

## TABLE OF CONTENTS

## INTRODUCTION: MIND OVER DATA

- P 2. Humans ask “why?” Causal inference is taking that question seriously.
- P.2. Causation is indicated by words such as “cause”, “preventing”, “policy” and “should”
- P 3. “Fundamental gap between vocabulary ... of causal questions, and ... scientific theories”
- P 3. No way to show that the barometer reading tracks (is caused by) atmospheric pressure.
- P 5. Galton and Pearson failed in answering causal questions using cross-generation data.
- P 5. “They declared [causal] questions off limits and turned to develop ... statistics.”
- P 5. “In Statistics 101, every student learns to chant, ‘Correlation is not causation.’ “ “With good reason, Rooster’s crow is highly correlated with ..., but does not cause the sunrise.”
- P 5. “Unfortunately statistics has fetishized this commonsense observation.  
It tells us that correlation is not causation, but it does not tell us what causation is.”
- P 6. “A shining exception was path analysis, invented by geneticist Sewall Wright in the 1920s.”
- P 6. “Much of the data-centric history haunts us today.”
- P 6. “One hundred years ago, the question of whether cigarette smoking causes a health hazard would have been considered unscientific.”
- P 7. “To me, the change is nothing short of a revolution. I dare to call it the Causal Revolution”
- P 7. “The calculus of causation consists of two languages: causal diagrams, to express what we know, and a symbolic language, resembling Algebra, to express what we want to know.”
- P 7. “Causal diagrams simply dot & arrow pictures. Dots... variables; arrows, causal relations.
- P 8. “The DO operator signifies that we are dealing with an [active] intervention...”
- P 9. “Counterfactual language”
- P 10. “Counterfactuals are the building blocks of moral behavior, as well as scientific thought.”
- P 11. “A causal reasoning model will give machines the ability to reflect on their mistakes, to pinpoint weaknesses... To function as moral entities, and to converse naturally with humans...”
- P 17. “Chapter 1 assembles the three steps of observation, intervention and counterfactuals in a Ladder of Causation, the central metaphor of this book.”
- P 18. “Chapter 2 tells how ... statistics inflicted causal blindness on itself. “
- P 18. Chapter 3 introduces, evaluates and rejects Bayesian networks.
- P 18. Chapter 4 talks about the major contribution of statistics to human knowledge: RCT.
- P 18. Chapter 5 talks about the smoking lung-cancer debate.
- P 19. Chapter 6 talks about paradoxes....
- P 19. Chapters 7-9 talk about the Ladder of Causation.
- P 21. “Sum up the message of the book.... You are better than your data.”

## Chapter 1 The Ladder of Causation

- P 24. In the last 50,000 years something unique happened which some call the Cognitive Revolution and others ... call the Great Leap Forward.
- P 25. I call this phenomenon the "super-evolutionary speedup".
- P 25. In his book *Sapiens*, Harari posits that our ancestor's capacity to imagine non-existent things was the key to everything, for it allowed them to communicate better.
- P 25. Their newly acquired causal imagination enabled them to do many things more efficiently
- P 26. *Figure 1.1 Perceived causes of a successful mammoth hunt.*
- P 26. These causal diagrams are the computation core of the "causal inference engine".
- P. 27: THREE LEVELS OF CAUSATION
- P 27. My research has taught me that a causal learner must master at least three distinct levels of cognitive ability: seeing, doing and imagining.
- P 27. Seeing or observing entails detection of regularities ... and is shared by many animals...
- P 27. Doing entails predicting the effects of deliberate alterations to produce desired outcome
- P 27. Only a small handful of species have demonstrated elements of this skill.
- P 27. Yet even tool users do not necessarily possess a "theory" of their tool that tells them why it works and what to do if it doesn't work. For that you need to have achieved a level of understanding that permits imagining.
- P 27. I cannot prove this [historically], but I can prove mathematically that the three levels differ fundamentally, each unleashing capabilities that the ones below it do not.
- P 28: *Figure 1.2 The Ladder of Causation: Association, Intervention and Counterfactuals.*
- P 29. The first rung of the ladder calls for predictions based on passive observations.
- P 29. The proportion of toothpaste buyers that bought floss, known as a conditional probability, measures the degree of association between buying toothpaste and buying floss.
- P 29. Statisticians have developed many ... methods to reduce data and identify associations.
- P 29. Statistics alone cannot tell which is the cause and which is the effect: toothpaste or floss.
- P 30. ... It may not matter. Good predictions need not have good explanations.
- P 30. The public believes that "strong AI", machines that think like humans, is just around the corner. In reality, nothing could be further from the truth.
- P 30. Deep learning has given us machines with truly impressive abilities but no intelligence.
- P 31. Intervention ranks higher than association b/c it involves not just seeing but changing what it. We cannot answer questions about interventions with passively collected data.
- P 32. A very direct way to predict the result of an intervention is to experiment with it ...
- P 32. Successful predictions of interventions can be made even without an experiment.
- P 32. The defining query of the 2<sup>nd</sup> rung of the Ladder of Causation is "what if we do ....?"
- P 32. Another question at the 2<sup>nd</sup> level of causation is "How?" This requires a causal model.
- P 33. My headache is gone. Why? These queries take us to the top rung of the Ladder of Causation: the level of counterfactuals, because to answer them we must go back in time, change history and ask, "What would have happened if I had not taken the aspirin?"
- P 33. Data [facts] cannot tell us what will happen in a counterfactual or imaginary world...
- P 33. Yet the human mind makes such explanation-seeking inferences reliably and repeatably.
- P 33. The laws of Physics can be interpreted as counterfactual assertions.
- P. 35: *Figure 1.3: The Lion Man of Stadel Cave*
- P. 35 The Lion man is different: a creature of pure imagination.
- P 35. The Lion Man is a precursor of every philosophical theory, scientific discovery and technological innovation....

