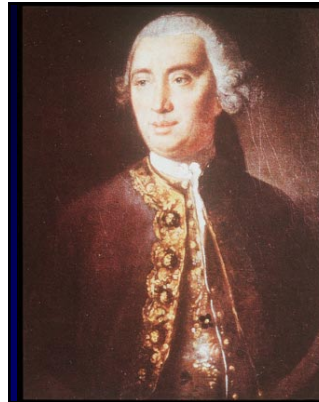


REASONING WITH CAUSE AND EFFECT

Judea Pearl
University of California
Los Angeles



David Hume
(1711–1776)

HUME'S LEGACY

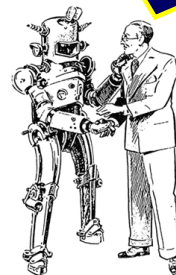
1. Analytical vs. empirical claims
2. Causal claims are empirical
3. All empirical claims originate from experience.

THE TWO RIDDLES OF CAUSATION

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



"Easy, man! that hurts!"



The Art of
Causal Mentoring

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

WHAT'S IN A CAUSAL MODEL?

Oracle that assigns truth value to causal sentences:

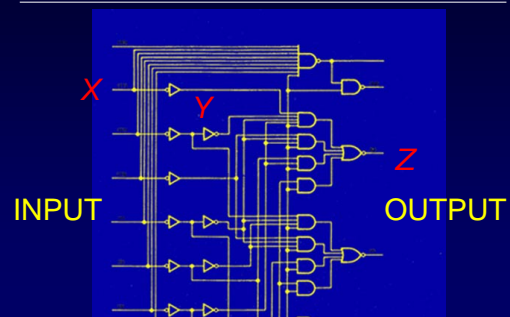
Action sentences: B if we do A .

Counterfactuals: $\neg B \Rightarrow B$ if it were A .

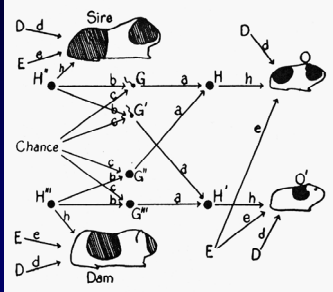
Explanation: B occurred **because** of A .

Optional: with what **probability**?

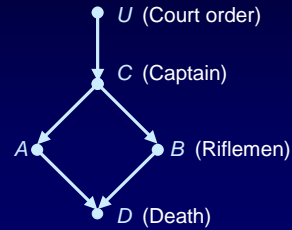
CAUSAL MODELS WHY THEY ARE NEEDED



GENETIC MODELS (S. WRIGHT, 1920)

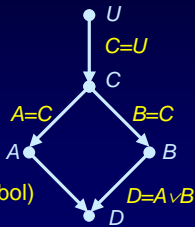


CAUSAL MODELS AT WORK (The impatient firing-squad)



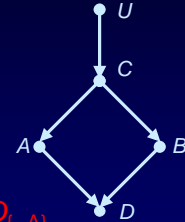
CAUSAL MODELS AT WORK (Glossary)

- U: Court orders the execution
- C: Captain gives a signal
- A: Rifleman-A shoots
- B: Rifleman-B shoots
- D: Prisoner dies



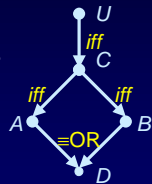
SENTENCES TO BE EVALUATED

- 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: Caused(A, D)



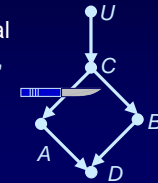
STANDARD MODEL FOR STANDARD QUERIES

- 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

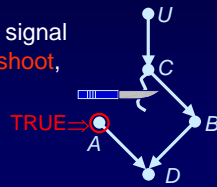
S4. (action):

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:

$\neg C \Rightarrow \neg B_A$



MUTILATION IN SYMBOLIC CAUSAL MODELS

Model M_A (Modify $A=C$):

$C = U$	(U)
$A = C$	(A)
$B = C$	(B)
$D = A \vee B$	(D)



Facts: $\neg C$

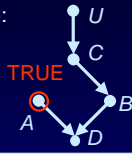
Conclusions: ?

