

THE MATHEMATICS OF CAUSE AND EFFECT

Judea Pearl
University of California
Los Angeles

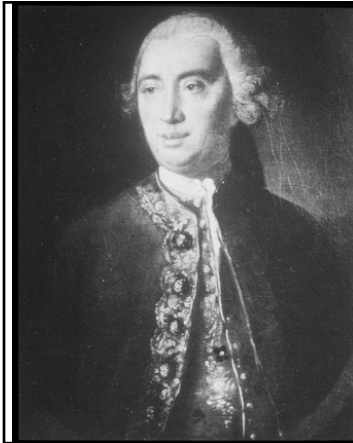
ANTIQUITY TO ROBOTICS

"I would rather discover one causal relation than be King of Persia"

Democritus (430-380 BC)

Development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment (during the Renaissance).

A. Einstein, April 23, 1953



David Hume
(1711–1776)

"I would rather discover one causal law than be King of Persia."

Democritus (460-370 B.C.)

"Development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment (during the Renaissance)."

A. Einstein, April 23, 1953

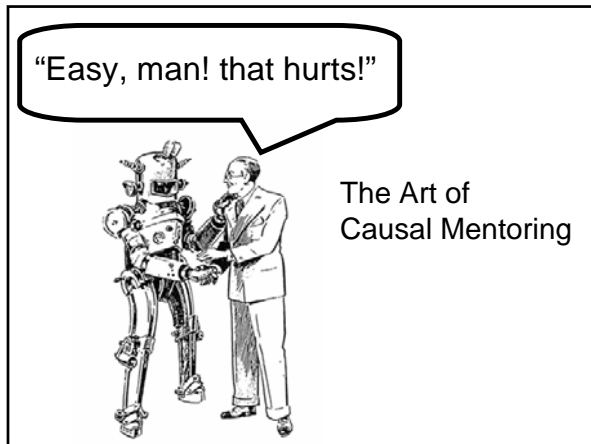
HUME'S LEGACY

1. Analytical vs. empirical claims
2. Causal claims are empirical
3. All empirical claims originate from experience.

THE TWO RIDDLES OF CAUSATION

- What empirical evidence legitimizes a cause-effect connection?
- What inferences can be drawn from causal information? and how?





OLD RIDDLES IN NEW DRESS

1. How should a robot acquire causal information from the environment?
2. How should a robot process causal information received from its creator-programmer?

CAUSATION AS A PROGRAMMER'S NIGHTMARE

Input:

1. “If the grass is wet, then it rained”
2. “if we break this bottle, the grass will get wet”

Output:

“If we break this bottle, then it rained”

CAUSATION AS A PROGRAMMER'S NIGHTMARE (Cont.) (Lin, 1995)

Input:

1. A suitcase will open iff both locks are open.
2. The right lock is open

Query:

What if we open the left lock?

Output:

The right lock might get closed.

THE BASIC PRINCIPLES

Causation = encoding of behavior under interventions

Interventions = surgeries on mechanisms

Mechanisms = stable functional relationships
= equations + graphs

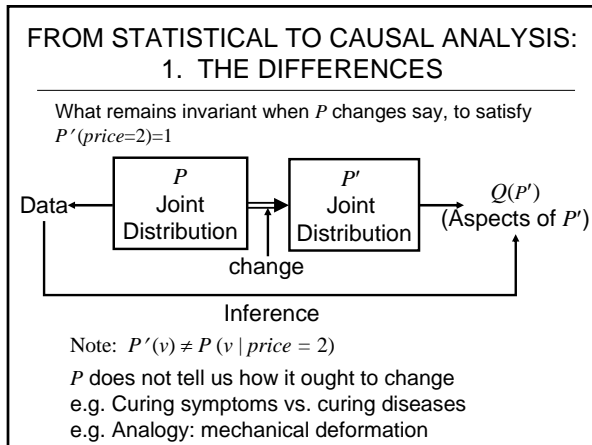
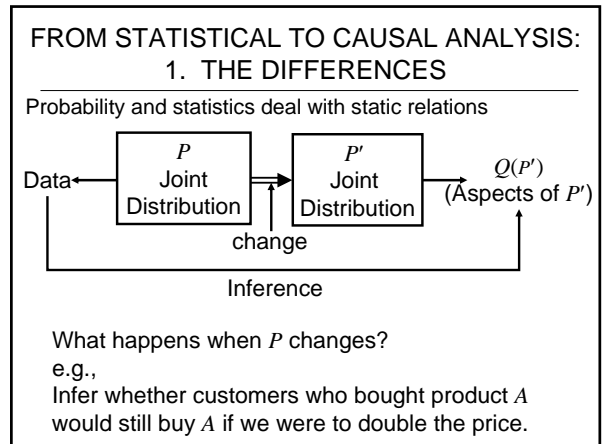
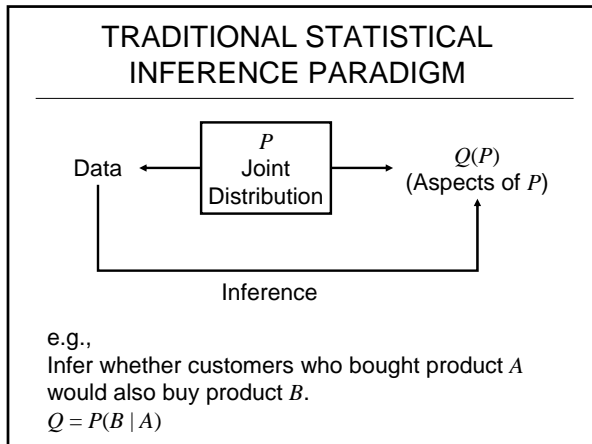
CAUSATION AS A PROGRAMMER'S NIGHTMARE

