REVISED
3:03 pm, Mar 21, 2012

harvey

THE ALGORITHMIZATION OF CAUSES AND COUNTERFACTUALS

Judea Pearl
UCLA

Harvey Prize Lecture, Technion
March 29, 2012

OUTLINE

1. From Bayesian Networks to causes and interventions
2. From causes to counterfactuals
3. Applications
   a. Man-machine communication
   b. Re-tooling the empirical sciences

BAYESIAN NETWORKS A PERSONAL HISTORY

1. Reading comprehension (Rumelhart 1976)
2. Trees (1982)
3. Polytrees (Kim, 1983)
4. Loop-breaking (1985)
5. *d*-separation / graphoids (Paz, Verma, Geiger 1985- -)
6. Belief propagation (1988)
7. Applications (1985 - -)
8. From beliefs to actions (1991)

TOWARD AN INTERACTIVE MODEL OF READING
David E. Rumelhart

CENTER FOR HUMAN INFORMATION PROCESSING

UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093

CHIP 56
MARCH 1976

Rumelhart (1976)
Figure 3

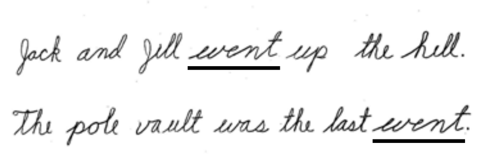


Figure 3 The dependence of letter perception of context. (After Nash-Weber, 1975.)

Rumelhart (1976)
Figure 10

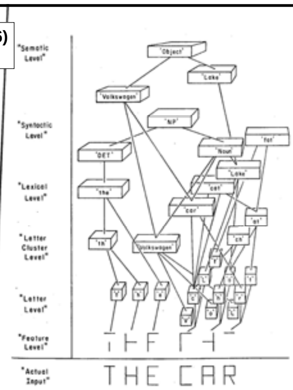


Figure 10 The message center well into the processing sequence.

Pearl (1982), (Belief Propagation in trees)

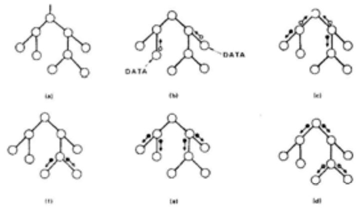


Figure 2

Properties of the Updating Scheme

1. The local computations required by the proposed scheme are efficient in both storage and time. For an m-ary tree with n states per node, each processor should store n^2+mn+2n real numbers, and perform $2n^2$+mn+2n multiplications per update. These expressions are on the order of the number of rules which each variable invokes.

Kim & Pearl (1983)
Explaining a way

data which we shall call D_R and ...vely. These data are defined as the observations and prior beliefs obtained only at the boundaries of network. Likewise, every node A partitions the graph into two parts: above A, G ., and below A, G_A, representing the data set D and D respectively. Figure 1 shows the causal network representing Mr. Holmes' belief structure.

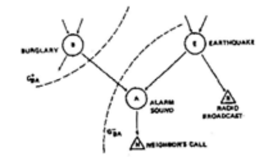


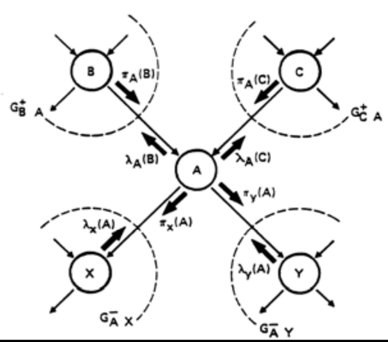
Figure 1 : Mr. Holmes' Belief Structure

III STRUCTURAL ASSUMPTIONS OF INDEPENDENCE

The likelihood of the various states of a variable X would, in general, depend on the entire data observed so far. However, the existence of only one path from G to X implies that the

which m... priori. certaint... this ... (See Fi... The i... to work... are vie... their ... coupled... burglari... earthqu... likeliho... the oc... relation... 2-b), a... (5) P

BELIEF PROPAGATION IN POLYTREES



BAYESIAN NETWORKS: A MODEL OF SELF-ACTIVATED MEMORY FOR EVIDENTIAL REASONING*

Bayes Net (1985)

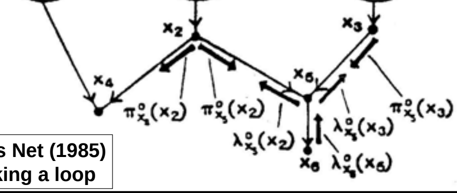
Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024
(judea@UCLA-locus)
(213) 825-3243

Topics: Memory Models
Belief Systems
Inference Mechanisms
Knowledge Representation

