Modeling the Causal Structure of the History of Science

Osvaldo Pessoa Jr.

Abstract. This paper is an overview of an approach in the philosophy of science of constructing causal models of the history of science. Units of scientific knowledge, called "advances", are taken to be related by causal connections, which are modeled in computers by probability distribution functions. Advances are taken to have varying "causal strengths" through time. The approach suggests that it would be interesting to develop a causal model for scientific reasoning. A discussion of counterfactual histories of science is made, with a classification of three types of counterfactual analyses: (i) in economic and technologic history, (ii) in the history of science and mathematics, and (iii) in social history and evolutionary biology.

1 The Model: Advances Connected by Causal Relations

This paper is part of a project of developing a computational model that describes the history of science. Such a representation stays close to the narrative of the historian of science, who writes about ideas, discoveries, instruments, theories, etc., each of which exerts influences, in differing degrees, on the appearance and confirmation of other scientific advances. These units of scientific knowledge, which are explicitly or tacitly passed among scientists, will be called *advances* (even though they might not be a positive contribution to the progress of science). The prototype of an advance is an idea, but there are other types of theoretical advances, such as explanations, laws, problems, theory development, as well as experimental advances, such as data, experiments, and instruments. Other advances include the comparison

---

Osvaldo Pessoa Jr.
Department of Philosophy, FFLCH, University of São Paulo, São Paulo, Brazil
e-mail: `opessoa@usp.br`

L. Magnani et al. (Eds.): Model-Based Reasoning in Science & Technology, SCI 314, pp. 643–654.
springerlink.com

between theory and experiment, methodological theses, metaphysical assertions, projects, tacit knowledge, etc.

Advances are connected in certain ways: they influence the *appearance* of other advances, and they also affect the *degree of acceptance* of other advances. In the present approach, such a connection is taken to be a *causal* relation, not a logical one. For example, the construction of the thermopile by Nobili & Melloni in 1830 was essential for the discovery of polarization of radiant heat by James Forbes in 1836: without the thermopile, Forbes would not have discovered polarization at that moment. The thermopile may therefore be considered a "cause" of the proposal that "radiant heat may be polarized", in the sense expressed by the so-called counterfactual definition of causality: if the cause had not occurred, then the effect would not have existed (in the case of a necessary connection), or the probability of its occurrence would have been different.

Causal relations in social systems are always complicated, and one can rarely single out a necessary and sufficient condition. A cause is better represented as an "INUS condition" [9], which amounts to saying, in the above example, that many other causes acted together with the thermopile to lead Forbes to his discovery, and that probably another sufficient set of conditions (not including the thermopile, but perhaps a more sensitive mercury thermometer) could have led to his discovery.

The use of computation to model causation in the history of science has also been explored by Gerd Graßhoff & Michael May, from a different perspective. They investigate how a scientist deploys causal reasoning to construct a causal model of their subject matter, such as done by Krebs & Henseleit for the biochemical pathway underlying the urea cycle [5].

2 Probabilistic Causal Relations Express Possible Histories

Another weakening of these causal relations is that a set of conditions can at best increase the *probability* that a scientist will arrive at a certain advance in a certain interval of time. The great number of causal influences that act haphazardly on a scientist, but cannot be accounted for by the model, are considered as "noise" or random fluctuations, the dispersion of which is encompassed by probability distribution functions.

Figure 1 is an example of how a causal connection may be modeled by a probability distribution function. Advance A1 is Newton's famous experiments with sunlight and prisms, which he reported in 1672. Such investigation was a necessary condition for the discovery of advance A2, that the solar spectrum has dark lines, discovered independently by William Wollaston (1802) and Joseph Fraunhofer (1814). To express the conditional probability of A2, given A1, one may use a gamma distribution, with mean value given by 1808 (the mean between 1802 and 1814) and the standard deviation (the