- P. 35 As shown in Figure 1.2, the characteristic queries of the 3<sup>rd</sup> rung of the Ladder of Causation are "What if I had done X?" and "Why?" Both involve comparing the actual world to a counterfactual world.
- P 36: THE MINI TURING TEST
- P 36 The Turing Test: A computer could be called a thinking machine if an ordinary human could not tell whether he was talking with a human or a computer.
- P 36 Turing predicted that in 50 years' time it will be possible to program computers to make them play the imitation game so well that an average interrogator will not have more than a 70% chance of making the right identification within five minutes of questioning.
- P 37 In 25 years ..., not a single program has fooled all of the judges or even half of them.
- P 37 Turing suggested a strategy: produce one which simulated the child's [mind].
- P 37 Pearl: "I think Turing was on to something."
- P 37 How can machines acquire causal knowledge? This is still a major challenge.
- P 37 We can answer a slightly less ambitious question: How can machines (and people) represent causal knowledge in a way that would enable them to access the necessary information swiftly, answer questions correctly, and do it with ease...?
- P 37. I call this the mini-Turing test. It is mini for two reasons: 1st, it is confined to causal reasoning..., 2nd we allow contestants to encode the story in any convenient representation.
- P 38 The question of representation must precede the question of acquisition.
- P 38 One major contribution of AI to cognition has been the paradigm "Representation first, acquisition second"
- P 38 People commonly claim that it [the mini Turing test] can be defeated by cheating.
- P 38 Searle's challenge has one flaw: cheating is not easy: In fact, it is impossible.
- P 38 We could ask roughly 30 million possible queries.....
- P 39 Humans must have some compact representation of the information...
- P 39 Such representation not only exists but has childlike simplicity: a causal diagram.
- P 39 *Figure 1.4. Causal diagram for the firing squad example.*
- P 39-43 **Firing squad example:** details and explication of counterfactuals...
- P 44-46 **Smallpox inoculation example:**
- P 47: ON PROBABILITIES AND CAUSATION
- P 47 Philosophers have tried to define causation in terms of probability.
- P 47. What prevented attempts from succeeding was not the idea itself, but the way it was articulated formally. Probabilities lie on the first rung of the Ladder of Causation and cannot ever answer queries on the second or third rung. Any attempt to "define" causation in terms of seemingly simpler first-rung concepts must fail.
- P 48 [Work in Progress]
- P 49 the notion of probability raising [causation] cannot be expressed in terms of probabilities.
- P 49 The proper way to rescue the probability-raising idea is with the *do-operator*.
- P 50 [Philosophers] have gotten over this blunder [probability language].
- P 50 Unfortunately similar ideas are pursued in economics: Granger causality & autocorrelation
- P 50-51 Pearl: I have a confession to make: I made the same mistake (Bayesian Networks)
- P 50-51 **Bayesian network details**
- P 51 Fortunately Bayesian Networks required only two slight twists to climb to the top.