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

MUTILATION IN SYMBOLIC CAUSAL MODELS

Model M_A (Modify $A=C$):

$C = U$	(U)
$A = C$	(A)
$B = C$	(B)
$D = A \vee B$	(D)



Facts: $\neg C$

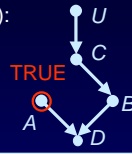
Conclusions: ?

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

MUTILATION IN SYMBOLIC CAUSAL MODELS

Model M_A (Modify $A=C$):

$C = U$	(U)
$A = C$	(A)
$B = C$	(B)
$D = A \vee B$	(D)



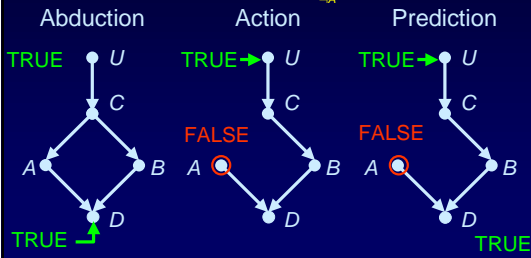
Facts: $\neg C$

Conclusions: $A, D, \neg B, \neg U, \neg C$

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

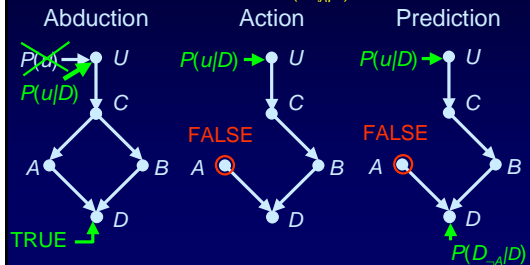
3-STEPS TO COMPUTING COUNTERFACTUALS

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. $P(D_{\neg A}|D) = ?$



SYMBOLIC EVALUATION OF COUNTERFACTUALS

Prove: $D \Rightarrow D_{\neg A}$

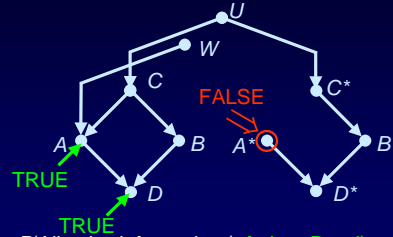
Combined Theory:

$C^* = U$	$C = U$	(U)
$\neg A^*$	$A = C$	(C)
$B^* = C^*$	$A = C$	(A)
$D^* = A^* \vee B^*$	$B = C$	(B)
	$D = A \vee B$	(D)

Facts: D

Conclusions: $U, A, B, C, D, \neg A^*, C^*, B^*, D^*$

PROBABILITY OF COUNTERFACTUALS THE TWIN NETWORK



$P(\text{Alive had A not shot} \mid \text{A shot, Dead}) =$
 $P(\neg D)$ in model $\langle M_{\neg A}, P(u, w \mid A, D) \rangle =$
 $P(\neg D^* \mid D)$ in twin-network

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 $\{U \times V \setminus V_i \rightarrow V_i\}$
 $v_i = f_i(p a_i, u_i)$

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

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

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:

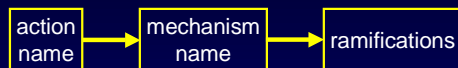
Will the output get higher – if we replace the transistor?

Counterfactual query:

Would the output be higher – had the transistor been replaced?

WHY CAUSALITY? FROM MECHANISMS TO MODALITY

Causality-free specification:



Causal specification:



Prerequisite: one-to-one correspondence between variables and mechanisms

SURGERY IN STRIPS STYLE

Action: $do(V_i = v^*)$

Current state: $V_i(u) = v$

DELETE-LIST

ADD-LIST

$V_i = v$
+ ramifications

$V_i = v^*$
+ ramifications

MECHANISM DELETE-LIST

MECHANISM ADD-LIST

$v_i = f_i(p a_i, u_i)$

$f_i(\cdot) = v^*$

MID-STORY OUTLINE

Background:

From Hume to robotics

Semantics and principles:

Causal models, Surgeries,
Actions and Counterfactuals

Applications I:

Evaluating Actions and Plans
from Data and Theories

Applications II:

