Croatian Journal of Philosophy
Vol. V, No. 14, 2005

Causal Models in the History of Science

OSVALDO PESSOA JR.
Universidade de São Paulo

The investigation of a method for postulating counterfactual histories of science has led to the development of a theory of science based on general units of knowledge, which are called "advances". Advances are passed on from scientist to scientist, and may be seen as "causing" the appearance of other advances. This results in networks which may be analyzed in terms of probabilistic causal models, which are readily encodable in computer language. The probability for a set of advances to give rise to another advance is taken to be invariant through history, but depends on a typical time span which is an inverse function of the degree of development of science. Examples are given from the early science of magnetism and from the 19th century physics.

1. *Introduction*

Historians of science occasionally make remarks concerning what could have happened if a certain fortuitous event had been different. For example, *if* Sadi Carnot had not had an untimely death, and had published his 1926 calculation of the mechanical equivalent of heat, then the principle of conservation of energy might have been anticipated by about twenty years. This type of assertion is "counterfactual", since it refers to a possible situation that did not turn out to be a fact. Human beings have an excellent ability for imagining counterfactual scenarios (which certainly follows from the advantage that this capacity gave us in biological evolution). Could this ability for postulating counterfactual histories be put on firmer bases by some methodology?

That was the topic of a previous work (Pessoa [2001]), in which possible histories were examined, beginning from the factual situation of physics in 1800, and leading to the discovery of quantum theory. It was argued that there would be four more probable paths and a few others with quite less probability. (If instead of considering 1800 as the initial date one had chosen say 1600, then the possible paths to quantum physics should be quite different.)

One of the motivations for this type of study is the intimate connection that exists between postulating counterfactual histories and postulating historical causes. That the discovery of the voltaic pile in 1800 was a necessary cause for Ampère's discovery of the electrodynamical law implies the counterfactual assertion that if the pile had not been discovered before 1820, Ampère would not have made his discovery.

This connection between causes and counterfactuals may be used to overcome the resistance historians have for accepting the postulation of counterfactual histories (a justified resistance, since it is difficult enough to uncover factual history). To do this, one can describe the history of science in terms of "causal models". If this be done in an adequate way, all the information about counterfactual histories will be contained in the causal description.

Causal models have been much studied in the last fifteen years (Pearl [2000]), and the problem addressed is how to infer causal relations from a collection of data (which furnish correlations) and experiments of "intervention". In the present study, one cannot infer causal relations, since history usually happens only once (except in cases of independent discoveries) and it is not possible to intervene in it. We will therefore employ only the notation that is used in causal models and in the analysis of certain structures which form in a network of causal relations.

2. *Advances: Units of Knowledge*

One consequence of the study of counterfactual histories was the development of a theory of science based on the notion of "advance", which are units of knowledge passed on from one scientist to another. There are theoretical advances, such as ideas, formulations of problems, laws, and explanations, recognition of similarities and distinctions, theoretical derivations, theory-data comparison, etc. There are also more experimental advances, such as data acquisition, development of experimental techniques, etc. One may also include motivations, tacit knowledge, and the use of methodological rules. Each scientist assimilates a set of advances, selects some of them, rejects others, even if temporarily, combines two or more advances, etc. Anything that the scientist does, is announced to other scientists, and contributes to the change of a scientific field is characterized as an advance.

In the present context, the word "advance" should not be necessarily understood as a step in the right direction, as conveyed by the usual meaning of the term; it may also apply to steps in a "wrong" direction. Synonyms that might be more adequate are "achievement" (Kuhn [1970], 169), "contribution" (Holton [1973]) or "novelty" (Hugh Lacey, private communication).

Fig. 1 represents some theoretical advances in the 19th century physics, including the "type" of advance. Diagrams of this sort are occasionally used to represent the development of a scientific field, like the diagrams from the 1950's which appear in Holton ([1973], 416–20). In an initial study of types of advances (Pessoa [2000]), around 60 types were grouped in the following ten "classes" of types: experimental techniques, experimental data, theory development, theoretical work, concepts & definitions, laws & facts, data-theory comparison, explanations, problems & criticisms, and motivations & values. In addition to these, more general cultural manifestations also play a causal role in science, especially in early science, as will be indicated in section 7.

