

WHAT IS CAUSAL INFERENCE AND WHERE IS DATA SCIENCE GOING?

Judea Pearl
UCLA
Computer Science and Statistics

IJCAI Conference, Vienna
July 28, 2022

1

OUTLINE

1. The Big Data Revolution: Success, limitations, and tensions
2. The Causal Revolution: From fitting to understanding
3. The ladder of causation
4. The seven wisdoms of causal thinking
5. Future directions:
 - a. Automated scientists
 - b. From population to individual decisions

2

DATA SCIENCE – A CLASH OF TWO PARADIGMS

1. The data-centric paradigm

- How best to fit the data so as to maximize some “**goodness of fit**” measure over the same data.

2. The scientific paradigm

- What should the **world** be like before I can answer a research question about that **world**?

Alternatively: Extract understanding from Data

3

WHAT **CAPABILITIES** DOES DEEP UNDERSTANDING ENTAIL?

A state of knowledge evoking a sensation of “**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 QUESTIONS NEEDING UNDERSTANDING

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.

$$Y = aX \quad \text{vs.} \quad Y \leftarrow aX$$

5

SEWALL WRIGHT – CAUSALITY’S FIRST FORMAL VOICE

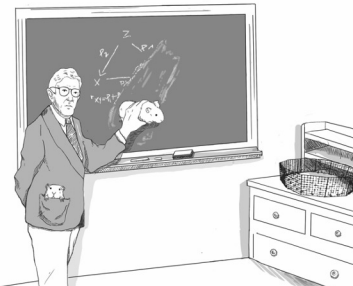


Figure 2.6. Sewall Wright was the first person to develop a mathematical method for answering causal questions from data, known as path diagrams. His love of mathematics surrendered only to his passion for guinea pigs.

6

SEWALL WRIGHT – CAUSALITY'S FIRST FORMAL VOICE

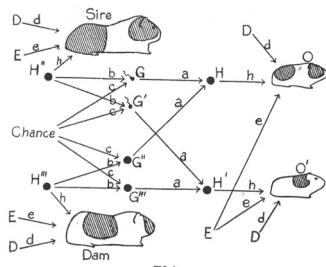


Figure 2.7. Sewall Wright's first path diagram, illustrating the factors leading to coat color in guinea pigs. D = developmental factors (after conception, before birth), E = environmental factors (after birth), G = genetic factors from each individual parent, H = combined hereditary factors from both parents, O, Oc = 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

WHY WAS WRIGHT ATTACKED? THE STATISTICS PARADIGM 1834–2022

- “The object of statistical methods is the reduction of data” (Fisher 1922).
- Statistical concepts are those expressible in terms of **joint distribution of observed variables**.
- All others are: “substantive matter,” “domain dependent,” “metaphysical,” “ad hockery,” i.e., outside the province of statistics, **ruling out all interesting questions**.
- Slow awakening since Neyman (1923) and Rubin (1974).
- Traditional Statistics Education = Causalophobia

8

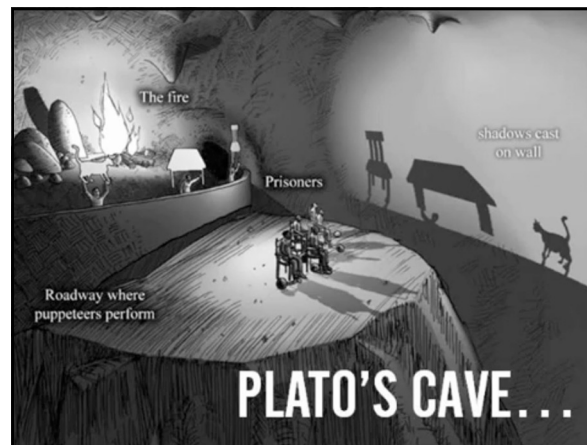
WHY IS STAT SO SEDUCTIVE? AND WHAT'S MISSING FROM IT?