Input:

1. “If the grass is wet, then it rained”
2. “if we break this bottle, the grass will get wet”

Output:

“If we break this bottle, then it rained”



- ### FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES (CONT)
-
1. Causal and statistical concepts do not mix.

CAUSAL Spurious correlation Randomization Confounding / Effect Instrument Holding constant Explanatory variables	STATISTICAL Regression Association / Independence "Controlling for" / Conditioning Odd and risk ratios Collapsibility
---	---
 - 2.
 - 3.
 - 4.

- ### FROM STATISTICAL TO CAUSAL ANALYSIS: 2. MENTAL BARRIERS
-
1. Causal and statistical concepts do not mix.

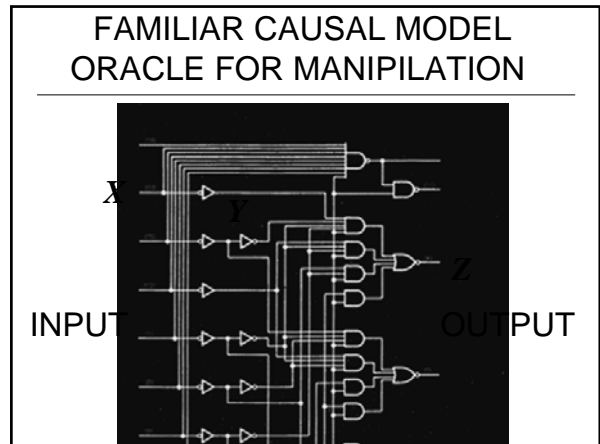
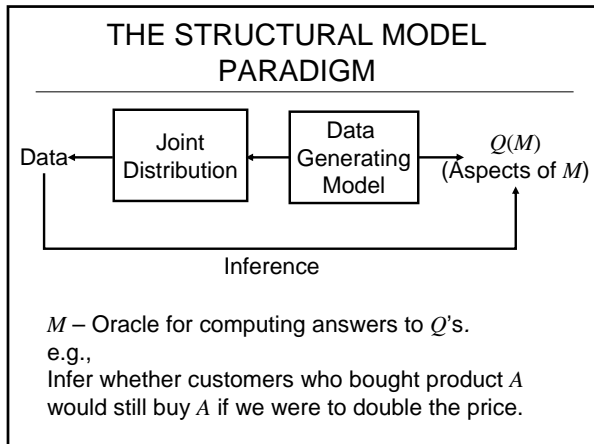
CAUSAL Spurious correlation Randomization Confounding / Effect Instrument Holding constant Explanatory variables	STATISTICAL Regression Association / Independence "Controlling for" / Conditioning Odd and risk ratios Collapsibility
---	---
 2. No causes in – no causes out (Cartwright, 1989)

statistical assumptions + data causal assumptions	}	\Rightarrow causal conclusions
--	---	----------------------------------
 3. Causal assumptions cannot be expressed in the mathematical language of standard statistics.
 - 4.

- ### FROM STATISTICAL TO CAUSAL ANALYSIS: 2. MENTAL BARRIERS
-
1. Causal and statistical concepts do not mix.