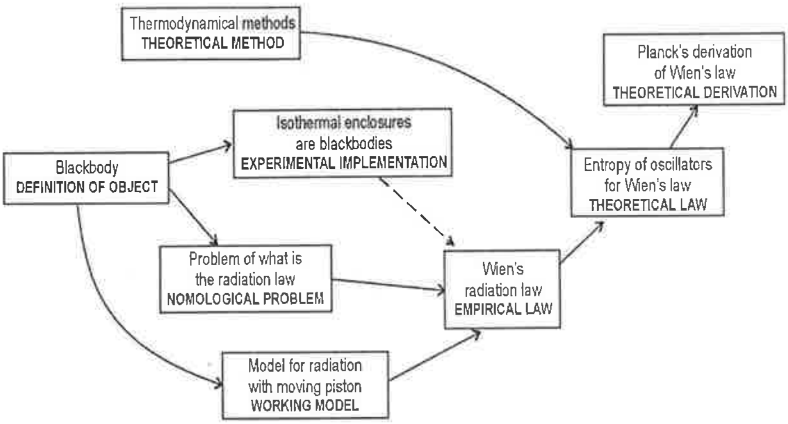

Figure 1. Some examples of advances,
with an indication of the "types", in boldface.

How does the concept of "advance" compare with similar terms in the literature? The prototype of advance is an "idea", and much has been written about the history of ideas, the impact of ideas on society, etc. The generalization of the concept of idea, which is a unit of thought or language, to other cultural expressions is also well explored in the social sciences, which deal with the transmission of "customs", "habits", "practices", "mores", etc. The set of human ideas and practices may be classified as "cultural manifestations". A scientific advance may thus be considered a unit of culture shared by scientists. However, we have included in the definition of advance only the cultural manifestations of epistemic nature, leaving out characteristics of rhetoric, style of work, etc.

In the context of evolutionary biology, the concept of "meme" appeared a few decades ago, being defined as a unit of cultural transmission or a unit of imitation (Dawkins [1976] 1989, 192), an idea, behavior, style, or use which spreads from person to person within a culture. A meme allegedly takes part in a process of evolution, obeying natural selection, in a way analogous to what happens with a gene. Genes and memes would be "replicators", which are *copied* between individuals, may be submitted to *variation* along this transmission, and are *selected* by the environment.

For our purposes, we don't have to take a position on the role of memes in human evolution. Accepting the definition of meme, one may say that advances are cognitive memes passed on between scientists. Similar notions (restricted to the transmission of ideas) also appear in the field of evolutionary epistemology (Campbell [1974]).

3. *Causal Relations between Advances*

When an advance is attained, it influences the appearance of other advances, either for the scientist who came up with it or, after the result is announced, to other scientists. Looking at this process from the perspective of a metascientist (i.e. a scientist of science), what may be observed is that an advance generally contributes as a cause to the appearance of another advance. In Fig. 1, these causal relations between advances are represented by arrows, expressing the influence that one advance had on another.

As is usual in the domain of human sciences, there will be a very large number of events which causally influence the appearance of an advance. One of the tasks of the historian of science is to neglect "weak" causes and point out what are the "strong" historical causes. To do this, he will point out "internalist" causes, which we have called advances, and "externalist" causes, originating from the scientist's social environment. There is no doubt that social causes are important for explaining the progress of science, but we chose to temporarily leave out such causes from our analysis, to facilitate our study. Thus, *social factors* will be left out from the set of cultural manifestations considered as "advances", although their inclusion is plausible and interesting in future work.

Another important question to be emphasized is whether the relations between advances are not only *logical*, rather than following material causation. This takes us back to the classical distinction between the context of discovery and the context of justification. As metascientists, we are concerned with the causes for the appearance of advances (which include the social and psychological environment of the scientist), with the context of discovery. However, scientists are reasonably rational beings, who after their discoveries usually elaborate a justification, which involves the conversion of material causes (with which we are concerned in this paper) to logical or theoretical relations. In this way, the order of facts transforms, with certain modifications, into the order of reasons.