## Chapter 2 From Buccaneers to Guinea Pigs: The Genesis of Causal Inference

P 53 On Feb 9, 1877 at the Royal Institute, Galton demonstrates the Galton Board.

P 54 The Galton Board is simply a visual demonstration of Laplace's theorem.

P 55 To Galton the quincunx was a model for the inheritance of ... genetic traits: a causal model.

P 55 Suppose we doubled the number of rows [add 2<sup>nd</sup> generation]. The bell-curve would get wider.

P 55 This is not what happens. In fact the width of the distribution of heights stays relatively constant.

P 56 Sons of tall men tend to be taller than average – but not as tall as their fathers.

P 56 Galton first called this "reversion" and later "reversion toward mediocrity."

P 56 Phenomenon of regression to the mean is ubiquitous in all facets of life, education & business

P 56 Galton thought he had stumbled onto a law of heredity [Do] rather than a law of statistics {See}

P 56 He believed the regression to the mean must have some cause.

P 56 In his ... lecture, he illustrated his point. After passing through the first array of pegs, the balls passed through sloping chutes that moved them closer to the center. This kept a constant width.

P 57 *Figure 2.1 The Galton board: a) 1 gen., b) 2 gen w/o reversion, c) 2 gen with reversion*

P 57 Thus, Galton conjectured, reversion to toward the mean was a physical process.

P 57 "The process of reversion cooperates with the general law of deviation" Galton told his audience

P 57 In 1877, [he pursued] a causal explanation and thought regression to mean was a causal process.

P 57 He was mistaken but he was far from alone. People make the same mistake to this day.

P 58 Sophomore slump does not need a causal explanation. It will happen ... by the laws of chance.

P 58 Modern statistical explanation is quite simple. See Daniel Kahneman in *Thinking Fast & Slow*.

P 58 Kahneman: "Success = talent + luck. Great success = a little more talent + a lot of luck"

P 58 By 1899, Galton had figured this out. In the process he took the first huge step toward divorcing statistics from causation. His reasoning is subtle. It is the newborn discipline of statistics uttering its first cry. He noticed when he plotted height vs. forearm length, the same phenomenon occurred.

P 58 Clearly height is not a cause of forearm length or vice-versa. If anything both are caused by genetic inheritance. Galton started using a new word for this kind of relationship: co-related.

P 58 Eventually, he opted for the more normal English word "correlated."

P 58 Later he realized a...more startling fact: in generational comparisons the order could be reversed

P 58 Fathers of sons also revert to the mean. Once Galton realized this he had to give up the idea of a causal explanation for reversion to the mean. This realization must have been paradoxical at first.

P 58 Tall dads usually have shorter sons & tall sons usually have shorter dads. How can both be true?

P 58 How can a son be both taller and shorter than his father?

P 59 *Figure 2.2: The scatter plot shows a data set of heights.*

P 59 We are not talking about an individual father and son but about two populations.

P 59 The population of father son pairs with 6' fathers is not the same as population with 5' 11" sons.

P 59 Every father in first is by definition 6' tall. The 2<sup>nd</sup> group will have more fathers that are shorter.

P 60 **Explaining correlation, prediction and the two regression lines.**

P 61 where regression to the mean is concerned, there is no difference between cause and effect.

P 61 *Figure 2.3. Galton's regression lines.*

P 62 The law of regression apply when we correlate two different quantities and scale properly...

P 62 For the first time Galton's idea of correlation gave an objective measure, independent of human judgment or interpretation, of how two variables are related to one another.

P 62 Galton's disciple Karl Pearson later derived a formula for the slope of the (properly rescaled) regression line & called it the correlation coefficient. This is still the first number that statisticians ... compute when they want to know the how strongly two different variables in a data set are correlated.