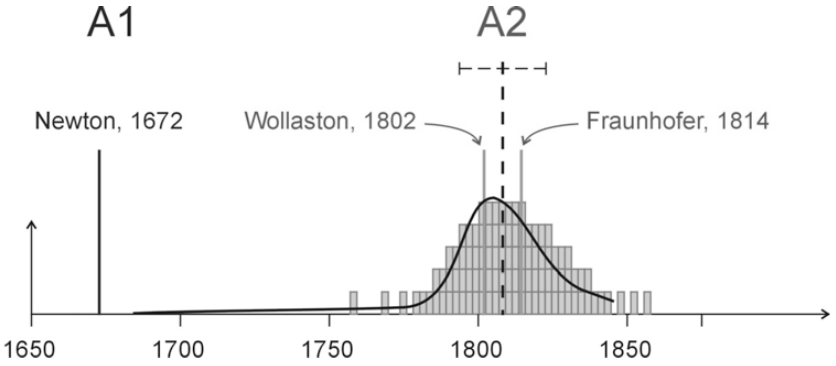

Fig. 1 Probability distribution for the appearance of advance A2, given A1.

half-width of the curve) determined by the spread of the dates of the independent discoveries.

One way to interpret such a distribution function is to think in terms of possible histories of science. Imagine one hundred worlds, created (say in 1673) from the actual world, but with small random changes (for a more detailed recipe for doing this, see [12]). In each of these possible worlds, in which A1 is given, how much time would it take for advance A2 to arise? It is natural to suppose that the time intervals would not be exactly the same, but would be distributed according to a certain curve. In Fig. 1, each possible scenario is represented by a small rectangle, placed in the year in which A2 would appear. The resulting histogram is supposed to be an approximation to the associated distribution function.

3 Causal Structure of Episodes in the History of Science

The present approach to modeling the history of science is being implemented in the Lisp-like Scheme programming language. Computer programs don't provide actual thinking and intuition, but they allow the storage of detailed information concerning the relations between advances and their causal strengths, and allow simulations to be run, which we hope might help to test different metatheoretical theses about the development of science.

Most of our historical studies has focused on the fields of optical spectroscopy and thermal radiation in the 19th century. The ultimate aim is to represent in detail the beginnings of quantum physics. In a preliminary study of the possible paths leading to the birth of the old quantum theory [10], it was suggested that there would be four main paths, the most probable not being the actual one (in the field of thermal radiation), but in the field of optical effects. A simple causal model helped to organize the study, but the conclusion was reached "intuitively", and should be qualified and refined with a more detailed causal model.

Independent discoveries offer interesting material for comparing possible histories of science. We have examined the origins of the science of magnetism in China and in Europe, up to around the 5th century, and constructed a single causal model which accounts for why the rudimentary magnetic compass was developed in China but not in Europe, based on different initial probabilities for the existence of divination techniques in both regions [11]. This followed the account of the historian of science Joseph Needham, for whom it was the widespread use of such techniques in China that allowed the directive property of lodestone to be discovered in the East. In this model, probabilities were assigned by identifying the "empirical time span" between two advances (involving the actual years in which the advances arose) with the mean of the associated distribution function $f(t)$ (such as the one in Fig. 1).

Another example illustrating the causal structure of an episode in the history of science is given in Fig. 2, which represents the actual paths leading to the independent discovery of the principle of spectral reversal, in 19th century spectroscopy (the associated probabilities are not represented in the figure). This principle states that a medium which absorbs well certain spectral lines will also emit well these lines. For example, when sodium gas is excited by an electric arc, it emits several spectral lines, especially the yellow D double lines, which appear strongly on the lower right corner of Fig. 2, in spectrum (a). On the other hand, when sodium gas intercepts light coming from another source, such as the sun, it absorbs strongly the D lines. The principle that good emitters are good absorbers, for each wavelength of light, was discovered independently by Foucault (1848), Ångström (1853), and Kirchhoff (1859), while the latter expressed such a principle as a mathematical law. Foucault's and Kirchhoff's path to discovery involved the curious observation that the sun's dark D lines get even darker as they pass through sodium gas. This is related to the phenomenon of self-reversal, depicted in spectrum (c) of Fig. 2, when the increasing intensity of the bright D line emission leads to a paradoxical darkening of the line, which is explained by the absorption of this line by the surrounding cooler sodium gas.