Finding Explanations and
Single-event Causation

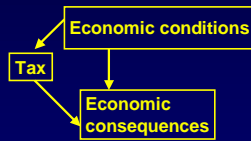
APPLICATIONS

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

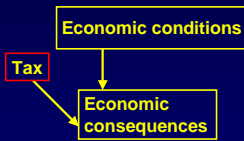
INTERVENTION AS SURGERY

Example: Policy analysis

Model underlying data



Model for policy evaluation



PREDICTING THE EFFECTS OF POLICIES

1. Surgeon General (1964):



$$P(c | do(s)) \approx P(c | s)$$

2. Tobacco Industry:



$$P(c | do(s)) = P(c)$$

3. Combined:



$$P(c | do(s)) = \text{noncomputable}$$

PREDICTING THE EFFECTS OF POLICIES

1. Surgeon General (1964):



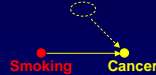
$$P(c | do(s)) \approx P(c | s)$$

2. Tobacco Industry:



$$P(c | do(s)) = P(c)$$

3. Combined:



$$P(c | do(s)) = \text{noncomputable}$$

PREDICTING THE EFFECTS OF POLICIES

1. Surgeon General (1964):



$$P(c | do(s)) \approx P(c | s)$$

2. Tobacco Industry:



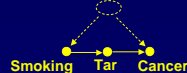
$$P(c | do(s)) = P(c)$$

3. Combined:




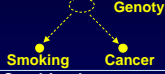
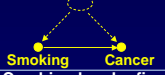
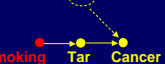
$$P(c | do(s)) = \text{noncomputable}$$

4. Combined and refined:



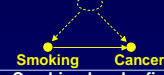



$$P(c | do(s)) = \text{computable}$$

PREDICTING THE EFFECTS OF POLICIES

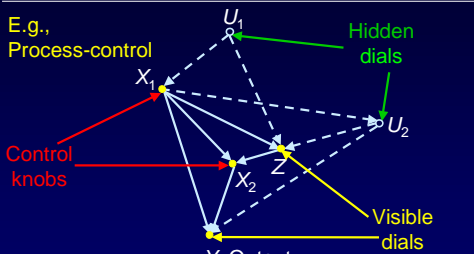
1. Surgeon General (1964): $P(c | do(s)) \approx P(c | s)$

2. Tobacco Industry: $P(c | do(s)) = P(c)$

3. Combined: $P(c | do(s)) = \text{noncomputable}$

4. Combined and refined: $P(c | do(s)) = \text{computable}$


PREDICTING THE EFFECTS OF POLICIES

1. Surgeon General (1964): $P(c | do(s)) \approx P(c | s)$

2. Tobacco Industry: $P(c | do(s)) = P(c)$

3. Combined: $P(c | do(s)) = \text{noncomputable}$

4. Combined and refined: $P(c | do(s)) = \text{computable}$


LEARNING TO ACT BY WATCHING OTHER ACTORS

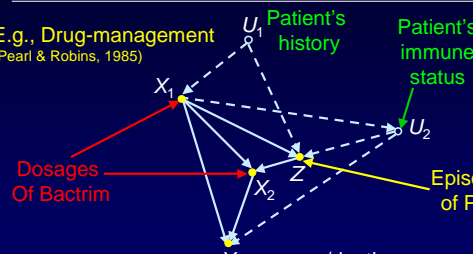
E.g., Process-control



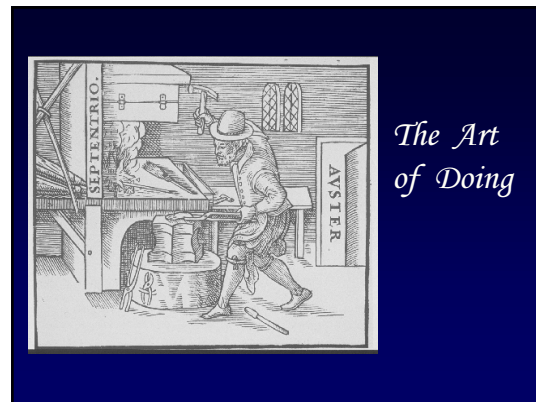
Problem: Find the effect of $(do(x_1), do(x_2))$ on Y , from data on X_1, Z, X_2 and Y .