P 62 For Pearson, especially, the slippery old concepts of cause & effect seemed outdated & unscientific compared to mathematically clear and precise concept of a correlation coefficient

P 63 GALTON AND THE ABANDONED QUEST

- P 63 It is an irony of history that Galton started in search of causation and ended up discovering correlation, a relationship that is oblivious of causation.
- P 63 The first sacrifice on the altar of correlation was Galton's elaborate machinery to explain the stability of the population's genetic endowment.
- P 63 Galton simply abandoned the effort and turned his attention to the siren song of correlation.
- P 63 Stigler: "In supreme irony, what started out as an attempt to mathematize the framework of the *Origin of the Species* ended with the essence of that great work being discarded as unnecessary!"
- P 63 Pearl: Looking back on Galton's machine in light of causal diagrams, the first thing I notice is that the machine was wrongly constructed. Ever-growing dispersion should never have been there.
- P 63 This [ever growing dispersion] stands in blatant contradiction to Kahneman's equations.
- P 64 According to Kahneman's equations generation 2 does not inherit luck from generation 1.
- P 64 Luck, by its very definition, is a transitory experience; no impact on future generations.
- P 64 *Figure 2.4: Two models of inheritance: a) Galton board model, b) a genetic model with luck.*
- P 64. In Figure 2-4a, success is transmitted across generations and luck variations accumulate... This is perhaps natural if "success" is equated to wealth or eminence.
- P 64 In Figure 2-4b, luck affects each generation independently; no way of affecting later generations
- P 65 Both models are compatible with the bell-shaped distributions of heights.
- P 65 First is incompatible with stability... Second we only need to explain the stability of success.
- P 65 That stability, now called the Hardy –Weinberg equilibrium (1908).
- P 65 They used a causal model – the Mendelian theory of inheritance.
- P 65 Galton was led astray by his beautiful but flawed causal model and later, having discovered the beauty of correlation, came to believe that causality was no longer needed.
- P 66 There is simply no other way to understand how statistics became a model-blind data-reduction enterprise, except by putting on our causal lens and retelling the stories of Galton and Pearson in the light of the new science of cause and effect.
- P 66 PEARSON: THE WRATH OF THE ZEALOT
- P 66 It remained to ... Karl Pearson, to complete the task of expunging causation from statistics.
- P 66 Pearson: "I felt like a buccaneer of Drake's days: not quite pirates but with decidedly piratical tendencies. I interpreted Galton to mean there was a category broader than causation, namely correlation, of which causation was only the limit and that this new conception of correlation brought psychology, anthropology, medicine and sociology in large part in the field of mathematical treatment. It was Galton who first freed me from the prejudice that sound mathematics could only be applied to natural phenomena under the category of causation."
- P 66 Pearson Grammar of Science (1892): That a certain sequence has occurred and reoccurred in the past is a matter of experience to which we give expression in the concept causation... Science in no case can demonstrate any inherent necessity in a sequence nor prove with absolute certainty that it must be repeated."
- P 66 Pearl: To summarize, causation for Pearson is only a matter of repetition and, in the determinate sense, can never be proven. As for causality in a non-deterministic world, Pearson was even more dismissive: "the ultimate scientific statement of description of the relation between two things can always be thrown back upon .. a contingency table." In other words, data is all there is to science. Full stop. In this view, the notions of intervention and counterfactuals discussed in Chapter 1 do not exist and the lowest rung of the Ladder of Causation is all that is needed for doing science.
- P 67 The mental leap from Galton to Pearson is breathtaking and indeed worthy of a buccaneer. Galton had proved only that one phenomenon – regression to the mean -- did not require a causal explanation. Now Pearson was completely removing causation from science. What made him take this leap?
- P 67 Porter: Pearson belonged to a philosophical school call positivism which holds that the universe is a product of human thought and that science is only a description of those thoughts.
- P 67-69 **Pearson's ascendance and dominance of early statistics.**