The same principle of spectral reversion was also discovered by Balfour Stewart (1858) in the field of infrared radiation, which at the time was called "radiant heat", and both he and Kirchhoff generalized the principle to both visible light and infrared radiation, after it became well accepted that both are essentially the same form of radiation, differing only in their wavelengths. A study of how the rate of development of these two fields was influenced by different sets of technological advances is presented in [13].

Fig. 2 illustrates how different pathways of actual discovery can be represented by causal models, involving conjunctions and disjunctions of paths. Further complications must be introduced to represent the causal strengths of advances, which varies with time (see section 5). The model does not include counterfactual scenarios.

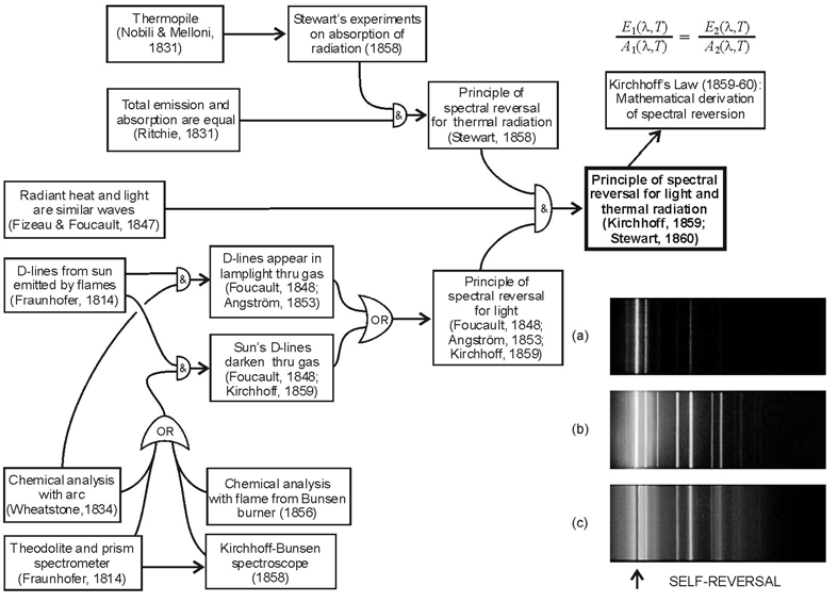

Fig. 2 Simplified causal model for the history of the discovery of the principle of spectral reversal.

There is a clear similarity between the possible histories presented here and the "investigative pathways" explored by David Gooding [3], Frederic Holmes [6], Andrea Loettgers [8], and other historians involved in the model-based reasoning approach to science. The difference involves the scale adopted by each approach.

One may classify the different scales of study in the history of science into at least five groups. (5) Global theses about the scientific institution, involving spans of hundreds of years, all of the world, and all of the scientific fields. (4) General views about the scientific change in a certain field (as done by Kuhn, Lakatos, and Laudan), usually involving decades and a whole civilization. (3) Study of a historical case, such as Darwin's ideas or science in Scotland, etc., involving months or years, and a few institutions. (2) Focus on the procedures adopted by scientists to bring about an advance; this is the scale of ethnomethodological studies, involving hours: "Faraday did this, and then that, etc." (1) Microcognition: cognitive details in the mind of the scientist, involving seconds: studies in this scale are still incipient.

The present approach to causal models in the history of science focuses mainly scale 3, while the investigative pathways are closer to scale 4. The next section will address scale 1.

4 Causal Model of Scientific Reasoning

When a scientist derives a new theoretical result, such a result is usually presented as a logical inference based on other advances. Although the connection between these advances is presented as a logical relation, a consideration of the actual circumstances of the derivation will point out which of the advances are the causes (being previously known), and which one is the effect (the new result). When a scientist justifies a result in deductive form, there are at least two possibilities for the causal history of the result: either the premisses are the actual causes of the conclusion (so the scientist actually discovered the conclusion by deductive inference from the premisses), or the conclusion was previously accepted by the scientist and led him to formulate a premiss as an explanatory hypothesis, in an abductive inference.

