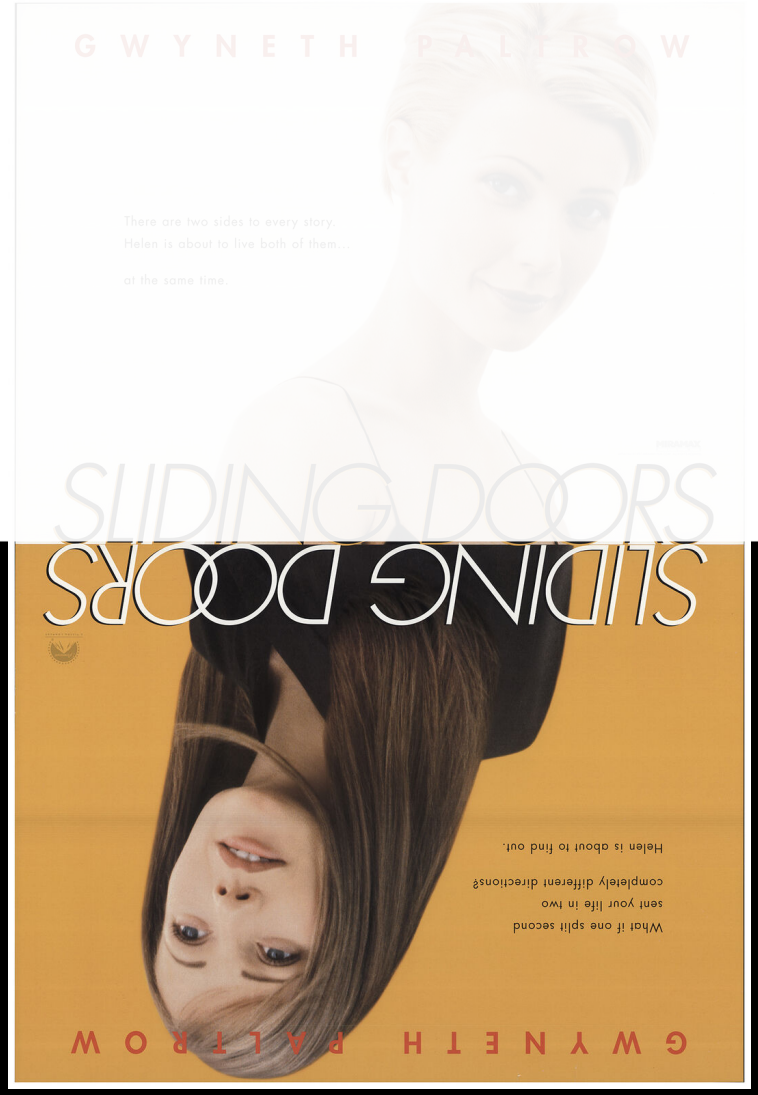
And some conditions x

We'd like to make statements about

$$p(y' \mid x')$$

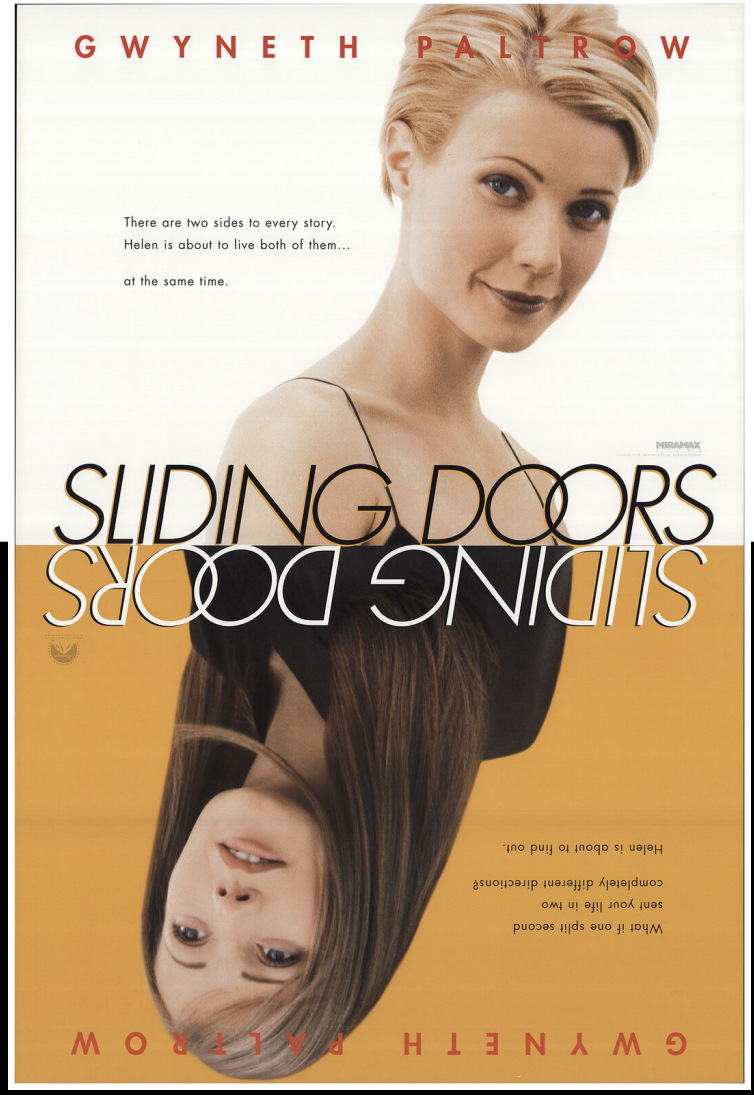
In one timeline she fails, and in the other she succeeds. When she succeeds, she dies.