- P 69 There were cracks in Pearson's edifice of causality-free science. Pearson wrote of "spurious correlation": a concept impossible to make sense of without making some reference to causation.
- P 69 To Pearson, causation is just a "fetish amidst the inscrutable arena of modern science."
- P 69 This [view] put him in an awkward position when he has to explain why one correlation is meaningful and another is "spurious".
- P 69 Pearson noticed it is relatively easy to find correlations that are just plain silly. 1) a nations per capita chocolate consumption and its number of Nobel prize winners. 2) England's mortality rate in a given year and the percentage of marriages [that were] conducted that year in the Church of England. 3) When two populations are aggregated. For Paris catacomb skulls, the correlation between length and breadth was negligible for males and females done separately but was 0.2 when combined.
- P 70 *Figure 2.5 Karl Pearson with a skull from the Paris Catacombs*
- P 71 This [aggregate correlation] is a case of more-general phenomena called Simpson's paradox.
- P 71 Pearson: "For those who persist in looking upon all correlations as cause and effect, the fact that correlation can be produced between two quite uncorrelated characters A and B by taking an artificial mixture of two closely allied races, must com rather as a shock"
- P 71 Pearl: Looking at the same example through the lens of causality, we say, 'What a missed opportunity.' ... such examples might have spurred a talented scientist to think about the reasons for his shock and develop a science to predict when spurious correlations appear.
- P 71 Pearson's only guidance to his followers is that an "artificial" mixture is bad.
- P 71 Pearson's followers did not all follow in lockstep behind him. Yule (initially in the hardline camp) changed his mind when he needed to explain poverty conditions in London. In 1899, he studied whether "out-relief" developed to a pauper's home (vs. a poor-house) increased the rate of poverty. Data showed that districts with more out-relief had a higher poverty rate. Yule realized the correlation was possibly spurious: these districts might also have more elderly who tend to be poorer.
- P 72 When Yule compared districts with equal proportions of elderly, the correlation remained. This emboldened him to say that the increased poverty rate was due to out-relief. Afterward he wrote in a footnote, "strictly speaking, for 'due to' read 'associated with.'" This set the pattern for generations of scientists after him. They would think "due to" and say "associated with".
- P 72 SEWALL, WRIGHT, GUINEA PICS, AND PATH DIAGRAMS
- P 72 Wright came to Harvard to study genetics, at the time one of the hottest topics in science b/c Mendel's theory of dominant and recessive genes had just been rediscovered. Wright's advisor had identified 8 hereditary factors that affected fur color in rabbits. Wright was to do it for guinea pigs.
- P 73 Wright's guinea pigs were the springboard to his whole career and his whole theory of evolution.
- P 73 The inheritance of coat color in guinea pigs stubbornly refused to play by Mendelian rules. It proved virtually impossible to breed an all-white or all-colored guinea pig. Wright began to doubt that genetics alone governed the amount of white and postulated that "developmental factors" in the womb was causing some of the variation.
- P 74 *Figure 2.6. Sewall Wright was the first person to develop a mathematical method for answering casual question from data, known as path diagrams.*
- P 74 He showed that if we knew the casual quantities in Figure 2.7, we could predict the correlations in the data by a simple graphical rule. This rule sets up a bridge from the deep hidden rule of causation to the surface world of correlations. It was the first bridge ever built between causality & probability, the first crossing of the barrier between rung two and rung one on the Ladder of Causation.
- P 75 *Figure 2.7 Sewall Wright's first path diagram. (1920)*
- P 75 Having built this bridge, Wright could travel backward over it, *from* the correlations measured in the data *to* the hidden causal quantities. He did this by solving algebraic equations.
- P 75 This [solving equations] turned out to be revolutionary because it was the first proof that the mantra "Correlation does not imply causation" should give way to "Some correlations do imply causation."

