

The Book of Why
The New Science of Cause and Effect
by Judea Pearl and Dana Mackenzie

reviewed by Lisa R. Goldberg*

forthcoming in *Notices of the American Mathematical Society*

October 25, 2018[†]

Judea Pearl is on a mission to change the way we interpret data. An eminent Professor of Computer Science, Pearl has documented his research and opinions in scholarly books and papers. Now, he has made his ideas accessible to a broad audience in *The Book of Why: The New Science of Cause and Effect*, co-authored with science writer Dana Mackenzie. With the release of this historically grounded and thought provoking book, Pearl leaps from the ivory tower into the real world.

The Book of Why takes aim at perceived limitations of *observational studies*, whose underlying data are found in nature and not controlled by researchers. Many believe that an observational study can elucidate association but not cause and effect. It cannot tell you why.

Perhaps the most famous example concerns the impact of smoking on health. By the mid 1950s, researchers had established a strong association

*Lisa Goldberg is a co-Director of the Consortium for Data Analytics in Risk and Adjunct Professor Economics and Statistics at University of California, Berkeley, She is Director of Research at Aperio Group, LLC.

[†]This review has benefitted from dialogs with David Aldous, Bob Anderson, Wachi Bandera, Jeff Bohn, Brad DeLong, Peng Ding, Tingyue Gan, Nate Jensen, Barry Mazur, Liz Michaels, LaDene Otsuki, Caroline Ribet, Ken Ribet, Stephanie Ribet, Cosma Shalizi, Alex Shkolnik, Philip Stark, Lee Wilkinson and the attendees of the University of California, Berkeley Statistics Department social lunch group. Thanks to Nick Jewell for informing me about scientific studies on the relationship between exercise and cholesterol, which enhanced my appreciation of *The Book of Why*.

between smoking and lung cancer. Only in 1984, however, did the US government mandate the phrase “smoking causes lung cancer.” The holdup was the specter of a latent factor, perhaps something genetic, that might cause both lung cancer and a craving for tobacco. If the latent factor were responsible for lung cancer, limiting cigarette smoking would not prevent the disease. Naturally, tobacco companies were fond of this explanation, but it was also advocated by the prominent statistician Ronald A. Fisher, co-inventor of the so-called gold standard of experimentation, the Randomized Controlled Trial (RCT).

Subjects in an RCT on smoking and lung cancer would have been assigned to smoke or not on the flip of a coin. The study had the potential to disqualify a latent factor as the primary cause of lung cancer and elevate cigarettes to the leading suspect. Since a smoking RCT would have been unethical, however, researchers made do with observational studies showing association, and demurred on the question of cause and effect for decades.

Was the problem simply that the tools available in the 1950s and 1960s were too limited in scope? Pearl addresses that question in his three-step *Ladder of Causation*, which organizes inferential methods in terms of the problems they can solve. The bottom rung is for model-free statistical methods that rely strictly on association or correlation. The middle rung is for interventions that allow for the measurement of cause and effect in observational studies. The top rung is for counterfactual analysis, the exploration of alternative realities.

Early scientific inquiries about the relationship between smoking and lung cancer relied on the bottom rung, model-free statistical methods whose modern analogs dominate the analysis of observational studies today. In one of *The Book of Why's* many wonderful historical anecdotes, the predominance of these methods is traced to the work of Francis Galton, who discovered the principle of regression to the mean in an attempt to understand the process that drives heredity of human characteristics. Regression to the mean involves association, and this led Galton and his disciple, Karl Pearson, to conclude that association was more central to science than causation.

Pearl places deep learning and other modern data mining tools on the bottom rung of the Ladder of Causation. Bottom rung methods include Alpha Go, the deep learning program defeated the world's best human Go players in 2015 and 2016, [1]. For the benefit of those who remember the ancient times before data mining changed everything, he explains,

The successes of deep learning have been truly remarkable and have caught many of us by surprise. Nevertheless, deep learning has succeeded primarily by showing that certain questions or tasks we thought were difficult are in fact not.

The issue is that algorithms, unlike three-year-olds, do as they are told, but in order to create an algorithm capable of causal reasoning,

...we have to teach the computer how to selectively break the rules of logic. Computers are not good at breaking rules, a skill at which children excel.

