



Pearl, Judea and Mackenzie, Dana: The book of why: the new science of cause and effect (2018)

Basic Books, New York. ISBN 9780465097609 (hardcover)

Karamjit S. Gill¹

© Springer-Verlag London Ltd., part of Springer Nature 2020

‘The Book of Why’ by Judea Pearl and Dana Mackenzie is for you if you are engaged in the pursuit of Data Science, artificial intelligence, cognition, deep learning; or curious about “correlation is not causation”, the ‘ladder of causation’, confounding bias, Bayesian rule and “do” calculus; or wondering whether a robot with real humanlike intelligence would be created if causality were integrated into artificial intelligence; and are interested in exploring whether counterfactuals and mediation can lead to better prediction, explanations, and diagnosis of treatments and health care strategies.

The authors sum up the message of their book when they say that: “Data do not understand causes and effects, humans do”. The advice to big data and model free enthusiasts is that you may be able to tease out all the information from data but it will ‘never’ answer even a simple question of causation, “What is the relative importance of various causes?” In doing so, the authors take issue with the false belief that ‘the answers to all scientific problems reside in the data, to be unveiled through clever data mining tricks’. Although the data-centric belief still prevails under such umbrellas as ‘data science’, ‘data economy’, and ‘data-centric intelligence’, they posit that data alone cannot make up for lack of scientific knowledge.

In recognizing the ‘dumbness’ of data, the authors argue for a shift from data-driven science to cause–effect relationships. They call it the ‘causal revolution’ which is rooted in the rigour of ‘calculus of causation’. 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. One of the ‘crowning achievements’ of the ‘Causal Revolution’, they say, has been

to explain how to predict the effects of an intervention without actually enacting it. Causal reasoning, when it involves retrospective thinking of scientific questions, is termed counterfactuals. Another gem of the causal revolution, Pearl says, is that in many cases we can use algorithms to emulate human retrospective thinking and produce answers to the observed counterfactual world.

Counterfactual reasoning, which deals with what-ifs, is seen as the building blocks of moral behaviour, as well as scientific thought. The ability to reflect on one’s past actions and envision alternative scenarios is the basis of free will and social responsibility. Pearl gives us an insight into his commitment to the ‘Causal revolution’ when he says that his departure on causal inference from his AI colleagues resides in his view of the AI world, that “you do not really understand a topic until you can teach it to a mechanical robot”. Moreover, this view resides in his deep conviction that “language shapes our thoughts. You cannot answer a question that you have no words for”. For Pearl, a major challenge remains, ‘how can machines acquire causal knowledge?’ Deep learning and data mining algorithms devoid of causal models can only fit a function to data, can only interpret data through a causal ‘model-blind’ lens, forgetting that human intuition and human intelligence are rooted in causal and not statistical logic. Consequently these algorithms lack adaptability to deal with prediction in the uncertain world.

Pearl takes the reader on a historical journey beginning with the foundation of the ‘holy grail’ of data and objectivity that is rooted in the 1834 founding charter of the Statistical Society of London. This said that ‘data were to receive priority in all cases over opinions and interpretations. Data are objective; opinions are subjective’. In the course of the journey, we learn about Francis Galton’s pursuit of the inheritance of stature and genetic traits, Karl Pearson’s passion for correlation and causality-free science, Sewall Wright’s construction of a path to the bridge between causality and

✉ Karamjit S. Gill
editoraisoc@yahoo.co.uk

¹ University of Brighton, Brighton, UK

probability. We also learn Wright's defense of the scientific method and interpretation of data and of the resistance to 'complete objectivity' of data and the allure of "model-free" methods to objectivity. This struggle for objectivity—the idea of reasoning exclusively from data and experiment—has been part of the way that science has defined itself ever since Galileo. Although the Bayesian rule proved a superior tool in predicting complex phenomenon such as weather forecasting and tracking enemy submarines, the influence of 'prior belief' vanishes as the size of data increases, leaving a 'single conclusion in the end'. Pearl, however, notes that the 'subjective' component in 'causal information' does not necessarily diminish over time, regardless how "big" the data are. This long lasting aspect of 'causal subjectivity', Pearl suggests, may be the reason for the refusal by the advocates of scientific objectivity to 'accept the inevitability of relying on subjective causal information'.

The authors note that the deficiency of data driven objectivity was implicitly evident by the early 1980s when research into rule based expert systems had worked itself into a cul-de-sac, and discovered that hard-and-fast rules can rarely capture real-life knowledge. It was not just that expert systems struggled with making inferences from uncertain knowledge, it was also that the computer could not replicate the inferential process of a human, expert because human experts themselves were not able to articulate their thinking in the rules provided by the system. Although there was an attempt to offer approaches of "fuzzy logic", "belief functions" and "certainty factor" to deal with uncertainty, they suffered from a common flaw—they modeled the expert and not the world. Pearl says that this is not to say that we do not recognise the role of big data techniques and machine learning in making inferences from past behaviour of a set of individuals of similar characteristics, for example, in personalised medicine. Although this inference making may help us to overcome the problem of dimensionality and screen off the irrelevant characteristics in making sense and articulating substantive assumption about how the world operates, but it does not provide us with a clue of how to draw a model of the real world and its operations. Given that big data enable us to access large amounts of data and an enormous number of studies from different locations, one of the interesting challenges is how to combine the data and results from remote and disparate studies and translate them to new

populations that may be different even in ways we have not anticipated. The authors explore this issue of 'transportability' through a symbiosis of big data and causal inference, for example, by identifying potential disparities in the data generating process, and recalibrating them between the two different environments. Furthermore, this recalibration may deal with the population selection bias in ways that may have adverse affects on the validity and transportability of data and its results. The challenge then is how to exploit the symbiosis of Big Data and causal logic to turn this adversity to opportunity and transcend the paralysis of the culture of "external validity", that is preoccupied with the 'categorising of external threats rather than fighting them'.

The authors ask the most basic question: can the machine pass the test of human intelligence without envisioning and contrasting alternative machine realities to the current existing human realities? And why do such natural and intuitive questions 'reside beyond the reach of the most advanced reasoning systems of the time'? They point out that until machines could be taught to understand the 'cause and effects' of our world, machine learning could not move beyond the 'shades of grey' predictions of the Bayesian culture. In response to their own question, "Should we make machines that think?" the authors say that 'Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation'. And to their question, "Can we make machines that are capable of distinguishing good from evil?", the book ends up with a positivist vision of artificial intelligence when the authors say that once we have a thinking machine, "it would be a wonderful companion for our species, and would truly qualify as AI's first and best gift to humanity". Whether we subscribe to or are apprehensive of this futurist view of the thinking machine, 'The Book of Why' should be a must read to get a deep insight into the debates on Data Science and Deep Learning of our AI times.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.