When she fails, she lives.



NEITHER GWEN CAN KNOW HOW THE OTHER GWEN TURNED OUT

In one timeline she fails, and in the other she succeeds. When she succeeds, she dies. When she fails, she lives.



We observe an outcome y

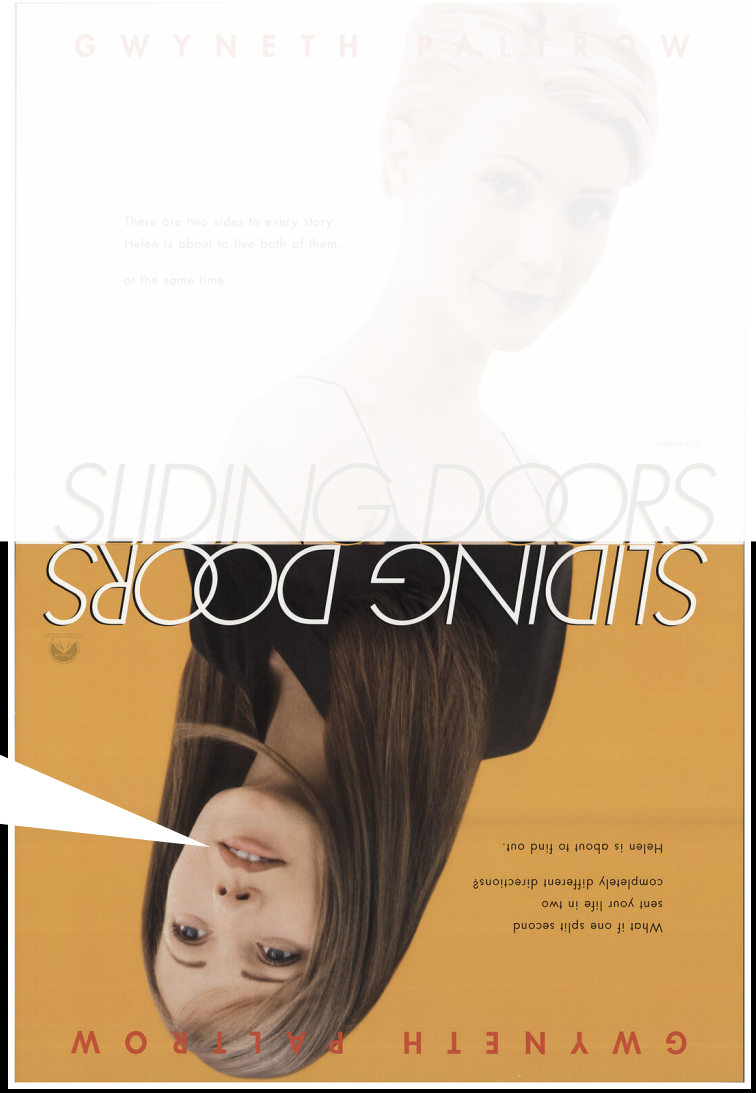
And some conditions x

We'd like to make statements about

$$p(y_x \mid x', y')$$

Would I be alive if I had caught the tube that day?

Given I didn't catch the tube and am definitely alive!



WHAT IS A COUNTERFACTUAL?