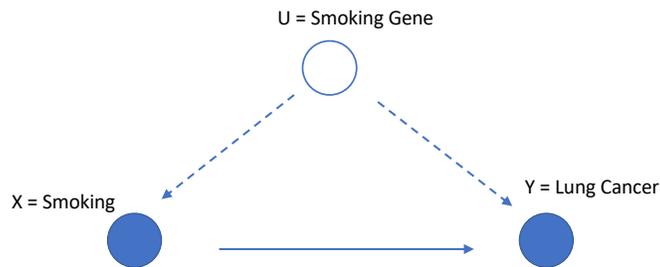


Figure 1: Causal model of assumed relationships among smoking, lung cancer, and a smoking gene.

Methods for extracting causal conclusions from observational studies are on the middle rung of Pearl’s Ladder of Causation, and they are expressed in a mathematical language that extends classical statistics and emphasizes graphical models,

Various options exist for causal models: causal diagrams, structural equations, logical statements, and so forth. I am strongly sold on causal diagrams for nearly all applications, primarily due to their transparency but also due to the explicit answers they provide to many of the questions we wish to ask.

The use of graphical models to determine cause and effect in observational studies was pioneered by Sewall Wright, whose work on the effects of birth weight, litter size, length of gestation period and other variables on the weight of a 33-day-old guinea pig is in [4]. Pearl relates Wright’s persistence in response to the cold reception his work received from the scientific community.

My admiration for Wright’s precision is second only to my admiration for his courage and determination. Imagine the situation in 1921. A self-taught mathematician faces the hegemony of the statistical establishment alone. They tell him “Your method is based on a complete misapprehension of the nature of causality in the scientific sense.” And he retorts, “Not so! My method is important and goes beyond anything you can generate.”

Pearl defines a *causal model* to be a directed acyclic graph that can be paired with data to produce quantitative causal estimates. The graph embodies the structural relationships that a researcher assumes are driving empirical results. The structure of the graphical model, including the identification of vertices as mediators, confounders, or colliders, guides experimental design through the identification of minimal sets of control variables. Modern expositions on graphical cause and effect models are [2] and [3].

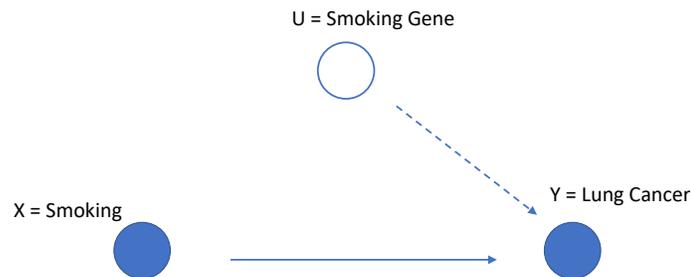


Figure 2: Mutated causal model facilitating the calculation of the effect of smoking on lung cancer. The arrow from the confounding smoking gene to the act of smoking is deleted.

Within this framework, Pearl defines the *do operator*, which isolates the impact of a single variable from other effects. The probability of Y do X , $P[Y|\text{do}(X)]$, is not the same thing as the conditional probability of Y given X . Rather $P[Y|\text{do}(X)]$ is estimated in a mutated causal model, from which arrows pointing into the assumed cause are removed. *Confounding* is the difference between $P[Y|\text{do}(X)]$ and $P[Y|X]$. In the 1950s, researchers were

after the former but could estimate only the latter in observational studies. That was Ronald A. Fisher's point.

Figure 1 depicts a simplified relationship between smoking and lung cancer. Directed edges represent assumed causal relationships, and the smoking gene is represented by an empty circle, indicating that variable was not observable when the connection between smoking and cancer was in question. Filled circles represent quantities that could be measured, like rates of smoking and lung cancer in a population. Figure 2 shows the mutated causal model that isolates the impact of smoking on lung cancer.

The conclusion that smoking causes lung cancer was eventually reached without appealing to a causal model. A crush of evidence, including the powerful sensitivity analysis developed in [5], ultimately swayed opinion. Pearl argues that his methods, had they been available, might have resolved the issue sooner. Pearl illustrates his point in a hypothetical setting where smoking causes cancer only by depositing tar in lungs. The corresponding causal diagram is shown in Figure 3. His *front door formula* corrects for the confounding of the unobservable smoking gene without ever mentioning it. The bias-corrected impact of smoking, X , on lung cancer, Y can be expressed

$$P[Y|\text{do}(X)] = \sum_z P[Z|X] \sum_{x'} P[y|x', z]P[x'].$$

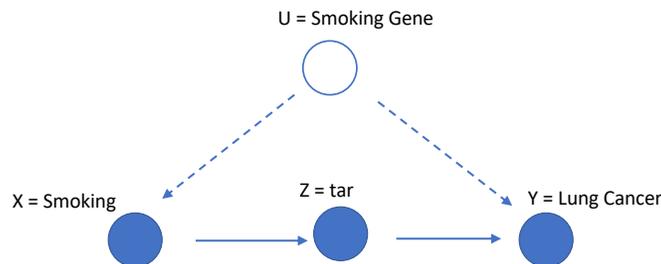


Figure 3: Pearl's front door formula corrects for bias due to latent variables in certain examples.

