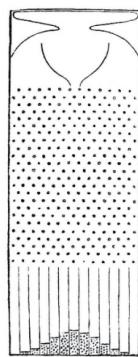
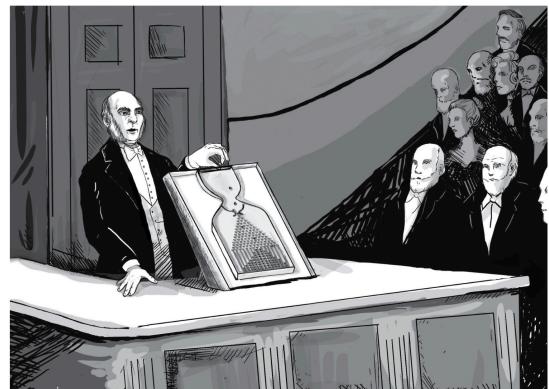
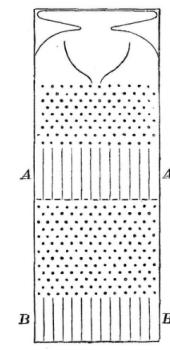


Is Artificial Intelligence dependable?

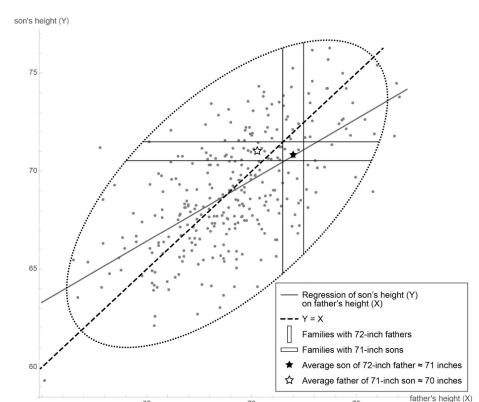
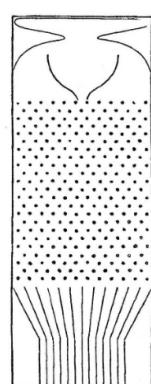
1

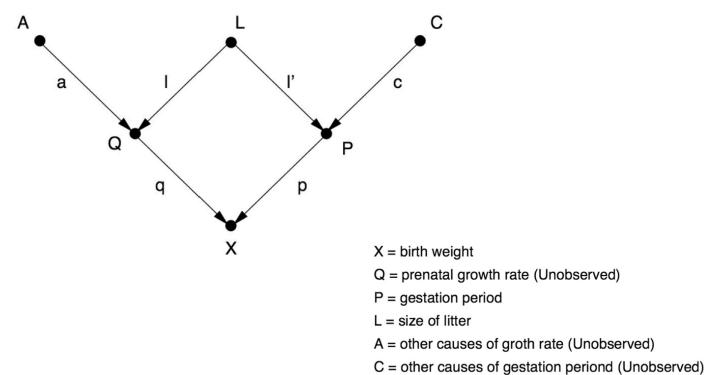
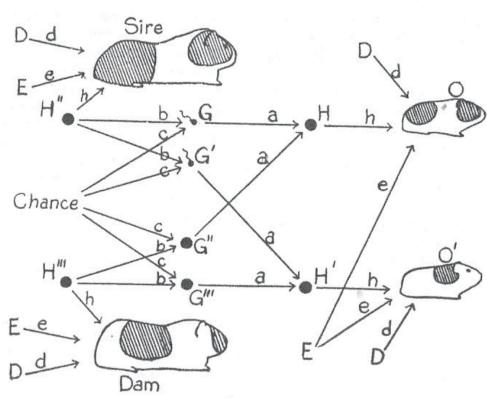
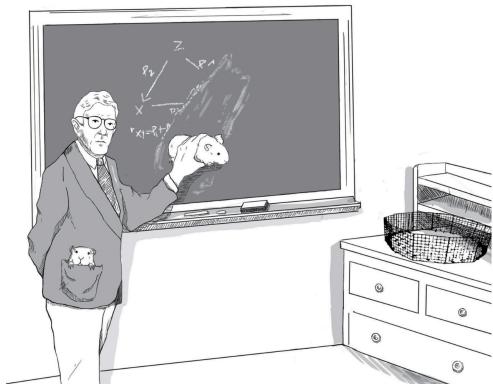
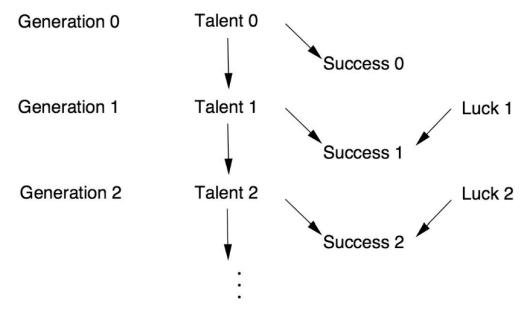
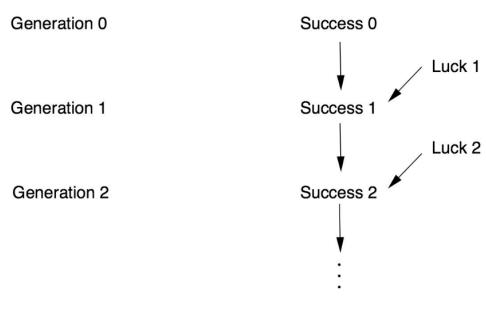


