

The Book of Why: The New Science of Cause and Effect

Pearl, J., & Mackenzie, D. (2018). New York, NY: Basic Books.

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
<http://www.jmde.com>

Steve Powell

ProMENTE Social Research, Sarajevo

This is a review of “The Book of Why”, (Pearl & Mackenzie, 2018), the first book for a general readership by Turing Prize winner Judea Pearl, one of the parents both of Artificial Intelligence and of the “Causal Revolution” in statistics. The present review is more extensive than most book reviews because of the fundamental significance of this book and its subject matter to the evaluation community.

Causality matters to evaluators. We need to understand how elements of a project or program are supposed to work: how interventions made on some things might contribute to differences made to other things, which themselves have further consequences. As such, we as evaluators are concerned with causality. Also when we try to judge in retrospect what in fact contributed to what, we are dealing with causality.

To be sure, most evaluators have abandoned, or would like to have the freedom to abandon, overly-mechanistic and linear models of how interventions work. The theories of change we (would like to) deal with are perhaps structured more like diagrams of organisms or ecosystems than like diagrams of machines, and they may have non-linear, partial, probabilistic connections with feedback loops, and may have vague and/or missing information. But they are networks of putative causal connections all the same. Even when someone says, “this is a complex system, it is very hard to predict its behavior, we need to identify key leverage points to influence its development in such-and-such a way”, they are still expressing a causal theory, albeit a fuzzy one, that tweaking this might influence that. So, the concept of causality in one form or another is essential to evaluation.

The central premise of the book is that we have been lacking almost any well-established, formal way to talk about causality – to express causal questions or answers, to write down the equation for “smoking causes cancer” or the question “what was the causal contribution of this intervention to that effect?” The place most empirical social scientists have turned to in the search for such a language is statistics. But for nearly a century, discourse around causality was dominated by the thousand-pound gorillas of classical, associationist statistics, originally in the persons of R.A. Fisher and Karl Pearson, who told us that firstly correlation is not causation, and secondly that there is no such thing as evidence for causation except in the case of a Randomised Controlled Trial (RCT), and thirdly that even when we have an RCT there is no way to explain what happened to individuals within that very trial; explanations were restricted to the collective. Consequently, we were told that in cases where no trial exists there can be no causal explanations, and where no trial can ever be conducted (perhaps for ethical reasons) there will never be any causal explanations. So, essentially, except in the rare and special case of reporting the result of a randomised controlled trial, the word “cause” was taboo.

Some fringe disciplines did flourish at the edge of the jungle away from these thousand-pound gorillas, giving us fragments of languages to talk about causation - Fuzzy Sets and Fuzzy Causal Maps (Kosko, 1986; Zadeh, 1973), Contribution Analysis (Mayne J, 2001), Process Tracing – see Collier (2011) – amongst others. But no alternative paradigm was strong enough to unseat the

associationist gorillas. Evaluation and the social sciences suffered as a result.

In practice, of course, we pose causal questions and give causal explanations all the time, very often on an ad-hoc basis. Children learn to do it in their very first years (even though they never conduct randomised controlled trials), just as they learn to use mathematical language. In later years, they will learn how to formalise mathematical questions and calculate their answers using special symbols such as the “+” symbol for addition. But until recently, no-one was taught how to formalise causal questions or how to calculate or express the answers.

Now all that is changing. A causal revolution is shaking the jungle, thanks primarily to the work of statistician Donald Rubin and, above all, to philosopher-statistician Judea Pearl. This is not a fringe skirmish or a dissenting footnote but a root-and-branch rewriting of how we formulate and gather evidence for causal statements, which is being felt in disciplines from epidemiology to climate science.

Most evaluators are not aware of this revolution. The present review attempts to contribute to spreading the good news to the evaluation community.

