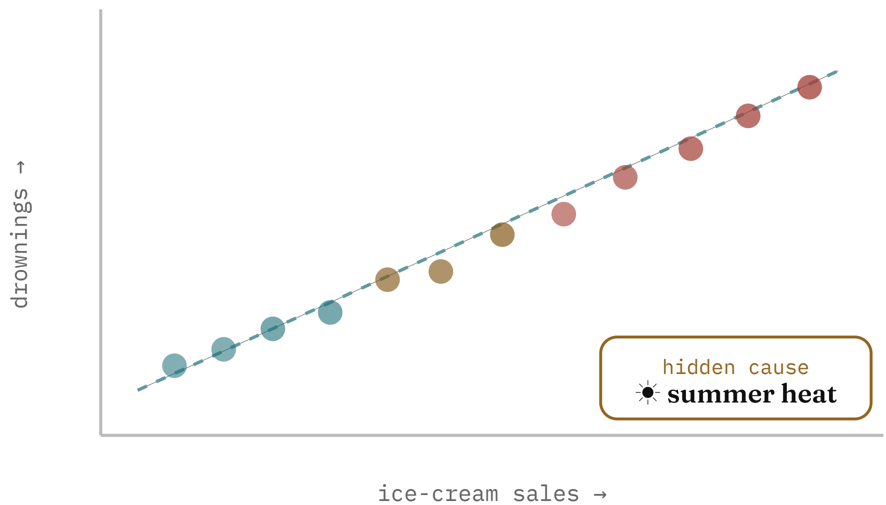
BLOCK I · FOUNDATIONS OF KNOWLEDGE & REASONING · DAY 05 / 180

Causation

Ice cream and drowning rise and fall together. Nobody drowned because of a sundae. So what is the difference between a pattern and a cause?



The dots march upward together — a textbook correlation. Their color is the variable the chart never shows: the season. Heat sells cones *and* sends people to the water.

Every summer, two numbers climb in lockstep. As ice-cream sales rise, so do drownings; as the cones stop selling in autumn, the drownings taper off too. Plot them and you get a clean, confident upward line — the kind of correlation that would make a careless analyst reach for a headline. Ban ice cream, save lives. And yet you already know, in your bones, that this is nonsense. No one has ever drowned *because* of a sundae.

What you know in your bones is one of the hardest things to write down in science. There is a third character lurking off-stage – **summer** – and it is pulling both strings at once. Heat drives people to buy ice cream, and heat drives people into lakes and oceans where some of them drown. The two visible numbers dance together only because an invisible one is conducting. Today is about the machinery for catching that invisible conductor – and the modern discovery that *causation* is not just a stronger correlation. It lives on a higher rung of a ladder you cannot climb by staring at data alone.

Four days in, the toolkit is starting to interlock. **Day 1** warned us about beliefs that are true but only *by luck* – the stopped clock. A *spurious correlation*[1] is exactly that at population scale: a number that's "right" for entirely the wrong reason. **Day 2** introduced Hume and the problem of induction; today we meet the *same Hume*, because his attack on causation and his attack on induction are one and the same blade. **Day 3** sorted reasoning into deduction, induction, and abduction – and causal discovery, we'll see, is abduction with teeth: inference to the best causal explanation. And **Day 4** gave us the do-versus-see distinction's parent, probability: today's punchline, $P(y|do(x)) \neq P(y|x)$, is the most important inequality in the course so far.

THE OLDEST PROBLEM

Hume kicks out the leg

Start where the trouble starts. In 1739, in *A Treatise of Human Nature*, David Hume asked a deceptively simple question: when one billiard ball strikes another and the second rolls away, *what exactly do you see?* You see the first ball move. You see them touch. You see the second ball move. What you never see – look as hard as you like – is the **causing** itself: the necessary connection, the hidden force, the little arrow of *because* linking the two events.