To be presented at
the 7th Conference of
the Cognitive Science Society
University of California, Irvine
August 15-17, 1985

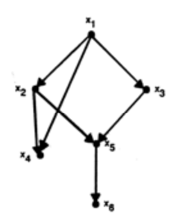
$$\pi^0_{x_6}(x_5) = \sum_{x_2, x_3 = 0,1} P(x_5 \mid x_2, x_3)\pi^0_{x_5}(x_2)\pi^0_{x_5}(x_3)$$

$$\pi^1_{x_6}(x_5) = \sum_{x_2, x_3 = 0,1} P(x_5 \mid x_2, x_3)\pi^1_{x_5}(x_2)\pi^1_{x_5}(x_3)$$



Bayes Net (1985)
Breaking a loop

BAYES NET – CONSTRUCTION AND *d*-SEPARATION



Simple construction:

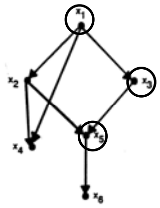
$$P(x_1, \ldots, x_n) = \prod_{i=1,\ldots,n} P(x_i \mid x_1, \ldots, x_{i-1})$$
$$= \prod_{i=1,\ldots,n} P(x_i \mid pa_i)$$

Qualitative judgment:
conditional independence

$$(X_i \perp\!\!\!\perp \text{PRED}_i \mid pa_i)$$

$$P(x_1, x_2, x_3, x_4, x_5, x_6) = P(x_6 \mid x_5) P(x_5 \mid x_2, x_3)$$
$$P(x_4 \mid x_1, x_2) P(x_3 \mid x_1) P(x_2 \mid x_1) P(x_1)$$

BAYES NET – CONSTRUCTION AND d-SEPARATION (cont)



$(X_i \perp\!\!\!\perp \text{PRED}_i \mid pa_i)$ conditional independence (by construction)

What else?

d-separation: $X \perp\!\!\!\perp Y \mid Z$ if every path from X to Y is d-separated by Z (Verma, Geiger & Pearl, 1987).

Normal valve: tail-to-tail, $x_1 \leftarrow (x_2) \rightarrow x_3$

Normal valve: head-to-tail, $x_1 \rightarrow (x_2) \rightarrow x_3$ or $x_1 \leftarrow (x_2) \leftarrow x_3$

Abnormal valve: head-to-head, $x_1 \rightarrow (x_2) \leftarrow x_3$

A path is d-separated by Z if all its valves are opened by Z.

CONDITIONAL INDEPENDENCE, GRAPH SEPARATION, AND GRAPHOIDS (Dawid, Spohn, Pearl, Paz, Geiger)

The Graphoid Axioms:

- *Symmetry:*
  $I(X, Z, Y) \Longleftrightarrow I(Y, Z, X)$ **(3.6*a*)**
- *Decomposition:*
  $I(X, Z, Y \cup W) \Longrightarrow I(X, Z, Y) \;\&\; I(X, Z, W)$ **(3.6*b*)**
- *Weak Union:*
  $I(X, Z, Y \cup W) \Longrightarrow I(X, Z \cup W, Y)$ **(3.6*c*)**
- *Contraction:*
  $I(X, Z, Y) \;\&\; I(X, Z \cup Y, W) \Longrightarrow I(X, Z, Y \cup W)$. **(3.6*d*)**

If P is strictly positive, then a fifth condition holds:

- *Intersection:*
  $I(X, Z \cup W, Y) \;\&\; I(X, Z \cup Y, W) \Longrightarrow I(X, Z, Y \cup W)$. **(3.6*e*)**

GRAPHICAL INTERPRETATION OF THE AXIOMS GOVERNING CONDITIONAL INDEPENDENCE

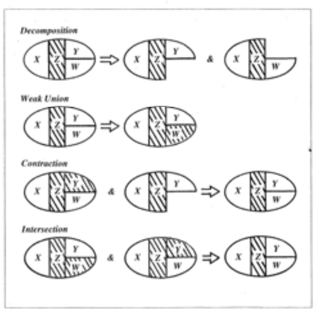


Figure 3.1. *Graphical interpretation of the axioms governing conditional independence.*

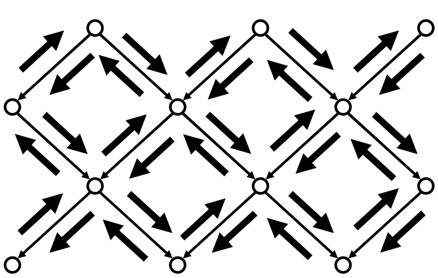
LOOP-BREAKING TECHNIQUES (1985 – 1990)

