# THE SCIENCE OF CAUSE AND EFFECT: FROM DEEP LEARNING TO DEEP UNDERSTANDING

Judea Pearl
UCLA
@yudapearl

University of California Irvine
February 25, 2022

1

## OUTLINE

1. The scientific paradigm
   - The two fundamental laws of CI
   - The ladder of Causation
   - *Do*-calculus: The Algebra of Interventions
2. What the science adds to Machine Learning
   - Combining data with prior causal knowledge
   - Seven Pillars of Causal Wisdom
   - Future horizons
     - Personalized Decision Making
     - Social Intelligence

2

2

## DATA SCIENCE – A CLASH OF TWO PARADIGMS

1. The scientific paradigm
   - What should the world be like before I can answer my research question?
2. The data-centric paradigm
   - How best to fit the data so as to maximize success on the training set.

3

3

## WHAT CAPABILITIES DOES DEEP UNDERSTANDING REPRESENT?

A state of knowledge evoking a sensation of "understanding" or "being in control."

1. Predict future events from past/present observations
2. Predict consequence of contemplated actions
3. Provide explanations of unanticipated events
4. Imagine alternative worlds or "Roads not Taken"
5. Design new experiments, seek new observations (attention, curiosity, and conjectures)

4

## TYPICAL CAUSAL QUESTIONS

1. How effective is a given treatment in preventing a disease?
2. Was it the new tax break that caused our sales to go up? Or our marketing campaign?
3. What is the annual health-care costs attributed to obesity?
4. Can hiring records prove an employer guilty of sex discrimination?
5. I am about to quit my job, will I regret it?
   - Unarticulatable in the standard grammar of science.

$Y = aX$ vs. $Y \leftarrow aX$

5

## SEWALL WRIGHT – CAUSALITY'S FIRST FORMAL VOICE

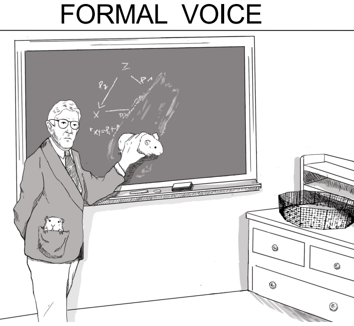


Figure 2.6. Sewall Wright was the first person to develop a mathematical method for answering causal questions from data, known as path diagrams. His love of mathematics surrendered only to his passion for guinea pigs.

6

## SEWALL WRIGHT – CAUSALITY'S FIRST FORMAL VOICE

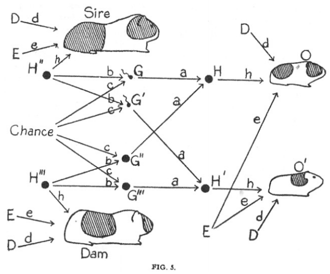



Figure 2.7. Sewall Wright's first path diagram, illustrating the factors leading to coat color in guinea pigs. D = developmental factors (after conception, before birth), E = environmental factors (after birth), G = genetic factors from each individual parent, H = combined hereditary factors from both parents, O, O¢ = offspring. The objective of analysis was to estimate the strength of the effects of D, E, H (written as d, e, h in the diagram). (Source: Sewall Wright, Proceedings of the National Academy of Sciences [1920], 320–332.)

7

## FROM SEWALL WRIGHT TO MODERN CAUSAL MODELS

SCM – Structural Causal Model M
1. The oracle of all causal statements
2. Rooted in human intuition

The world as a society of listeners:

$$v_i := f_i(v_1, v_2, \ldots, v_n; u_i)$$

:= Assignment operator, non algebraic.
$U_i$ – exogenous factors,
$G$ – Causal Graph, an abstraction of $M$

8

## THE TWO FUNDAMENTAL LAWS OF CAUSAL INFERENCE

1. The Law of Counterfactuals (and Interventions)

$$\boxed{Y_x(u) = Y_{M_x}(u)}$$

($M$ generates and evaluates all counterfactuals.)

2. The Law of Conditional Independence ($d$-separation)

$$\boxed{(X \text{ sep } Y \mid Z)_{G(M)} \Rightarrow (X \perp\!\!\!\perp Y \mid Z)_{P(v)}}$$

(Separation in the model ⇒ independence in the distribution.)