All we ever actually observe, Hume argued, is *constant conjunction*[2]: events of this type are reliably followed by events of that type. Add that the cause comes first (priority) and that the two events touch in space and time (contiguity), and you have everything experience delivers. The sense of *necessity* – the feeling that the second ball *had* to move – is not out in the world at all. It is a habit of mind, a customary expectation built up by repetition and then projected back onto the world like a film onto a screen. Hume found two definitions of "cause" tangled together in our heads: one about the world (constant conjunction) and one about us (the mind's practiced leap from one to the other).

This should feel familiar. It *is* the **problem of induction from Day 2**, wearing a different coat. If causation is just "this has always been followed by that," then claiming the next collision will behave like the last is precisely the unprovable bet on the uniformity of nature that Hume showed can never be justified non-circularly. Causation and induction are the same wound. For two centuries philosophy basically picked at it.

WHY "THE CEMENT OF THE UNIVERSE"?

The phrase gets attached to Hume, but it is better treated as Mackie's title and image: **J. L. Mackie's 1974 book** is called *The Cement of the Universe*. The irony Hume relished remains: this "cement" is the one thing we can never see. We infer the glue only from the fact that the bricks keep ending up stuck together.

FOUR REPAIRS

Saying what "more" there is

If causation is *more* than constant conjunction, the obvious move is to say what the "more" is. The twentieth century produced several serious answers – different ways to finish the sentence "C causes E means...". They are not simple rivals so much as lenses that modern causal modeling keeps borrowing from.

LEWIS · 1973

Counterfactual[3]**.** C causes E means: *had C not happened, E would not have happened.* Cash it out with "possible worlds" – picture the nearest world where C is absent and check whether E still occurs. Clean and intuitive, but it has to wrestle with backup causes and double-killings (preemption and overdetermination).

REICHENBACH · SUPPES · CARTWRIGHT

Probabilistic. Causes *raise the probability* of their effects. Reichenbach's *common-cause principle*[4]: if A and B are correlated but neither causes the other, a shared cause C "screens them off" – hold C fixed and the correlation dissolves. (Summer, exactly.)

WOODWARD · 2003

Interventionist. C causes E if *wiggling C – and only C – would change E.* No human needed: an *intervention*[5] is a surgical nudge, so volcanoes cause ash even with no one to push the button. This is the philosophical twin of Pearl's machinery, arriving in the same era: below, Pearl turns the "wiggle C only" idea into `do(C)` and graph surgery.

THE THROUGH-LINE

One question, four lenses. Counterfactual = "what if it hadn't?"; probabilistic = "does it raise the odds, holding rivals fixed?"; interventional = "what changes if I wiggle it?". Pearl's framework, next, gives all three a shared grammar.

Notice how **Cartwright (1979)** sharpened the probabilistic story, because her fix is the hinge of the whole day. Causes raise the probability of effects, yes – *but only inside a "causally homogeneous" background*, with all the other causes held fixed. Forget that proviso and you walk straight into the most beautiful trap in statistics.

THE TRAP

Simpson's paradox: when the numbers literally reverse

Here is a fact that sounds impossible until you've seen it: a treatment can be **better for small stones, better for large stones, and yet worse overall.** Not "looks worse"

— is, on the pooled numbers, worse. The real kidney-stone dataset below lets you watch the bars flip.

The mechanism is always a lurking variable that's distributed unevenly between the groups you're comparing. In the kidney-stone data (Charig et al., *BMJ*, 1986), surgeons gave the gentler treatment **B** mostly to the *easy* small stones and reserved old-fashioned open surgery **A** for the *hard* big ones. So B's overall success rate is flattered by an easier caseload. Split by stone size — hold the *confounder*[6] fixed, exactly as Cartwright demanded — and A pulls ahead in both rooms.

The Reversal Machine below turns that into one picture. Each square is one patient and never moves. Color says which treatment they received; row says whether the case was small-stone or large-stone; fill says whether the treatment worked. The pooled view compares the two treatment columns. The conditioned view keeps the same picture but compares inside each row first. That is the foundation of **conditioning**: compare like with like before you call anything a cause.