Judea Pearl and his Contribution

Judea Pearl started his career as an engineer. In the 1980s he was working on problems in artificial intelligence. He wanted to be able to teach an artificially intelligent system to navigate and interact with the real world. He invented Bayesian Networks in order to facilitate the processing of probabilistic rather than only deterministic information (some of his work powers the phone in your pocket, along with many other innovations). More specifically, although the idea of using probabilistic information was not new, in practice, probabilistic knowledge was encoded in huge probability tables in which the association of every variable with every other was stored, and these required just too much computing power to work with. Pearl’s contribution was to show how to use network diagrams to break down these tables into much more manageable units.

But Pearl was still frustrated because these Bayesian Networks could not encode causal information. They could not explain why making the cock crow does not cause the sun to rise, even though the two events are highly correlated: if one occurs, the other is highly probable. Classical statistics does not give us any convincing way to say

anything about the direction or underlying nature of this connection - and Pearl and his colleagues had no way to store this information in an artificially intelligent system. But, he reasoned, if children can learn to understand and store causal information, machines can too.

Pearl went back to the work of the geneticist Sewall Wright who invented Path Analysis as a way to encode causal information using diagrams in the early 1920s (Wright, 1921). These causal diagrams represent connections prior to the probabilistic information we actually observe. They explain and go beyond our observations.

Wright’s work was vigorously and explicitly discouraged by the thousand-pound gorillas of Pearson and Fisher. They stuck almost religiously to the positivistic ideology that all knowledge is sensory information, and sensory information can not encode causal connections.

The authors also refer to the work of Barbara Burks (1926), who may have preceded Wright with the use of causal diagrams in particular in the study of mediation, but the uptake of her work suffered under the twin pressures of mainstream statistics and the prejudices against women in science in the early and mid-twentieth century.

Causal diagrams are easy to understand and are not so different in principle from the logic models and theories of change, with all their various weaknesses, which evaluators and program staff use all the time. Essentially, a causal arrow from “fire” to “smoke” says “intervening on the variable ‘fire’ will do something to the variable ‘smoke’”. As a by-product, this intervention may also make “smoke” more probable; but this depends also on other circumstances (other variables affecting “smoke”). The arrow in a Pearlian diagram is about causal contribution, not about probability: it explains the probabilities. It remains a stable and valid piece of knowledge even if in a particular instance someone sucks away the smoke with a vacuum cleaner.

By around 2000, Pearl had succeeded in providing a formal way to store causal information and solve causal problems using a combination of diagrams and mathematical expressions. This work is summarised in his outstanding contribution (Pearl, 2000). Unfortunately, even a reader with some familiarity with graph theory, statistics and formal logic will probably find it a difficult read. Pearl’s web page does list some public presentations which are more accessible, but only with the publication of “The Book of Why” in 2018 does the ordinary reader finally have access to some of Pearl’s most consequential ideas.

What Pearl and his co-workers have done is to break the taboo imposed by classical statistics on explicitly causal language, whether graphical or written, in mainstream empirical social science.

What is in the Book?

Pearl presents a ladder of knowledge with three levels: Association, Causation and Counterfactuals, differentiated in terms of the kind of statements which can be asked and answered at each level. The chapters of the book are loosely structured around these levels.

Level One (Association)

Classical and Bayesian statistics are situated on this level. From a Bayesian perspective, typical questions at this level are of the form:

How would seeing X change my belief in Y?

So, for example, seeing smoke raises the probability of fire:

$$P(\text{fire}) \mid \text{see}(\text{smoke}) > P(\text{fire}) \mid \text{see}(\text{not-smoke})$$

This “see()” operator is not necessary in classical statistics but it is shown here to distinguish it from the do() operator which is introduced in Level Two. (“P(x)” means “the probability of x”, and “|” means “given that”.)

Information at this level has, on its own, no causal implications. If, and only if, we are in possession of the appropriate causal model (see Level Two) can we use it to answer Level Two queries about causation or even Level Three queries about counterfactuals. Therefore, according to Pearl, “data-driven” approaches in machine learning and “Big Data” will forever remain at this first level if they do not also integrate pre-existing causal knowledge about the domains in question.