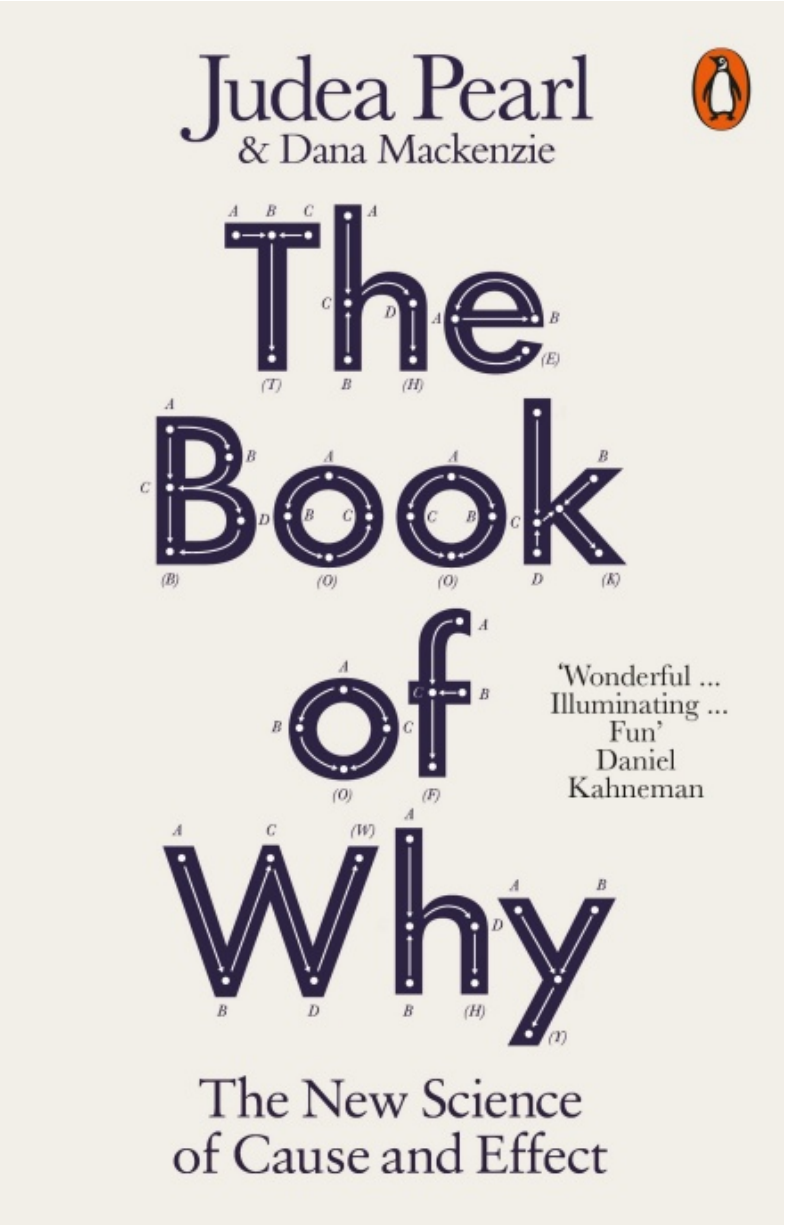
They are “what-if X” scenarios & imply “but-for X” causation.

WHO CARES?

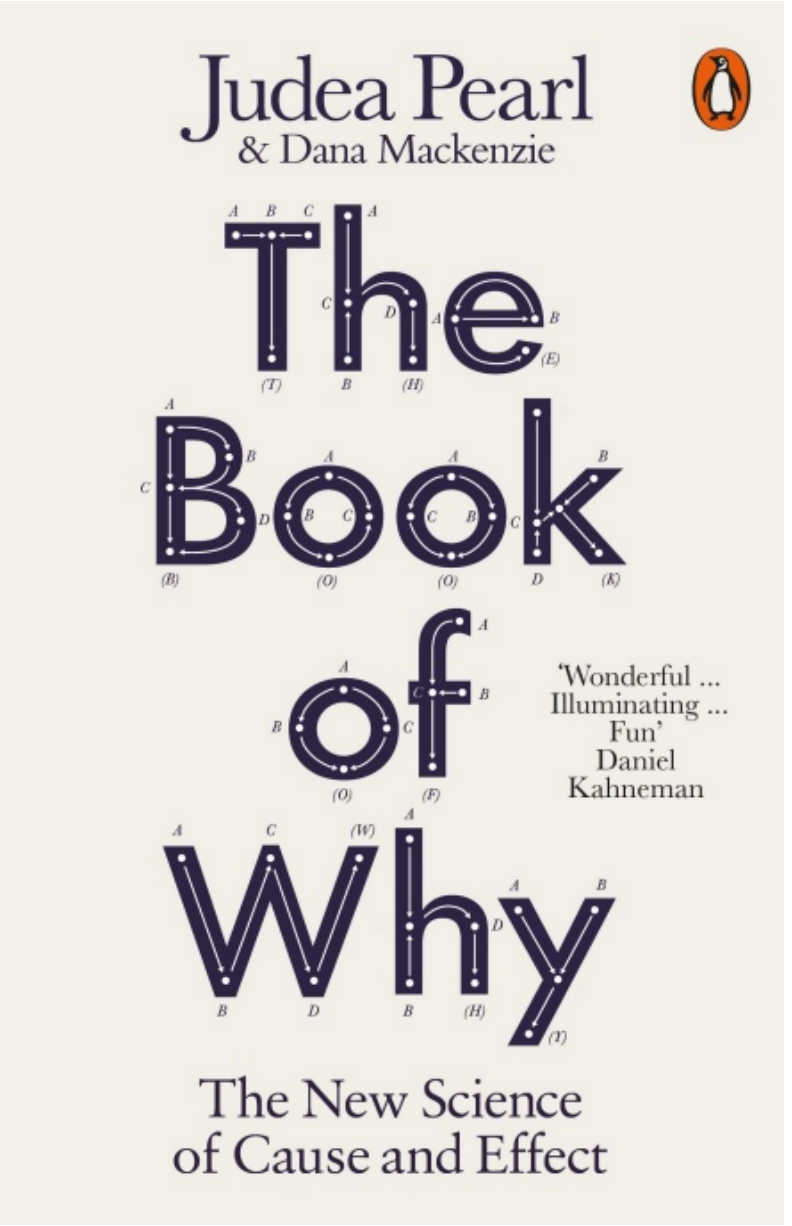
You should! They are fundamental to human reasoning.

