Introduction

Federica Russo and Jon Williamson

Causality and probability in the sciences

Towards the end of the nineteenth century Karl Pearson noted that a probabilistic dependence between two variables does not necessarily imply that the two variables are causally connected (Pearson, 1897). This led Pearson, in the third (1911) edition of his book Pearson (1892), to argue that talk of cause and effect should be eradicated from science in favour of talk of probabilistic dependence (in the form of contingency tables). Around the same time, Bertrand Russell threw his intellectual weight behind this purge of causality (Russell, 1913). In the same vein, Ernst Mach (1905) argued that causality, understood as a way to explain phenomena, should be replaced by the concept of *relation*, which is a way to merely *describe* phenomena. These attacks had a profound influence on much of twentieth century science. Although scientists continued to reason causally—e.g., to find causes of phenomena, to devise experiments to measure interventions, to inform policy decisions—explicit mention of causality met with disapproval.

Then came the 1980s. As explained below, causal methods developed in Artificial Intelligence (AI) in the 1980s helped to rehabilitate the concept of cause. While causality was no less controversial from a philosophical perspective, new formalisms for handling causality and probability together helped mathematise the notion of cause. Fig. 1 and Fig. 2 show the resulting transformation. A search of the *Web of Science* databases for papers whose titles include a word beginning with 'caus-' (e.g., 'causality', 'causation', 'causal') revealed a stark increase in the numbers of such papers after about 1990. This was so for the *Science Citation Index Expanded* (SCIE) database, which deals mainly with physical, biological and computational sciences, and for the *Social Sciences Citation Index* (SSCI) database. The growth in papers on causality in the *Arts and Humanities Citation Index* (AHCI), which covers philosophy, is rather more gradual. (Of course, the volume of all academic papers increased markedly in this period. In an effort to compensate for this general growth, for each of the three databases Fig. 2 portrays the yearly number of papers involving causal terms divided by the yearly number of papers with an author whose name begins with the letter 'J', a rather arbitrary indicator of the general volume of papers in the database in question.)

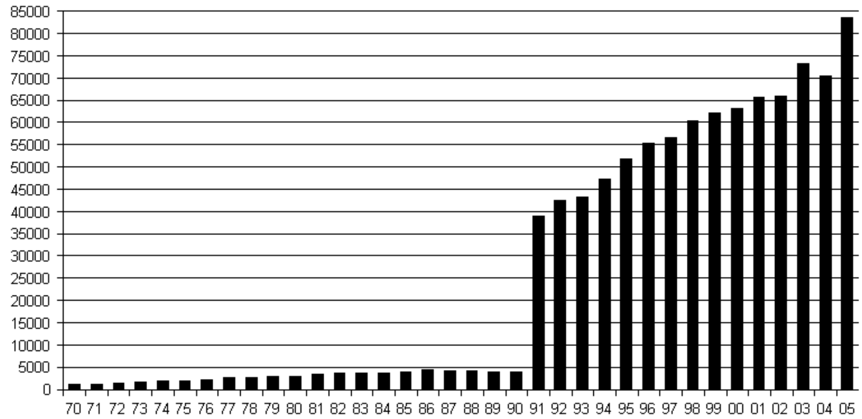

Figure 1. Total yearly numbers of papers involving causal terms.

Given this rehabilitation of causal talk it is all the more important to further our understanding of the notion of cause. As a step in this direction, the papers in this volume seek to shed light on the relationship between causality and probability. Methodologically, the work presented here is *science-driven*: the papers seek to learn lessons about causality and probability motivated by actual scientific practice.