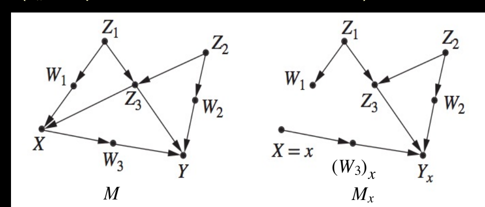
9

## SCM: AN ORACLE FOR COUNTERFACTUALS

1. The Law of Counterfactuals (and Interventions)

$$\boxed{Y_x(u) = Y_{M_x}(u)}$$

($Y_x$ is equal to $Y$ in a mutilated model $M_x$)




10

## COUNTERFACTUALS ARE EMBARRASSINGLY SIMPLE

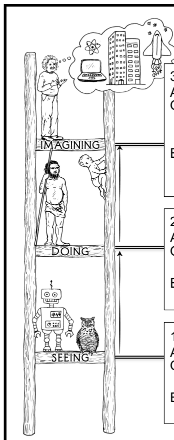
Definition:
The sentence: "$Y$ would be $y$ (in situation $u$), had $X$ been $x$," denoted $Y_x(u) = y$, means:
$y$ is the solution for $Y$ in a mutilated model $M_x$, with input $U=u$.

The Fundamental Equation of Counterfactuals:

$$\boxed{Y_x(u) = Y_{M_x}(u)}$$

11

## 3-LEVEL HIERARCHY

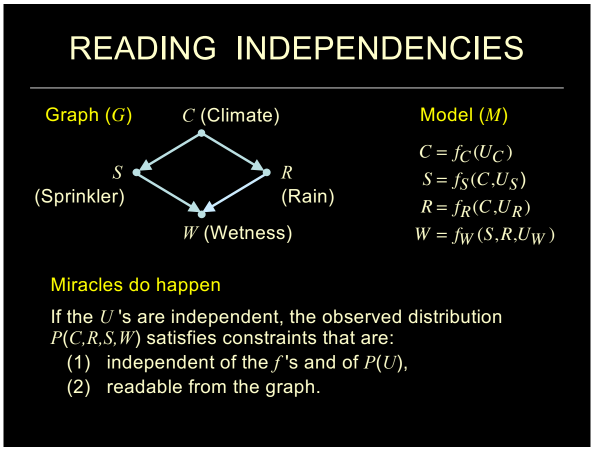



3. COUNTERFACTUALS
ACTIVITY: Imagining, Retrospection, Understanding
QUESTIONS: *What if I had done . . . ? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)
EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked the last 2 years?

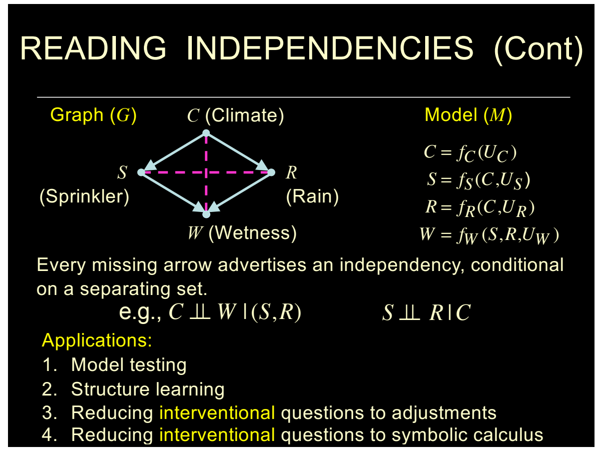
2. INTERVENTION
ACTIVITY: Doing, Intervening
QUESTIONS: *What if I do . . . ? How?*
(What would Y be if I do X?)
EXAMPLES: If I take aspirin, will my headache be cured? What if we ban cigarettes?

1. ASSOCIATION
ACTIVITY: Seeing, Observing
QUESTIONS: *What if I see . . . ?*
(How would seeing X change my belief in Y?)
EXAMPLES: What does a symptom tell me about a disease? What does a survey tell us about the election results?