1. Conditioning (1985)
2. Stochastic simulation (1987)
3. Tree clustering (Spiegelhalter & Lauritzen 1986)
4. Node elimination (Shachter 1986)

Problems:

- Time exponential in tree-width (Dechter 1987)
- Autonomy is lost

BELIEF PROPAGATION WHEN THERE ARE LOOPS



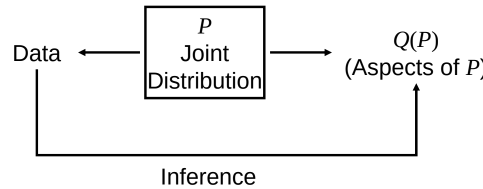
APPLICATIONS OF BAYESIAN NETWORKS

1. Medical Diagnosis
2. Clinical Decision Support
3. Complex Genetic Models
4. Crime Risk Factors Analysis
5. Spatial Dynamics in Geography
6. Inference Problems in Forensic Science
7. Conservation of a Threatened Bird
8. Classifiers for Modelling of Mineral Potential
9. Student Modelling
10. Sensor Validation
11. An Information Retrieval System
12. Reliability Analysis of Systems
13. Terrorism Risk Management
14. Credit-Rating of Companies
15. Classification of Wines
16. Pavement and Bridge Management
17. Complex Industrial Process Operation
18. Probability of Default for Large Corporates
19. Risk Management in Robotics

FROM BAYESIAN NETS TO CAUSES AND COUNTERFACTUALS

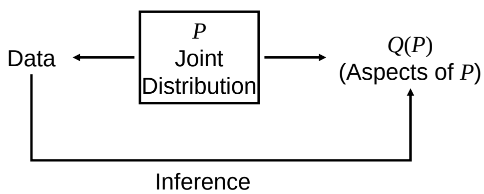
1. Distinctions and mental barriers
2. Effects of interventions
3. *do*-calculus
4. Graphical identifiability
5. Why counterfactuals
6. The causal hierarchy
   What / What-if / Why
7. Applications:
   Evaluation / Mediation / Generalization

BEYOND EVIDENCE, BELIEF, AND STATISTICS



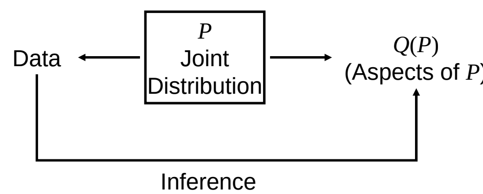
e.g.,
Infer whether customers who bought product ***A*** would also buy product ***B***.
$Q = P(B \mid A)$

STATISTICS 1ST LIMITATION INTERVENTION



e.g.,
Infer whether customers who bought product *A* would buy product *B*. If we double the price
$Q = P(B \mid A, do\ (price = 2p_1))$ Not an aspect of *P*.

STATISTICS 2ND LIMITATION RETROSPECTION



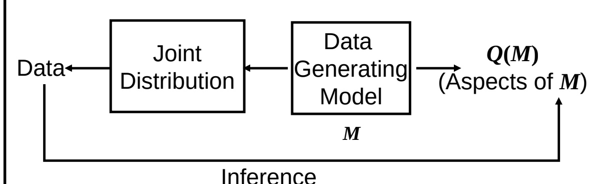
e.g.,
Infer whether Joe who bought product *A* would have bought *A*, had we doubled the price
$Q = P(A_{p_2} \mid A_{p_1})$ Not an aspect of *P*.

THE CAUSAL HIERARCHY

1. Associational (Statistical, Evidential)
   e.g., What if I see $X{=}x$?
2. Interventional (Experimental, Causal)
   e.g., What if I do $X{=}x$?
3. Retrospectional (Counterfactual, token)
   e.g., What if I hadn't done $X{=}x$?

No mixing:
No claim at layer *i* without assumptions from layer *i* or higher.

THE STRUCTURAL MODEL PARADIGM



M – Invariant strategy (mechanism, recipe, law, protocol, counterfactual) by which Nature assigns values to variables in the analysis.

"Think Nature, not experiment!" Why counterfactuals?

WHY PHYSICS NEEDS SPECIAL MATHEMATICS

Scientific Equations (e.g., Hooke's Law) are non-algebraic

e.g., Length (Y) equals a constant (2) times the weight (X)

Correct notation:

Process information	The solution
$Y := 2X$	$X = 1$
~~$X = 1$~~	$Y = 2$

Had X been 3, Y would be 6.
If we raise X to 3, Y would be 6.
Must "wipe out" $X = 1$.