Concerning the distinction between how science *is* and how it *should be*, the present approach (for the time being) sets aside the questions of justification and concentrates on the question of how science *is* (or was) and how it *can be* (or could have been).

4. *Advances in Counterfactual Histories*

As already mentioned, our interest in scientific advances appeared when we tried to develop a method for postulating "counterfactual histories" of science, that is, possible histories that did not occur (Pessoa [2001]). One way of constructing a counterfactual history is to set up the network of causal influences between advances for the factual history of a scientific field, and then postulate a *reordering of the same advances*. For example, one may suppose that an advance of experimental technique, like the construction of a "bolometer" by Langley in 1881, took place ten years before it actually did, or ten years after, and then imagine the consequences that this would have for the appearance of other advances.

This kind of intellectual exercise presupposes that the advances remain more or less the same from one possible history to another. Thus, the suggested analysis of counterfactual histories involves *units that conserve their meaning* in different possible histories (when the order of the advances is altered, usually new advances have to be postulated, as exemplified in Pessoa [2000], section 6d). This "objectivist" assumption may be criticized from a holistic perspective, if one considers that the meaning of a concept depends on the other concepts that take part of a scientific theory. Notwithstanding, this assumption is a useful "first approximation" for postulating a class of counterfactual histories that is quite "close" to the actual history.

Another problem involving the notion of advance is the *distinction between its definition and its degree of acceptance*. Consider a theoretical advance that makes an assertion about reality. Whether this assertion is a speculation, a hypothesis or a confirmed thesis should not affect the definition of the advance, or its truth value. For example, the advance that is usually called "hypothesis of energy quantization" should be stated without reference to its degree of acceptance, for instance as follows: "the energy of microscopic oscillators is quantized". This separation between the definition of an advance and its degree of confirmation (or acceptance) is more important when we suppose that the advance is conserved in the passage from one possible history to another, since what is a hypothesis in one history may be a derived thesis in another.

5. *Causal Models*

The representation of the development of a scientific field, illustrated in Fig. 1, involves variables – the advances – which are connected by causal relations. This is a graphical example of a "causal model", a subject that has been intensively studied in the last decades (see for instance Pearl [2000]). The problem put forth is to establish what are the causal relations between a set of variables. To do this, it is not enough to observe the statistical behavior of the variables, but one also has to perform actual experiments, which involve an intervention (control, manipulation) on the variables, cutting the causal links they have with conditioning factors (Pearl [2000], 42–3, 348; Woodward [2001]).

Causal statements which refer to a certain event involve what is called "singular" or "actual" cause, as opposed to a "general" or "generic" cause, typical of law-like statements.

Considering singular causes, a first classic distinction (Mill, 1843) is between a necessary cause and a sufficient cause. An event C is a *necessary cause* of an event E if, in the absence of C, E would not have occurred. Event C is a *sufficient cause* of E if the occurrence of C guarantees the occurrence of E. John Stuart Mill already noticed that no cause is truly sufficient or necessary for the occurrence of its effect (Pearl [2001], 313).

This observation was explored in 1965 by John L. Mackie, who proposed a logical criterion which would correspond to our intuition that "C is cause of E". This would occur if C were "an *insufficient* but *necessary* part of a condition that is itself *unnecessary* but *sufficient* for the result". The initials of the four italicized terms reads INUS, which names this criterion (Pearl [2000], 314; Cartwright [1989], 25–7).