The present approach sees a scientist as a very complex cognitive machine that receives a large number of advances (with changing degrees of acceptance) as causal inputs and generates new advances, which will causally affect himself and other scientists. Although the present approach should impose no requirements on how human beings think, it would be interesting for the completeness of the programme if the human mind could be modeled in strict causal terms. This would satisfy a certain "causal closure" of the world, but this expression should allow for the possibility that truly stochastic, nondeterministic events could occur in nature (the existence of such events is an open question in the philosophy of physics). The present author would love to give at least a sketch of the project of describing scientific reasoning (especially abduction) in causal mechanical terms, but he has yet no clue of how this could be done, although many contributions to the "model-based reasoning" community seem to be relevant for this Hobbesian dream.

5 Causal Strength of an Advance

One must also take into account the "strength" of the causal relation. The time interval between the appearance of the first advance (the cause A1) and of the second (the effect A2) is an indication of this strength: the shorter the time, the stronger the cause. Another aspect of this concept of causal strength is that it is a measure of the degree of acceptance of the advance, and it varies with time, as scientists discuss its merits. If the advance is an idea, this discussion might involve debating its degree of confirmation, which affects the degree of acceptance of the idea. If the advance is a new instrument, different scientists must investigate its performance, which then affects how trustworthy are its measurements. If the advance is a problem, then its strength reflects how many scientists are concerned with it.

The *causal strength* of an advance may thus be defined as the potentiality that it may influence the appearance of other advances, or that it may affect the causal strength of other advances (mediated, of course, by the brains and hands of scientists, and by their social and institutional interactions) [14]. When working with causal models in the history of science, an advance should always be considered together with an estimate of its causal strength.

Fig. 3 shows graphically how the causal strength (represented by the thickness of the “strip”) of a few advances in 19th-century research on optical spectroscopy and radiant heat developed through time. Advances that are successful usually start out with little support and gradually become widely accepted, such as depicted in strip (d). Some don’t have a monotonic growth, such as the thesis that the dark lines in the solar spectrum originate in the solar atmosphere (strip c). This view was suggested by David Brewster and others in the early 1830’s. However, during the solar eclipse of 1836, James Forbes concluded, from his observation of the spectra arising from the solar corona, that the dark lines in the solar spectrum do not arise in the sun’s atmosphere. Brewster & Gladstone repeated this negative point on the eve of Kirchhoff’s discoveries, who in 1859 argued convincingly that the dark lines of the solar spectrum are not caused by the earth’s atmosphere, but originate from the presence of those substances in the glowing solar atmosphere.

Many advances have their causal strength going to zero, such as the thesis that radiant heat is of a different nature from visible light, after 1872 (strip b in Fig. 3).

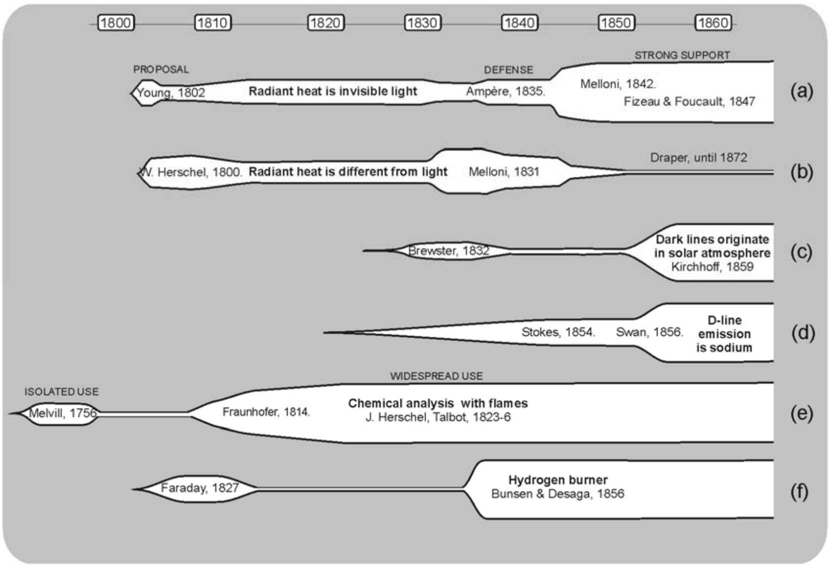

Fig. 3 Change in casual strength through time, for a few advances in 19th-century research on optical spectroscopy and radiant heat.