STRATEGIES

"It's one thing to say 'smoking causes cancer' but another to say that my uncle Joe, who smoked a pack a day for thirty years, would have been alive had he not smoked. The difference is both obvious and profound: none of the people who, like Uncle Joe, smoked for thirty years and died can ever be observed in the alternate world where they did not smoke for thirty years."



"It's one thing to say 'smoking causes cancer' but another to say that my uncle Joe, who smoked a pack a day for thirty years, would have been alive had he not smoked. The difference is both obvious and profound: none of the people who, like Uncle Joe, smoked for thirty years and died can ever be observed in the alternate world where they did not smoke for thirty years."



HOW CAN WE "KNOW" THINGS?

HOW CAN WE "KNOW" THINGS?

SEEING Observed relationship between effect and cause

Is Gwen alive when she misses her train?

Do unequal societies have cities with really different sizes?

HOW CAN WE "KNOW" THINGS?

SEEING $p(y \mid x)$

Is Gwen alive when she misses her train?

Do unequal societies have cities with really different sizes?

HOW CAN WE "KNOW" THINGS?

SEEING $p(y \mid x)$

Is Gwen alive when she misses her train?

Do unequal societies have cities with really different sizes?

DOING Predictive relationship between effect and cause

If Gwen catches the train, will she live?

If we reduce inequality in society, will city size converge?

HOW CAN WE "KNOW" THINGS?

SEEING $p(y \mid x)$

Is Gwen alive when she misses her train?

Do unequal societies have cities with really different sizes?

DOING $p(y \mid \mathbf{do}(x))$

If Gwen catches the train, will she live?

If we reduce inequality in society, will city size converge?

HOW CAN WE "KNOW" THINGS?

SEEING $p(y \mid x)$

Is Gwen alive when she misses her train?

Do unequal societies have cities with really different sizes?

DOING $p(y \mid \mathbf{do}(x))$

If Gwen catches the train, will she live?

If we reduce inequality in society, will city size converge?

IMAGINING Predictive relationship between alternative cause/effect

If Gwen had caught the train, would she be alive today?

If society had not been so unequal, would city size converge?

THE LADDER OF CAUSATION — PEARL (2016)

HOW CAN WE "KNOW" THINGS?

SEEING $p(y \mid x)$

Is Gwen alive when she misses her train?

Do unequal societies have cities with really different sizes?

DOING $p(y \mid \mathbf{do}(x))$

If Gwen catches the train, will she live?

If we reduce inequality in society, will city size converge?

IMAGINING $p(y_x \mid x', y')$

If Gwen had caught the train, would she be alive today?

If society had not been so unequal, would city size converge?

HOW CAN WE "KNOW" THINGS?

SEEING $p(y \mid x)$

Is Gwen alive when she misses her train?

Do unequal societies have cities with really different sizes?

DOING $p(y \mid \mathbf{do}(x))$

If Gwen catches the train, will she live?

If we reduce inequality in society, will city size converge?

IMAGINING $p(y_x \mid x', y')$

If Gwen had caught the train, would she be alive today?

If society had not been so unequal, would city size converge?

WHY FOCUS ON IMAGINING?

IMAGINING $p(y_x \mid x', y')$

If Gwen had caught the train, would she be alive today?

If society had not been so unequal, would city size converge?

WHY FOCUS ON IMAGINING?

"If we have a model that can answer counterfactual questions, we can also answer questions about interventions and observations."

IMAGINING $p(y_x \mid x', y')$

If Gwen had caught the train, would she be alive today?

If society had not been so unequal, would city size converge?

"If we have a model that can answer counterfactual questions, we can also answer questions about interventions and observations."

$p(y_x \mid x', y')$

We saw: Gwen missed the train, she's alive.
We model: If Gwen had caught the train, would she have lived?

$p(y \mid \mathbf{do}(x))$

We saw: Gwen missed the train, she's alive.
We model: If Gwen catches the train from now on, will she live?

"If we have a model that can answer counterfactual questions, we can also answer questions about interventions and observations."

$p(y_x \mid x', y')$

We saw: Gwen missed the train, she's alive.
We model: If Gwen had caught the train, would she have lived?

$p(y \mid \mathbf{do}(x))$

We saw: Gwen missed the train, she's alive.
We model: If Gwen catches the train from now on, will she live?

If we know the counterfactual, we can answer the intervention:

When we set conditions for the counterfactual, we **set them to right now!**

"If we have a model that can answer counterfactual questions, we can also answer questions about interventions and observations."

$p(y_x \mid x', y')$

We saw: Gwen missed the train, she's alive.
We model: If Gwen had caught the train, would she have lived?

$p(y \mid x)$

We saw: Gwen missed the train, she's alive.
We model: When Gwen caught the train, did she live?

"If we have a model that can answer counterfactual questions, we can also answer questions about interventions and observations."

$p(y_x \mid x', y')$

We saw: Gwen missed the train, she's alive.
We model: If Gwen had caught the train, would she have lived?

$p(y \mid x)$

We saw: Gwen missed the train, she's alive.
We model: When Gwen caught the train, did she live?