The Book of Why draws from a substantial body of academic literature, which I explored in order to get a more complete picture of Pearl's

work. From a mathematical perspective, an important application is Nicholas Christakis and James Fowler's 2007 study described in [6] arguing that obesity is contagious. The attention-grabbing claim was controversial because the mechanism of social contagion is hard to pin down, and because the study was observational. In their paper, Christakis and Fowler upgraded an observed association, clusters of obese individuals in a social network, to the assertion that obese individuals cause their friends, and friends of their friends, to become obese. It is difficult to comprehend the complex web of assumptions, arguments and data that comprise this study. It is also difficult to comprehend its nuanced refutations by Russell Lyons [7] and by Cosma Shalizi and Andrew Thomas [8], which appeared in 2011. There is a moment of clarity, however, in the commentary by Shalizi and Thomas, when they cite Pearl's theorem about *non-identifiability* in particular graphical models. Using Pearl's results, Shalizi and Thomas show that in the social network that Christakis and Fowler studied, it is impossible to disentangle contagion, the propagation of obesity via friendship, from the shared inclinations that led the friendship to be formed in the first place.

The top rung of the Ladder of Causation concerns counterfactuals, which Michael Lewis brought to the attention of the world with his best selling book, *The Undoing Project*, [9]. Lewis tells the story of Israeli psychologists Daniel Kahneman and Amos Tversky, experts in human error, who fundamentally changed our understanding of how we make decisions. Pearl draws on the work of Kahneman and Tversky in *The Book of Why*, and Pearl's approach to analyzing counterfactuals might be best explained in terms of a question that Kahneman and Tversky posed in their study [10] of how we explore alternative realities.

How close did Hitler's scientists come to developing the atom bomb in World War II? If they had developed it in February 1945, would the outcome of the war been different?

-The Simulation Heuristic

Pearl's approach to this question includes the *probability of necessity* for Germany and its allies to have won World II had they developed the atom bomb in 1945, given our historical knowledge that they did not have an atomic bomb in February 1945 and lost the war. If Y denotes Germany winning or losing the war (0 or 1) and X denotes Germany having the bomb in 1945 or not having it (0 or 1), the probability of necessity can be expressed

in the language of potential outcomes,

$$P[Y_{X=0} = 0 \mid X = 1, Y = 1].$$

Dual to the *probability of sufficiency*, the probability of necessity mirrors the legal notion of “but-for” causation as in: but for its failure to build an atomic bomb by February 1945, Germany would probably have won the war. Pearl applies the same type of reasoning to generate transparent statements regarding climate change. Was anthropogenic global warming responsible for the 2003 heat wave in Europe? We’ve all heard that while global warming due to human activity tends to raise the probability of extreme heat waves, it is not possible to attribute any particular event to this activity. According to Pearl and a team of climate scientists, the response can be framed differently: there is a 90% chance that the 2003 heat wave in Europe would not have occurred in the absence of anthropogenic global warming, [11].

This formulation of the impact of anthropogenic global warming on the earth is strong and clear, but is it correct? According to the principle of garbage-in-garbage-out, results based on a causal model are no better than its underlying assumptions. These assumptions can represent a researcher’s knowledge and experience. However, many scholars are concerned that model assumptions represent researcher bias, or are simply unexamined. David Freedman emphasizes this in [12], and as he wrote more recently in [13],

Assumptions behind models are rarely articulated, let alone defended. The problem is exacerbated because journals tend to favor a mild degree of novelty in statistical procedures. Modeling, the search for significance, the preference for novelty, and the lack of interest in assumptions—these norms are likely to generate a flood of non-reproducible results.

–Oasis or Mirage

Causal models can be used to work backwards from conclusions we favor to supporting assumptions. Our tendency to reason in the service of our prior beliefs is a favorite topic of moral psychologist, Jonathan Haidt, author of *The Righteous Mind* [14], who wrote about “the emotional dog and its rational tail.” Or as Udney Yule explained in [15],

Now I suppose it is possible, given a little ingenuity and good will, to rationalize very nearly anything.

–1926 presidential address to the Royal Statistical Society

Concern about the impact of biases and preconceptions on empirical studies is growing, and it comes from sources as diverse as Professor of Medicine John Ioannides, who explained why most published research findings are false [16], comedian John Oliver, who warned us to be skeptical when we hear the phrase “studies show” [17], and former *New Yorker* writer Jonah Lehrer, who wrote about the problems with empirical science in [18] but was later discredited for representing stuff he made up as fact.