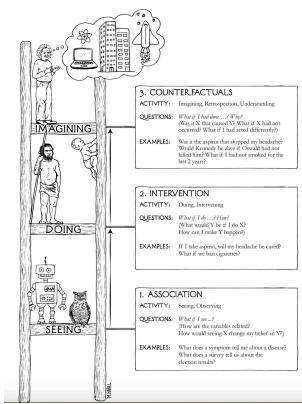
3



4

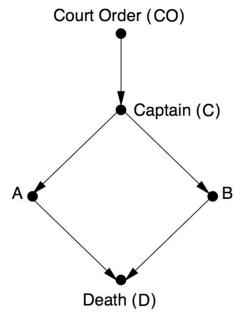






Example of a Causal Diagram

- A prisoner is about to be executed by a firing squad.
- Chain of events:
 - Court orders execution
 - Captain signals the soldiers on the firing squad (A and B) to fire.
 - A and B only fire on command
 - If either of them shoots, the prisoner dies.

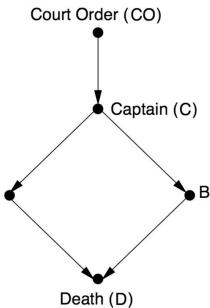


14

Answering Questions on Association

- Questions on Association — What one fact tells us about another.
- If the prisoner is dead, does that mean the court order was given?

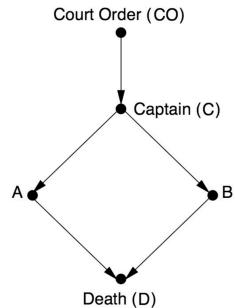
Trace the rules behind each arrow.



15

Answering Questions on Intervention

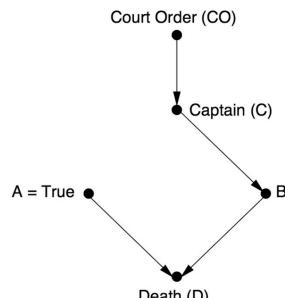
- What if soldier A decides on his own initiative to fire, without waiting for the captain's command?
- The computer (causal reasoner) needs to make an event happen.



16

Answering Questions on Intervention

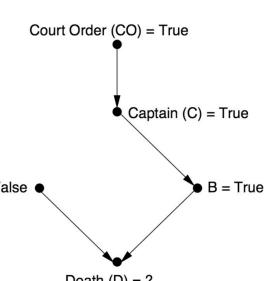
- Erase all the arrows leading into the intervened variable A.
- Set that variable manually to its prescribed value.
- Evaluate the intervention using the resulting causal diagram.
 - Prisoner will be dead
 - B did not shoot.



17

Answering Questions on Counterfactual Reasoning

- Suppose the prisoner is lying dead on the ground. From this we conclude that A shot B shot, the captain gave the signal, the court gave the order.
- But what if A had decided not to shoot? Would the prisoner be alive?