If we know the counterfactual, we can answer the observation:
The observational conditions are the **grounds for the counterfactual!**

WHAT IS A COUNTERFACTUAL?

They are “what-if X” scenarios & imply “but-for X” causation.

WHO CARES?

You should! They are fundamental to human reasoning.

STRATEGIES

Causality is about theory, so get theoretical before empirical.

HEURISTIC

Use counterfactual thinking to make assumptions about processes clear.

ANALYTIC

Use counterfactual analysis to do causal inference.

HEURISTIC

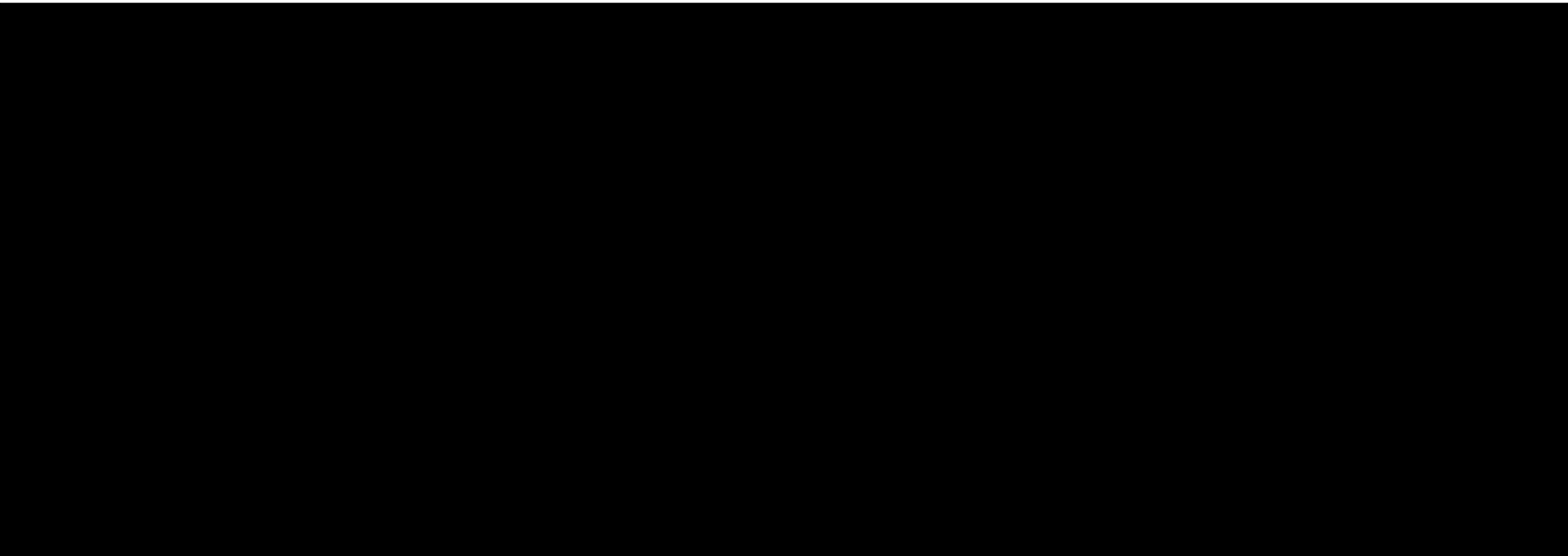
Use counterfactual thinking to make assumptions about processes clear.

ANALYTIC

Use counterfactual analysis to do causal inference.

HEURISTIC

Asthma is caused by exposure to air pollution.

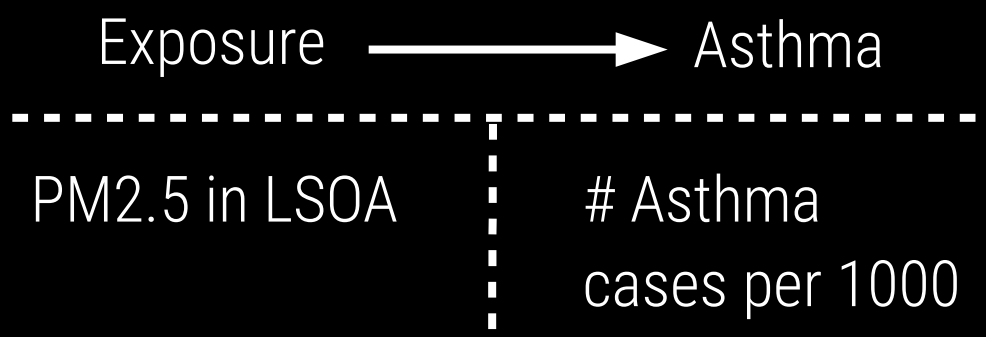


HEURISTIC

Asthma is caused by exposure to air pollution.