LEARNING TO ACT BY WATCHING OTHER ACTORS

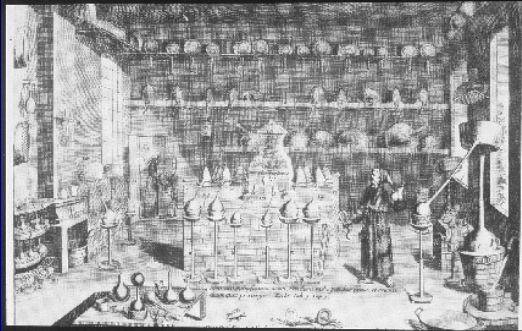
E.g., Drug-management (Pearl & Robins, 1985)



Solution: $P(y | do(x_1), do(x_2)) = \sum_z P(y | z, x_1, x_2) P(z | x_1)$



Combining Seeing and Doing



NEEDED: ALGEBRA OF DOING

Available: algebra of **seeing**

e.g., What is the chance it rained if we **see** the grass wet?

$$P(\text{rain} | \text{wet}) = ? \quad \{= P(\text{wet} | \text{rain}) \frac{P(\text{rain})}{P(\text{wet})}\}$$

Needed: algebra of **doing**

e.g., What is the chance it rained if we **make** the grass wet?

$$P(\text{rain} | \text{do}(\text{wet})) = ? \quad \{= P(\text{rain})\}$$

RULES OF CAUSAL CALCULUS

Rule 1: Ignoring observations

$$P(y | \text{do}(x), z, w) = P(y | \text{do}(x), w)$$

if $(Y \perp\!\!\!\perp Z | X, W)_{G_{\bar{z}}}$

Rule 2: Action/observation exchange

$$P(y | \text{do}(x), \text{do}(z), w) = P(y | \text{do}(x), z, w)$$

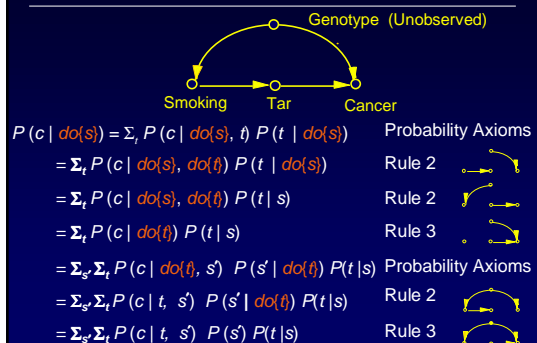
if $(Y \perp\!\!\!\perp Z | X, W)_{G_{\bar{z}}}$

Rule 3: Ignoring actions

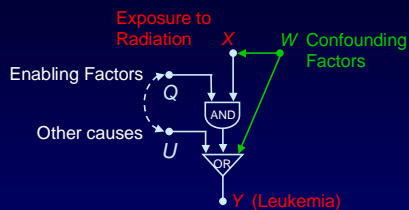
$$P(y | \text{do}(x), \text{do}(z), w) = P(y | \text{do}(x), w)$$

if $(Y \perp\!\!\!\perp Z | X, W)_{G_{\bar{z}, \bar{w}}}$

DERIVATION IN CAUSAL CALCULUS



LEGAL ATTRIBUTION: WHEN IS A DISEASE DUE TO EXPOSURE?



BUT-FOR criterion: $PN = P(Y_x \neq y | X = x, Y = y) > 0.5$

Q. When is PN identifiable from $P(x, y)$?

A. No confounding + monotonicity

$$PN = [P(y | x) - P(y' | x')] / P(y | x) + \text{correction}$$


APPLICATIONS-II

4. Finding explanations for reported events
5. Generating verbal explanations
6. Understanding causal talk
7. Formulating theories of causal thinking



Causal Explanation

"She handed me the fruit and I ate"



Causal Explanation

"She handed me the fruit and I ate"

"The serpent deceived me, and I ate"

ACTUAL CAUSATION AND THE COUNTERFACTUAL TEST

"We may define a **cause** to be an object followed by another,...., where, if the first object **had not been**, the second never had existed."

Hume, *Enquiry*, 1748

Lewis (1973): "x **CAUSED** y" if x and y are true, and y is false in the closest non-x-world.