1. Stat adopted **probability** as its mother tongue – a mighty belief-updating **oracle**.
2. Any probability function $P(X, Y, Z, \dots)$ tells us not only $P(Y=y)$, but also how $P(Y=y)$ ought to change when **new evidence** $X=x$ is received.
3. Alas, it can't tell us how $P(Y=y)$ ought to change when a **new intervention** $do(X=x)$ is enacted.
4. Example: Y = Life expectancy, X = Exercise

$$P(x, y) \text{ dictates } P(y|x) = P(x, y) / \sum_y P(x, y)$$

$$P(x, y) \text{ silent about } P(y|do(x)) \text{ or } P(y|x')$$

9



10

FROM STATISTICAL TO CAUSAL ANALYSIS: 2. THE SHARP BOUNDARY

1. Causal and associational concepts do not mix.

CAUSAL	ASSOCIATIONAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
“Holding constant” / “Fixing”	“Controlling for” / Conditioning
Confounding / Effect	Odds and risk ratios
Instrumental variable	Collapsibility / Granger causality
Ignorability / Exogeneity	Propensity score
- 2.
- 3.
- 4.

11

FROM STATISTICAL TO CAUSAL ANALYSIS: 3. THE NEW ORACLE

1. Causal and associational concepts do not mix.

CAUSAL	ASSOCIATIONAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
“Holding constant” / “Fixing”	“Controlling for” / Conditioning
Confounding / Effect	Odds and risk ratios
Instrumental variable	Collapsibility / Granger causality
Ignorability / Exogeneity	Propensity score
2. **No causes in – no causes out** (Cartwright, 1989)

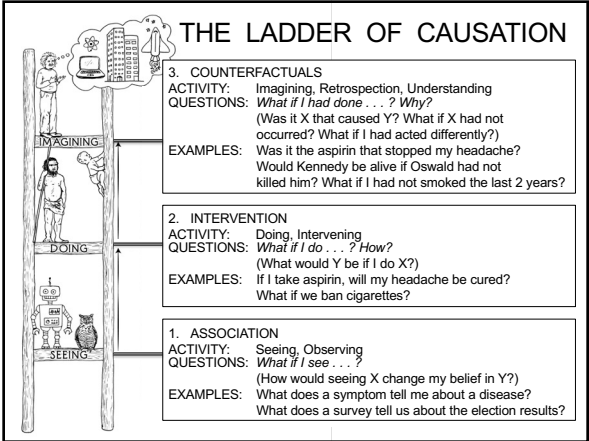
$$\left. \begin{array}{l} \text{data} \\ \text{causal assumptions (or experiments)} \end{array} \right\} \Rightarrow \text{causal conclusions}$$
3. Causal assumptions cannot be expressed in the mathematical language of standard statistics.
4. **Non-standard mathematics (New Oracle):**
 - a) Structural Causal Model (Wright, 1920; $X \rightarrow Y, Y = f(X)$)
 - b) Related attempts (Neyman-Rubin (Y_X), Lewis ($x \boxrightarrow Y$))

12

THE CAUSAL REVOLUTION

- 1. "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)
- 2. From liability to respectability
 - JSM 2003 – 13 papers
 - JSM 2013 – 130 papers
 - 2022 – dozens of causality-specific workshops and conferences
- 3. Fun, profit, education, and Mind

13



14

THE SEVEN WISDOMS (TOOLS) OF CAUSAL INFERENCE

- Tool 1:** Encoding causal information in transparent and testable way
- Tool 2:** Predicting the effects of actions and policies
- Tool 3:** Computing counterfactuals and finding causes of effects
(**attribution, explanation, susceptibility**)
- Tool 4:** Computing direct and indirect effects (Mediation)
(**discrimination, inequities, fairness**)
- Tool 5:** Integrating data from diverse sources
(**fusion, transportability, transfer-learning**)
- Tool 6:** Recovering from missing Data
- Tool 7:** Causal Discovery