HEURISTIC

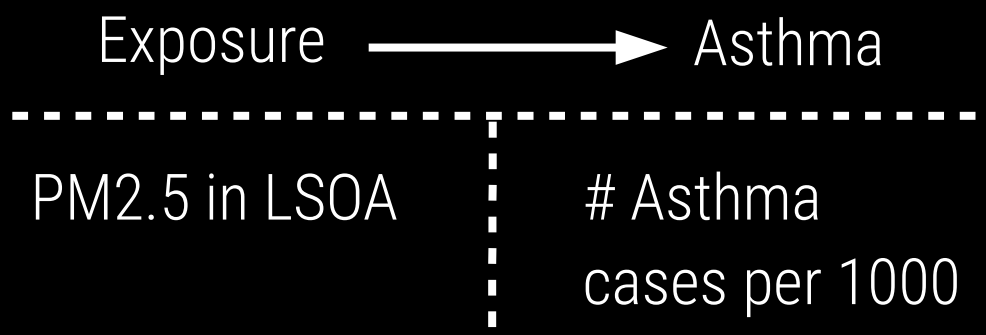
Asthma is caused by exposure to air pollution.



IF THIS IS TRUE

ARE THERE ANY PLACES
WITH IDENTICAL EXPOSURE
BUT DIFFERENT ASTHMA?
WHY WOULD THEY DIFFER?

Asthma is caused by exposure to air pollution.



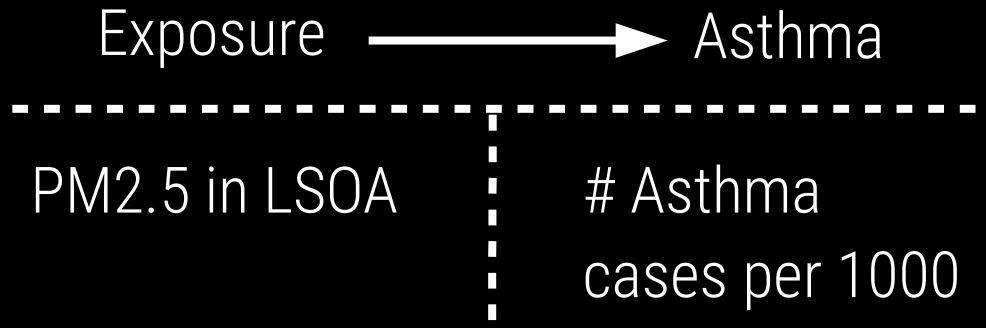
IF THIS IS TRUE

ARE THERE ANY PLACES
WITH IDENTICAL EXPOSURE
BUT DIFFERENT ASTHMA?
WHY WOULD THEY DIFFER?

Asthma is caused by exposure to air pollution.

PM2.5 may not be exposure!

- Ozone, PM10, NO_x



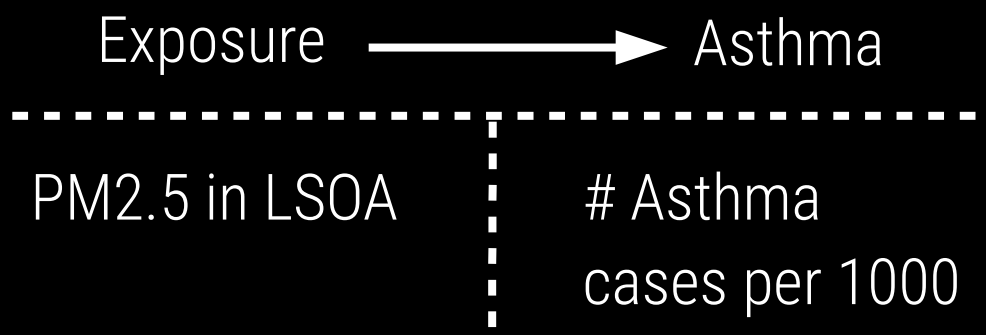
IF THIS IS TRUE

ARE THERE ANY PLACES
WITH IDENTICAL EXPOSURE
BUT DIFFERENT ASTHMA?
WHY WOULD THEY DIFFER?

Asthma is caused by exposure to air pollution.

LSOA may not be exposure!

- a cyclist commuting daily from a "clean" location is exposed on their ride daily. And, the exposure is intense.
- a builder on a construction site is exposed at their workplace, independently of env. pollutants



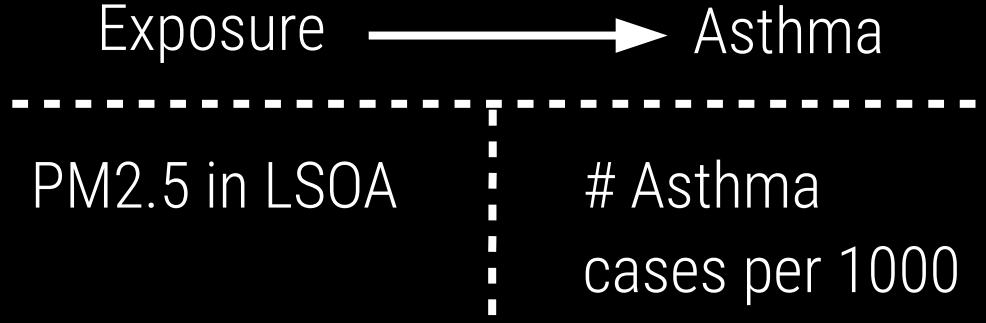
IF THIS IS TRUE

ARE THERE ANY PLACES
WITH IDENTICAL EXPOSURE
BUT DIFFERENT ASTHMA?
WHY WOULD THEY DIFFER?

