THEORETICAL IMPEDIMENTS TO MACHINE LEARNING
WITH SEVEN SPARKS FROM THE CAUSAL REVOLUTION

Judea Pearl
University of California, Los Angeles
judea@cs.ucla.edu

CAUSAL MODELS AND THE COGNITIVE REVOLUTION

• 10,000 years ago, human beings accounted for less than a tenth of 1 percent of all vertebrate life on planet Earth. Today, that percentage, including livestock and pets, is in the neighborhood of 98! (Daniel Dennett, 2006)
• What Happened?
• What computational facility did humans acquire 10,000 years ago that they did not possess before?

COUNTERFACTUALS: THE HOMOSAPIENS’ SECRET

3-LEVEL HIERARCHY

1. COUNTERFACTUALS
   ACTIVITY: Imaging, Retrospection
   QUESTIONS: What if I had done . . . ? Why?
   (Was it X that caused Y? What if I had not occurred? What if I had acted differently?)
   EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked the last 2 years?

2. INTERVENTION
   ACTIVITY: Doing, Interventing
   QUESTIONS: What if I do . . . ? Why?
   (What would Y be if I do X?)
   EXAMPLES: If I take aspirin, will my headache be cured?

3. ASSOCIATION
   ACTIVITY: Seeing, Observing
   QUESTIONS: What if I see . . . ?
   (How are the variables related?)
   EXAMPLES: What does a symptom tell me about a disease? What does a survey tell us about the election results?

PREDICTION, INTERVENTION, AND COUNTERFACTUALS

Questions:
1. What is the expected value of the demand Q if the price is reported to be \( P = p_0 \)?
   \( E[Q | P = p_0] \)
2. What is the expected value of the demand Q if the price is set to \( P = p_0 \)?
   \( E[Q | do(P = p_0)] \)
3. Given that the current price is \( P = p_0 \), what would the expected value of the demand Q have been if we were to set the price at \( P = p_1 \)?
   \( E[Q | do(P = p_0)] \)
**THE STRUCTURAL CAUSAL MODEL (SCM)**

**A BI-LINGUAL LOGIC FOR CAUSAL INFERENCE**

Causal Diagram

(To specify what we know – Assumptions)

Counterfactual Language

(To specify what we wish to know – Queries)

\[ P(X, Y, Z) \]

\[ P(Y \mid \text{do}(x)) \]

\[ \text{Causal Inference} \]

\[ \hat{\theta}(Q) \]

\[ \hat{\theta} \]

\[ \text{estimand} \]

**THE SEVEN PILLARS**

Pillar 1: Transparency and Testability of Causal Assumptions

Pillar 2: The control of confounding

Pillar 3: Counterfactuals Algorithmization

Pillar 4: Mediation Analysis and the Assessment of Direct and Indirect Effects

Pillar 5: External Validity and Sample Selection Bias

Pillar 6: Missing Data (Karthika Mohan, 2017)

Pillar 7: Causal Discovery

---

**PILLAR 1: MEANINGFUL COMPACT REPRESENTATION FOR CAUSAL ASSUMPTIONS**

Task: Represent causal knowledge in compact, transparent, and testable way.

• Are the assumption plausible? Sufficient?
• Are the assumptions compatible with the available data? If not, which needs repair?

Result: Transparency and testability galore

Graphical criteria tell us, for any pattern of paths, what pattern of dependencies we should expect in the data.

---

**PILLAR 2: THE CONTROL OF CONFOUNDING**

Problem: Determine if a desired causal relation can be estimated from data and how.

Solution: The menace of Confounding has been demystified and “deconfounded”

• "back-door" – reduces covariate selection to a game
• "front door" – extends it beyond adjustment
• *do-calculus* – predicts the effect of policy interventions whenever feasible

---

**PILLAR 3: THE ALGORITHMIZATION OF COUNTERFACTUALS**

Task: Given (Model + Data), determine what Joe’s salary would be had he had one more year of education.

Solution: Algorithms have been developed for determining if/how the probability of any counterfactual sentence is estimable from experimental or observational studies, or combination thereof.

How?

• Every model determines the truth value of every counterfactual by a toy-like “surgery” procedure.
• Corollary: “Causes of effect” formalized

---

**PILLAR 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.

Result: The graphical representation of counterfactuals tells us when direct and indirect effects are estimable from data, and, if so, how necessary (or sufficient) mediation is for the effect.
PILLAR 5: TRANSFER LEARNING, EXTERNAL VALIDITY, AND SAMPLE SELECTION BIAS

Task: A machine trained in one environment finds that environmental conditions changed. When/how can it amortize past learning to the new environment?

Solution: Complete formal solution obtained through the do-calculus and “selection diagrams” (Bareinboim et al., 2016)

Lesson: Ancient threats disarmed by working solutions.

PILLAR 6: MISSING DATA (Mohan, 2015)

Problem: Given data corrupted by missing values and a model of what causes missingness. Determine when relations of interest can be estimated consistently “as if no data were missing.”

Results: Graphical criteria unveil when estimability is possible, when it is not, and how.

Corollaries:
• When the missingness model is testable and when it is not.
• When model-blind estimators can yield consistent estimation and when they cannot.
• All results are query specific.
• Missing data is a causal problem.

CONCLUSIONS

• Model-blind approaches to AI impose intrinsic limitations on the cognitive tasks that they can perform.
• The seven tasks described, exemplify what can be done with models that cannot be done without, regardless how big the data.
• DATA SCIENCE is only as much of a science as it facilitates the interpretation of data -- a two-body problem involving both data and reality.
• DATA SCIENCE lacking a model of reality may be statistics but hardly a science.
• Human-level AI cannot emerge from model-blind learning machines.

THANK YOU

Joint work with:
Elias Bareinboim
Karthika Mohan
Ilya Shpitser
Jin Tian
Many more . . .

Refs: http://bayes.cs.ucla.edu/jp_home.html