As an example of the plausibility of this definition, let us consider the causes that led Planck to formulate his radiation law e. He deployed thermodynamic methods (C_1), Wien's law (C_2), Rayleigh's law (C_3), and the data of Rubens & Kurlbaum (C_4), among other things. The set $\{C_1, C_2, C_3, C_4\}$ is a sufficient condition for e, but not necessary, since one can imagine scenarios in which Planck's law would have been derived from other causes. For example, the counterfactual scenario which begins with the discovery of the law of specific heats of solids (C_5), together with the thermodynamic methods (C_1) (Pessoa [2000], 189–90). However, considering the aforementioned set, Wien's law (C_2) is a necessary part of this condition (since without it Planck would not have arrived at his law), but of course it is not a sufficient part, since there are three other parts (C_1, C_3, C_4). Thus, according to the INUS criterion, Wien's law was a cause of Planck's law. Another way of phrasing this criterion is to consider the cause a "necessary element of a sufficient set" (Pearl [2000], 314).

In Fig. 2, the causal relations of this example are represented graphically, by means of structural diagrams (directed acyclic graphs), with nodes standing for variables (advances) and arrows standing for causal dependences between the variables. This kind of representation is called a "causal model", in a broad sense. For Pearl [2000], 203, a causal model is a mathematical description of a set of variables v_i, by means of a set of functions f_i the arguments of which are other endogenous variables a_i and also exogenous variables u_i (represented stochastically): $v_i = f_i\ (a_i, u_i)$. Alternatively, one may use a probabilistic representation which makes use of Bayes' theorem for calculating conditional probabilities in light of new evidence (such methods rival with classical statistics, generating much methodological discussion; see Howson & Urbach [1993]). This second type of causal model may be more promising for our purposes.

Notice, in Fig. 2, the use of the logical conjunction operator "&". When arrows point to the same advance without this conjunction symbol, it is implicitly assumed that the logical operation is a disjunction.

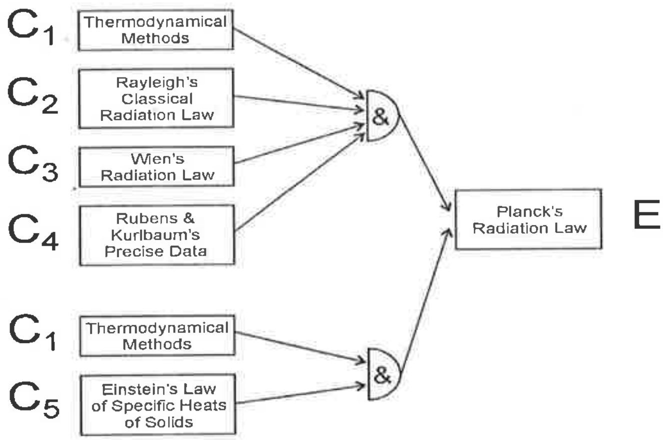

Figure 2: Situation in which there are two possible histories (each one sufficient) for the production of an advance E.

An additional issue to be examined involves the notion of "asymmetry between cause and effect", or the "unidirectionality of causation" (Bunge [1959], ch. 6). A cause, like the motion of an elevator, produces an effect, like the movement of the pointer in the hotel lobby, indicating the floor the elevator is in. If someone tries to intervene turning the pointer, he will not be able to control the elevator. Control of the effect does not alter the cause, although the converse is true.

In our study, however, by contemplating the possibility of counterfactual histories, we have admitted the possibility of an effect producing a cause. For example, historically, Planck's radiation law (E) was one of the causes of Einstein's law of specific heat (C_5). Notwithstanding, one may postulate a counterfactual scenario represented by the lower portion of Fig. 2, in which the order is inverted. In other words, if we leave the level of singular factual causes and put ourselves at the level of possible causes, we violate the asymmetry between cause and effect. Such a violation of asymmetry is not uncommon in physics, for example, when the energies involved in the cause and the effect are comparable.

6. *Attribution of Probabilities in Causal Models*

In the history of science, many times the conditions for the appearance of an advance are given, but still the advance does not occur. This indicates that the relation between causes and effect is *probabilistic*. As an illustration, let us consider examples from the ancient science of magnetism, which are interesting because the beginnings of science in China and in the West occurred quite independently, thus constituting two different factual possible historical paths (Needham [1962]). Given that the precise compass (for example, the thread suspension magnetized needle) was developed (call this advance "$A5$"), there would be a certain probability that magnetic declination (the deviation of the needle from the astronomical north, call it "$A8$") would be discovered in a certain interval of time Δt. This probability may be written as $p_{\Delta t}(A8/A5)$.