WHY CAUSALITY NEEDS SPECIAL MATHEMATICS

Scientific Equations (e.g., Hooke's Law) are non-algebraic

e.g., Length (Y) equals a constant (2) times the weight (X)

Correct notation:
(or)

Process information	The solution
$Y \leftarrow 2X$	$X = 1$
~~$X = 1$~~	$Y = 2$

Had X been 3, Y would be 6.
If we raise X to 3, Y would be 6.
Must "wipe out" $X = 1$.

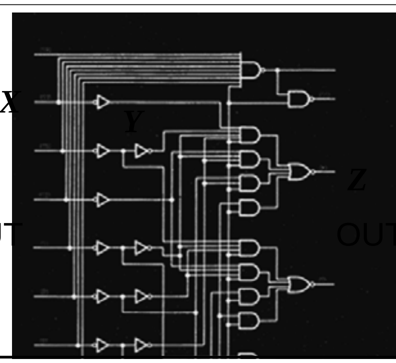
MATHEMATICAL EXTRAPOLATION: THE WORLD AS A COLLECTION OF SPRINGS

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

- $V = \{V_1,...,V_n\}$ are endogeneas variables
- $U = \{U_1,...,U_m\}$ are background variables
- $F = \{f_1,..., f_n\}$ are functions determining V, $v_i = f_i(v, u)$ e.g., $y = \alpha + \beta x + u_Y$
- $P(u)$ is a distribution over U

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

FAMILIAR CAUSAL MODEL ORACLE FOR COUNTERFACTUALS



THE TRAUMA OF DE-CROWNING AN ORACLE

- Ten years ago, when I began writing *Probabilistic Reasoning in Intelligent Systems* (1988), I was working within the empiricist tradition. In this tradition, probabilistic relationships constitute the foundations of human knowledge, whereas causality simply provides useful ways of abbreviating and organizing intricate patterns of probabilistic relationships.

- Today, I take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world.

(Pearl, *Causality*, 2000)

COUNTERFACTUALS ARE EMBARRASINGLY SIMPLE

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.

The Fundamental Equation of Counterfactuals:

$$\boxed{Y_X(u) = Y_{M_X}(u)}$$

COUNTERFACTUALS ARE EMBARRASINGLY SIMPLE

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:

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

In particular:

$$P(y \mid do(x)) \triangleq P(Y_x = y) = \sum_{u: Y_x(u)=y} P(u)$$

$$P(Y_{x'} = y' \mid x, y) = \sum_{u: Y_{x'}(u)=y'} P(u \mid x, y)$$

THE FUNDAMENTAL THEOREM OF CAUSAL INFERENCE

Causal Markov Theorem:
Any distribution generated by Markovian structural model M (recursive, with independent disturbances) can be factorized as

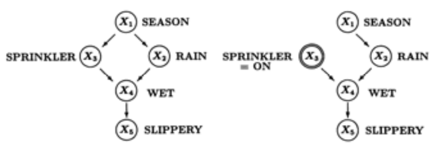
$$P(v_1, v_2, ..., v_n) = \prod_i P(v_i \mid pa_i)$$

Where pa_i are the (values of) the parents of V_i in the causal diagram associated with M.

Corollary: (Truncated factorization, Manipulation Theorem)
The distribution generated by an intervention $do(X=x)$ (in a Markovian model M) is given by the truncated factorization

$$P(v_1, v_2, ..., v_n \mid do(x)) = \prod_{i \mid V_i \notin X} P(v_i \mid pa_i) \Big|_{X=x}$$

SEEING VS. DOING



$$P(x_1, ..., x_n) = \prod_i P(x_i \mid pa_i)$$

$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1)P(x_4 \mid x_2, x_3)P(x_5 \mid x_4)$$

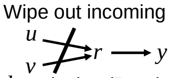
Effect of turning the sprinkler ON

$$P_{X_3=\text{ON}}(x_1, x_2, x_4, x_5) = P(x_1)P(x_2 \mid x_1)P(x_4 \mid x_2, X_3 = \text{ON})P(x_5 \mid x_4)$$

$$\neq P(x_1, x_2, x_4, X_5 \mid X_3 = \text{ON})$$

THE EVOLUTION OF CAUSAL CALCULUS

- Haavelmo's surgery (1943)
  Add adjustable force $(+g_i)$
- Strotz and Wold surgery (1960). "Wipe out" the equation
  $r_i = u_i + v_i$, and replace it with $r_i = \text{constant}$
- Graphical surgery (Spirtes et al., 1993; Pearl, 1993).
  Wipe out incoming arrows to r



$$P(u, v, r, y) = P(u)P(v)P(r \mid u, v)P(y \mid r)$$

- *do*-calculus (Pearl, 1994)
  $P(Y = y \mid do(r))$ new operator
- Structural counterfactuals (Balke and Pearl, 1995)
  $Y_r(u) = Y(u)$ in the r-mutilated model
- Unification with Neyman-Rubin $Y_x(u)$ and Lewis (1973)