Asthma is caused by exposure to air pollution.

Effect of exposure might vary!

- Rich LSOAs might have “protection”



HEURISTIC

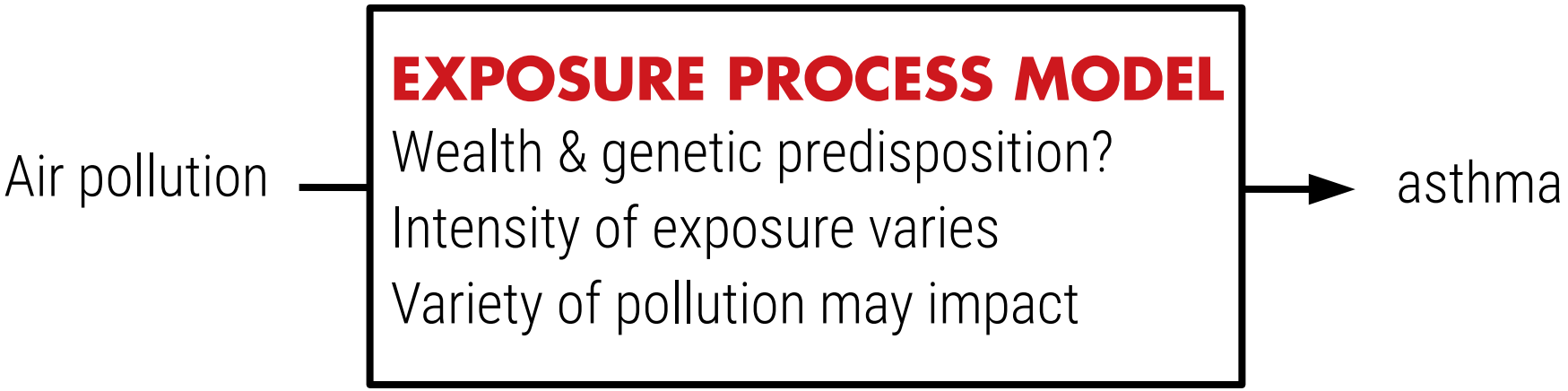
IF THIS THEORY IS TRUE,
What would prove me wrong?

ANALYTIC

Use counterfactual analysis
to do causal inference.

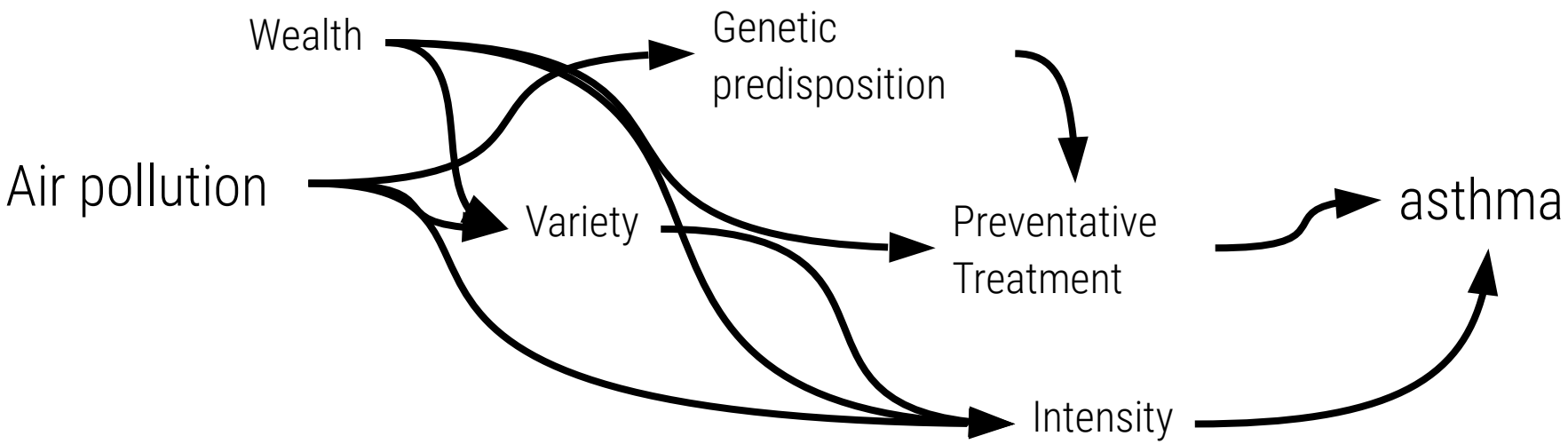
ANALYTIC

Use counterfactual analysis to do causal inference.