Structural interpretation:

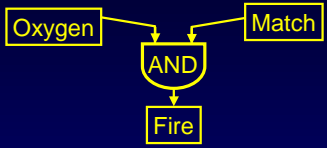
- (i) $X(u)=x$
- (ii) $Y(u)=y$
- (iii) $Y_{x'}(u) \neq y$ for $x' \neq x$

PROBLEMS WITH THE COUNTERFACTUAL TEST

1. **NECESSITY** –
Ignores aspects of sufficiency (Production)
Fails in presence of other causes (Overdetermination)
2. **COARSENESS** –
Ignores structure of intervening mechanisms.
Fails when other causes are preempted (Preemption)

SOLUTION:
Supplement counterfactual test with **Sustenance**

THE IMPORTANCE OF SUFFICIENCY (PRODUCTION)



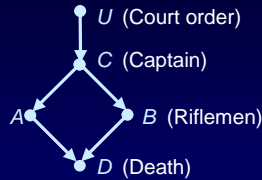
```

graph TD
    O[Oxygen] --> AND{AND}
    M[Match] --> AND
    AND --> F[Fire]
  
```

Observation: Fire broke out.
Question: Why is oxygen an awkward explanation?
Answer: Because Oxygen is (usually) not sufficient

$P(\text{Oxygen is sufficient}) = P(\text{Match is lighted}) = \text{low}$
 $P(\text{Match is sufficient}) = P(\text{Oxygen present}) = \text{high}$

OVERDETERMINATION: HOW THE COUNTERFACTUAL TEST FAILS?

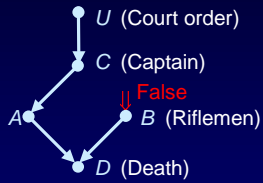


```

graph TD
    U((U (Court order))) --> C((C (Captain)))
    C --> A((A))
    C --> B((B (Riflemen)))
    A --> D((D (Death)))
    B --> D
  
```

Observation: Dead prisoner with two bullets.
Query: Was A a cause of death?
Answer: Yes, A **sustains** D against B.

OVERDETERMINATION: HOW THE SUSTENANCE TEST SUCCEEDS?



Observation: Dead prisoner with two bullets.
 Query: Was A a cause of death?
 Answer: Yes, A **sustains** D against B.

NUANCES IN CAUSAL TALK

y **depends** on x (in u)
 $X(u)=x, Y(u)=y, Y_{x'}(u)=y'$

x can **produce** y (in u)
 $X(u)=x', Y(u)=y', Y_x(u)=y$

x **sustains** y relative to $W=w'$
 $X(u)=x, Y(u)=y, Y_{xw'}(u)=y, Y_{x'w'}(u)=y'$

NUANCES IN CAUSAL TALK

y **depends** on x (in u)
 $X(u)=x, Y(u)=y, Y_{x'}(u)=y'$

x can produce y (in u)
 $X(u)=x', Y(u)=y', Y_x(u)=y$

x **sustains** y relative to $W=w'$
 $X(u)=x, Y(u)=y, Y_{xw'}(u)=y, Y_{x'w'}(u)=y'$

x caused y,
 necessary for,
 responsible for,
 y due to x,
 y attributed to x.

NUANCES IN CAUSAL TALK

y depends on x (in u)
 $X(u)=x, Y(u)=y, Y_{x'}(u)=y'$

x can **produce** y (in u)
 $X(u)=x', Y(u)=y', Y_x(u)=y$

x sustains y relative to $W=w'$
 $X(u)=x, Y(u)=y, Y_{xw'}(u)=y, Y_{x'w'}(u)=y'$

x causes y,
 sufficient for,
 enables,
 triggers,
 brings about,
 activates,
 responds to,
 susceptible to.

NUANCES IN CAUSAL TALK

y depends on x (in u)
 $X(u)=x, Y(u)=y, Y_{x'}(u)=y'$

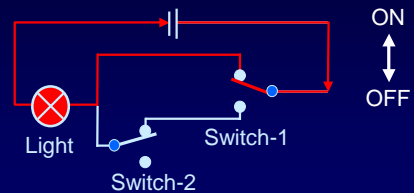
x can produce y (in u)
 $X(u)=x', Y(u)=y', Y_x(u)=y$

x **sustains** y relative to $W=w'$
 $X(u)=x, Y(u)=y, Y_{xw'}(u)=y, Y_{x'w'}(u)=y'$

maintain,
 protect,
 uphold,
 keep up,
 back up,
 prolong,
 support,
 rests on.