Level Two (Causation)

Pearl’s approach to encoding causal information combines causal diagrams and ordinary mathematical and statistical expressions, extended to include the “do()” operator.

Pearl’s causal diagrams. Pearl’s causal diagrams are in essence no different than those introduced by Burks and White, consisting of symbols for different variables connected by arrows

representing explicitly causal, not correlational, links.

Each link in a diagram corresponds to a written expression of the form

$$\text{smoke} = f(\text{fire}, \dots)$$

in which the value of one variable is expressed as a (mathematical) function of one or more other variables. It is important to note that these are not equations in the normal sense, in particular because it is not usually possible to reverse them: the sentence above says that smoke is causally dependent on fire, and not vice versa.

Some evaluators might be about to stop reading at this point because it might seem that this approach is only useful for things which can be measured precisely in numbers. Far from it: although many of Pearl’s examples do assume linear, numerical models, Pearl and his collaborators have shown that this approach is general enough to include any functional relationships expressing how variables of any type are influenced by others (this generalization is due to a student of Pearl, Thomas Verma). For example, we might believe that the acceptance of a new policy depends in part in some (as yet indeterminate) way on its ability to inspire young people. Although the authors do not explicitly say so, there is nothing to stop us writing this down:

$$\text{acceptance of new policy} = f(\text{ability of policy to inspire young people}, \dots)$$

and drawing a corresponding arrow in our causal diagram even before we know with any precision how we are going to measure these variables or how we are going to formulate their relationship. We have certainly made no assumptions that this relationship should be deterministic or linear. Yet we have already started Pearlian theory-building. For example, we can predict the results of some Level One observations, such as that we will not, in general, find zero association between some measure of acceptance and some other measure of the ability of the policy to inspire young people.

Perhaps the most important task facing a researcher is to find out about the causal influence of one variable, X, on another variable, Y, while excluding any influences on Y not actually due to the causal effect of X. Pearl’s most substantial contribution has been to identify several important rules for distinguishing causal from non-causal paths in a causal diagram just by looking at its structure, as explained very briefly in the following paragraph.

Causal influences of X on Y can be conveyed by any path in which all the arrows point from X to Y. Pearl explains that non-causal influences can potentially be conveyed via a “back-door” path from X to Y: any path that starts with an arrow pointing towards X. For example, the path $X \leftarrow B \rightarrow Y$ is a (short) back-door path from X to Y. B is the researcher’s nightmare: a confounding variable which, if not controlled for, makes it look like X is influencing Y when in fact both are being influenced, at least partly, by B. But how can a researcher know which variables need controlling for, when the intervening path(s) might involve several variables? Any such path presents a risk that X and Y might be confounded, i.e. they might be associated in ways which are not due to X’s causal influence over Y. (The possibility of such paths is the key reason to demand randomisation of X – to break the influence of any other variables on X by ensuring that X is controlled only by the throw of dice.) Pearl explains under what conditions non-causal information can flow, or not flow, down a back-door path. A path from B to D is blocked by a configuration like $B \rightarrow C \leftarrow D$ (called a “collider”); controlling for C unblocks this junction and allows information to flow through it. Conversely, “chain”

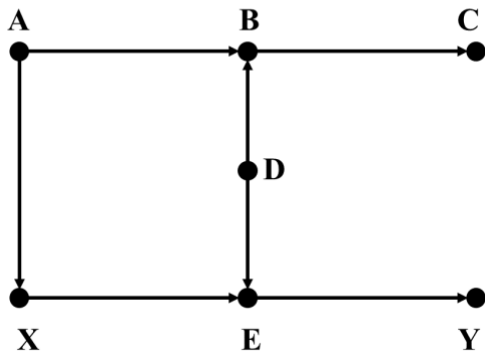


Figure 1. Example of a Causal Network (Adapted from Pearl and Mackenzie, 2018).

junctions like $B \rightarrow C \rightarrow D$ or “forks” like $B \leftarrow C \rightarrow D$, when uncontrolled, allow information flow - which can be blocked by controlling for C. For example, in Figure 1 there is a (back-door) path $X \leftarrow A \rightarrow B \leftarrow D \rightarrow E \rightarrow Y$ which is already blocked by the collider at B. Controlling for B would be a disaster because it would open this back-door route - contrary to the instinct of many statisticians to control for any variable which seems relevant.