This volume

Causality and probability in AI

Researchers in AI have been responsible for many of the key developments in causal modelling since the early 1980s. Broadly speaking this line of research is motivated by the following question: given a dataset containing a series of past observations, what is the most appropriate causal model of the domain in question? This question led to the development of *causal nets*, also called *causally interpreted Bayesian nets* (Pearl, 1988; Williamson, 2005), which represent qualitative causal connections by means of a directed acyclic graph (DAG) and which represent a joint probability distribution by means of the probability distribution of each variable conditional on its direct causes together with an independence assumption, the *Causal Markov Condition*, which says that each variable is probabilistically independent of its non-

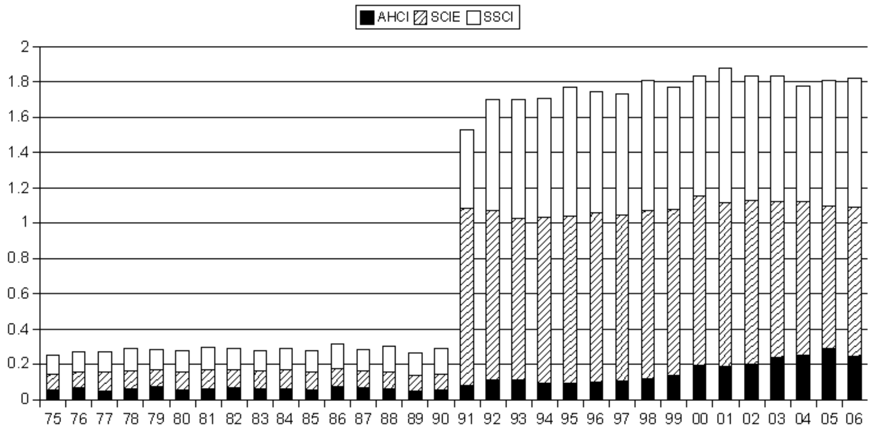

Figure 2. Papers involving causal terms, according to citation database and adjusted by overall volume.

effects conditional on its direct causes. Closely related are *structural equation models*, which can be thought of as causal nets with functional equations instead of conditional probability distributions (Spirtes et al., 1993; Pearl, 2000). Causal nets and structural equation models are often called *graphical causal models*.

In 'An integral approach to causal inference with latent variables', Sam Maes, Stijn Meganck and Philippe Leray contribute to the development of the causal net framework. As originally devised, causal net learning algorithms produce the causal net on the variables of the given dataset that best fits the dataset. Now it may be that a probabilistic dependence amongst two variables in the dataset is induced, not by causal relationships amongst the variables in the dataset, but because the two variables are effects of a common cause that is not measured in the dataset. Such an unmeasured common cause is called a *latent variable*. The question thus arises as to which causal net, involving latent variables as well as variables in the dataset, best fits the dataset. Causal nets have been extended in various ways in order to try to answer this question, and this chapter develops the *semi-Markovian causal model* approach. It also extends the causal net framework in another direction, by allowing experimental as well as observational data to help determine the appropriate causal model.

Alex Freitas, Ken McGarry and Elon Correa forge an interesting connection between causal nets and knowledge discovery in 'Integrating Bayesian networks and Simpson's paradox in data mining'. In AI, knowledge discovery has two types of application. It may be that an artificial agent needs to extract generalities from data that will be useful to it in its activities: an automated system for drug synthesis may need to determine laws about the structure of molecules from experimental and observation data, for example, in order to synthesise molecules of the appropriate shape. By and large, although useful to the machine in question, these laws are of less interest to human chemists (though as Gillies (1996) points out, this is not always the case). On the other hand, artificial agents often work less autonomously, liaising with humans: for instance a financial system may need to sift through past data to offer general investment advice to human investors. In this second type of application it is important that the knowledge gleaned from the data be of interest to the humans. This chapter exploits Simpson's paradox, a conundrum that often crops up in the literature on probabilistic causality, to help isolate interesting pieces of knowledge. The authors also argue that Simpson's paradox can be exploited to help learn causal nets from data.