The interval of time considered could have any value, but it would be interesting to choose a "typical" value, which would reflect the stage of development of science. In other words, for each time and location in the history of science, represented by an year τ and a place λ, we will take a typical value Δt (τ, λ) which expresses the average number of years, after the appearance of a sufficient cause, that it takes for the corresponding effect to be produced. For example, in the Chinese 12th century, one might stipulate that Δt be one century: Δt(1120,China) = 100 years. If we estimate, for this time and place, that the probability for $A8$ to be produced after the appearance of $A5$, in 100 years, is .9 (that is, $p_{\Delta t}(A8/A5) = .9$), then we may ask how long it would take for $A8$ to be produced from $A5$ in another time and place, like France in the beginning of the 19th century. In this case, we might estimate that such a production would have a .9 probability of occurring not in 100 years, but maybe in 10, in such a way that Δt(1820,France) = 10 years. By defining, in this manner, the typical interval of time as an inverse function of the degree of development of science in a time and a place, we leave invariant, through history, the conditional probability involving two causally connected events.

With this assumption, we are implicitly assuming that the degree of development of a science affects equally the production of all types of advances. A quantitative notion of degree of development – or at least a graphical representation of it – has been suggested by different authors, such as Needham's comparison of the rate of development of European and Chinese science (Needham [1970], 414) or Rescher's analysis of scientific progress (Rescher [1978]).

In the history of science, as in most of the social sciences, a certain event E is generally conditioned by a large number of causal factors C_i. In the aforementioned example, the development of the precise compass was *necessary* for the discovery of magnetic declination. However, there were several different forms of the precision compass, so if each one be taken individually, each of them would not be necessary, although each one would be sufficient, in conjunction with other factors and with a certain probability, for the appearance of the effect. Each of these paths, which compose a disjunction of causal connections, corresponds to a possible history.

We mentioned that the notion of "sufficient cause" can only be understood if associated to a certain probability smaller than 1, in a certain time interval Δt. If we call this cause "strong", then there would be many "weaker" causes affecting the probability of occurrence of the effect. For example, above we gave the rough estimate that $p_{\Delta t}(A8/A5) = .9$, that is, given only the invention of the precise compass, the probability that magnetic declination be discovered in a typical interval of time would be .9. However, it so happened that the Taoist cultural practice of "geomancy" (the art of adapting the residences of the living and the tombs of the dead so as to cooperate and harmonize with the local currents of the cosmic breath; see NEEDHAM [1962], 239–42) accelerated the exploration of the directive properties of the compass and contributed positively for the discovery of declination. Representing this practice by G, such positive contribution may be expressed by an increase in probability: $p_{\Delta t}(A8/A5\&G) = .95$.

A similar situation is depicted in the following diagram, which shows how an advance is represented in the computational environment of the "SCHEME" program used to store historical information and run simulations based on them. The overall "list" (which starts and ends with a parenthesis) includes, after the name of the advance, a description of it (in quotes), its causal influences (with probabilities) and the type of the advance. The information concerning who discovered or used an advance is stored separately.

```
(define Directive_property_of_lodestone
  '("If lodestone is free to move in the horizontal
     plane, it will tend to align along a fixed
     direction."
  (((Lodestone_effect) .2)
   ((Lodestone_effect Floating_lodestone_needle) .95))
   observed-fact))
```

It can also happen that the existence of an advance contributes negatively to the appearance of another. An example taken from the history of astronomy is the negative influence that the Greek notion of "perfection of the cosmic sphere" had on the observation of sunspots, which were only recognized in the West by Galileo, but which were already known to the Chinese in the 1st century A.D. These examples indicate a difficulty for the estimation of probabilities in the causal models of history of science. Composition of causes takes place in a non-linear manner, so that the action of a set of causes which are practically sufficient for the production of an effect may be blocked by the appearance of an additional factor.