6 Counterfactual Histories of Science

We have extracted from the historian's discourse a model of science in which advances are connected by probabilistic causal relations. By defining such causation in counterfactual terms, we have automatically introduced the controversial notion of "counterfactual" or "virtual history". A counterfactual situation is a *possible* situation that did not happen. Is it really necessary to introduce counterfactual possibilities in a causal description of history? One can always choose to avoid counterfactual statements, but it can be argued that every causal statement implicitly implies counterfactual scenarios. For example, if one asserts that the main cause of the decline of science in France around the 1830's was its centralized organizational structure, then one is implicitly asserting that *had* such a structure *been* transformed into a more decentralized structure, as in the German countries, then French science *would have* thrived.

Counterfactual scenarios in history are always speculative, but so is the postulation of causes. In the hard sciences, a causal statement may be tested by exploring the different outcomes of an experiment for each value of the parameters controlled by the scientist. These experiments map out the possibilities of outcomes of the experimental situation, so it might be said that the possible histories (describing the measurement outcomes of the experiment) are all *actual*, since the "history" of the experimental situation repeats itself many times. In the case of social history, repetitions of sets of conditions are quite rare (but not so rare in the history of science), so attributions of causes are difficult to test, remaining speculative, just like counterfactual assertions.

Still, human beings have a very good intuition for imagining counterfactual situations [16], as well as for imagining causes, which is connected to the evolutionary fitness value for predicting the future. Historians of science frequently make counterfactual assertions, such as Harry Woolf's ([17],p. 628) comment that the pioneer of flame spectroscopy, Thomas Melvill, "was clearly on the road to major discovery in science" (such as the dark lines in the solar spectrum), had he not died prematurely in 1753, at the age of 27. Such assertions are usually made in a marginal way, but recently more attention has been given to counterfactual assertions in the history of science (see [15]).

The counterfactual histories to be sketched in our approach are very close to factual history, and much of the research investigates the delay or anticipation of an advance. Throughout the possible histories that we have postulated, each advance maintains its identity (i.e., we neglect changes of meaning due to different contexts); what changes is the order in which they appear (their causal path).

A counterfactual scenario is a possible situation that did not actually happen. But what is a "possible" situation? For our purposes, we will not be concerned with logical possibilities, as is common in the metaphysics of possible worlds, but with what has be called "temporal possibilities" (or "causal possibilities"). We start by considering that our future is "open", and the different future possibilities are partially dependent on our choices and on random events in the physical world. (If the universe were strictly deterministic, then there would be only one temporally possible scenario for the future, and only one possible history of science.)

Granted this, we can define a possible scenario as *a future possibility at some instant* t_0 *of the past*. According to this definition, a counterfactual history must be defined in relation to a branching time t_0 in the past (the time when the counterfactual situation "branched off" from the actual history of science). The probability attributed to a counterfactual state of affairs usually changes according to the branching time being considered.

One might ask whether it would be causally possible that bacteria were discovered on Earth without the use of optical microscopes. Suppose that there were no way of producing glass on Earth; then it is plausible to speculate that bacteria would have been discovered by some other path, not involving optical microscopes. However, there is no instant t_0 in the past from which a possible world without glass could branch off (unless, maybe, if we go back to a time close to the Big Bang). Therefore, such a scenario is not causally possible, although it is physically possible (in the sense that it doesn't violate any law of physics) and logically possible.

The notion of a "tree of possible histories" is useful in philosophy of science for clarifying different conceptions of scientific progress, such as the more traditional one of convergence to the truth (Popper, etc.) and the more relativist conception of selection of the fittest theory (Kuhn) (see [12]).

7 Counterfactual Scenarios in Different Fields

There are at least three different types of counterfactual analyses that may be done in the historical sciences. The most fruitful one comes from the field of economic history, starting with the work of Robert William Fogel [2] on railroads and the economic growth of the United States in the 19th century. There was a traditional conception that the railroads were indispensable to the American progress in the 19th century, i.e., they were a necessary cause for this progress. Fogel examined this thesis, and calculated in detail the costs and the efficiency of other alternatives, and concluded that if railroad technology were not available at the time, there was an equally efficient alternative which was transportation in waterways. According to his calculations, the gross national product that the United States in fact attained in January 1, 1890, would have been reached without railroads (but with waterways) only three months later! The option for waterways would make use of the navigable rivers and lakes, the canals already built, and also many new canals. The industrial regions that would develop would be partially different from the ones that have in fact developed in our actual world.