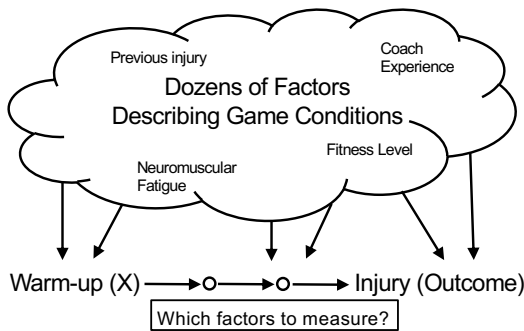
15

THE SEVEN WISDOMS (TOOLS) OF CAUSAL INFERENCE

- Tool 1:** Encoding causal information in transparent and testable way
- Tool 2:** Predicting the effects of actions and policies
- Tool 3:** Computing counterfactuals and finding causes of effects
(**attribution, explanation, susceptibility**)
- Tool 4:** Computing direct and indirect effects (Mediation)
(**discrimination, inequities, fairness**)
- Tool 5:** Integrating data from diverse sources
(**fusion, transportability, transfer-learning**)
- Tool 6:** Recovering from missing Data
- Tool 7:** Causal Discovery

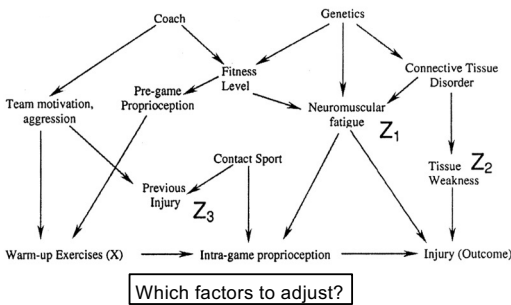
16

EFFECT OF WARM-UP ON INJURY
(After Shrier & Platt, 2008)



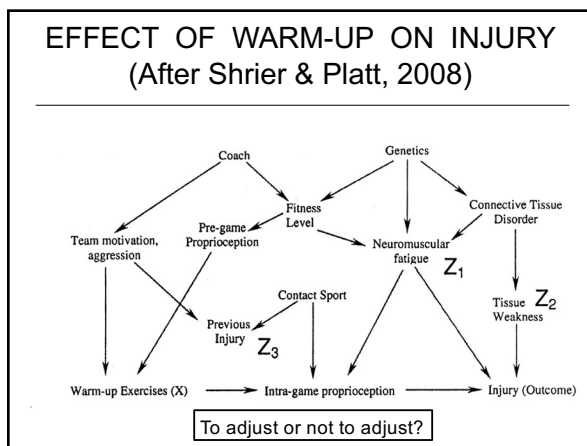
17

EFFECT OF WARM-UP ON INJURY
(After Shrier & Platt, 2008)



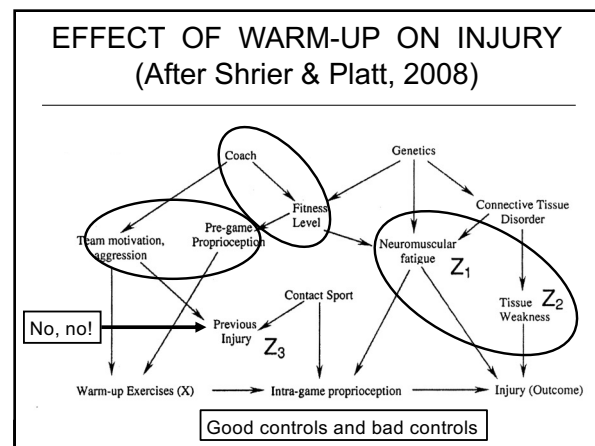
18

EFFECT OF WARM-UP ON INJURY (After Shrier & Platt, 2008)