Using the “back-door” rule it is possible to just look at the structure of a diagram to identify potential confounding and “de-confounding” variables; treating these variables correctly will allow a researcher to make predictions about the result of an intervention X on Y without performing it. In short, to understand the causal influence of X on Y free from confounding influences. Later in the book, the authors also introduce the “front-door” rule which potentially allows researchers to control for arbitrary confounding variables.

The “do operator”. Level Two also introduces the “do operator” which allows us to say things like this:

$$P(\text{smoke}) \mid \text{do}(\text{fire}) > P(\text{smoke}) \mid \text{do}(\text{not-fire})$$

which translates as “The probability of smoke given an intervention which makes fire happen is greater than the probability of smoke given an intervention which makes fire not happen”.

The “do operator” features in three axioms (which can perhaps be seen as collectively defining the operator). The authors use them to show, amongst other things, how to calculate effects of an intervention X on variables several links away from it in an arbitrarily complicated causal network.

‘Do(X)’ is essentially different from its cousin ‘see(X)’. Making smoke (via some method which is independent of lighting the original fire, e.g. by using a smoke machine) does nothing to the probability of fire:

$$P(\text{fire}) \mid \text{do}(\text{smoke}) = P(\text{fire}) \mid \text{do}(\text{not-smoke})$$

which can be contrasted with:

$$P(\text{fire}) \mid \text{see}(\text{smoke}) > P(\text{fire}) \mid \text{see}(\text{not-smoke})$$

as we saw above.

We know all this, provided we possess the appropriate causal information expressed in this diagram:

$$\text{fire} \rightarrow \text{smoke}$$

In particular we know that when we apply a do operator to a variable, we remove all the arrows pointing to that variable in the corresponding causal diagram – what Pearl calls “doing surgery” on the diagram, an insight he attributes to Peter Spirtes (Spirtes, Glymour, and Scheines, 2000) – a procedure which neatly removes any possible back-door paths which might allow non-causal influences to flow from X.

This approach defines the effect of an intervention on a variable downstream of the intervention in terms of a comparison between the actual value of the variable given the intervention and the alternative value which it would take without the intervention. This comparative understanding of effects goes back at least to Daniel Lewis (1974). But Lewis's approach was based on comparing the actual world with a similar world which is "just different enough" to provide an appropriate contrast. He struggled to define just how different is different enough, for the same reason that statisticians have struggled to know which variables to control for and which not to control for in causal analysis. Pearl's formulation avoids this problem because applying the "do operator" allow us potentially to calculate not only the actual and contrasting values of the variable in question but also of all the surrounding variables too, free of any back-door influences.

The authors also explain the special status of RCTs: they are that mode of knowledge-gathering which is closest to the "do operator": arbitrarily (i.e. free of other influences) "wiggling" one variable within a causal network to see what happens to variables downstream. Crucially, the authors show how in the right situations we can predict the results of various interventions, including RCTs, from observational, intervention-free data; and at Level Three they show how to make causal (counterfactual) statements which even RCTs cannot generate.

Level Three (Counterfactuals)

Theories are useful for making predictions about what will happen, but also for explaining how things happened in the past.