Causality and probability in the physical sciences

Bertrand Russell (1913) maintained that the physical sciences do not appeal to the concept of cause, but instead deal with functional equations. These days, causal claims are not considered to be a world apart from functional equations; a structural equation model, for example, is a type of causal model. The question then arises as to how the functional equations of physics might be given a causal interpretation. Thus the ideal gas law, $PV = nRT$ (where P is pressure, V is volume, n is the number of moles of gas, R is the universal gas constant, and T is temperature), is typically taken to encapsulate a number of causal relationships, with each of P, V and T causally dependent on the other two.

The functional equations of quantum mechanics are often thought to pose a special problem for causality, with several philosophers arguing that the Einstein-Podolsky-Rosen thought experiment can not be interpreted causally. Mauricio Suárez, in 'Causal inference in quantum mechanics: a reassessment', takes these philosophers to task. Suárez puts forward five different causal models that could be taken to underlie this experiment. The problem isn't so much that quantum mechanics is incompatible with causality, but rather that the most plausible model (model II in this chapter) forces a reassessment of the relationship between causality and probability. Under this model, the Causal Markov Condition, often taken to be an invariant or even defining feature of causality, fails. If so, the Causal Markov Condition may have to be relegated to the status of a *default* rule (Williamson, 2005).

Causality and probability in the social sciences

Karl Popper (1934) put forward a hypothetico-deductive account of how one should discover causal relationships: first hypothesise a causal law, then deduce its consequences, rejecting the law if these predictions are not borne out. In economics, the Cowles Commission, founded by Alfred Cowles in 1932 with the aim of promoting a mathematical approach to economic theory, suggested that economic theory be used to provide the hypothesis and statistics be used to determine the parameters of a causal model and to derive predictions.

In contrast, the approaches to causal modelling developed in AI open up the possibility of an inductivist approach to discovery: from a dataset of past observations, directly induce causal laws (Spirtes et al., 1993). In 'Mediating between causes and probabilities: the use of graphical models in econometrics', Alessio Moneta argues that the graphical models of AI are better seen as involving aspects of both hypothetico-deductivism and inductivism. Moneta takes seriously the idea that the assumptions behind causal models—such as the Causal Markov Condition—need not hold invariably and should be treated as default working assumptions. In which case, the models yielded by inductive methods can at best be viewed as tentative hypotheses, in need of further testing.

Stephen LeRoy compares the approach of the Cowles Commission with more recent developments in 'Causality in economics'. LeRoy argues that graphical causal models are a point of departure from the more traditional approach in economics due to Herbert Simon, who began his economic career in the Cowles Commission. This is because the functional equations of the traditional approach do not admit a straightforward causal interpretation by treating a variable on the left-hand side of the equation as the effect and those variables on the right-hand side as its direct causes, while structural equation models as defined by Pearl (2000) do admit such an interpretation. LeRoy favours the traditional Cowles approach.

Damien Fennell, in 'Causality, mechanisms and modularity: structural models in econometrics', takes this comparison a step further. Fennell argues that the graphical model approach differs from Simon's analysis in another respect: advocates of the graphical model approach tend to assume *modularity*, i.e., that one can intervene to fix the value of any variable without changing the values of its causes in the model and without changing the nature of the causal relationships in the model (a so-called *divine intervention* or *perfect intervention*), while Simon makes no such assumption. Modularity does not always hold—Fennell argues that this gives another reason to prefer the traditional Cowles approach over the graphical causal model approach.

In 'Time series, nonsense correlations and the principle of the common cause', Julian Reiss discusses a further way in which the Causal Markov Condition can fail. The Causal Markov Condition implies the *Principle of the Common Cause*, which says that if two variables are probabilistically dependent then either one causes the other or they are effects of a common cause. But it is well known that time series give rise to correlations that admit no such causal explanation. Yule (1926), for instance, cited a correlation between the proportion of Church of England marriages and the mortality rate, in the years 1866-1911; both are decreasing but for different reasons, not because one causes the other or because they are effects of a common cause. Advocates of graphical causal models often try to hold out against such counterexamples to the Causal Markov Condition, either by maintaining that these counterexamples dissolve under closer scrutiny, or by claiming that they do not make a practical difference on the use of graphical models. Reiss argues against both these moves and concludes that the Principle of the Common Cause is a fallible assumption.