What allows economic calculations of plausible counterfactual scenarios is the possibility of making reasonably accurate quantitative predictions about the future. For example, the government may open a bid for a contract on an alternative form of energy, so different engineering projects may be presented, each with a possible scenario for the future. After one of them is chosen and implemented, the non-realized projects will have become counterfactual histories (since they were future possibilities at a time in the past). These counterfactual scenarios will be more accurate than the original projects, since hindsight includes information about how the circumstances actually evolved. These two elements, *predictability* and *hindsight*, make counterfactual assessments quite plausible in economic and technologic history.

A second type of counterfactual analysis is done in the history of science and mathematics. Here, the postulation of counterfactual scenarios is less accurate than in economic history, since there is no way of predicting the future of science, contrary to what happens, to a certain extent, in engineering, technology, and economics. One may predict situations related to science policy, but one cannot predict what new discoveries will be made.

However, there is distinguishing feature in the development of science and mathematics which is its *objectivity.* To put it in simple terms, natural science is an attempt to mirror reality, so this reality (which is invariant across the possible worlds) constrains the appearance of scientific advances. In more general terms, without such a commitment to scientific realism (but only to objectivity), there are “attractors” in science, mathematics, and technology (be it reality, consistency, subjective categories, material determinations, or whatever) which constrain the formulation of these disciplines. In almost all causally possible worlds, branching say after the year 1800, scientists would have discovered that the molecule responsible for inheritance has the structure of a double helix, so in this sense there is a common attractor acting on these possible histories of science.

With the advantage of *hindsight* and of the present knowledge of the field, we now know (to a large extent) what the scientists of the past were close to discovering. This allows us to imagine to what consequences slight modifications in the circumstances and choices surrounding the scientists could lead. We may conjecture what could be the different possible paths leading to a discovery, such as the quantization of energy [10]. We can investigate what consequences would have arisen if an advance appeared before or after the time it actually appeared.

But what would be the use of postulating counterfactual histories of science, of generating them with the help of a computer? Without postulating counterfactual scenarios, a lot could be done with detailed causal models, such as testing different metatheoretical theses. But if we were able to generate counterfactual scenarios that are plausible to the historian’s intuitions, that would indicate that the theory of science behind these models is well constructed, and that is the ultimate aim of the present project: to contribute to a testable theory of science.

A third type of counterfactual analysis occurs in social, political, and cultural history, in the approach known as “virtual history”. Here, however, the constraints are much weaker than in the two previous types: one does not have an economical rationality which allows to predict with some detail the collective choices of the agents, and neither a strong attractor as in science, mathematics, and technology. For example, what would have happened if the shot that killed John F. Kennedy had missed him? Our knowledge of human behavior tells us, for sure, that he would have immediately taken cover, and then left Houston, but what next? The number of possible scenarios increases immensely. A few events, such as the presidential election of 1964, would seem predictable: in this counterfactual scenario, Kennedy would have a high probability of being reelected. But after that, would the United States remain at war against Vietnam? Many have given their opinion, but there is no consensus (see [7]). The best one could do would be to attribute a probability around 1/2 for each alternative, but that would lead nowhere, since subsequent events would also be unpredictable.