"Patient X died after having taken drug Y, but if she had not taken drug Y, would she have survived?" It is not trivial to interpret such a sentence. We cannot unthinkingly apply Level Two logic. On both Level Two and Level Three, we base a causal explanation on the twin fact that X leads to Y, and that non-X leads to non-X. But when looking at the past, one of those two tracks is contradicted by what actually happened. We need additional machinery to complete Level Three calculations and avoid running into flat-out contradictions. It is not sufficient to merely "forget" the fact that she actually did take it, because that fact may have had other causal implications which are important for our calculation. Thus, this "counterfactual" sentence seems to be about a world in which the patient both did and did not take the drug. Purely

associative statistics has serious trouble dealing with this kind of counterfactual. Not even RCTs can help: "No experiment in the world can deny treatment to an already treated person" (p. 33). The authors acknowledge Donald Rubin as being the first statistician to grasp the bull by the horns and provide a way to formulate and even answer such questions. They argue however that Rubin's approach (Rosenbaum and Rubin, 1983) is fundamentally flawed because it tries to get from Level One (observation) to Level Three (counterfactuals) using data alone, without explicitly acknowledging the need for a relevant causal model. Pearl's own solution for Level Three is however not as intuitive as the pure "do-calculus" of Level Two, involving as it does the use of counterfactual subscripts and several steps of calculation.

Level Three is revealed also to be the proper home of some quite familiar phenomena like mediation and direct versus indirect effects. The authors cast new light on Simpson's Paradox (Blyth, 1972) as a problem of direct versus indirect effects which can only be properly understood in an explicitly causal framework. They also discuss necessary and sufficient conditions as one important special case of counterfactual argumentation. They present methods for estimating "the probability of sufficiency, defined as the probability of the presence of an active causal process capable of producing the effect, and the probability of necessity, defined as the probability that no alternative process that could also produce the effect is present." Although the concepts of necessary and sufficient conditions probably sound familiar to most evaluators, the authors illustrate with attempts to attribute specific extreme weather events to climate change, how confusing they can be even to trained scientists.

What Can Evaluators Learn from the Book?

Being Bayesian about Causal Information

It is bad science to pretend that we start from a tabula rasa in assembling and interpreting evidence. The authors argue persuasively that when R. A. Fisher, as a Grandmaster of classical statistics, tried to do this as part of his support for tobacco companies during the debates about smoking and cancer, he was culpable in the unnecessary deaths of many. Even before the re-emergence of causal thinking, a Bayesian would have argued that a belief

in zero connection between two variables nearly always represents a substantive and extreme position: the correct starting point is always the experts' best bet, given substantial incentives, based on the best evidence. Pearl simply extends this familiar Bayesian approach from associational to causal statements.

The astute reader will have noticed that Pearl's methods for getting from Level One to Level Two and Level Three always involve the use of pre-existing causal knowledge. This knowledge may go beyond the knowledge that was input, but if we really had no causal beliefs, we would never be able to start our way up the ladder – except, arguably, by conducting RCTs. But Pearl persuades us to think like Bayes: our problem is not, given that we have no causal beliefs, how to get some. Rather, our problem is that, given that we already have a multitude of more or less well-founded causal beliefs, how we can leverage associational data and also manipulate our surroundings to improve and extend those beliefs. Claiming that we have no relevant causal beliefs is an extreme position which can make us, in certain circumstances, culpable. This message should speak very loudly to evaluators. It is never good enough if an evaluator says “the evidence certainly backs up what most experts believe about this causal connection, but we could not afford an RCT, and since correlation is not causation, we cannot really draw any conclusions.” Instead, evaluators should initiate a substantive discussion to establish the best available causal model and use that to work out which additional data is available or obtainable which could help revise or improve it further. The causal paths in this revised model, not the null hypothesis, should then guide us in answering the causal questions in our evaluation report.

Pearl argues that in (at least) this sense, science can never be objective because it essentially involves making causal claims which go “beyond the data.” Causal information cannot be sucked up or “data mined” out of our observational surroundings by any purely data-driven process. Those who long ago tired of “qualitative/quantitative” arguments in the philosophy of science should note that Pearl's position comes from an analysis of the properties and use of mathematical statements, not out of any interest in proving membership of one tribe or another.