The previous papers focus on methodological aspects in the fields of economics and econometrics. In 'Conceptual tools for causal analysis in the social sciences', Erik Weber attempts to provide social scientists with useful tools for causal analysis. Weber distinguishes two different tasks of a philosophical investigation into causality. On the one hand, a conceptual analysis develops a definition of causality to adequately represent our everyday causal talk. On the other, we can develop a set of concepts that are supposed to help scientists. Weber confines his discussion to the second task. As a conceptual pluralist, Weber puts forward three concepts for social scientists. The first is the causal relation at the population level, as defined by standard average effect theories; the second is the causal interaction at the individual level, and the third is the specification of spontaneous preservation that takes place after causal interactions. The most original part of the paper consists in the modification of Salmon's concept of causal interaction and the definition of spontaneous preservation. In this way, the concept of causal interaction, originally thought for physics, is now also well suited to the social sciences.

Causality and probability in the biomedical sciences

When trying to assess frameworks for causal modelling, two key questions arise. First, what notion of cause is employed in the model? Causality can be interpreted in a variety of ways—e.g., mechanistic, probabilistic, counterfactual, agency, epistemic—and the choice of interpretation can have a bearing on whether the modelling assumptions hold. Second, what notion of probability is employed in the model? Probability also admits of a variety of interpretations—e.g., frequency, propensity, chance, classical, logical and several Bayesian interpretations—and modelling assumptions such as the Causal Markov Condition depend on the chosen notion of probability as well as that of causality.

In 'Interpreting probability in causal models for cancer', Federica Russo and Jon Williamson argue that cancer epidemiology is distinctive in that it is concerned with both generic and single-case probabilities, since it deals with both causal laws and with particular patient diagnoses and prognoses. Consequently, we claim, it requires a twin interpretation of probability: generic probabilistic claims should be given a frequency interpretation while single-case claims should be interpreted using degrees of belief. In particular, objective Bayesianism turns out to be the most appropriate interpretation in the single-case. If we are right, then this has an important consequence for modelling: the Causal Markov Condition can be proved to hold by default under an objective Bayesian interpretation (Williamson, 2005), so graphical causal modelling becomes a plausible methodology in cancer epidemiology.

The choice of statistical framework can have crucial repercussions. A good framework can lead to perspicuous models with well-articulated claims and assumptions, while a poor framework can obfuscate the problem in hand and even lead to situations in which no model in the framework adequately captures what is going on. Bert Leuridan, in 'Galton's blinding glasses: modern statistics hiding causal structure in early theories of inheritance', argues that the statistical context in which Francis Galton worked prevented him from finding the causal story behind inheritance. This is because Galton's choice of model was constrained by having to account for occurrences of the normal distribution and of regression towards the mean. In contrast Gregor Mendel, who made little use of statistical theory, hit upon essentially the right causal picture. The lesson, Leuridan maintains, is that we should be very careful in our use of contemporary frameworks such as graphical causal models: we should test their assumptions and be aware of their possible blinding influence.

Vanessa Didelez and Nuala A. Sheehan advocate the graphical causal modelling framework in 'Mendelian randomisation: why epidemiology needs a formal language for causality'. They show that when it is not possible, either in principle or for practical reasons, to perfectly intervene on a variable, one can instead use observational data and instrumental variables to decide causal claims, and that this procedure can be nicely represented via the graphical causal model approach. Didelez and Sheehan argue that by casting a problem in this formal language one can clarify the key causal questions that one is trying to answer, and isolate the conditions that make such causal inferences possible. Thus graphical causal models can be illuminating rather than blinding.