The graphical approach to causal inference that Pearl favors has been influential, but it is not the only approach. Many researchers rely on the Neyman (or Neyman–Rubin) potential outcomes model, which is discussed in [19], [20], [21] and [22]. In the language of medical randomized control trials, a researcher using this model tries to quantify the difference in impact between treatment and no treatment on subjects in an observational study. Propensity scores are matched in an attempt to balance inequities between treated and untreated subjects. Since no subject can be both treated and untreated, however, the required estimate of impact can be formulated as a missing value problem. Other researchers favor models by James Heckman, [23] whose concept of “fixing” resembles, superficially at least, the do operator that Pearl uses. Those who enjoy scholarly disputes may look to Andrew Gelman’s blog, [24] and [25], for back-and-forth between Pearl and Rubin disciples (Rubin himself does not seem to participate—in that forum, at least) or to the tributes written by Pearl [26] and Heckman [27] to the reclusive Nobel Laureate, Trygve Haavelmo, who pioneered causal inference in economics in the 1940 in [28] and [29]. These dialogs have been contentious at times, and they bring to mind Sayre’s law, which says that academic politics is the most vicious and bitter form of politics because the stakes are so low. It is this reviewer’s opinion that the differences between these approaches to causal inference are far less important than their similarities. Support for this includes a construction by Thomas Richard and James Robins incorporating counterfactuals into graphical cause-and-effect models, thereby unifying various threads of the causal inference literature, [30].

Late one afternoon in July 2018, Pearl’s co-author Dana Mackenzie spoke on causal inference at UC Berkeley’s Simons Institute. His presentation was in the first person singular from Pearl’s perspective, the same voice used in *The Book of Why*, and it concluded with an image of the first self-driving car to kill a pedestrian. According to a report [31] by the National Transportation Safety Board (NTSB), the car recognized an object in its path six seconds prior to the fatal collision. With a lead time of a second and a

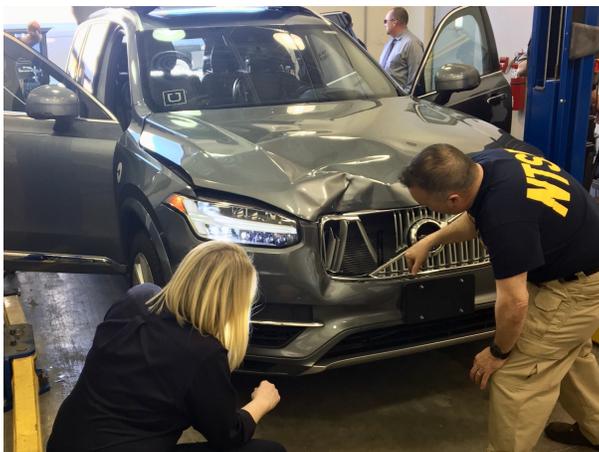


Figure 4: National Transportation Safety Board Inspectors examining the self-driving Uber that killed a pedestrian in Tempe, Arizona on March 18, 2018.

half, the car identified the object as a pedestrian. When the car attempted to engage its emergency braking system, nothing happened. The NTSB report states that engineers had disabled the system in response to a preponderance of false positives in test runs.

The engineers were right, of course, that frequent, abrupt stops render a self-driving car useless. Mackenzie gently and optimistically suggested that endowing the car with a causal model that can make nuanced judgments about pedestrian intent might help. If this were to lead to safer and smarter self-driving cars, it would not be the first time that Pearl's ideas would have led to better technology. His foundational work on Bayesian networks has been incorporated into cell phone technology, spam filters, bio-monitoring and many other applications of practical importance.

Professor Judea Pearl has given us an elegant, powerful, controversial theory of causality. How can he give his theory the best shot at changing the way we interpret data? There is no recipe for doing this, but teaming up with science writer and teacher, Dana Mackenzie, a scholar in his own right, was a pretty good idea.