AXIOMS OF STRUCTURAL COUNTERFACTUALS

(Galles & Pearl, Halpern, 1998)

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

ILLUSTRATING THE WORKING OF CAUSAL CALCULUS

Model 1 / Model 2 (Linear version)

$Y = f(W_3, Z_3, W_2, u)$ $\quad X = g(W_1, Z_3, u'')$

$W_3 = g_3(X, u'_3)$ $\quad W_1 = g'_1(Z_1, u'_1)$

$Z_3 = f_3(Z_1, Z_2, u_3)$ $\quad Z_1 = f_1(u_1)$

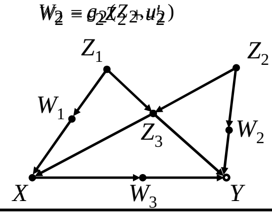
$W_2 = g_2(Z_2, u_2)$ $\quad Z_2 = f_2(u_2)$

$Y = aW_3 + bZ_3 + cW_2 + u$ $\quad X = t_1W_1 + t_2Z_3 + u''$

$W_3 = c_3X + u'_3$ $\quad W_1 = a'_1Z_1 + u'_1$

$Z_3 = a_3Z_1 + b_3Z_2 + u_3$ $\quad Z_1 = u_1$

$W_2 = c_2Z_2 + u_2$ $\quad Z_2 = u_2$



U's are mutually independent

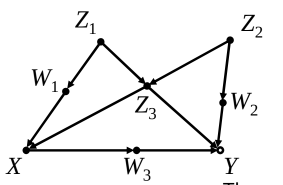
WHAT ARE THE TESTABLE IMPLICATIONS OF THE MODEL?

Model 1 $Y = f(W_3, Z_3, W_2, u)$ $X = g(W_1, Z_3, u")$

$W_3 = g_3(X, u_3'')$ $W_1 = g_1(Z_1, u_1')$

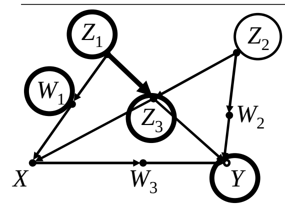
$Z_3 = f_3(Z_1, Z_2, u_3)$ $Z_1 = f_1(u_1)$

$W_2 = g_2(Z_2, u_2')$ $Z_2 = f_2(u_2)$



Missing edges:
$Z_1 - Z_2, Z_1 - Y, Z_2 - X \ldots$
Separating sets:
$\{0\}, \{X, Z_2, Z_3\}, \{Z_1, Z_3\} \ldots$
Testable implications (FINITE!): / Zero partial correlations:
$r_{Z_1 Z_2} = 0$ and $r_{\{YZ_1 \cdot XZ_2Z_3\}} = 0$
$Z_1 \perp\!\!\!\perp Y \mid \{X_1, Z_2, Z_3\}$,
$Z_2 \perp\!\!\!\perp X \mid \{Z_1, Z_3\}$.
These imply *all* misspecification tests

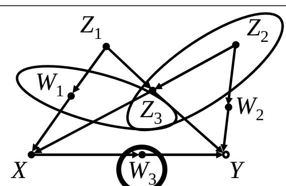
IMPLIED MISSPECIFICATION TESTS



Question 4:
If we regress Z_1 on all other variables in the model, which regression coefficient will be zero?
Answer: All but these three
Question 5:
If we regress Z_1 on Z_3 and W_1, which regression coefficient will change if we add Y as a regressor?

Answer:
The coefficient of Z_3 will change and the coefficient of W_1 will remain invariant. Non-invariance may not be misspecification.

CAUSAL CALCULUS AS AN ORACLE FOR INTERVENTIONS

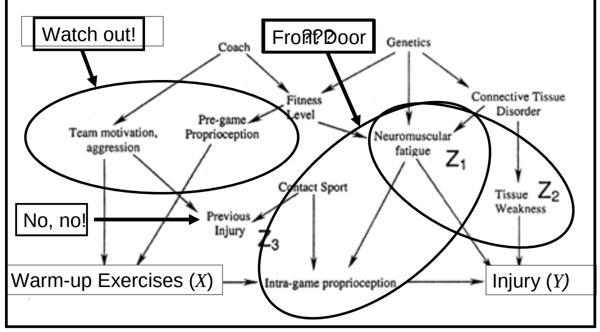


Suppose we wish to estimate the average causal effect of X on Y

$$ACE = P(Y = y \mid do(X = 1)) - P(Y = y \mid do(X = 0)).$$

- Which subsets of variables need to be adjusted to obtain an unbiased estimate of ACE?
- Is there a single variable that, if measured, would allow an unbiased estimate of ACE?