Another important causal structure is the conjunction of necessary causes. Such a conjunction is very common in a mature science (such as the 19th century physics), but not so much in early science. Let us illustrate this situation with one of the few examples from the early science of magnetism. The construction of the "compass card" (the amalgamation of a compass with a wind-rose) required the simultaneous presence of the dry-pivoted compass (call it "$A9$") and the nautical use of the compass ("$A6$"). Accepting the validity of this example, and representing the compass card by the letter "C", this situation could be roughly expressed in the following way:

$$p_{\Delta t}(C/A6\&A9) = .95 \ ; \ p_{\Delta t}(C/A6) = .1 \ ; \ p_{\Delta t}(C/A9) = .1 \ .$$

A very common example of conjunction of causes occurs when an explanation is given for a certain phenomenon. According to the present approach, an explanation is "caused" by the phenomenon to be explained (for example, the attractive properties of lodestone) and by the general theory within which the explanation is given (for example, ancient atomism). To say that an explanation is "caused" by a phenomenon and by a general theory may sound strange, but we should remember that the notion of causality is being applied to a domain (history of science) which contains rational agents (the scientists). These agents make use of different sorts of methods of inference to arrive at new results, and such a procedure is usually understood as following a "order of reasons" which does not make use of the notion of causality. However, if we look at the history from the outside, there is undoubtedly a certain ordering of advances, and the relation between them may be considered a probabilistic causal connection. We will leave aside the interesting question of the ultimate nature of this causal connection.

7. *Example of a Description by means of a Causal Model*

We will now give a fuller example of the attribution of probabilities within a causal model, making use of the two factual but independent histories of the early science of magnetism in China and in Europe. The most striking difference between these two possible histories was the discovery of the compass (or the directive property of lodestone) in China, but not in Europe. The main causal factor responsible for this difference, according to Needham [1962], was the widespread presence, in the Chinese culture of the 1st century, of divinatory arts involving floating needles and boards on which pieces were thrown, while in Europe such techniques had much less importance (although they did exist, for example in the island of Samothrace).

We have reconstructed these events according to the following possible history: the lodestone effect and the existence of a method of divination based on a greased needle of iron floating in water led to the use of rudimentary floating lodestone needles. This practical situation would have led to the discovery of the directive property of lodestone (its ability to align along the north-south axis). After that, a dry compass was developed, in the shape of the Great Dipper constellation, resembling a spoon, to be used in a diviner's board.

To represent this causal description, we introduce the following notation and rough probabilities:

$A1$: Lodestone effect	$p_{-200,\text{China}}(A1) = 1$; $p_{-200,\text{Europe}}(A1) = 1$;
$A3$: Directive property of lodestone	$p_{-200,\text{China}}(D) = 1$; $p_{-200,\text{Europe}}(D) = 0$;
F: Floating lodestone needle	$p_{-200,\text{China}}(T) = 1$; $p_{-200,\text{Europe}}(T) = 0$;
D: Divination by floating needle	$p_{\Delta t}(F/A1) = .1$; $p_{\Delta t}(F/A1\&D) = .9$; $\Delta t = 400$ yrs.
T: Diviner's board	$p_{\Delta t}(A3/A1) = .2$; $p_{\Delta t}(A3/A1\&F) = .95$;
B: Spoon shaped lodestone compass	$p_{\Delta t}(B/A3) = .1$; $p_{\Delta t}(B/T) = .1$; $p_{\Delta t}(B/A3\&T) = .9$.

The importance of divination techniques in China, in contrast to the situation in Europe, is expressed in a simplified way by stipulating a prior probability of 1 for the divination techniques D and T in China around the year 200 B.C., while the corresponding prior probabilities in Europe are fixed to 0. It is also assumed that the lodestone effect was known in both cultures around this time. A typical time interval of Dt = 400 years was considered, for both China and Europe around 200 b.C. With these and the other stipulations of the preceding chart, which are graphically represented in Fig. 3, one can calculate that the probability for the compass being discovered in the 400 years following 200 b.C. was high in China and low in Europe.