References

- [1] D. Silver, J. S. K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, “Mastering the game of go without human knowledge,” *Nature*, vol. 550, pp. 354–359, 2017.
- [2] J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge University Press, second ed., 2009.
- [3] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction and Search*. The MIT Press, 2000.
- [4] S. Wright, “Correlation and causation,” *Journal of Agricultural Research*, vol. 20, no. 7, pp. 557–585, 1921.
- [5] J. Cornfield, W. Haenszel, E. C. Hammond, M. B. Shimkin, and E. L. Wynder, “Smoking and lung cancer: recent evidence and a discussion of some questions,” *Journal of the National Cancer Institute*, vol. 22, no. 1, pp. 173–203, 1959.
- [6] N. A. Christakis and J. H. Fowler, “The spread of obesity in a large social network over 32 years,” *The New England Journal of Medicine*, vol. 357, no. 4, pp. 370–379, 2007.
- [7] R. Lyons, “The spread of evidence-poor medicine via flawed social-network analysis,” *Statistics, Politics, and Policy*, vol. 2, no. 1, pp. DOI: 10.2202/2151–7509.1024, 2011.
- [8] C. R. Shalizi and A. C. Thomas, “Homophily and contagion are generically confounded in observational social network studies,” *Sociological Methods & Research*, vol. 40, no. 2, pp. 211–239, 2011.
- [9] M. Lewis, *The Undoing Project: A Friendship That Changed Our Minds*. W.W. Norton and Company, 2016.
- [10] D. Kahneman and A. Tversky, “The simulation heuristic,” in *Judgment under Uncertainty: Heuristics and Biases* (D. Kahneman, P. Slovic, and A. Tversky, eds.), pp. 201–208, Cambridge University Press, 1982.

- [11] A. Hannart, J. Pearl, F. Otto, P. Naveu, and M. Ghil, “Causal counterfactual theory for the attribution of weather and climate-related events,” *Bulletin of the American Meteorological Society*, vol. 97, pp. 99–110, 2016.
- [12] D. A. Freedman, “Statistical models and shoe leather,” *Sociological Methodology*, vol. 21, pp. 291–313, 1991.
- [13] D. A. Freedman, “Oasis or mirage?,” *Chance*, vol. 21, no. 1, pp. 59–61, 2009.
- [14] J. Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Vintage, 2013.
- [15] U. Yule, “Why do we sometimes get nonsense-correlations between time-series?—a study insampling and the nature of time-series,” *Royal Statistical Society*, vol. 89, no. 1, 1926.
- [16] J. P. A. Ioannidis, “Why most published research findings are false,” *PLoS Med*, vol. 2, no. 8, p. <https://doi.org/10.1371/journal.pmed.0020124>, 2005.
- [17] J. Oliver, “Scientific studies: Last week tonight with John Oliver (hbo),” May 2016.
- [18] J. Lehrer, “The truth wears off,” *The New Yorker*, December 2010.
- [19] J. Neyman, “Sur les applications de la theorie des probabilities aux experiences agricoles: Essaies des principes.,” *Statistical Science*, vol. 5, pp. 463–472, 1923, 1990. 1923 manuscript translated by D.M. Dabrowska and T.P. Speed.
- [20] D. B. Rubin, “Estimating causal effects of treatments in randomized and non-randomized studies,” *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.
- [21] D. B. Rubin, “Causal inference using potential outcomes,” *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 322–331, 2005.

- [22] J. Sekhon, “The Neyman-Rubin model of causal inference and estimation via matching methods,” in *The Oxford Handbook of Political Methodology* (J. M. Box-Steffensmeier, H. E. Brady, and D. Collier, eds.), Oxford Handbooks Online, Oxford University Press, 2008.
- [23] J. J. Heckman, “The scientific model of causality,” *Sociological Methodology*, vol. 35, pp. 1–97, 2005.
- [24] A. Gelman, “Resolving disputes between J. Pearl and D. Rubin on causal inference,” July 2009.
- [25] A. Gelman, “Judea Pearl overview on causal inference, and more general thoughts on the reexpression of existing methods by considering their implicit assumptions,” 2014.
- [26] J. Pearl, “Trygve Haavelmo and the emergence of causal calculus,” *Econometric Theory*, vol. 31, no. 1, pp. 152–179, 2015.
- [27] J. Heckman and R. Pinto, “Causal analysis after Haavelmo,” *Econometric Theory*, vol. 31, no. 1, pp. 115–151, 2015.
- [28] T. Haavelmo, “The statistical implications of a system of simultaneous equations,” *Econometrica*, vol. 11, no. 1, pp. 1–12, 1943.
- [29] T. Haavelmo, “The probability approach in econometrics,” *Econometrica*, vol. 12, no. Supplement, pp. iii–iv+1–115, 1944.
- [30] T. S. Richardson and J. M. Robins, “Single world intervention graphs (SWIGS): A unification of the counterfactual and graphical approaches to causality.” April 2013.
- [31] NTSB, “Preliminary report released for crash involving pedestrian, uber technologies, inc., test vehicle,” May 2018.