Kidney-Stone Simpson Reversal

VIEW	TREATMENT A	TREATMENT B	APPARENT WINNER
All patients	273 / 350 = 78.0%	289 / 350 = 82.6%	B
Small stones	81 / 87 = 93.1%	234 / 270 = 86.7%	A
Large stones	192 / 263 = 73.0%	55 / 80 = 68.8%	A

The pooled comparison reverses because B was used much more often on easier small-stone cases.

The lesson lands hard: **you cannot read causation off a table of numbers.** The very same digits support opposite conclusions depending on a variable that may not even be in the spreadsheet. Which raises the question that organizes everything after this – if the data alone won't tell you, what *will*? The answer, from a computer scientist who spent the 1980s building machines that reason under uncertainty, is that you need to add something the data doesn't contain: a **model of which arrows point where.**

THE CAUSAL REVOLUTION

Pearl's ladder, and the verb that changed everything

Judea Pearl won the 2011 Turing Award – often called computing's Nobel – "for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning." His central image, popularized in *The Book of Why* (2018), is a *Ladder of Causation*[7] with three rungs, each demanding a stronger kind of question and a stronger kind of evidence than the one below. Most of statistics and nearly all of machine learning, Pearl likes to needle, never leaves the bottom rung.

Climb it yourself:

Pearl's Ladder of Causation

RUNG	VERB	NOTATION	QUESTION
1. Association	Seeing	`P(Y | X)`	How do outcomes differ among cases where X is observed?
2. Intervention	Doing	`P(Y	do(X))`	What would happen if X were set by intervention?
3. Counterfactuals	Imagining	$P(Y_x \mid X', Y')$	What would have happened in a different world, given what actually happened?

In the last line, Y_x means "Y in the world where X is set to x"; `X'` and `Y'` are the actual facts you condition on.

The do-operator: seeing is not doing

Here is the conceptual hinge of the entire modern field, and it is worth saying slowly. There are two very different things you can do with a variable X.

You can **condition**[8] on it – written `P(Y | X = x)`. This means: among all the cases where X *happened* to equal x, what's the distribution of Y? You're filtering an existing population. This is *seeing*.

Or you can **intervene** – written, in Pearl's notation, `P(Y | do(X = x))`. This means: reach in, *force* X to equal x for everyone, snipping X away from whatever usually causes it, and then watch Y. This is *doing* – it's what a randomized experiment performs.

When there's a confounder, these two numbers come apart, and the gap between them *is* the bias. Among people who *happen* to buy lots of ice cream, drowning really is more common (because they're the summer people) — so the *seeing* quantity is high. But if you *force* a random sample of people to buy ice cream in, say, evenly distributed weather, drownings don't budge — the *doing* quantity is flat. The machine below lets you turn a confounder up and down and watch the two quantities diverge.

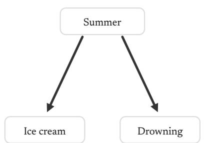
Seeing Is Not Doing

In the toy world, the observed association equals the true direct effect plus a bias term from summer. With the default settings, the true direct effect is +0.05, but the observed association is much larger because summer pushes both ice-cream buying and swimming deaths upward.

QUANTITY	WHAT IT ASKS	DEFAULT READING
P(drown | buy) - P(drown | no buy)	Seeing: compare existing buyers with non-buyers.	True effect plus confounding bias.
P(drown | do(buy)) - P(drown | do(no buy))	Doing: set buying directly and sever its usual causes.	Only the true direct effect.

That severed arrow in the right-hand graph is the do-operator made visual. Intervening doesn't just *look* at X — it **deletes the arrows pointing into X** and replaces them with your hand. The summer→ice-cream link is cut, so summer can no longer use ice cream as a backdoor route to fake an effect on drowning. What survives is only the real thing.

Pearl gave us a grammar for doing this on paper. A *structural causal model*[9] draws the variables as a *directed acyclic graph*[10] – boxes and arrows, no loops – and three named patterns do most of the work. A *fork* (X ← Z → Y) is a confounder; block it by conditioning on Z. A *chain* (X → Z → Y) is a mediator. And a *collider* (X → Z ← Y) is the trap: X and Y are unrelated until you condition on their shared effect Z, at which point a phantom correlation springs into being. This is why "controlling for everything you can measure" is not cautious but reckless – condition on a collider and you *manufacture* the very bias you were trying to remove.