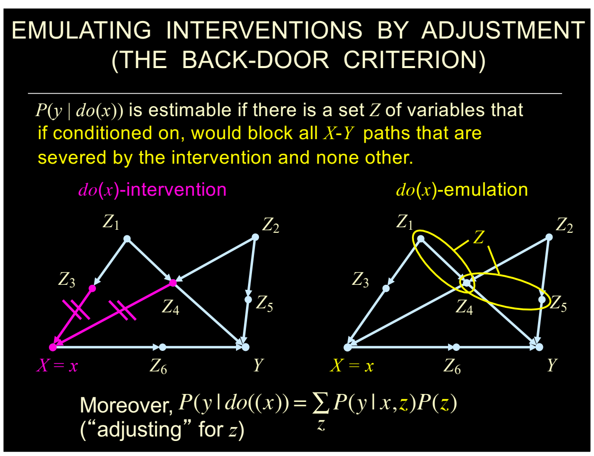
12



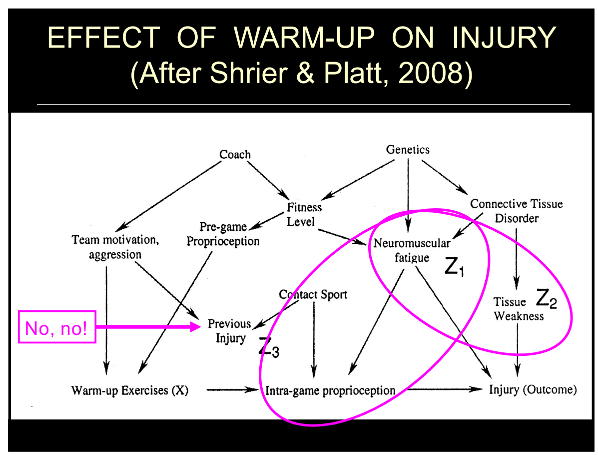

13



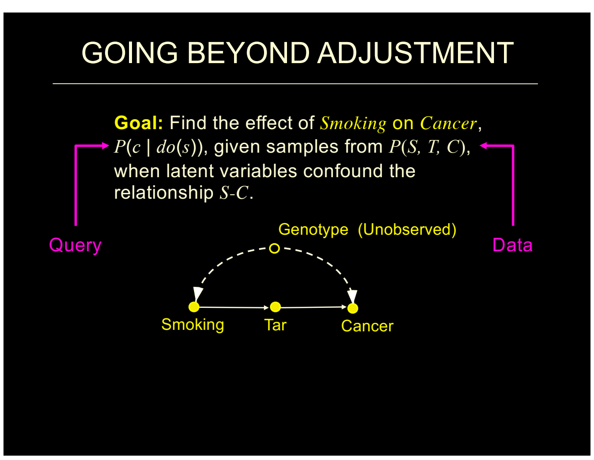

14



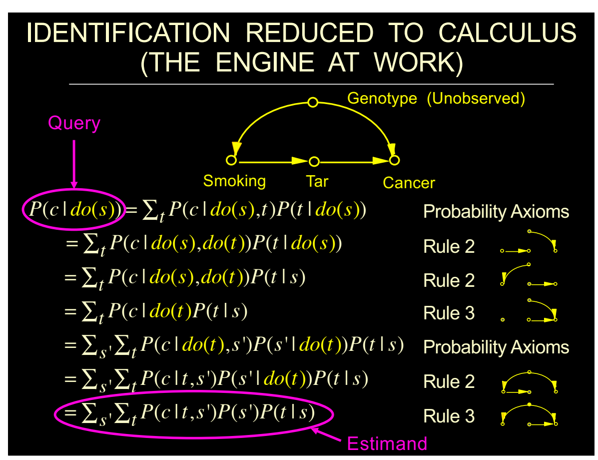

15




16




17




18

## *DO*-CALCULUS
## THE ALGEBRA OF INTERVENTIONS

The following transformations are valid for every interventional distribution generated by a structural causal model $M$:

Rule 1: Ignoring observations
$$P(y \mid do(x), z, w) = P(y \mid do(x), w), \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$$

Rule 2: Action/observation exchange
$$P(y \mid do(x), do(z), w) = P(y \mid do(x), z, w), \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$$

Rule 3: Ignoring actions
$$P(y \mid do(x), do(z), w) = P(y \mid do(x), w), \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\,\overline{Z(W)}}}$$

19

## OUTLINE

1. The scientific paradigm
   - The two fundamental laws of CI
   - The ladder of Causation
   - *Do*-calculus: The Algebra of Interventions
2. What the science adds to Machine Learning
   - Combining data with prior causal knowledge
   - Seven Pillars of Causal Wisdom
   - Future horizons
     - Personalized Decision Making
     - Social Intelligence

20

20

## THE DATA FUSION PROBLEM

The general problem

- How to combine results of several experimental and observational studies, each conducted on a different population and under a different set of conditions,
- so as to construct a valid estimate of effect size in yet a new population, unmatched by any of those studied.