PREEMPTION: HOW THE COUNTERFACTUAL TEST FAILS

Which switch is the **actual cause** of light? $S_1!$



Deceiving symmetry: $Light = S_1 \vee S_2$

PREEMPTION: HOW THE COUNTERFACTUAL TEST FAILS

Which switch is the **actual cause** of light? $S_1!$

Deceiving symmetry: $Light = S_1 \vee S_2$

PREEMPTION: HOW THE COUNTERFACTUAL TEST FAILS

Which switch is the **actual cause** of light? $S_1!$

Deceiving symmetry: $Light = S_1 \vee S_2$

PREEMPTION: HOW THE COUNTERFACTUAL TEST FAILS

Which switch is the **actual cause** of light? $S_1!$

Deceiving symmetry: $Light = S_1 \vee S_2$

PREEMPTION: HOW THE COUNTERFACTUAL TEST FAILS

Which switch is the **actual cause** of light? $S_1!$

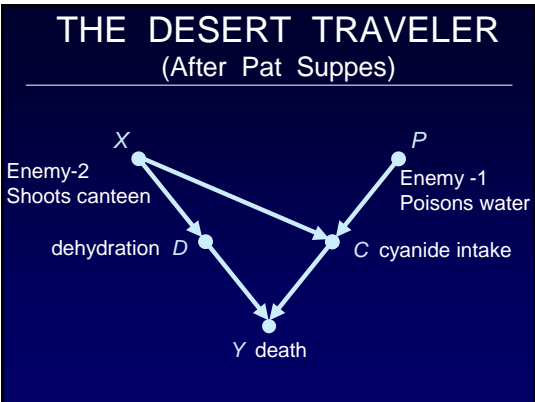
Deceiving symmetry: $Light = S_1 \vee S_2$

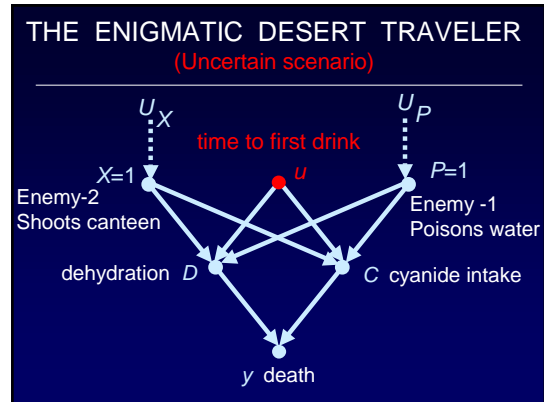
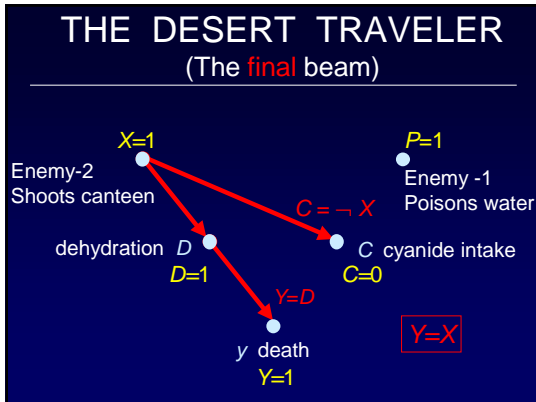
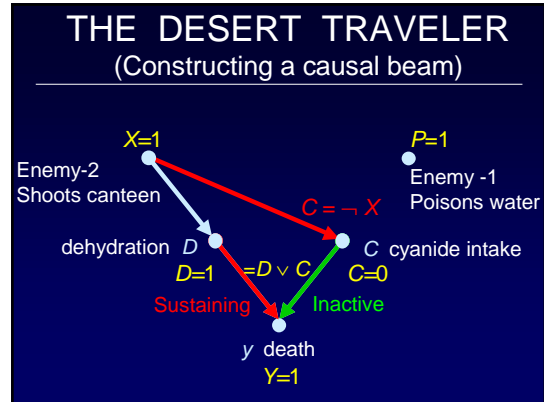
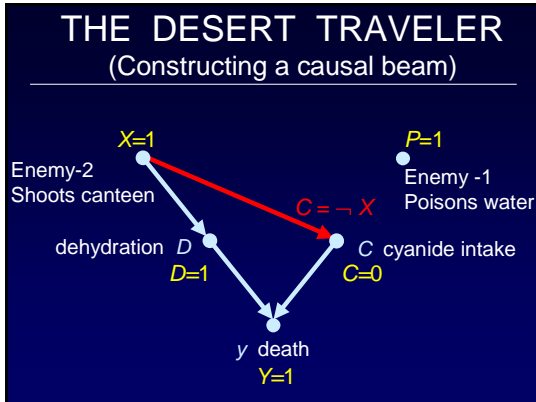
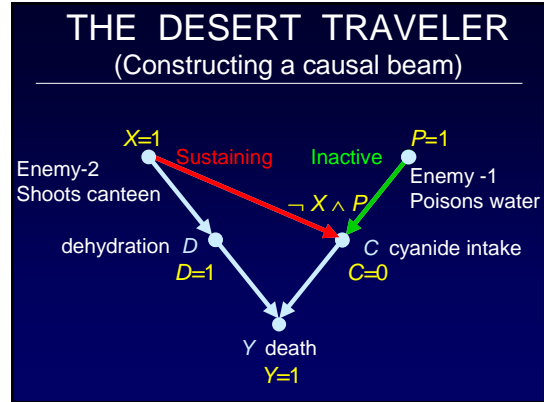
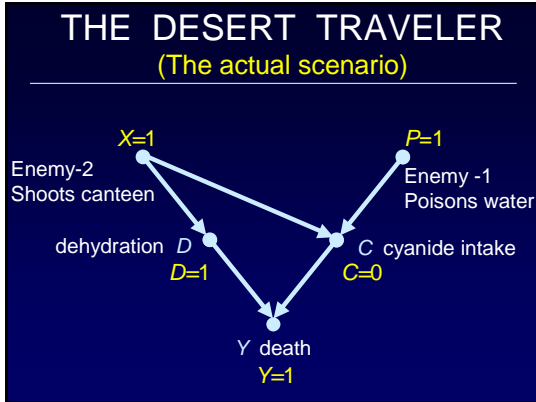
CAUSAL BEAM