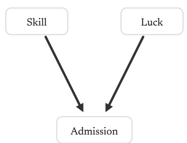


FORK / CONFOUNDER

Summer changes both exposure and outcome. Condition on summer before comparing.

CHAIN / MEDIATOR

Tar carries part of smoking's effect. Conditioning on tar changes the question.



COLLIDER

Skill and luck meet at admission. Conditioning on admission opens a fake relationship.

THE FRONT-DOOR TRICK

Pearl's *front-door criterion*[11] is one of the framework's neatest moves: sometimes you can estimate the effect of smoking on cancer *even with an unmeasured genetic confounder*, as long as you can measure a complete mediator in between (say, tar deposits in the lungs). The calculation has three moves: estimate how smoking changes tar; estimate how tar changes cancer while accounting for smoking; then average those pieces over the observed smoking mix. The hidden genetic confounder affects smoking and cancer, but it does not enter the measured smoking → tar and tar → cancer pieces in the final front-door expression. It is a genuine "get a causal answer from observational data" move — but only because you supplied the *graph* that says tar is a full mediator. No free lunch: the assumptions just moved from the spreadsheet into the diagram.

THE FRONTIER · 2026

Three live edges — and the hype filter

Now the question that has launched a thousand papers and at least one industry: **can you infer causation from observation alone?** The honest 2026 answer is a precise "partly — and there's a proven wall." Each claim below is tagged for how much weight it can bear.

Edge 01 [ESTABLISHED] [ESTABLISHED]

The two theorems that fence the field

This is the most solid ground in the whole day — not empirical findings that could be overturned, but *mathematical proofs*. First, **do-calculus is complete.** Pearl's three rewrite rules turn `do()` expressions into ordinary probabilities whenever that's possible at all; and Shpitser & Pearl (2006) and, independently, Huang & Valtorta (2006) proved that if the rules *cannot* eliminate the do-operator, then **no method can** — the effect is genuinely unidentifiable from observational data plus that graph. Pearl called this closing "the chapter of nonparametric identification." A clean, permanent answer to "when can seeing substitute for doing?"

WHAT A DO-CALCULUS REWRITE LOOKS LIKE

- **Delete an irrelevant observation:** if Z already blocks every open path from X to Y, then learning X adds no more information: `P(Y | X, Z) = P(Y | Z)`.
- **Swap an action for an observation:** after the right adjustment, setting X and observing X can answer the same question: `P(Y | do(X), Z) = P(Y | X, Z)`.
- **Delete an irrelevant action:** if intervening on Z cannot reach Y once X is fixed, drop it: `P(Y | do(X), do(Z)) = P(Y | do(X))`.

Those are toy examples, not the full formal side-conditions. The point is the flavor: the graph licenses exactly which symbols may be erased or exchanged.

Second, the wall on the other side: **the Markov-equivalence ceiling.** Using only the conditional-independence patterns in observational data, certain different causal graphs are *provably indistinguishable*. X→Y→Z, X←Y←Z, and X←Y→Z all imply the exact same single fact ("X and Z are independent once you know Y"), so no amount of that data can tell them apart. They form a *Markov equivalence class*[12]. Only the collider X→Y←Z stands out, because X and Z are independent in the raw data but become dependent after you condition on their shared effect Y. The takeaway is stark: **observation alone, assumption-free, can never deliver a unique causal graph** – only a class of candidates. Both results are `[ESTABLISHED]` in the strongest sense available: theorems.