CAUSAL Spurious correlation Randomization Confounding / Effect Instrument Holding constant Explanatory variables	STATISTICAL Regression Association / Independence "Controlling for" / Conditioning Odd and risk ratios Collapsibility
---	---
 2. No causes in – no causes out (Cartwright, 1989)

statistical assumptions + data causal assumptions	}	\Rightarrow causal conclusions
--	---	----------------------------------
 3. Causal assumptions cannot be expressed in the mathematical language of standard statistics.
 4. Non-standard mathematics:
 - a) Structural equation models (Wright, 1920; Simon, 1960)
 - b) Counterfactuals (Neyman-Rubin (Y_x), Lewis ($x \boxrightarrow Y$))

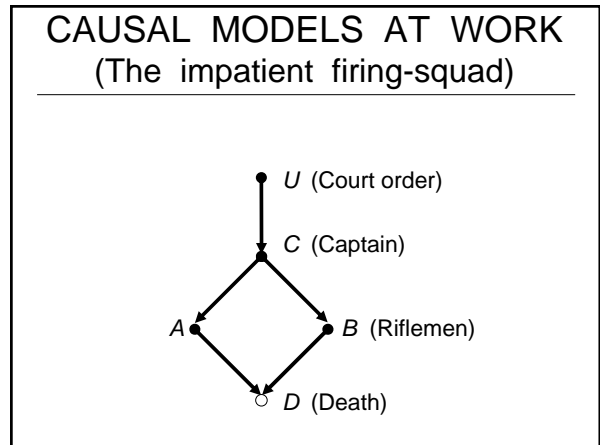


STRUCTURAL CAUSAL MODELS

Definition: A structural causal model is a 4-tuple $\langle V, U, F, P(u) \rangle$, where

- $V = \{V_1, \dots, V_n\}$ are observable variables
- $U = \{U_1, \dots, U_m\}$ are background variables
- $F = \{f_1, \dots, f_n\}$ are functions determining V ,
 $v_i = f_i(v, u)$
- $P(u)$ is a distribution over U

$P(u)$ and F induce a distribution $P(v)$ over observable variables



CAUSAL MODELS AT WORK (Glossary)

U : Court orders the execution
 C : Captain gives a signal
 A : Rifleman-A shoots
 B : Rifleman-B shoots
 D : Prisoner dies
 $=$: Functional Equality (new symbol)

SENTENCES TO BE EVALUATED

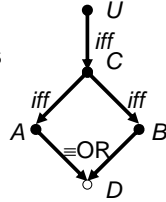
S1. prediction: $\neg A \Rightarrow \neg D$
 S2. abduction: $\neg D \Rightarrow \neg C$
 S3. transduction: $A \Rightarrow B$
 S4. action: $\neg C \Rightarrow D_A$
 S5. counterfactual: $D \Rightarrow D_{\{\neg A\}}$
 S6. explanation: $\text{Caused}(A, D)$

STANDARD MODEL FOR STANDARD QUERIES

S1. (prediction): If rifleman-A shot, the prisoner is dead,
 $A \Rightarrow D$

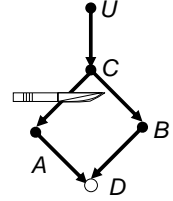
S2. (abduction): If the prisoner is alive, then the Captain did not signal,
 $\neg D \Rightarrow \neg C$

S3. (transduction): If rifleman-A shot, then B shot as well,
 $A \Rightarrow B$



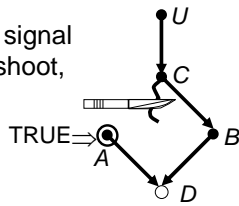
WHY CAUSAL MODELS? GUIDE FOR SURGERY

S4. (action):
 If the captain gave no signal and Mr. A decides to shoot, the prisoner will die:
 $\neg C \Rightarrow D_A$
 and B will not shoot:
 $\neg C \Rightarrow \neg B_A$



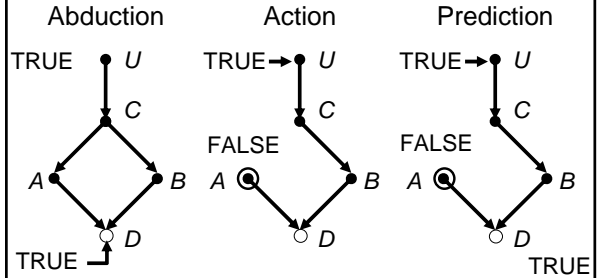
WHY CAUSAL MODELS? GUIDE FOR SURGERY

S4. (action):
 If the captain gave no signal and Mr. A decides to shoot, the prisoner will die:
 $\neg C \Rightarrow D_A$
 and B will not shoot:
 $\neg C \Rightarrow \neg B_A$