- P 76 In a random population of guinea pigs, 45% of variation in color was due to heredity (58% was developmental). In a highly inbred family, only 3% was due to heredity (92%) was developmental.
- P 76 **Details on reading and interpreting the first path diagram**
- P 77 Each arrow is accompanied by a small letter. These letters, called path coefficients, represent the strength of the causal effects that Wright wanted to solve for. Roughly speaking, a path coefficient represents the amount of variability in the target variable accounted for by the source variable.
- P 77 This interpretation (amount of variation explained by a variable) was reasonable at the time. The modern causal interpretation is different: the path coefficients represent the results of a hypothetical intervention on the source variable.
- P 77 Pearl: I want to emphasize that the path diagram is not just a pretty picture; it is a powerful computation device because the rule for computing correlations (the bridge from rung two to rung one) involved tracing the paths that connect two variables to each other and multiplying the coefficient encountered along the way.
- P 77 Omitted arrows actual convey more significant assumptions than those that are present. An omitted arrow restricts the causal effect to zero, while a present arrow remains agnostic about the magnitude of the effect.
- P 77 Wright's paper was a tour de force... one of the landmarks of 20<sup>th</sup> century biology. Certainly it is a landmark for the history of causality. Figure 2.7 is the first causal diagram ever published, the first step of 20<sup>th</sup> century science onto the second rung of the Ladder of Causation. And not a tentative step but a bold and decisive one!
- P 78 The following year Wright published a much more general paper called "Correlation and Causation" that explained how path diagrams worked in other settings.
- P 78 The reaction he [Wright] got must have stunned him. A rebuttal by Henry Niles, a student of Raymond Pearl who in turn was a student of Karl Pearson: the godfather of statistics.
- P 78 Pearl: Academia is full of genteel savagery. I have seldom seen a criticism as save as Nile's. Niles disparages Wright's entire methodology.
- P 78 Niles: "We conclude that philosophically the basis of the method of path coefficients is faulty, while practically the result of applying it where it can be checked prove it to be wholly unreliable."
- P 78 Pearl: His [Niles's] paper is very important to us as historians of causation. First, it faithfully reflects the attitude of his generation toward causation and the total grip that his mentor, Karl Pearson, had on scientific thinking of his time. Second we continue to hear Niles's objections today.
- P 78 Pearl: Of course at times we don't know the entire web of relationships between their variables. Wright argued that we can use the diagrams in exploratory mode; we can postulate certain causal relationships and work out the predicted correlations between variables. If they contradict the data, we have evidence that the relationships we assumed were false.
- P 79 This way [exploratory] of using path diagrams, rediscovered in 1953 by Herbert Simon, inspired much work in the social sciences.
- P 78 Wright made one point clear: you cannot draw causal conclusions without some causal hypotheses.
- P 78 Pearl: You cannot answer a question on rung two of the Ladder of Causation using only data collected from rung one.
- P 79 Pearl: People ask me, "Doesn't that make causal reasoning circular? Aren't you assuming what you want to prove?" The answer is No. Extracting the non-obvious from the obvious is not circular – it is a scientific triumph and deserves to be hailed as such.
- P 79 Wright's contribution is unique because the information leading up to the conclusion (of 42% heritability) resided in two distinct, almost incompatible mathematical languages: the language of diagrams on the one side, and that of data on the other. The heretical idea of marrying qualitative "arrow information" to quantitative "data information" (two foreign languages) was one of the miracles that first attracted me, as a computer scientist, to this enterprise.
- P 79 Many people still make Niles's mistake of thinking that the goal of causal analysis is to prove that X is a cause of Y or to find the cause of Y from scratch: the problem of causal discovery.

- P 79 In contrast, the focus of Wright's research, as well as this book, is representing plausible causal knowledge in some mathematical language, combining it with empirical data, and answering causal queries that are of practical value.
- P 80 Wright understood from the very beginning that causal discovery was much more difficult and perhaps impossible.
- P 80 In his response to Niles, he [Wright] writes, "The writer [Wright] has never made the preposterous claim that the theory of path coefficients provides a general formula for the deduction of causal relations. He wishes to submit that the combination of knowledge of correlations with knowledge of causal relations to obtain certain results, is a different thing from the deduction of causal relations from correlation implied by Niles' statement."
- P 80 E PUR SI MUOVE (AND YET IT MOVES)
- P 80 Pearl: I cannot contain myself from expressing my sheer admiration for the precision of Wright's words in the quote ending the previous section which have not gone stale in the 90 years since he first articulated them and which essentially defined the new paradigm of modern causal analysis.
- P 80 My admiration for Wright's precision is second only to my admiration for his courage and determination. Wright did not blink. Only philosophers had dared to express an opinion on the nature of causation. Where did Wright get his inner conviction?
- P 81 Wright didn't have even one theorem to lean on. Scientists had abandoned causation so Wright could not fall back on any theoretical framework. Nor could he rely on authorities.
- P 81 But one solace Wright had, and one sign that he was on the right path, must have been his understanding that he could answer questions that cannot be answered in any other way.
- P 82 Another beautiful example asks "How much a guinea pig's birth weight will be affected if it spends one more day in the womb. I want to examine Wright's answer.
- P 82 We cannot answer Wright's question directly, because we can't weigh a guinea pig in the womb. We can compare the birth weights of guinea pigs that spend (say) 66 days gestating with those that spend 67 days. On average, 5.66# more at birth. We might naively suppose 5.6#/day in the womb.
- P 82 "Wrong," says Wright. Pups are usually born later for a reason: they have fewer litter mates. A pup with only two siblings will weigh more on day 66 than a pup with four siblings.
- P 82 Thus, the difference in birth weights has two causes, and we want to disentangle them. How much of the 5.66# per day is due to spending an additional day in utero and how much is due to having fewer siblings to compete with?
- P 82 *Figure 2.8 Causal (path) diagram for birth-weight example*
- P 82-84 **Path analysis of pre-birth weight gain per day: Details**
- P 84 Pearl: Why didn't you [Wright's colleagues] pay attention?
- P 84 Crowe suggests one reason: "path analysis doesn't lend itself to 'canned programs'. The user has to have a hypothesis and must devise an appropriate diagram of multiple causal sequences."
- P 85 Pearl: Path analysis requires scientific thinking, as does every exercise in causal inference. Statistics, as frequently practiced, discourages it [scientific thinking] and encourages "canned" procedures instead. Scientists will always prefer routine calculations on data to methods that challenge their scientific knowledge.
- P 85 R. A. Fisher, the undisputed high priest of statistics in the generation after Galton and Pearson, described this difference succinctly. In 1925, he wrote "Statistics may be regarded as ... the study of methods of the reduction of data." Pay attention to the words "methods", "reduction" and "data."
- P 85 Wright abhorred the idea of statistics as merely a collection of methods; Fisher embraced it.
- P 85 Causal analysis is definitely not just about data. We must incorporate some understanding of the process that produces the data, and then we get something that was not in the data to begin with.
- P 85 Wright's biographer, William Provine, noted that Fisher considered Wright his enemy. The real focus of the Fisher-Wright rivalry was not path analysis but evolutionary biology.
- P 85 In the 1960s, a group of social scientists rediscovered path analysis as a method of predicting the effect of social and educational policies.
- P 86 The fate of path analysis followed different trajectories: each leading to a betrayal of Wright's ideas.