18

Potential outcomes (counterfactuals)

An example

Employee (u)	EX (u)	ED(u)	S0(u)	S1(u)	S2(u)
Alice	6	0	81000	?	?
Bert	9	1	?	92500	?
Caroline	9	2	?	?	97000
David	8	1	?	91000	?
Ernest	12	1	?	100000	?
Frances	13	0	97000	?	?

Let's predict Alice's **potential salary** if she had a graduate degree

20

Apply linear regression

Employee (u)	EX (u)	ED(u)	S0(u)	S1(u)	S2(u)
Alice	6	0	81000	?	?
Bert	9	1	?	92500	?
Caroline	9	2	?	?	97000
David	8	1	?	91000	?
Ernest	12	1	?	100000	?
Frances	13	0	97000	?	?

$$S(\text{Alice}) = 65000 + 2500 \times EX + 5000 \times ED$$

$$S(\text{Alice}) = 65000 + 2500 \times 6 + 5000 \times 1 = 85000$$

21

Apply linear regression

Employee (u)	EX (u)	ED(u)	S0(u)	S1(u)	S2(u)
Alice	6	0	81000	?	?
Bert	9	1	?	92500	?
Caroline	9	2	?	?	97000
David	8	1	?	91000	?
Ernest	12	1	?	100000	?
Frances	13	0	97000	?	?

$$S(\text{Alice}) = 65000 + 2500 \times 6 + 5000 \times 1 = 85000$$

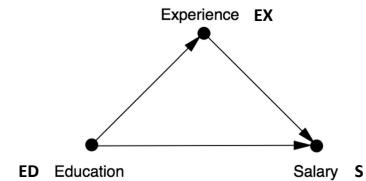
No method based only on data can answer counterfactual questions!

22

- We need to use causal modelling !
- Note that experience is likely to depend on education.
 - Employees who got an extra educational degree took four years of their lives to do so.

23

- We need to use causal modelling !
- Note that experience is likely to depend on education.
 - Employees who got an extra educational degree took four years of their lives to do so.



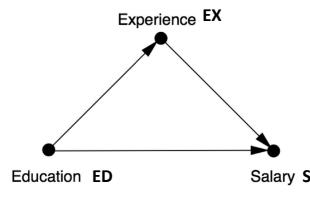
24

- We will write structural equations for Salary and Experience.

- $S = f_s(EX, ED)$
- $S = f_s(EX, ED, U_s)$ (still better)

$$S = 65000 + 2500 \times EX + 5000 \times ED + U_s$$

- $EX = 10 - 4 \times ED + U_{EX}$



25

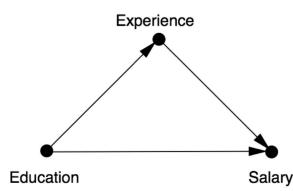
Step 1

- We observe from data that $EX(Alice)=6$ and $ED(Alice)=0$
- We substitute these values into
 $S = 65000 + 2500 \times EX + 5000 \times ED + U_s$ and
 $EX = 10 - 4 \times ED + U_{EX}$
to obtain $U_s(Alice) = 1000$ and $U_{EX}(Alice) = -4$

26

Step 2 (Perform intervention)

- Use the do-operator to erase the arrows pointing to the variable that is being set to a counterfactual value (Education) and set Alice's education to a college degree (Education = 1).



27

Step 3

- Update the model to reflect the new information
 $U_s = 1000$
 $U_{EX} = -4$
 $ED = 1$

$$S = 65000 + 2500 \times EX + 5000 \times ED + U_s \quad \text{and}$$

$$EX = 10 - 4 \times ED + U_{EX}$$

$$EX = 10 - 4 - 4 = 2 \text{ years}$$

$$S = 65000 + 2500 \times 2 + 5000 \times 1 + 1000 = 76000$$

28

Conclusions

- Counterfactual outcome we obtain from structural method follows logically from the assumptions displayed in the model.
- The answer obtained from data-driven method tends to depict spurious correlations because it leaves important modelling assumptions unaccounted for.
- There is no such thing as free lunch!

29