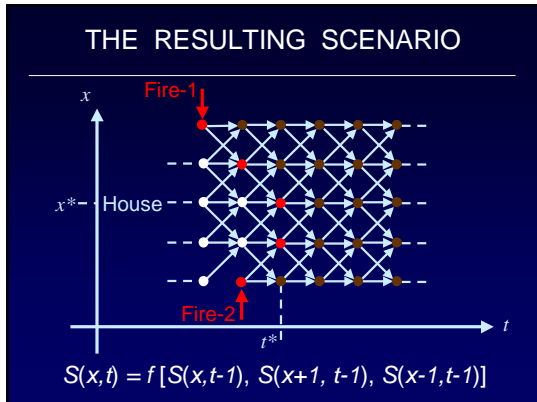
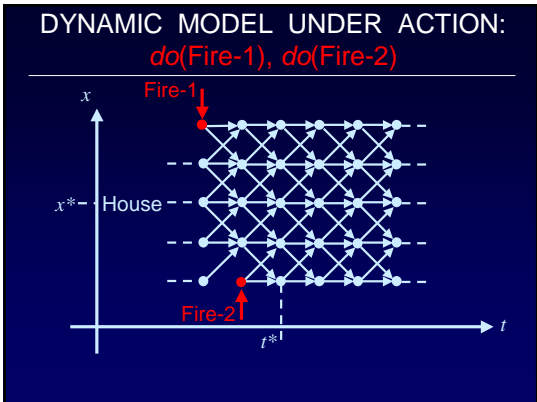
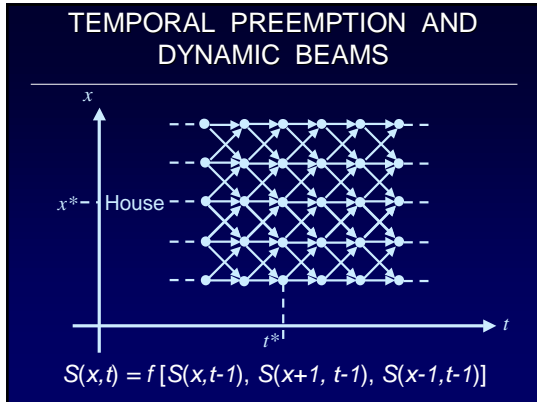
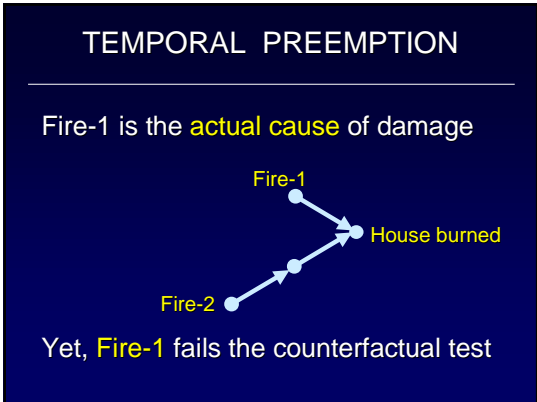
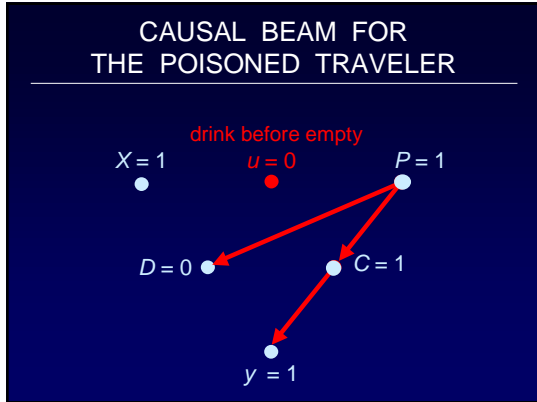
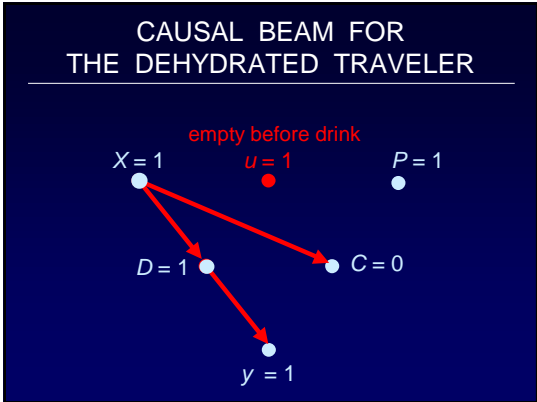
Locally sustaining sub-process