- P 86 Sociologists renamed path analysis as structural equation modeling (SEM), embraced diagrams and used them extensively until 1970 when a computer package called LISREL automated the calculation of path coefficients (in some cases). In the 1980s a public challenge by Freedman to explain their assumptions went unanswered, and some leading SEM experts even disavowed that Ems had anything to do with causality.
- P 86 In Economics, the algebraic part of path analysis became known as simultaneous equation models (no acronym). Economists essentially never used path diagrams and continue not to use them to this day, relying instead on numerical equations and matrix algebra. A direct consequence of this is that, since equations are non-directional, economists had no notational means to distinguish causal from regression equations and thus were unable to answer policy-related questions.
- P 86 In 1983, Wright himself was called back into the ring in the *American Journal of Human Genetics*. At this time, Wright was past 90 years old. It is both wonderful and tragic to read his essay. Wonderful because 60 years had passed. Tragic because his theory had advanced little since the 1920s. It was a critique of path analysis in the same journal by Samuel Karlin.
- P 87 Karlin objected to path analysis for a reason that Niles did not raise. It [path analysis] assumes that all relationships are linear. This assumption allows Wright to describe the causal relationships with a single number. Neither knew that a general nonlinear theory was just around the corner.
- P 87 Karlin's most interesting criticism was also the one he considered most important. One can adopt an essentially model-free approach... This approach emphasizes...robustness in interpreting results.
- P 87 In one sentence, Karlin articulates how little had changed from the days of Pearson and how much influence Pearson's ideology still had in 1983. He [Karlin} is saying that the data themselves already contains all the scientific wisdom; they need only be cajoled and massaged...into dispersing those pearls of wisdom. We could do just as well, if not better, with a "model-free" approach.
- P 88 If Pearson were alive today, he would say exactly this: the answers are all in the data.
- P 88 Wight, to his credit, understood the enormous stakes and stated in no uncertain terms, "Karlin et al., are urging an abandonment of the purpose of path analysis and evaluation of the relative importance of varying causes. There can be no such analysis without a model. Wright understood that he was defending the very essence of the scientific method and the interpretation of data.
- P 88 Pearl: I would give the same advice today to big-data, model-free enthusiasts. Of course, it is okay to tease out all the information that the data can provide, but let's ask how far this will get us. It will never get us beyond the first rung of the Ladder of Causation, and it will never answer even as simple a question as "What is the relative importance of various causes/" *E pur si muove!*
- P 88 FROM OBJECTIVITY TO SUBJECTIVITY: THE BAYESIAN CONNECTION
- P 88 One other theme in Wright's rebuttal may hint at another reason for the resistance of statisticians to causality. He repeatedly states that he did not want path analysis to become "stereotyped." According to Wright, "The un-stereotyped approach of path analysis differs profoundly from the stereotyped modes of description designed to avoid any departures from complete objectivity."
- P 88 What does this mean? First, path analysis should be based on the user's personal understanding of causal processes reflected in the causal diagram. Second, Wright traces the allure of "model-free" methods to their objectivity. This [objectivity] has been the holy grail of statisticians since day one—or since March 15, 1834 when the Statistical Society of London was founded. Its founding charter said that data were to receive priority in all cases over opinions and interpretations. Data are objective; opinions are subjective. This paradigm long predates Pearson. The struggle for objectivity – the idea of reasoning exclusively from data and experiment – has long been part of the way that science has defined itself ever since Galileo.
- P 89 Unlike correlation and most of the other tools of mainstream statistics, causal analysis requires the user to make a subjective commitment. She must draw a causal diagram that reflects her in her field of expertise... She must abandon the centuries-old dogma of objectivity for objectivity's sake. Where causation is concerned, a grain of wise subjectivity tells us more about the real world than any amount of objectivity.