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



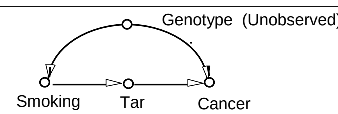
THE MACHINERY OF CAUSAL CALCULUS

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

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

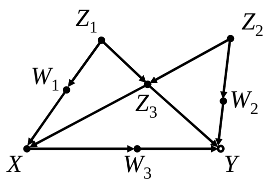
Rule 3: Ignoring actions
P(*y* | *do*{*x*}, *do*{*z*}, *w*) = *P*(*y* | *do*{*x*}, *w*)
if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\,\overline{Z(W)}}}$

DERIVATION IN CAUSAL CALCULUS



$P(c \mid do\{s\}) = \Sigma_t P(c \mid do\{s\}, t)\, P(t \mid do\{s\})$	Probability Axioms
$= \Sigma_t P(c \mid do\{s\}, do\{t\})\, P(t \mid do\{s\})$	Rule 2
$= \Sigma_t P(c \mid do\{s\}, do\{t\})\, P(t \mid s)$	Rule 2
$= \Sigma_t P(c \mid do\{t\})\, P(t \mid s)$	Rule 3
$= \Sigma_{s'} \Sigma_t P(c \mid do\{t\}, s')\, P(s' \mid do\{t\})\, P(t \mid s)$	Probability Axioms
$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s' \mid do\{t\})\, P(t \mid s)$	Rule 2
$= \Sigma_{s'} \Sigma_t P(c \mid t, s')\, P(s')\, P(t \mid s)$	Rule 3

WHAT ELSE CAN CAUSAL CALCULUS DO FOR US?



- Equivalent models
- Identifying Counterfactual queries (ETT, PC)
- Mediation
- Causes of Effects
- External validity

Finding instruments
Is there an instrumental variable for the $\boldsymbol{Z_3 \rightarrow Y}$ relationship?
Answer: No
Can we turn $\boldsymbol{Z_1}$ into an IV?
Answer: Yes, condition on $\boldsymbol{W_1}$.

THE PUZZLE OF COUNTERFACTUAL CONSENSUS

- Indicative: "If Oswald didn't kill Kennedy, someone else did,"
- Subjunctive: "If Oswald hadn't killed Kennedy, someone else would have."

(Adams 1975)

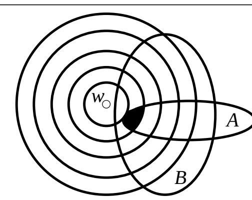
THE PUZZLING UBIQUITY OF COUNTERFACTUALS

Hume's Definition of "cause":
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 1748/1958, sec. VII).

Lewis's Definition of "cause":
"*A* has caused *B*" if "*B* would not have occurred if it were not for *A* (Lewis 1986).

- Why not define counterfactuals in terms of causes? (Pearl 2000)

STRUCTURAL AND SIMILARITY-BASED COUNTERFACTUALS



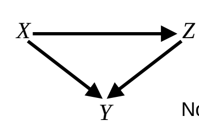
Lewis's account (1973):
The counterfactual "*B* if it were *A*" is true in a world w just in case *B* is true in all the closest *A*-worlds to w.
Structural account (1995):
"*Y* would be y if *X* were x" is true in situation u just in case

MEDIATION: A COUNTERFACTUAL TRIUMPH

Why decompose effects?

1. To understand how Nature works
2. To comply with legal requirements
3. To predict the effects of new type of interventions:
   Signal routing, rather than variable fixing

COUNTERFACTUAL DEFINITION OF INDIRECT EFFECTS



$z = f(x, u)$
$y = g(x, z, u)$

No Controlled Indirect Effect

Indirect Effect of X on Y: $IE(x_0, x_1; Y)$
The expected change in Y when we keep X constant, say at x_0, and let Z change to whatever value it would have attained had X changed to x_1.

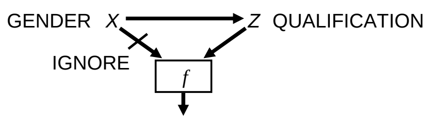
$$E[Y_{x_0 Z_{x_1}} - Y_{x_0}]$$

In linear models, $IE = TE - DE$

POLICY IMPLICATIONS OF INDIRECT EFFECTS

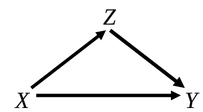
What is the indirect effect of X on Y?