Not Making Statistical Mistakes

This “Causal Revolution” is not just a philosophical shift but also has serious practical consequences. The book gives many examples of hair-raising potential flaws in classical statistical methods, for example in traditional approaches to mediation - as Keele (2015) has already warned evaluators. Anyone applying and interpreting statistical methods as basic as a two-way frequency table should be aware of Pearl's plea to always be explicitly aware of the underlying causal model. If you get the model wrong, you will probably go wrong in your choice of method and in your interpretation of the results.

A New Perspective on RCTs

As mentioned already, a researcher in possession of an appropriate causal diagram and the corresponding data can simulate or predict the result on one variable of intervening on another, even protected from the influence of an arbitrary confounder - precisely the feature which makes RCTs so powerful. This puts RCTs into their proper place: “Either we can view them as a special case of our inference engine, or we can view causal inference as a vast extension of RCTs” (p. 133).

Causal Thinking and Causal Illusions

The authors argue that members of our species think natively in causal terms, perhaps because this has proven a successful way to model the world. It is extremely difficult, and probably pointless, for us to comprehend purely associative information, and we will always try, whether sanctioned or not, to code associative information in terms of causal connections – even when these associations are in fact spurious. Pearl explains several famous statistical paradoxes such as the “Monty Hall Problem” in precisely this way - as our mis-application of causal thinking. This is yet another source of illusion to which our data-gathering processes may fall prey, and not only are we as evaluators vulnerable to it but so are our interviewees.

Mechanisms and Knowledge Nuggets

Pearl claims that we as evaluators, scientists and ordinary beings encode knowledge into chunks (discrete relationships between small groups of variables) and networks of such chunks. Properly

encoded causal information can often be “transported” from one setting to another even if the settings differ in some important aspects. We may be perfectly well aware that the truth and utility of the knowledge nugget “fire causes smoke” may fail in all kinds of ways - when combined with overlapping and possibly contradictory nuggets - yet it is out of this cacophony of competing knowledge fragments that we form a useful understanding of our world. This is perhaps the opposite of the claim “oh, but everything always depends on context.” So Pearl has a strong epistemological thesis - a claim about how we gain and encode knowledge in the form of causal chunks and networks of causal chunks. He further makes the corresponding psychological claim that this is close to the way our brains work as Homo Sapiens. This also explains how people manage to communicate with one another about causal effects - because we share, broadly speaking, causal models. Thirdly he makes a corresponding engineering claim that successful AI is compelled more or less to mimic this approach. Elsewhere, Pearl, as a realist, makes a fourth ontological claim corresponding to the first three: that the world is in fact organised in terms of quite stable mechanisms which are relatively autonomous. This fourth thesis will remind evaluators of Realist theory in evaluation (Pawson & Tilley, 1997). Pearl often uses the word “mechanism” to name these chunks in the world (though he has no specific word for the corresponding small theories such as “fire causes smoke” in which we encode our nuggets of knowledge about those mechanisms).

Evaluators may also try interpreting the links within project theories of change as “causal chunks” à la Pearl; and we may be interested in whether these chunks could be useful building blocks for understanding the cognitive maps which, in turn, interviewees and other stakeholders use to understand and act in their respective worlds.

Criticisms

More Engineering than Social Science?

Stephen West (2014), reviewing Pearl’s seminal work (2000), asks whether Pearl’s background in engineering explains why many of his examples seem too much like simple questions of “which button to press.” In real life, of course, it can be a challenge for evaluators to work out which button was actually pressed by an intervention. This is the issue of the “construct validity of the treatment” according to Campbell (Cook and Campbell, 1989).

While this is an important practical point, there seems to be nothing in Pearl’s framework which would prevent it being extended to cope with this criticism.

Donald Rubin offers an alternative framework, also not shy of dealing with causality, which is much more fully developed for the needs and concerns of social scientists in general and evaluators in particular. But Pearl is much clearer in underlining and illustrating the scale of the causal revolution and the importance of using our new freedom to explicitly adopt causal models.