Edge 02 `[PROMISING]` `[ESTABLISHED]`

Squeezing direction out of still data — and why experiments still reign

So is the ceiling the end? Not quite – you can climb over it by importing extra assumptions the bare independence-tests don't use. ***LiNGAM***[13] (Shimizu et al., *JMLR*, 2006) showed that if relationships are linear, there are no hidden confounders, and the noise is *non-Gaussian*, the full direction becomes identifiable – the asymmetry of non-Gaussian noise breaks the X→Y / Y→X tie. In the true direction, the leftover noise is independent of the cause; in the wrong direction, the residuals still carry a telltale dependence. ***Additive-noise models***[14] (Hoyer, Janzing, Mooij, Peters & Schölkopf, NeurIPS 2009) extended this to nonlinear cause-effect pairs. On the standard **Tübingen Cause-Effect Pairs**[15] benchmark – 108 pairs with known ground truth, like altitude→temperature – one strong system reported about **83% accuracy** (Mosaic, Wu & Fukumizu 2020), evidence that something "often thought impossible" is partly doable. But note the shape of the trick: you escape the ceiling only by *assuming* non-Gaussianity or additivity. Those assumptions are not testable from the same observational distribution alone – and the load-bearing one for the whole constraint-based program, *faithfulness*[16], can fail silently in finite samples. `[PROMISING]`

Which is why the gold standard remains brutally simple: **run the experiment.** A *randomized controlled trial*[17] physically performs `do(X)` by assigning X at random, severing every backdoor at a stroke. When you can't randomize, economics' *credibility revolution*[18] hunts for "natural experiments" that mimic randomization – instrumental variables, regression discontinuity, difference-in-differences. That program earned the **2021 Nobel** in economics for David Card, Joshua Angrist, and Guido Imbens (we'll spend real time here on **Day 152**). `[ESTABLISHED]`

Edge 03 `[PROMISING]` `[CONTESTED]`

Can the machines do it? Causal ML and the "causal parrot"

The hottest and haziest edge. *Causal representation learning*[19] (Schölkopf et al., *Proceedings of the IEEE*, May 2021) asks the deep question: classical causal discovery assumes the variables are handed to you, but the real world arrives as pixels and words. Can a network *learn* the high-level causal variables – and would that buy robustness to *distribution shift*[20], which today's models often lack? It's a serious, active program whose biggest promises remain [PROMISING], not yet delivered at scale.

Then the lightning rod: **can large language models reason about cause and effect?** Kıcıman, Ness, Sharma & Tan (2023) reported GPT-4 hitting **97% on the Tübingen pairwise causal-direction task** – a 13-point jump over the prior best – plus strong counterfactual scores, and argued memorization alone can't explain it. The rebuttal came fast: Zečević et al., "Causal Parrots" (TMLR, 2023), argued LLMs *talk* causality without *being* causal – they recite causal facts marinating in their training text rather than performing Pearl-style inference. The 2024–25 synthesis (e.g., Jin et al., ICLR 2024, "Can LLMs Infer Causation from Correlation?") splits the difference: models are strong causal-*knowledge retrievers* and often shine on Rung-1 questions, but the hard test is genuinely novel Rung-2/3 interventional and counterfactual structure. Recall the **Day 1 Gettier trap** in a new guise: an LLM that outputs a true causal claim for reasons that have nothing to do with the causal structure is right, but does it *know*? And recall **Day 3**: reciting a memorized fact isn't abduction. Verdict on "LLMs reason causally in Pearl's sense": [CONTESTED] Useful for *setting up* a causal analysis; not established as a causal reasoner. The open-source tooling underneath causal ML (DoWhy, EconML) is genuine and built on exactly the theorems above, but the sales pitch that causal AI will soon replace ordinary correlational ML is still ahead of the evidence.

OPEN QUESTIONS

What's genuinely unsettled

- **Is there a "right" theory of causation at all,** or are counterfactual, probabilistic, and interventionist accounts each capturing a different facet — with none reducible to the others? No analysis has escaped all the counterexamples (preemption, overdetermination, contextual unanimity).
- **How far can we trust faithfulness?** The assumption that real systems never have exactly-cancelling causal paths is convenient and untestable — and biology, with its feedback and homeostasis, may violate it routinely.
- **Can causal variables be learned from raw data** (pixels, language) rather than handed to the algorithm — and is that even well-posed, since the "right" carve-up of the world into variables may itself be perspective-dependent?
- **Do large models build internal causal world-models,** or only ever statistics of causal talk? The answer reaches straight into **Days 138–145** and the question of whether prediction can ever amount to understanding.
- **Where do the arrows come from?** Every method here needs *some* causal input — a graph, an assumption, an experiment. Hume's ghost still asks whether that input is ever read off the world, or always brought to it.