The effect of Gender on Hiring if sex discrimination is eliminated.



Deactivating a link – a new type of intervention

MEDIATION FORMULAS IN UNCONFOUNDED MODELS



$z = f(x, u_1)$
$y = g(x, z, u_2)$
u_1 independent of u_2

$$DE = \sum_z [E(Y \mid x_1, z) - E(Y \mid x_0, z)] P(z \mid x_0)$$

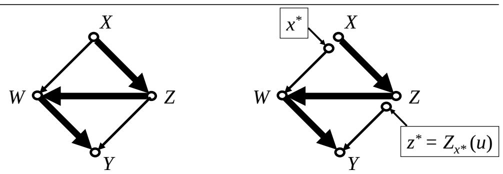
$$IE = \sum_z [E(Y \mid x_0, z)[P(z \mid x_1) - P(z \mid x_0)]$$

$$TE = E(Y \mid x_1) - E(Y \mid x_0) \qquad TE \neq DE + IE$$

$IE =$ Fraction of responses explained by mediation (sufficient)

$TE - DE =$ Fraction of responses owed to mediation (necessary)

GENERAL PATH-SPECIFIC EFFECTS (Def.)



Form a new model, M_g^*, specific to active subgraph g

$$f_i^*(pa_i, u; g) = f_i(pa_i(g), pa_i^*(\bar{g}), u)$$

Definition: g-specific effect

$$E_g(x, x^*; Y)_M = TE(x, x^*; Y)_{M_g^*}$$

Nonidentifiable even in some Markovian models (Avin, Shpitser, & Pearl, 2005)

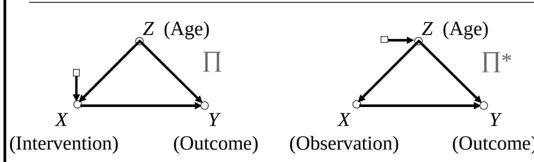
EXTERNAL VALIDITY
From Threats to Licenses

- "`External validity' asks the question of generalizability: To what population, settings, treatment variables, and measurement variables can this effect be generalized?" (Campbell 1963; Shadish, Cook, & Campbell 2002)
- "An experiment is said to have "external validity" if the distribution of outcomes realized by a treatment group is the same as the distribution of outcome that would be realized in an actual program." (Manski 2007)
- "A threat to external validity is an explanation of how you might be wrong in making a generalization." (Wikipedia 2011, after Trochin)
- "A license of validity is a set of theoretical assumptions that neutralizes all conceivable threats." (Anon 2011)

TRANSPORTABILITY ACROSS DOMAINS

1. A Theory of causal transportability
   When can causal relations learned from experiments be transferred to a different environment in which no experiment can be conducted?
2. A Theory of statistical transportability
   When can statistical information learned in one domain be transferred to a different domain in which
   a. only a subset of variables can be observed? Or,
   b. only a few samples are available?

MOTIVATION
WHAT CAN EXPERIMENTS IN LA TELL ABOUT NYC?



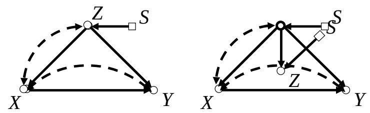
Experimental study in LA
Measured: $P(x, y, z)$
$P(y \mid do(x), z)$

Observational study in NYC
Measured:

Needed: $P^*(y \mid do(x)) =$? $= \sum_z P(y \mid do(x), z) P^*(z)$

Transport Formula (calibration): $F(P, P_{do}, P^*)$

TRANSPORT FORMULAS DEPEND ON THE STORY



$S \rightarrow$ Factors producing differences

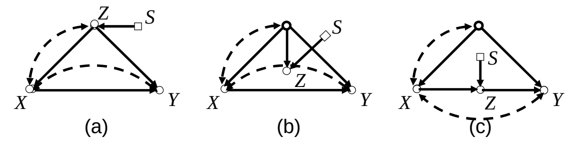
a) $\boldsymbol{Z}$ represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

b) $\boldsymbol{Z}$ represents language skill

$$P^*(y \mid do(x)) = ?P(y \mid do(x))$$

TRANSPORT FORMULAS DEPEND ON THE STORY



a) $\boldsymbol{Z}$ represents age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$