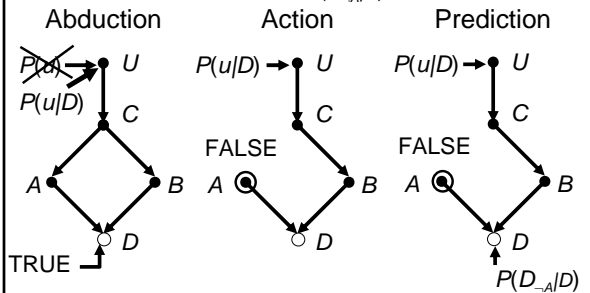
3-STEPS TO COMPUTING COUNTERFACTUALS

S5. If the prisoner is dead, he would still be dead if A were not to have shot. $D \Rightarrow D_{\neg A}$



COMPUTING PROBABILITIES OF COUNTERFACTUALS

P(S5). The prisoner is dead. How likely is it that he would be dead if A were not to have shot. $P(D_{\neg A}|D) = ?$



CAUSAL MODEL (FORMAL)

$M = \langle U, V, F \rangle$ or $\langle U, V, F, P(u) \rangle$

U - Background variables

V - Endogenous variables

F - Set of functions $\{U \times V \mid V_i \rightarrow V_i\}$
 $v_i = f_i(pa_i, u_i)$

Submodel: $M_x = \langle U, V, F_x \rangle$, representing $do(x)$

$F_x =$ Replaces equation for X with $X=x$

Actions and Counterfactuals:

$Y_x(u) =$ Solution of Y in M_x

$P(y \mid do(x)) \triangleq P(Y_x=y)$

HIERACHY OF GRAPHICAL MODELS

1. Bayes Networks
Encode: conditional independencies
Answer: Abduction, transduction, prediction
2. Causal Bayes Networks
Encode: interventional independencies
Answer: effects of interventions (and observations)
3. Functional Causal Networks
Encode: functional independencies
Answer: retrospective counterfactuals

WHY COUNTERFACTUALS?

Action queries are triggered by (modifiable) observations, demanding abductive step, i.e., counterfactual processing.

E.g., Troubleshooting

- | | |
|-----------------------|--|
| Observation: | The output is low |
| Action query: | Will the output get higher – if we replace the transistor? |
| Counterfactual query: | Would the output be higher – had the transistor been replaced? |

APPLICATIONS

1. Predicting effects of actions and policies
2. Learning causal relationships from assumptions and data
3. Troubleshooting physical systems and plans
4. Finding explanations for reported events
5. Generating verbal explanations
6. Understanding causal talk
7. Formulating theories of causal thinking

○ CAUSAL MODELS AND COUNTERFACTUALS

Definition:

The sentence: “ Y would be y (in situation u), had X been x ,” denoted $Y_x(u) = y$, means:

The solution for Y in a mutilated model M_x , (i.e., the equations for X replaced by $X = x$) with input $U = u$, is equal to y .

Joint probabilities of counterfactuals:

$$P(Y_x = y, Z_w = z) = \sum_{u: Y_x(u)=y, Z_w(u)=z} P(u)$$

The super-distribution P^* is derived from M .
Parsimonious, consistent, and transparent

AXIOMS OF STRUCTURAL COUNTERFACTUALS

$Y_x(u)=y$: Y would be y , had X been x (in state $U = u$)
(Galles, Pearl, Halpern, 1998):


1. Definiteness
 $\exists x \in X \text{ s.t. } X_y(u) = x$
2. Uniqueness
 $(X_y(u) = x) \& (X_{y'}(u) = x') \Rightarrow x = x'$
3. Effectiveness
 $X_{xw}(u) = x$
4. Composition (generalized consistency)
 $X_w(u) = x \Rightarrow Y_{wx}(u) = Y_y(u)$
5. Reversibility
 $(Y_{xw}(u) = y) \& (W_{xy}(u) = w) \Rightarrow Y_x(u) = y$

GRAPHICAL – COUNTERFACTUALS SYMBIOSIS

Every causal graph expresses counterfactuals assumptions,

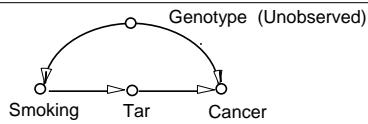
e.g., $X \rightarrow Y \rightarrow Z$

1. Missing arrows $X \rightarrow Z$ $Z_{x,y}(u) = Z_y(u)$

2. Missing arcs  $Y_x \perp\!\!\!\perp Z_y$

Assumptions are guaranteed consistency.
Assumptions are readable from the graph.