- P 89 In the above paragraph, I said "most of" the tools of statistics. There is one important exception: a branch of statistics called Bayesian statistics.
- P 90 Bayesian statistics give us an objective way of combining the observed evidence with our prior knowledge (or subjective belief) to obtain a revised belief.
- P 90 Unfortunately, the acceptance of Bayesian subjectivity in mainstream statistics did nothing to help the acceptance of causal subjectivity, the kind needed to specify a path diagram. Why? A grand linguistic barrier. Bayesian statisticians still use the language of probability, the native language of Galton and Pearson. The assumptions entering causal inference...require a richer language (path diagrams) that is foreign to Bayesians and frequentists alike.
- P 90 Moreover, the subjective component in causal information does not necessarily diminish over time, even as the amount of data increases. This is a terrifying prospect for advocates of scientific objectivity, which explains their refusal to accept the inevitability of relying on subjective causal information.

### Chapter 3: From Evidence to Causes: Reverend Bayes meets Dr. Holmes.

- P 93 Deduction: from Hypothesis to conclusion. Induction works in the opposite direction, from evidence to hypothesis.
- P 93 having induced several hypotheses, Holmes eliminated them one by one in order to deduce (by elimination) the correct one.
- P 93 In recent years experts in artificial intelligence (AI) have made considerable progress toward automating the process of reasoning from evidence to hypothesis and likewise from effect to cause. ...one of its basic tools, called Bayesian networks.
- P. 95 Bayesian networks, the machine reasoning tool that underlies the Bonaparte software, affect our lives in many ways. ...speech recognition, spam filters, weather forecasting, evaluation of potential oil wells, FDA approval process for medical devices.
- P. 95 In this chapter I will tell the story of Bayesian networks from their roots in the 18<sup>th</sup> century to their development in the 1980s.
- P. 95 They are related to causal diagrams in a simple ways: a causal diagram is a Bayesian network in which every arrow signifies a direct causal relation, or at least the possibility of one. Not all Bayesian networks are causal, and in many applications it does not matter.

#### REVEREND BAYES AND THE PROBLEM OF INVERSE PROBABILITY

- P 96 Fig 3.1 Title page of the Journal where Bayes' article was published.
- P 99 Suppose two-thirds order tea and half of the tea drinkers order scones. What fraction order both?
- 100 We could have observed that 5/12 of customers ordered scones and 4/5ths of these ordered tea. So what fraction ordered both?  $4/12 = 1/3$ . No coincidence that it came out the same.

	Scones		
Tea	No	Yes	ALL
No			
Yes		40	80
ALL		50	120

- 101  $P(S\&T) = P(S|T)*P(T) = P(T|S)*P(S)$  Bayes Rule
- 102 Mathematically that is all there is to Bayes rule. I have glossed over two profound objections: one philosophical, and the other practical.
- 102 philosophical: can we legitimately translate the expression "given that I know" into the language of probabilities? The expression, "given that I know" is epistemological and should be governed by the logic of knowledge, not that of frequencies and proportions.
- 103 1931 Harold Jeffries introduced the now standard vertical bar in  $P(S|T)$
- 103 The more miraculous the miracle, the more credible the hypothesis that explains its occurrence. '
- 103 Practically, we need a prior probability of the length (of the table). This variability, also known as subjectivity, is sometimes seen as a disadvantage in Bayesian inference.

Chapter 4: Confounding and De-confounding: Or, Slaying the Lurking Variable

Chapter 5: The Smoke-Filled Debate: Clearing the Air

P 175. "Cornfield condition"

P 176. Sensitivity analysis

P 179. There is much we can learn from observational data.

Smoking for Newborns.