ACTUAL CAUSATION
 "x is an actual cause of y" in scenario u, if x passes the following test:

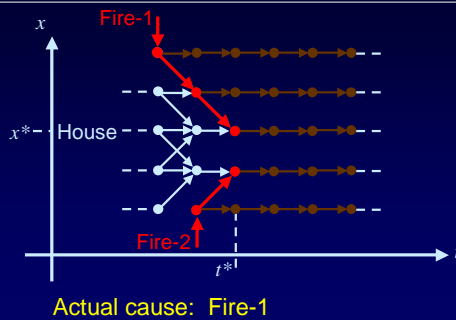
1. Construct a new model $Beam(u, w')$
 - 1.1 In each family, retain a subset of parents that minimally sustains the child
 - 1.2 Set the other parents to some value w'
2. Test if x is necessary for y in $Beam(u, w')$ for some w'







THE DYNAMIC BEAM



CONCLUSIONS

"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

CONCLUSIONS

"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

ACKNOWLEDGEMENT-I

Collaborators in Causality:

Alex Balke	Moisés Goldszmidt
David Chickering	Sander Greenland
Adnan Darwiche	David Heckerman
Rina Dechter	Jin Kim
Hector Geffner	Jamie Robins
David Galles	Tom Verma

ACKNOWLEDGEMENT-II

Influential ideas:

S. Wright (1920)	P. Spirtes, C. Glymour & R. Scheines (1993)
T. Haavelmo (1943)	P. Nayak (1994)
H. Simon (1953)	F. Lin (1995)
I.J. Good (1961)	D. Heckerman & R. Shachter (1995)
R. Strotz & H. Wold (1963)	N. Hall (1998)
D. Lewis (1973)	J. Halpern (1998)
R. Reiter (1987)	D. Michie (1998)
Y. Shoham (1988)	
M. Druzdel & H. Simon (1993)	