◆ THE DAY IN THREE SENTENCES

BIG IDEA

Causation is not a stronger correlation but a different *kind* of thing, living on higher rungs of Pearl's ladder — *doing* and *imagining* above mere *seeing* — so that `P(Y|do(X)) ≠ P(Y|X)` whenever a confounder lurks, and no amount of staring at observational data alone can close that gap without importing causal assumptions.

BEST ANALOGY

Ice cream and drowning marching upward together while summer, off-stage, pulls both strings — and the do-operator as a pair of scissors that cuts the strings into a variable so you can see what it really drives.

LIVE CONTROVERSY

Whether causation can be inferred from observation alone — provably "only up to a Markov-equivalence class" without extra assumptions, partly recoverable with them (LiNGAM, additive noise) — and the noisy 2026 fight over whether LLMs genuinely reason causally or merely parrot causal talk.

THREADS TODAY › information (a graph is the extra information data lacks; the do-operator quantifies evidence vs intervention) · computation (do-calculus as a complete *algorithm*; causal discovery as search) · emergence (causal structure as a higher-level layer over raw correlation) — with callbacks to Day 1 (true-by-luck), Day 2 (Hume), Day 3 (abduction), and Day 4 (P(y|x)).

TOMORROW → DAY 06

Statistics & the Art of Not Fooling Yourself

Today we saw how a confounder can flip a conclusion. Tomorrow we meet the subtler enemy: *yourself*. p-hacking a coin flip, the garden of forking paths, effect sizes versus the worship of "significance" — and the collider trap we just met returns as one of the easiest ways an honest researcher fools an honest audience. Bring today's instinct to ask, every time, "what's the variable that isn't on this chart?"

NOTES

1. A spurious correlation is a real statistical association caused by something other than a direct causal link between the two measured variables.
2. Constant conjunction means one kind of event has repeatedly been followed by another kind of event.
3. A counterfactual asks what would have happened if some fact had been different.
4. The common-cause principle says a surprising correlation often needs either a direct cause or a shared prior cause that explains it.
5. An intervention is an idealized change that sets one variable while blocking the usual causes of that variable.
6. A confounder is a third variable that influences both the apparent cause and the apparent effect, making them move together.
7. Pearl's ladder separates association, intervention, and counterfactual reasoning into increasingly demanding levels.
8. Conditioning means narrowing attention to cases where some variable already has a given value.
9. A structural causal model represents how variables are generated from other variables and from background noise.
10. A directed acyclic graph, or DAG, is a network of arrows with no directed loops.
11. The front-door criterion identifies a causal effect through a measured mediator when the direct cause-effect path is confounded.
12. A Markov equivalence class is a set of causal graphs that imply the same conditional independences, so observational data of that kind cannot choose among them.
13. LiNGAM is a causal-discovery method whose name abbreviates linear non-Gaussian acyclic model.
14. Additive-noise models assume the effect equals a function of the cause plus independent noise, an asymmetry that can sometimes identify direction.
15. The Tübingen benchmark is a collection of real-world variable pairs with expert-labeled causal directions.

16. Faithfulness says the independences in the data reflect graph structure, not exact cancellations among causal paths.
17. A randomized controlled trial assigns participants to conditions by chance so background causes are balanced on average.
18. The credibility revolution in economics moved emphasis from clever regressions to research designs that approximate randomized experiments: instruments, discontinuities, and policy changes that assign exposure as-if at random.
19. Causal representation learning tries to learn useful causal variables from raw data such as pixels, text, or sensor streams.
20. Distribution shift means the data a model sees after deployment differ from the data it learned from.