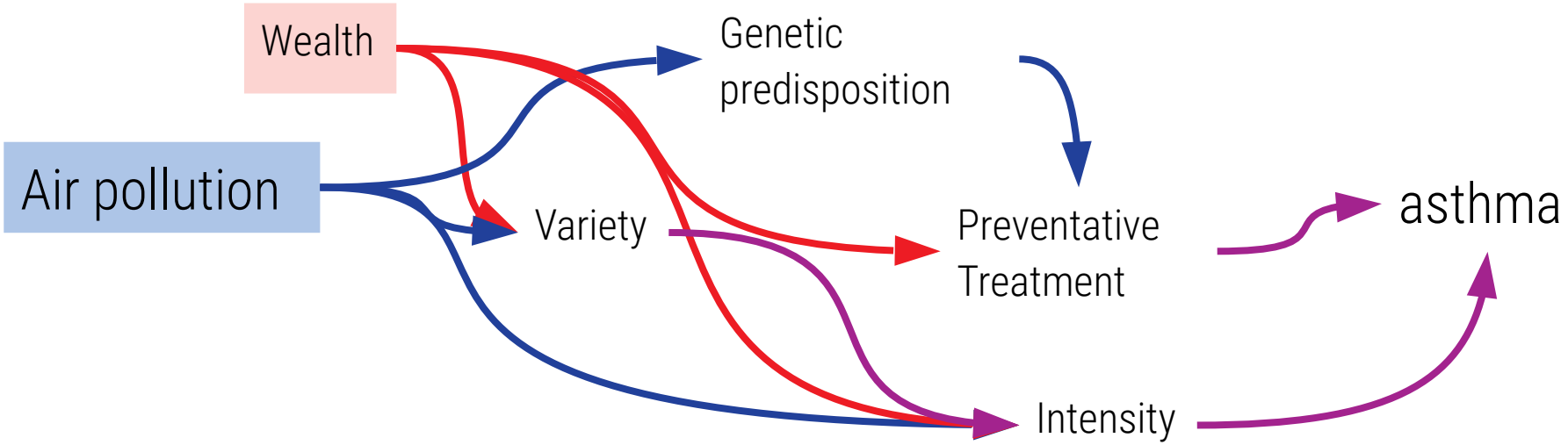
ANALYTIC

Use counterfactual analysis to do causal inference.



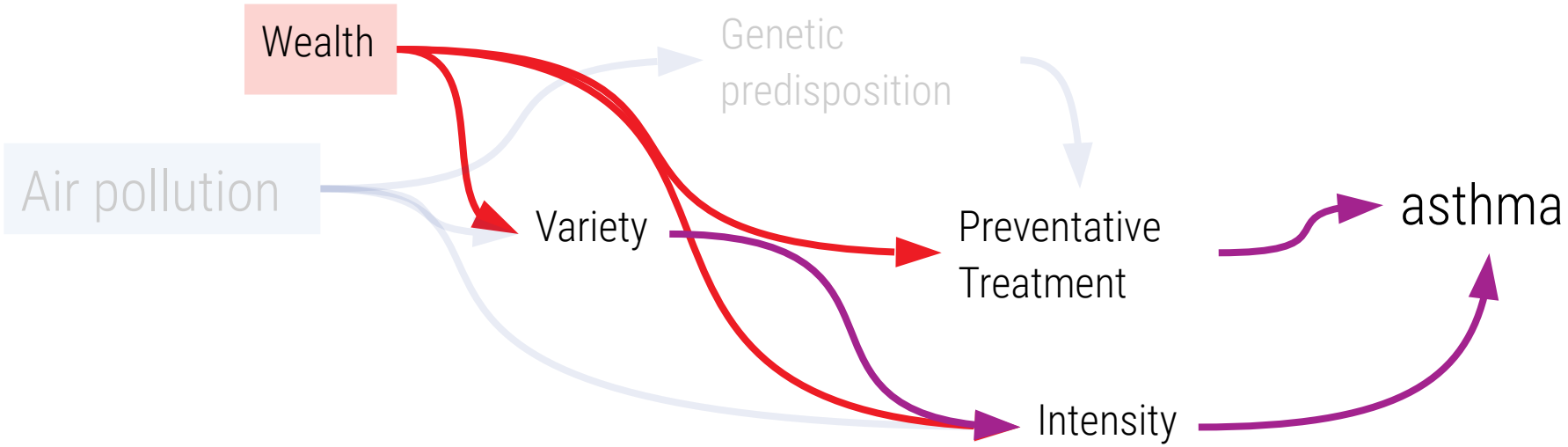
ANALYTIC

Use counterfactual analysis to do causal inference.



ANALYTIC

Use counterfactual analysis to do causal inference.



Changing wealth affects *more than just asthma outcomes!*

If individual A were £20kpa richer, she might get treatment, but also might change her commute or residence! **Regression misses this!**

ANALYTIC

Use counterfactual analysis to do causal inference.

HEURISTIC

IF THIS THEORY IS TRUE,
What would prove me wrong?

ANALYTIC

DIRECT & INDIRECT
Do causes interact/interfere?

WHAT IS A COUNTERFACTUAL?

They are “what-if X” scenarios & imply “but-for X” causation.

WHO CARES?

You should! They are fundamental to human reasoning.

STRATEGIES

Informal heuristics & formal analysis can use counterfactuals!

QUESTIONS ABOUT

THE LOGIC OF CAUSATION

COUNTERFACTUAL THINKING

LEVI JOHN WOLF

levi.john.wolf@bristol.ac.uk
calendly.com/ljwolf
@levijohnwolf
ljwolf.org