Much more could be said about virtual history, but let us consider a final case of counterfactual reasoning, which arises in biological evolution. Biologists such as Stephen Jay Gould [4, ch. 5], Stuart Kauffman and Richard Dawkins [1, pp. 482–93] have examined the question of how biological evolution on Earth would take place if the "tape of evolution" were run back to a moment of the past, and if random variation made living beings evolve in different directions. The consensus is that the species that would appear on earth would be quite different from the present ones, and what we define as the human species would not appear for branching times earlier than a few million years ago. The paleontologist Dale Russell and the geologist S. Conway Morris have speculated on what could have happened if a great meteor had not fallen on Earth 65 million years ago, extinguishing the dinosaurs. Maybe a descendent of the troodont would have become as intelligent as we are, and be doing philosophy of science by now. Notice, however, that in spite of the great divergence in variation (although there are constraints to this), it is reasonable to suppose that intelligent beings would eventually inhabit the Earth, which is an example of convergent evolution. One may say that environmental niches act as attractors to the development of biological structures, or "ecological types". The postulation of counterfactual evolutionary histories would depend on knowledge of what variations are possible and on how selective pressures act (a knowledge that has apparently already been achieved). However, the number of possible branches would be huge, contrary to the case of appearance of advances in the history of science, and to the rational possibilities in economic history, but similarly to virtual history and to the outcome of a sports game.

References

1. Dawkins, R.: The Ancestor's Tale. Weidenfeld & Nicolson, London (2004)
2. Fogel, R.W.: Railroads and American Economic Growth. Johns Hopkins Press, Baltimore (1964)

3. Gooding, D.: Mapping experiment as a learning process: How the first electromagnetic motor was invented. Science, Technology and Human Values 15(2), 165–201 (1990)
4. Gould, S.J.: Wonderful Life. W.W. Norton, New York (1989)
5. Graßhoff, G., May, M.: Hans Krebs' and Kurt Henseleit's laboratory notebooks and their discovery of the urea cycle – Reconstructed with computer models. In: Holmes, F.L., Renn, J., Rheinberger, H.-J. (eds.) Reworking the Bench, pp. 269–294. Kluwer, Dordrecht (2003)
6. Holmes, F.L.: Laboratory notebooks and investigative pathways. In: Holmes, F.L., Renn, J., Rheinberger, H.-J. (eds.) Reworking the Bench, pp. 295–308. Kluwer, Dordrecht (2003)
7. Kunz, D.: Camelot continued: What if John F. Kennedy had lived? In: Ferguson, N. (ed.) Virtual History, pp. 368–391. Picador, London (1997)
8. Loettgers, A.: Exploring concepts and boundaries of experimental practice in laboratory notebooks: Samuel Pierpont Langley and the mapping of the infrared region of the solar spectrum. In: Holmes, F.L., Renn, J., Rheinberger, H.-J. (eds.) Reworking the Bench, pp. 159–182. Kluwer, Dordrecht (2003)
9. Mackie, J.L.: Causes and conditions. American Philosophical Quarterly 2, 245–264 (1965)
10. Pessoa Jr., O.: Counterfactual histories: The beginning of quantum physics. Philosophy of Science (Proceedings) 68, S519–S530 (2001)
11. Pessoa Jr., O.: Computation of probabilities in causal models of history of science. Principia (Florianópolis, Brazil) 10(2), 109–124 (2006)
12. Pessoa Jr., O.: Scientific progress as expressed by tree diagrams of possible histories. In: Mortari, C.A., Dutra, L.H.A. (eds.) Anais do V Simpósio Internacional Principia, pp. 114–122. Núcleo de Estudos da Linguagem - Universidade Federal de Santa Catarina, Florianópolis (2009a)
13. Pessoa Jr., O.: Independent discoveries following different paths: The case of the law of spectral reversion (1848-59). In: Cattani, M.S.D., Crispino, L.C.B., Gomes, M.O.C., Santoro, A.F.S. (eds.) Trends in Physics, pp. 259–282. Livraria da Física, São Paulo (2009b)
14. Pessoa Jr., O.: The causal strength of scientific advances. Forthcoming. In: Videira, A.A.P., Krause, D. (eds.) Brazilian Studies in the History and Philosophy of Science. Boston Studies in the Philosophy of Science. Springer, New York (2010)
15. Radick, G.: Introduction: Why what if? Isis 99, 547–551 (2008)
16. Roese, N.J., Olson, J.M. (eds.): What Might Have Been: The Social Psychology of Counterfactual Thinking. Erlbaum, Mahwah (1995)
17. Woolf, H.: The beginnings of astronomical spectroscopy. In: Cohen, I.B., Taton, R. (eds.) L'Aventure de la Science, Mélanges Alexandre Koyré, Hermann, Paris, vol. 1, pp. 619–634 (1964)