21

## THE PROBLEM IN REAL LIFE

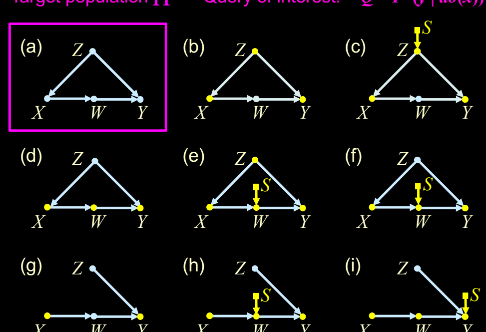
Target population $\prod^*$ Query of interest: $Q = P^*(y \mid do(x))$

| | | |
|---|---|---|
| (a) Arkansas<br>Survey data available | (b) New York<br>Survey data<br>Resembling target | (c) Los Angeles<br>Survey data<br>Younger population |
| (d) Boston<br>Age not recorded<br>Mostly successful lawyers | (e) San Francisco<br>High post-treatment blood pressure | (f) Texas<br>Mostly Spanish subjects<br>High attrition |
| (g) Toronto<br>Randomized trial<br>College students | (h) Utah<br>RCT, paid volunteers, unemployed | (i) Wyoming<br>RCT, young athletes |

22

## THE PROBLEM IN MATHEMATICS

Target population $\prod^*$ Query of interest: $Q = P^*(y \mid do(x))$




23

## SUMMARY OF
## TRANSPORTABILITY RESULTS

- Nonparametric transportability of experimental results from multiple environments can be determined provided that commonalities and differences are encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The algorithm is complete.

24

## RECOVERING FROM SELECTION BIAS

Query: Find $P(y \mid do(x))$

Data: $P(y \mid do(x), z, S=1)$ from study

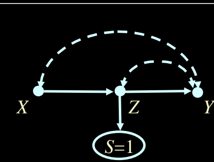
$P(y, x, z)$ from survey

Theorem:
A query $Q$ can be recovered from selection biased data iff $Q$ can be transformed, using $do$-calculus to a form in which:
(i) All $do$-expressions are conditioned on $S$ = 1
(ii) No $do$-free expression is conditioned on $S$ = 1

25

## RECOVERING FROM SELECTION BIAS

Example:




$$P(y \mid do(x)) = \sum_z P(y \mid do(x), z) P(z \mid do(x))$$
$$= \sum_z P(y \mid do(x), z) P(z \mid x) \quad \text{(Rule 2)}$$
$$= \sum_z P(y \mid do(x), z, S=1) P(z \mid x) \quad \text{(Rule 1)}$$

26

## PERSONALIZED MEDICINE

Counterfactual analysis permits us to take population data and estimate the probability that a given individual $u$ would benefit (or be harmed) by a given treatment $X$, as opposed to the average recovery rate in the subpopulation resembling the individual.

$$PNS = P(Y(1) - Y(0) = 1) \mid C(u))$$
$$= P[Y(1) = 1, Y(0) = 0 \mid C(u)]$$

vs.

$$ATE = E[Y(1) - Y(0) \mid C(u)]$$

Example: no effect vs cure and kill

27

## HOW PERSONALIZED TREATMENT EFFECT IS ESTIMATED?

- PNS is not identifiable, but can be bounded
- The bounds improve by combination and may become point estimate for certain combination
- Why both experimental and observational studies are needed?
- Example, "but for" test for personal liability
- Recent developments

28

## ATTRIBUTION

- Your Honor! My client (Mr. A) died BECAUSE he used this drug.

- 

29

## ATTRIBUTION

- Your Honor! My client (Mr. A) died BECAUSE he used this drug.

- Court to decide if it is MORE PROBABLE THAN NOT that Mr. A would be alive BUT FOR the drug!
- $PN = P(alive_{\{no\ drugs\}} \mid dead, drug) \geq 0.50$

30

## CAN FREQUENCY DATA DETERMINE LIABILITY?

Sometimes:

When $PN$ is bounded above 0.50.


- WITH PROBABILITY ONE $1 \le PN \le 1$
- Combined data tell more that each study alone

31

## IDENTIFYING "PATIENTS IN NEED"

Counterfactual: Patients susceptible to treatment.

$PNS$ = Probability that a patient with characteristics $c$ will improve IF AND ONLY IF treated.

$$PNS = P(Y(1) = 1, Y(0) = 0 \mid C = c)$$

Experimental and observational studies provide informative bounds on $PNS$.

Going from group data to individual behavior requires counterfactual logic.

32

## IDENTIFYING "PATIENTS IN NEED"

Counterfactual: Patients susceptible to treatment.