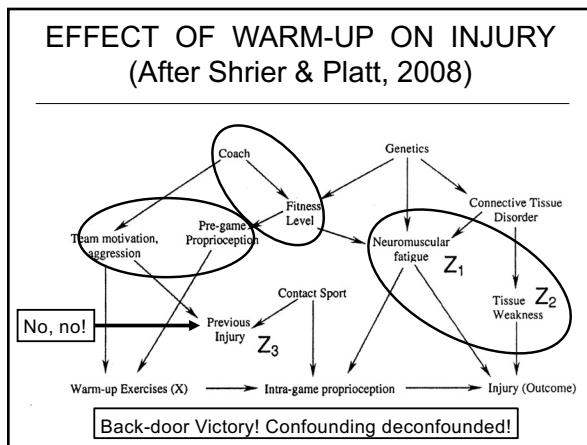
19

EFFECT OF WARM-UP ON INJURY (After Shrier & Platt, 2008)



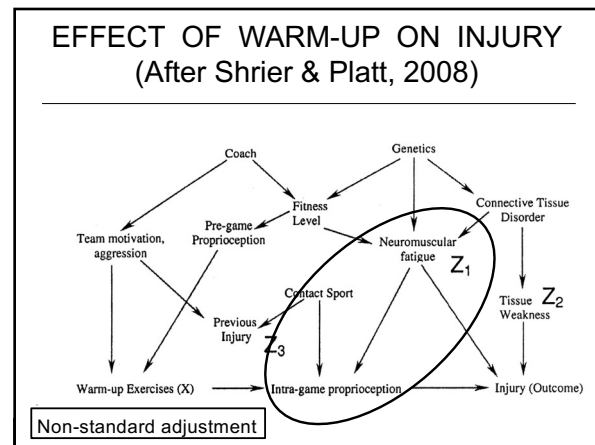
20

EFFECT OF WARM-UP ON INJURY (After Shrier & Platt, 2008)



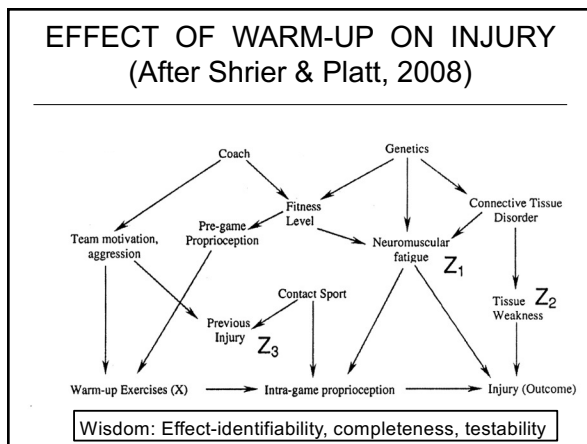
21

EFFECT OF WARM-UP ON INJURY (After Shrier & Platt, 2008)



22

EFFECT OF WARM-UP ON INJURY (After Shrier & Platt, 2008)



23

THE SEVEN WISDOMS (TOOLS) OF CAUSAL INFERENCE

- Tool 1:** Encoding causal information in transparent and testable way
- Tool 2:** Predicting the effects of actions and policies
- Tool 3:** Computing counterfactuals and finding causes of effects
(**attribution, explanation, susceptibility**)
- Tool 4:** Computing direct and indirect effects (Mediation)
(**discrimination, inequities, fairness**)
- Tool 5:** Integrating data from diverse sources
(**fusion, transportability, transfer-learning**)
- Tool 6:** Recovering from missing Data
- Tool 7:** Causal Discovery

24

ATTRIBUTION (Rung-3)

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



•

25

ATTRIBUTION (Rung-3)

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



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

26

CAN WE COMPUTE COUNTERFACTUALS?

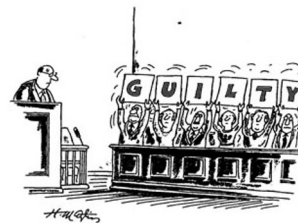
- **Yes!** If we know the **functions** behind the arrows, every counterfactual gets a truth value.
- If we don't, we can **bound** them using the **logic** of counterfactuals (Halpern & Pearl).
- The bounds improve when **combined data** are available and may narrow down sufficiently to reveal the **individual** responsibility.