“But We Knew that Already”

Many evaluators, even more than working social scientists, may be thinking “who cares about what statisticians say, we have been modestly working with causal statements and even causal networks and diagrams for years: just look at all our theories of change, for example.” And indeed, Pearl maybe overstates his claim about how comprehensively causal language was ever banned from statistics. In particular, the diagrams associated with Structural Equation Models (SEMs) certainly look as if they make causal claims and have regularly been understood in such a way.

Even so, the work of Pearl and his collaborators puts causal language on new foundations and opens up the possibility firstly of having the right arguments to oppose those who still say, for example “you can never get from correlation to causation.” Secondly, there is enormous potential to provide a more unified and systematic foundation for theories of change and causal reasoning in evaluation, especially if Pearl’s ideas can be more explicitly extended to cope with data and models which are fuzzy, uncertain, non-linear and incompletely formulated. There is however no easy Pearlian cook-book even for statisticians – although Pearl, Glymour and Jewell (2016) is a first step – let alone for evaluators, upon which such an extension could build.

What is more, when Pearl says “data” he is only ever thinking about correlations and the occasional experiment. Although he has a highly-developed framework for integrating this kind of data from different sources and contexts, he says nothing about how to integrate, for example, the opinion of a stakeholder which is explicitly presented as a causal model (“this leads to that”) without accompanying evidence, or an observation based on the thought processes around a single case.

There is plenty of work still to be done.

Need for Some Basic Definitions

Pearl is perhaps too much of a polymath, and in too much of a hurry to answer burning and practical questions, to provide a really satisfactory step-by-step exposition of his ideas from the ground up. Even his seminal work “Causality” (Pearl, 2000) has been criticised on this basis. Most importantly, it seems that the basic causal link (“Y listens to X”; we “wiggle X to influence Y”) is really a primitive notion defined by the three axioms about which nothing else can be said; yet at the same time he argues for the plausibility of the axioms on other grounds, which is not really consistent with an axiomatic approach. It is a pity that this book does not provide an annex in which the basic concepts covered are built up one by one from first principles. In particular, this might have made it easier to understand the concepts on Level Three.

How Does it Work as a Book?

Is this book going to become a New York Times bestseller like, say, Kahneman’s “Thinking, Fast and Slow” (2011)? The chapters of the book are mostly structured around a series of vignettes such as the story of Galton’s intriguing “quincunx,” an overview of Bayesian statistics, and a brief history of the battle to show a causal link between smoking and cancer, which the present review has mostly ignored but which include some engaging stories. But Pearl is in deadly earnest and many non-technical readers will find even this “general readership” book a little too hard to really work as entertainment. One point of encouragement is that the reader who gives up on Level Three - which is probably the hardest section of the book - will still take away plenty from it.

References

- Blyth, C. R. (1972). On Simpson’s paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67(338), 364–366.
- Burks, B. S. (1926). On the inadequacy of the partial and multiple correlation technique. *Journal of Educational Psychology*, 17(8), 532 - 540.
- Collier, D. (2011). Understanding process tracing. *PS: Political Science & Politics*, 44(4), 823–830.
- Cook, T.D., & Campbell, D.T. (1979). The design and conduct of true experiments and quasi-experiments in field settings. In R.T. Mowday & R.M. Steers, (Eds.), *Research in organizations: Issues and controversies*. Santa Monica, CA: Goodyear Publishing Company.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus; Giroux.
- Keele, L. (2015). Causal mediation analysis: Warning! Assumptions ahead. *American Journal of Evaluation*, 36(4), 500–513.
- Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, 24(1), 65–75.
- Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70(17), 556-567.
- Mayne J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *The Canadian Journal of Program Evaluation*, 16(1), 1–24.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge: Cambridge University Press.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Basic Books
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. West Sussex, UK: John Wiley & Sons.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, Massachusetts: MIT press.
- West, S. G., & Koch, T. (2014). *Restoring causal analysis to structural equation modeling. Review of causality: models, reasoning, and inference*, by Judea Pearl: New York, NY: Cambridge University Press.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20(7), 557–585.
- Zadeh, L. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(1), 28–44.