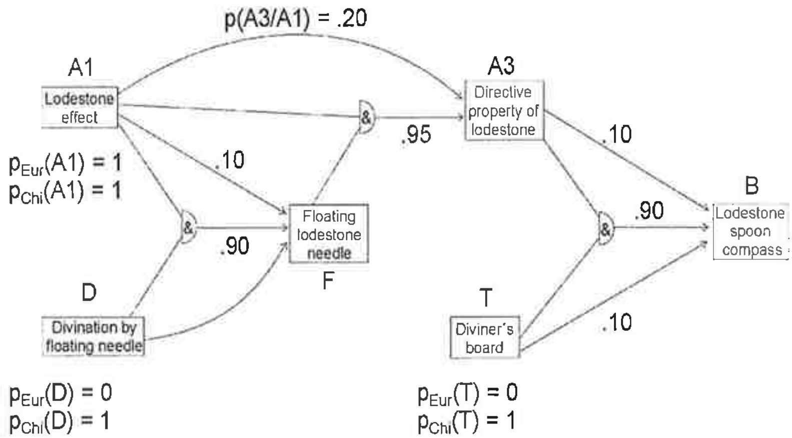

Figure 3: Causal model based on a reconstruction of the development of the lodestone compass in China.The difference between what occurred in Europe and China is expressed by different prior probabilities.

These numbers, of course, are based on very rough estimates, but the aim here is to give an example of the use of causal models in the description of science. The hope is that the piecemeal construction of such nets, with probabilities that express our historical intuitions in a rough way, might lead to interesting insights once the overall network is put to work, with the help of a computer.

Bibliography

Bunge, M. [1959], *Causality and Modern Science* (Cambridge: Harvard University Press). 3rd revised edition: New York: Dover, 1979.

Campbell, D.T. [1974), "Evolutionary Epistemology", in Schilpp, P.A. (ed.), *The Philosophy of Karl Popper*. Vol. 1. (La Salle: Open Court), 413–63.

Cartwright, N. [1989], *Nature's Capacities and their Measurement* (Oxford: Clarendon).

Dawkins, R. [1976]. [1989], *The Selfish Gene*. 2nd enlarged edition (Oxford: Oxford University Press).

Holton, G. [1973], "Models for Understanding the Growth of Research", in Holton, *Thematic Origins of Scientific Thought* (Cambridge: Harvard University Press), 397–430. Original version published in *Daedalus*, 1962.

Howson, C. & Urbach, P. [1993], *Scientific Reasoning: The Bayesian Approach*. 2nd edition (Chicago: Open Court).

Kuhn, T.S. [1970], *The Structure of Scientific Revolutions*. 2nd edition (Chicago: University of Chicago Press).

Mackie, J.L. [1965], "Causes and Conditions", *American Philosophical Quarterly*, *2*, 245–64. Republished in Kim, J. & Sosa, E. (eds.) [1999], *Metaphysics – An Anthology*. Oxford: Blackwell, 413–27.

Needham, J. [1962], *Science and Civilization in China*, vol. 4: "Physics and Physical Technology", part I: "Physics", in colaboration with W. Ling & K.G. Robinson (Cambridge: Cambridge University Press).

_______[1970], *Clerks and Craftsmen in China and the West* (Cambridge: Cambridge University Press).

Pearl, J. [2000], *Causality – Models, Reasoning and Inference* (Cambridge: Cambridge University Press).

Pessoa Jr., O. [2000], "Histórias Contrafactuais: O Surgimento da Física Quântica", *Estudos Avançados*, 14, 175–204.

_______[2001], "Counterfactual Histories: The Beginning of Quantum Physics", *Philosophy of Science* 68 (Proceedings), 519–530.

Rescher, N. [1978], *Scientific Progress* (Oxford: Blackwell).

Woodward, J. [2001], "Causation and Manipulability", *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/causation-manipulability/ .