27

CAN FREQUENCY DATA DETERMINE LIABILITY?

Sometimes:

When PN is bounded above 0.50.



- WITH PROBABILITY ONE $1 \leq PN \leq 1$
- Combined data tell more than each study alone

28

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.

29

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.

- Situation-specific decisions

30

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

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.

- Identify customers worthy of offer/recommendation

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.

- Characterize voters swayed by a slogan

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.

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

34

THE SEVEN WISDOMS (TOOLS) OF CAUSAL INFERENCE

Tool 1: Encoding causal information in transparent and testable way

Tool 2: Predicting the effects of actions and policies

Tool 3: Computing counterfactuals and finding causes of effects

(**attribution, explanation, susceptibility**)

Tool 4: Computing direct and indirect effects (Mediation)

(**discrimination, inequities, fairness**)

Tool 5: Integrating data from diverse sources

(**fusion, transportability, transfer-learning**)

Tool 6: Recovering from missing Data

Tool 7: Causal Discovery

35

TOOL 4: MEDIATION ANALYSIS – DIRECT AND INDIRECT EFFECTS

Task: Given {Data + Model}, Unveil and quantify the **mechanisms that transmit** changes from a cause to its effects.

Wisdom: Counterfactual analysis tells us when direct and indirect effects are **estimable** from data, and, if so, **how necessary** (or sufficient) mediation is for the effect.

36

COUNTERFACTUAL DEFINITION
OF DISCRIMINATION

“The central question in any employment-discrimination case is whether the employer **would have taken** the same **action** **had the employee** been of a different race (age, sex, religion, national origin, etc.) and everything else **had been** the same.”

(In Carson vs Bethlehem Steel Corp., 70 FEP Cases 921, 7th Cir. (1996).)

37

LEGAL IMPLICATIONS
OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?

What is the direct effect of X on Y ?

- $NDE(X,Y)$ = The expected change in Y **had X changed** and **had M been constant** at whatever value it attained before the change.
- Meditation formulas, identification, standardization

38

THE SEVEN WISDOMS (TOOLS)
OF CAUSAL INFERENCE

Tool 1: Encoding causal information in transparent and testable way

Tool 2: Predicting the effects of actions and policies

Tool 3: Computing counterfactuals and finding causes of effects
(**attribution, explanation, susceptibility**)

Tool 4: Computing direct and indirect effects (Mediation)
(**discrimination, inequities, fairness**)

Tool 5: Integrating data from diverse sources
(**fusion, transportability, transfer-learning**)

Tool 6: Recovering from missing Data

Tool 7: Causal Discovery

39

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.
- Subproblems: External validity, selection bias

40

THE PROBLEM IN REAL LIFE

Target population Π^*	Query of interest: $Q = P^*(y \mid do(x))$	
(a) Arkansas Survey data available	(b) New York Survey data Resembling target	(c) Los Angeles Survey data Younger population
(d) Boston Age not recorded Mostly successful lawyers	(e) San Francisco High post-treatment blood pressure	(f) Texas Mostly Spanish subjects High attrition
(g) Toronto Randomized trial College students	(h) Utah RCT, paid volunteers, unemployed	(i) Wyoming RCT, young athletes

41

THE PROBLEM IN MATHEMATICS

Target population Π^*	Query of interest: $Q = P^*(y \mid do(x))$	
(a)	(b)	(c)
(d)	(e)	(f)
(g)	(h)	(i)

42

WHAT CAN BE FUSED AND HOW?

- Experimental results from multiple sources **can be fused** provided that commonalities and differences are encoded in selection diagrams.
- When estimation is feasible, a fusion formula can be derived in **polynomial time**.
- The algorithm is **complete**.