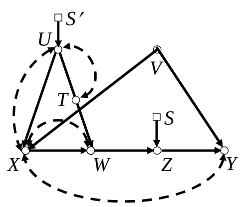
b) $\boldsymbol{Z}$ represents language skill

$$P^*(y \mid do(x)) = P(y \mid do(x))$$

c) $\boldsymbol{Z}$ represents a bio-marker

$$P^*(y \mid do(x)) = ?\sum_z P(y \mid do(x), z) P^*(z \mid x)$$

GOAL: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



INPUT: Annotated Causal Graph
$S \rightarrow$ Factors creating differences

OUTPUT:
1. Transportable or not?
2. Measurements to be taken in the experimental study
3. Measurements to be taken in the target population
4. A transport formula

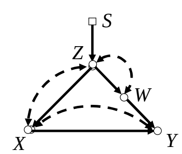
$$P^*(y \mid do(x)) = \sum P(y \mid do(x), z) \sum_w P^*(z \mid w) \sum_t P(w \mid do(x), t) P^*(t)$$

TRANSPORTABILITY REDUCED TO CALCULUS

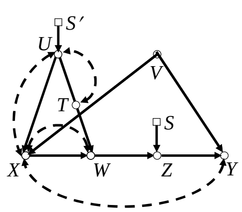
Theorem 1
Let $\boldsymbol{D}$ be the selection diagram characterizing Π and Π^*, and $\boldsymbol{S}$ a set of selection variables in $\boldsymbol{D}$.

A causal relation $\boldsymbol{R}$ is transportable from Π to Π^* if and only if $\boldsymbol{R}(\Pi^*)$ is reducible, using the rules of ***do***-calculus, to an expression in which $\boldsymbol{S}$ appears only as a conditioning variable in ***do***-free terms.

$$\begin{aligned} R(\Pi^*) &= P^*(y \mid do(x)) = P(y \mid do(x), s) \\ &= \sum_w P(y \mid do(x), s, w) P(w \mid do(x), s) \\ &= \sum_w P(y \mid do(x), w) P(w \mid s) \\ &= \sum_w P(y \mid do(x), w) P^*(w) \end{aligned}$$



RESULT: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



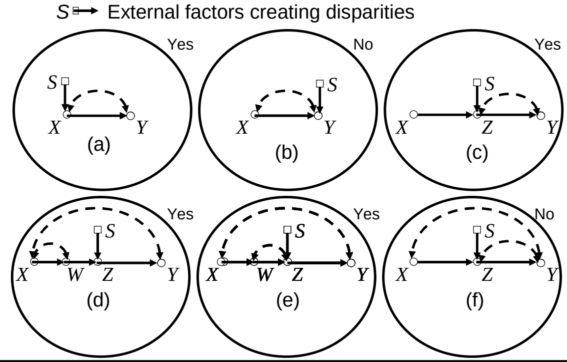
INPUT: Annotated Causal Graph
$S \rightarrow$ Factors creating differences

OUTPUT:
1. Transportable or not?
2. Measurements to be taken in the experimental study
3. Measurements to be taken in the target population
4. A transport formula

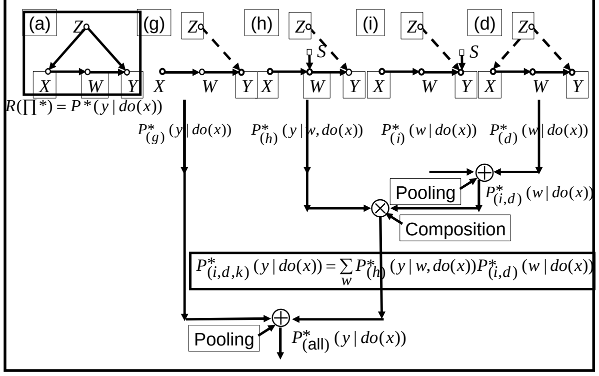
$$P^*(y \mid do(x)) = \sum P(y \mid do(x), z) \sum_w P^*(z \mid w) \sum_t P(w \mid do(x), t) P^*(t)$$

WHICH MODEL LICENSES THE TRANSPORT OF THE CAUSAL EFFECT $X \rightarrow Y$

$S \rightarrow$ External factors creating disparities



COMPOSITIONAL MULTI-SOURCE LEARNING



CONCLUSIONS

- Counterfactuals are the building blocks of scientific thought, free will and moral behavior.
- The algorithmization of counterfactuals thus brings us a step closer to achieving cooperative behavior among robots and humans.
- Moreover, the algorithmization of counterfactuals has benefited several problem areas in the empirical sciences, including policy evaluation, mediation analysis, generalizability, and credit and blame determination.

CONCLUSIONS

"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)."

(Albert Einstein, 1953)

Putting the two together, and basing causal inference on a formal system that is reducible to algorithmic implementation may well be the next development Western science is waiting for.

Thank you