Causality and probability

In philosophy, causality has been a central theme since the ancient Greeks and has never lost its appeal. However, two events in the sciences have radically changed, and indeed invigorated, the debate: the advent of probability theory and the discovery of indeterministic phenomena. These two events gave a new flavour to old questions and raised completely new ones. For instance, is causality essentially indeterministic or deterministic? If probabilities are at the heart of causal processes, what is the relation between causes and probabilities? If probabilities characterise physical and social processes, how are probabilities to be interpreted?

As a consequence of these changes, a notion of cause in terms of necessary and sufficient conditions has gradually been replaced by a probabilistic notion. Thus Hans Reichenbach (1956) put forward the Principle of the Common Cause and Patrick Suppes (1970) provided a probabilistic account of causality that was very general in scope. The development of graphical causal models in AI can be viewed as a continuation of this tradition.

In the 'The causal roots of probability', Marianne Belis investigates the relation between causality and probability, particularly focusing on the single case. Pursuing the idea that singular causes have ontological priority over Humean regularities, Belis argues that two concepts elucidate the sought-after relation: 'propensity' and 'capacity', borrowed from Popper (1959) and Cartwright (1989) respectively. Belis defends the propensity notion of probability in spite of the well-known difficulty of their measurement—a difficulty that often sees propensities labelled as 'metaphysical rather than scientific'. Instead, Belis argues that there is a way to measure propensities—that is through the algebraic sum of all the strengths exerted upon them. Moreover, propensities 'reveal the ontological roots of probability', i.e., the inner *causal* character of probability in the single case.

Andrea L'Episcopo, in 'Causality and the axiomatic probability calculus', narrows down on Phil Dowe's conserved quantity (CQ) theory and probabilistic theories of causality. The main claim of the paper is that any theory of causality should be evaluated in its proper domain of application. For this reason, argues L'Episcopo, many counterexamples and criticisms to the major accounts of causality are misdirected. This claim hinges upon two distinctions. The first is between an intuitive versus a physical notion of causality, and the second is between an empirical versus a conceptual analysis of causality. For instance, Dowe's CQ theory is an empirical analysis of the physical notion of causality, and therefore, L'Episcopo argues, causation by omission or prevention do not constitute genuine problems for it.

In 'Two probabilities of dysfunction and two kinds of chance' Françoise Longy draws our attention to the interesting case of the probability of dysfunctions of artifacts. The probability of dysfunction of artifacts may be given two different interpretations. Consider, for instance, the probability that a light bulb will burn out within five minutes after its first use. The first reading concerns the *physical* probability that this particular object (i.e., the *physical* object) with this particular physical structure will burn out within five minutes—let us call this probability DYSF-PHYS probability; the second concerns the probability that the object, as an artifact produced in such and such factory, will burn out within five minutes—let us call this probability DYSF-ART probability. According to Longy, the question is whether or not these are two different sorts of objective probability of dysfunction, interpretable as chances rooted in some particular feature of the world. Indeed they are. In particular, the DYSF-ART probability is rooted in a number of conditions that determine the object as an artifact but not as the specific physical make-up of the object. Longy supports this claim by relying on the concept of function as is developed in the biological sciences.

Causal pluralism

In recent decades many different views of causality have been proposed. Among the most influential are Suppes' probabilistic theory (Suppes, 1970), Lewis' counterfactual approach (Lewis, 1973), the Salmon-Dowe process theory (Salmon, 1998; Dowe, 2000), and agency and manipulability approaches (Menzies and Price, 1993; Woodward, 2003). These approaches provides us with a variety of accounts of what causality is. This raises the problem of whether causality is genuinely a plurality of different concepts or whether it is a single concept.

This abundance of accounts has made pluralism a fashionable stance. Pluralists argue that different concepts of cause fit different contexts and that, therefore, there is no real incompatibility between, say, the probabilistic and process approach, exactly because they employ different concepts of cause in different domains.