SOURCES

Sources & further reading

1. Hume, D. (1739–40). *A Treatise of Human Nature*, Book I, Part III; and (1748) *An Enquiry Concerning Human Understanding*, §VII. – constant conjunction; no observable necessary connection.
2. Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation.* Oxford University Press. – INUS conditions; the title's phrase.
3. Lewis, D. (1973). "Causation." *Journal of Philosophy* 70(17): 556–567. doi:10.2307/2025310. doi.org/10.2307/2025310 See also Lewis, *Counterfactuals* (Blackwell, 1973); revised "influence" account (2000).
4. Reichenbach, H. (1956). *The Direction of Time.* University of California Press. – the common-cause principle & screening off.
5. Suppes, P. (1970). *A Probabilistic Theory of Causality.* North-Holland. – prima facie vs spurious causes.
6. Cartwright, N. (1979). "Causal Laws and Effective Strategies." *Noûs* 13(4): 419–437. – probability-raising only within causally homogeneous contexts. overview
7. Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press. – the interventionist/manipulationist theory.
8. Charig, C. R., Webb, D. R., Payne, S. R. & Wickham, J. E. A. (1986). "Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy." *British Medical Journal* 292: 879–882. – the kidney-stone Simpson's-paradox data.

9. Simpson, E. H. (1951). "The Interpretation of Interaction in Contingency Tables." *JRSS B* 13: 238–241. Blyth, C. R. (1972), *JASA* 67: 364–366 (coins "Simpson's paradox"). Yule, G. U. (1903) on spurious correlation.

10. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge University Press. And Pearl, J. & Mackenzie, D. (2018). *The Book of Why.* Basic Books. — the Ladder of Causation; do-calculus; back-door/front-door.

11. Shpitser, I. & Pearl, J. (2006). "Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models." *AAAI.* & Huang, Y. & Valtorta, M. (2006). "Pearl's Calculus of Intervention Is Complete." *UAI.* — completeness of do-calculus (proven).

12. ACM (2012). 2011 A.M. Turing Award — Judea Pearl. amturing.acm.org/award_winners/pearl

13. Spirtes, P., Glymour, C. & Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd ed. MIT Press. — the PC and FCI algorithms; Markov equivalence.

14. Shimizu, S., Hoyer, P. O., Hyvärinen, A. & Kerminen, A. (2006). "A Linear Non-Gaussian Acyclic Model for Causal Discovery." *JMLR* 7: 2003–2030. reference code

15. Hoyer, P., Janzing, D., Mooij, J., Peters, J. & Schölkopf, B. (2009). "Nonlinear causal discovery with additive noise models." *NeurIPS.* Mooij, J. et al. (2016). "Distinguishing Cause from Effect Using Observational Data." *JMLR* 17(32): 1–102. Wu, P. & Fukumizu, K. (2020). "Causal Mosaic: Cause-Effect Inference via Nonlinear ICA and Ensemble Method." *Proceedings of Machine Learning Research* 108: 1157–1167. — additive-noise models, the Tübingen pairs benchmark, and the Mosaic result.

16. Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A. & Bengio, Y. (2021). "Toward Causal Representation Learning." *Proceedings of the IEEE* 109(5): 612–634. doi:10.1109/JPROC.2021.3058954.

17. Kıcıman, E., Ness, R., Sharma, A. & Tan, C. (2023). "Causal Reasoning and Large Language Models." arXiv:2305.00050; TMLR 2024. & Zečević, M., Willig, M., Dhami, D. S. & Kersting, K. (2023). "Causal Parrots: Large Language Models May Talk Causality But Are Not Causal." *TMLR.* arXiv:2308.13067. & Jin, Z. et al. (2024). "Can Large Language Models Infer Causation from Correlation?" *ICLR.*

18. The Royal Swedish Academy of Sciences (2021). The Sveriges Riksbank Prize in Economic Sciences — Card, Angrist & Imbens. nobelprize.org/prizes/economic-sciences/2021

19. Stanford Encyclopedia of Philosophy: "Causation" / "Counterfactual Theories of Causation" / "Causal Models" / "Probabilistic Causation." plato.stanford.edu/entries/causation-

counterfactual

END OF DAY 05 · 175 DESCENTS REMAIN