DERIVATION IN CAUSAL CALCULUS

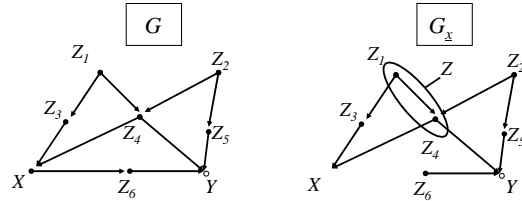


$$\begin{aligned}
 P(c | do(s)) &= \sum_t P(c | do(s), t) P(t | do(s)) && \text{Probability Axioms} \\
 &= \sum_t P(c | do(s), do(t)) P(t | do(s)) && \text{Rule 2} \\
 &= \sum_t P(c | do(s), do(t)) P(t | s) && \text{Rule 2} \\
 &= \sum_t P(c | do(t)) P(t | s) && \text{Rule 3} \\
 &= \sum_{s'} \sum_t P(c | do(t), s') P(s' | do(t)) P(t | s) && \text{Probability Axioms} \\
 &= \sum_{s'} \sum_t P(c | t, s') P(s' | do(t)) P(t | s) && \text{Rule 2} \\
 &= \sum_{s'} \sum_t P(c | t, s') P(s') P(t | s) && \text{Rule 3}
 \end{aligned}$$

THE BACK-DOOR CRITERION

Graphical test of identification

$P(y | do(x))$ is identifiable in G if there is a set Z of variables such that Z d -separates X from Y in $G_{\bar{x}}$.



Moreover, $P(y | do(x)) = \sum_z P(y | x, z) P(z)$
("adjusting" for Z)

RECENT RESULTS ON IDENTIFICATION

- do -calculus is complete
- Complete graphical criterion for identifying causal effects (Shpitser and Pearl, 2006).
- Complete graphical criterion for empirical testability of counterfactuals (Shpitser and Pearl, 2007).

DETERMINING THE CAUSES OF EFFECTS (The Attribution Problem)

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



DETERMINING THE CAUSES OF EFFECTS (The Attribution Problem)

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



- Court to decide if it is MORE PROBABLE THAN NOT that A would be alive BUT FOR the drug!
 $PN = P(? | A \text{ is dead, took the drug}) \geq 0.50$

THE PROBLEM

Semantical Problem:

1. What is the meaning of $PN(x, y)$:
"Probability that event y would not have occurred if it were not for event x , given that x and y did in fact occur."

THE PROBLEM

Semantical Problem:

1. What is the meaning of $PN(x,y)$:
 "Probability that event y would not have occurred if it were not for event x , given that x and y did in fact occur."

Answer:

$$PN(x, y) = P(Y_{x'} = y' | x, y)$$

Computable from M

THE PROBLEM

Semantical Problem:

1. What is the meaning of $PN(x,y)$:
 "Probability that event y would not have occurred if it were not for event x , given that x and y did in fact occur."

Analytical Problem:

2. Under what condition can $PN(x,y)$ be learned from statistical data, i.e., observational, experimental and combined.

TYPICAL THEOREMS

(Tian and Pearl, 2000)

- Bounds given combined nonexperimental and experimental data

$$\max \left\{ \frac{P(y) - P(y_{x'})}{P(x,y)} \right\} \leq PN \leq \min \left\{ \frac{P(y'_{x'})}{P(x,y)} \right\}$$

- Identifiability under monotonicity (Combined data)

$$PN = \frac{P(y/x) - P(y/x')}{P(y/x)} + \frac{P(y/x') - P(y'_{x'})}{P(x,y)}$$

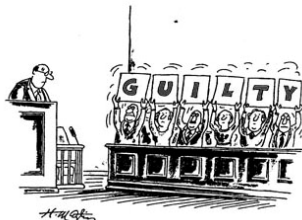
corrected Excess-Risk-Ratio

CAN FREQUENCY DATA DECIDE LEGAL RESPONSIBILITY?

	Experimental		Nonexperimental	
	$do(x)$	$do(x')$	x	x'
Deaths (y)	16	14	2	28
Survivals (y')	984	986	998	972
	1,000	1,000	1,000	1,000

- Nonexperimental data: drug usage predicts longer life
- Experimental data: drug has negligible effect on survival
- Plaintiff: Mr. A is special.
 1. He actually died
 2. He used the drug by choice
- Court to decide (given both data):
 Is it more probable than not that A would be alive but for the drug?

SOLUTION TO THE ATTRIBUTION PROBLEM



- WITH PROBABILITY ONE $1 \leq P(y'_{x'} | x, y) \leq 1$
- Combined data tell more than each study alone

CONCLUSIONS

Structural-model semantics, enriched with logic and graphs, provides:

- Complete formal basis for causal reasoning
- Powerful and friendly causal calculus
- Lays the foundations for asking more difficult questions: What is an action? What is free will? Should robots be programmed to have this illusion?