In 'Causal dualism: which position? Which argument?', Monika Dullstein focuses on Hall's recent argument for causal dualism. Hall argues that 'cause' has two different meanings: the first relates to the concept of production (causes are physically linked to and produce their effects), and the second to difference-making (causes are responsible for differences, either probabilistic or counterfactual, in the occurrence of their effects). The reason to defend this dualist position lies in the fact that, on the one hand, no production account can deal with negative causation, and, on the other, no difference-making account can deal with overdetermination. Dullstein points out that such a pluralist stance can be simply read as a way of thinking of causation, or, more interestingly, as a specific tenet about the metaphysics of causation. Dullstein then raises the question of whether Hall's argument for causal dualism can provide any reason to believe in *metaphysical* dualism. Her answer is that Hall's argument fails in this respect.

Amit Pundik, in 'Can one deny both causation by omission and causal pluralism? The case of legal causation' draws the philosophers' attention to the intricate case of causation in the law. This ambitious paper aims at establishing a number of points. In the first place, Pundik shows that in the legal context omissions are often regarded as genuine cases of causation. Secondly, that causation by omission and causal pluralism cannot be denied coherently. In other words, if we opt for causal pluralism and accept that law has its specific concept of causation, then this concept must include causation by omission. Thirdly, he tries to convince those who might dismiss legal causation as a genuine philosophical problem that, instead, this is an extremely relevant and interesting matter.

General frameworks for causal analysis

A philosophical account of causality can have different purposes. For instance, the metaphysics of causality investigates what causality in fact is; epistemology is interested in how we come to know about causal relations; and methodology explores new methods for causal reasoning and inference. Phil Dowe (2000) introduced the distinction alluded to above between a conceptual and an empirical analysis of causality. Conceptual analysis is concerned with our talk about causality, i.e., about the meaning of cause in ordinary language. Empirical analysis is instead concerned with the meaning of cause in science.

Friedel Weinert, in 'A conditional view of causality', puts forward a framework for causal analysis that fits both the natural and the social sciences. His conditional model is based on Mackie's INUS model, involving conditions that are Insufficient but Necessary components of a set of Unnecessary but Sufficient conditions. In Weinert's account, causal relations are facts about conditional dependencies between antecedent conditions and consequent conditions. To show the wide applicability of such a conditional view, Weinert analyses the Franck-Hertz experiment in quantum mechanics and Max Weber's adequate causation in the social sciences.

Aviezer Tucker, in 'The inference of common cause naturalized', draws our attention once more to the Principle of Common Cause, one of the most debated principles in the philosophy of causality. In this paper, Tucker shows that Reichenbach failed to deduce the principle from the second law of thermodynamics, and points out that much of the literature on the topic failed to distinguish between the inference that *some* common cause existed (without specifying the particular properties of such a common cause) and the inference that a *concrete* common cause existed (with a unique set of properties). He then proposes a naturalised theory of common causes, relying on examples from various disciplines—for instance, biology, linguistics and philology. The inference of the common cause involves three consecutive stages of comparisons: first, a comparison of likelihoods given some common cause, whose properties are unknown, and given separate causes; second, if the common cause makes the evidence more likely, five types of common cause hypotheses compete; and third, if it is possible to prove which of the five types is the most probable, scientists attempt to infer the actual properties of common causes.

In 'Contexts for causal models' Margherita Benzi distinguishes between two possible approaches to causal modelling. Whilst the first relies on the idea of an underlying omnicomprehensive causal network, the second relativises the construction of the causal model to the context of inquiry. Although the context-sensitivity of causal models generally attracts a consensus among philosophers and practising scientists, there is little agreement as to what this context exactly is. Benzi proposes a taxonomy of contexts that is based on the idea that different levels of analysis reflect an increasing specificity of the factors to be taken into account. At the lowest level, we find generic background knowledge; this is refined in the context of inquiry, where factors are selected according to the needs of the specific study at hand; the causal model further refines the context of inquiry and picks out the factors to include in the model; the last level involves those factors that are more directly relevant for the assessment of the causal relation. Benzi argues that approaches that relativise to the context of inquiry ought to be preferred even if the price to pay is to give up the hope of the reduction of causes to probabilities.