43

THE SEVEN WISDOMS (TOOLS) OF CAUSAL INFERENCE

- Tool 1:** Encoding causal information in transparent and testable way
Tool 2: Predicting the effects of actions and policies
Tool 3: Computing counterfactuals and finding causes of effects
 (attribution, explanation, susceptibility)
Tool 4: Computing direct and indirect effects (Mediation)
 (discrimination, inequities, fairness)
Tool 5: Integrating data from diverse sources
 (fusion, transportability, transfer-learning)
Tool 6: Recovering from missing Data
Tool 7: Causal Discovery

44

THE SEVEN WISDOMS (TOOLS) OF CAUSAL INFERENCE

- Tool 1:** Encoding causal information in transparent and testable way
Tool 2: Predicting the effects of actions and policies
Tool 3: Computing counterfactuals and finding causes of effects
 (attribution, explanation, susceptibility)
Tool 4: Computing direct and indirect effects (Mediation)
 (discrimination, inequities, fairness)
Tool 5: Integrating data from diverse sources
 (fusion, transportability, transfer-learning)
Tool 6: Recovering from missing Data
Tool 7: Causal Discovery

45

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)

46

CONCLUSIONS (cont.)

- Q. So What is Causal Inference?
- A. It's the leverage that elevates Data Science from Rung-1 of the Ladder to Rungs 2 and 3,
- i.e., from data-fitting to deep understanding.

47

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 . . .

48

For a trailer, click WHY on my home page.
For discussions: @yudapearl

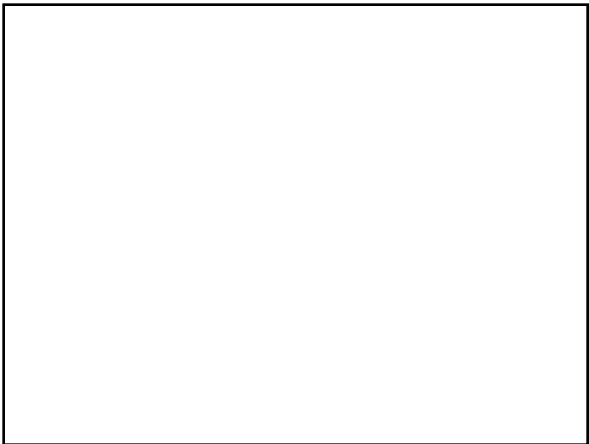
JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE

THE
BOOK OF
WHY

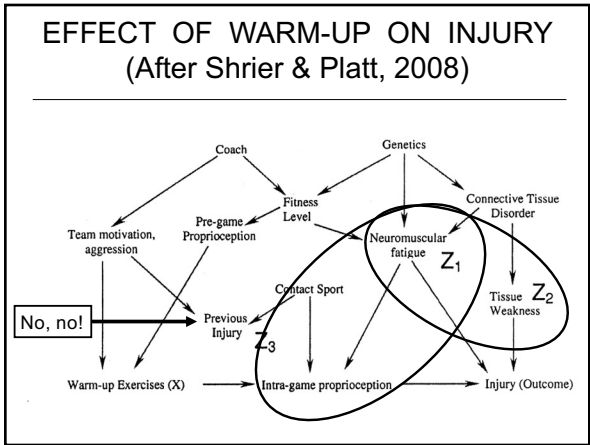
α  β

THE NEW SCIENCE
OF CAUSE AND EFFECT

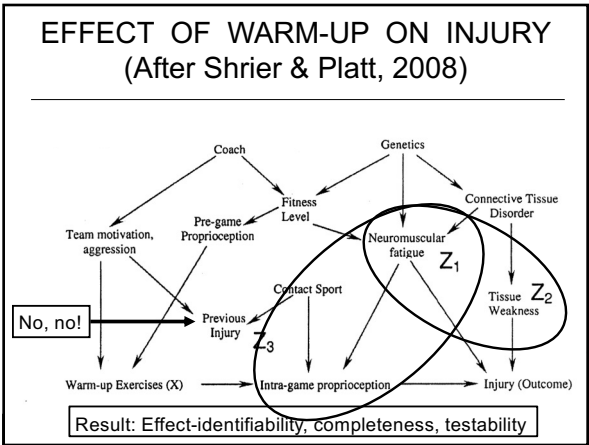
49



50



51



52