$PNS$ = Probability that a patient with characteristics $c$ will improve IF AND ONLY IF treated.

$$PNS = P(Y(1) = 1, Y(0) = 0 \mid C = c)$$

Experimental and observational studies provide informative bounds on $PNS$.

Going from group data to individual behavior requires counterfactual logic.

•Personalized medicine

33

## IDENTIFYING "PATIENTS IN NEED"

Counterfactual: Patients susceptible to treatment.

$PNS$ = Probability that a patient with characteristics $c$ will improve IF AND ONLY IF treated.

$$PNS = P(Y(1) = 1, Y(0) = 0 \mid C = c)$$

Experimental and observational studies provide informative bounds on $PNS$.

Going from group data to individual behavior requires counterfactual logic.

•Identify customers worthy of offer/recommendation

34

## IDENTIFYING "PATIENTS IN NEED"

Counterfactual: Patients susceptible to treatment.

$PNS$ = Probability that a patient with characteristics $c$ will improve IF AND ONLY IF treated.

$$PNS = P(Y(1) = 1, Y(0) = 0 \mid C = c)$$

Experimental and observational studies provide informative bounds on $PNS$.

Going from group data to individual behavior requires counterfactual logic.

•Characterize voters swayable by a slogan

35

## IDENTIFYING "PATIENTS IN NEED"

Counterfactual: Patients susceptible to treatment.

$PNS$ = Probability that a patient with characteristics $c$ will improve IF AND ONLY IF treated.

$$PNS = P(Y(1) = 1, Y(0) = 0 \mid C = c)$$

Experimental and observational studies provide informative bounds on $PNS$.

Going from group data to individual behavior requires counterfactual logic.

•Unit Selection: Li, Mueller and Pearl (2021)

36

## THE SEVEN PILLARS

Pillar 1: Transparency and Testability of Causal Assumptions
Pillar 2: Confounding deconfounded
Pillar 3: Counterfactuals Algorithmitized
Pillar 4: Mediation Analysis and the Assessment of Direct and Indirect Effects
Pillar 5: External Validity and Sample Selection Bias
Pillar 6: Missing Data (w/ Karthika Mohan, 2017)
Pillar 7: Causal Discovery

37

## BUT WHAT IF I DON'T HAVE A MODEL

1. Study SCM – COVID-19 can't wait
2. Study SCM – to help find one
3. Study SCM – to help use the one you find
4. Study SCM – to help explain your findings

Today, only 1 of every 1,000 DL students studies the science of cause and effect.

38

## THE NEXT TWO MISSIONS OF SCM:

Automated scientist

- Design of new experiments, seek new observations.
- Control of attention, simulated curiosity, conjectures generation.
- Laboratory for theories of scientific thinking.

Social Intelligence

- From deep understanding of a domain to the understanding of other agents.
- Natural communication among robots and man-machine involving: Trust, desires, responsibility, awareness, intension, motivation, …

39

## TYPICAL CHALLENGES – FORMALIZE "RESPONSIBILITY"

An agent is "morally responsible" for an outcome, if (Stanford Encyclopedia):

1. There is a causal connection between the agent's action and the outcome, and the agent had some control over the outcome.
2. The agent has knowledge of and is able to consider the consequences of its actions.
3. The agent is able to freely choose to act in certain way.

Note the importance of explicit vs. implicit knowledge.

40

## CONCLUSIONS

"More has been learned about causal inference in the last few decades than the sum total of everything that had been learned about it in all prior recorded history."
(Gary King, Harvard, 2014)

"The next revolution will be even more impactful upon realizing that data science is the science of interpreting reality, not of summarizing data."
(The Author, UCLA, 2022)

41

Paper available: http://ftp.cs.ucla.edu/pub/stat_ser/r475.pdf
Refs: http://bayes.cs.ucla.edu/jp_home.html

*Every science that has thriven has thriven upon its own symbols*
*~Augustus de Morgan (1864)*

## THANK YOU

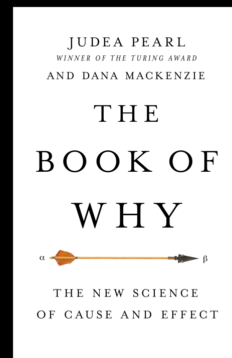
Joint work with:
Elias Bareinboim
Karthika Mohan
Ilya Shpitser
Jin Tian
Many more . . .

42

Time for a short commercial

43

For a trailer, click WHY on my home page.




44