As mentioned above, causally interpreted Bayesian nets have shed new light on causal inference in the last two decades. The major problem with Bayesian nets is that they will deliver successful results *if* some basic assumptions—e.g., the Causal Markov and Faithfulness conditions—are satisfied. In 'Causal inference. How can Bayesian nets can contribute?' Isabelle Drouet investigates the extent to which Bayesian nets algorithms satisfying those assumptions can be integrated into the traditional path analytic methodology. Drouet proposes a mixed methodology for causal inference, in which Bayesian nets algorithms are run only after good reasons that basic assumptions are satisfied have been provided.

Philip Dawid, in ‘Counterfactuals, hypothetical and potential responses: a philosophical examination of statistical causality’, runs a detailed and careful analysis of the frameworks and tools developed for causal inference by statisticians. In particular, he focuses on the potential response model and on decision theory. The potential response model (PR) is perhaps the dominant methodology. However, Dawid gives reasons to prefer the decision theoretic approach (DT). The reasons to prefer DT lie in the distinction between hypothetical and counterfactual queries. Hypothetical and counterfactuals, in turn, relate to the different tasks of inferring effects of causes or causes of effects, respectively. Dawid’s main criticism of PR, which includes structural equations, is that, although it can handle the inference of effects of causes, it faces serious troubles in inferring causes of effects, especially when only observational data is available. The superiority of DT is claimed on both philosophical and pragmatic grounds.

The anti-causal correlation-mongers have been deposed. Causality is no relic of a bygone age; the renewed interest in causality from the 1980s onward proves the importance of causal reasoning both for cognitive and action-oriented purposes. On the one hand, knowledge of causes is essential to the intellectual enterprise of understanding and explaining the world. On the other, the action-oriented goal—e.g., to guide and inform policies, prescribe treatments, etc.—needs a solid causal grounding, not the quicksands of mere correlation.

Federica Russo
Philosophy, University of Kent, UK.
f.russo@kent.ac.uk

Jon Williamson
Philosophy, University of Kent, UK.
j.williamson@kent.ac.uk

BIBLIOGRAPHY

Cartwright, N. (1989). *Nature’s capacities and their measurement.* Clarendon Press.
Dowe, P. (2000). *Physical causation.* Cambridge University Press.
Gillies, D. (1996). *Artificial intelligence and scientific method.* Oxford University Press, Oxford.
Lewis, D. K. (1973). Causation. In *Philosophical papers*, volume 2, pages 159–213. Oxford University Press (1986), Oxford.
Mach, E. (1905). *Knowledge and error.* Reidel Publishing Company.
Menzies, P. and Price, H. (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science*, 44:187–203.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann, San Mateo CA.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge University Press, Cambridge.
Pearson, K. (1892). *The grammar of science*. Black, London, third (1911) edition.
Pearson, K. (1897). Mathematical contributions to the theory of evolution.—on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, 60:489–498.
Popper, K. (1959). The propensity interpretation of probability. *British Journal for Philosophy of Science*, 10:25–42.
Popper, K. R. (1934). *The Logic of Scientific Discovery*. Routledge (1999), London. With new appendices of 1959.
Reichenbach, H. (1956). *The direction of time*. University of California Press.
Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society*, 13:1–26.
Salmon, W. C. (1998). *Causality and explanation*. Oxford University Press, Oxford.
Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction, and Search*. MIT Press, Cambridge MA, second (2000) edition.
Suppes, P. (1970). *A probabilistic theory of causality*. North Holland Publishing Company, Amsterdam.
Williamson, J. (2005). *Bayesian nets and causality: philosophical and computational foundations*. Oxford University Press, Oxford.
Woodward, J. (2003). *Making things happen: a theory of causal explanation*. Oxford University Press, Oxford.
Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89(1):1–63.