Methodological and Epistemic Differences between Historical Science and Experimental Science*
Author(s): Carol E. Cleland
Source: *Philosophy of Science*, Vol. 69, No. 3 (September 2002), pp. 447-451
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/10.1086/342455
Accessed: 26-09-2017 00:01 UTC

# Methodological and Epistemic Differences between Historical Science and Experimental Science*

Carol E. Cleland†‡
University of Colorado, Boulder

Experimental research is commonly held up as the paradigm of "good" science. Although experiment plays many roles in science, its classical role is testing hypotheses in controlled laboratory settings. Historical science (which includes work in geology, biology, and astronomy, as well as paleontology and archaeology) is sometimes held to be inferior on the grounds that its hypothesis cannot be tested by controlled laboratory experiments. Using contemporary examples from diverse scientific disciplines, this paper explores differences in practice between historical and experimental research vis-à-vis the testing of hypotheses. It rejects the claim that historical research is epistemically inferior. For as I argue, scientists engage in two very different patterns of evidential reasoning and, although there is overlap, one pattern predominates in historical research and the other pattern predominates in classical experimental research. I show that these different patterns of reasoning are grounded in an objective and remarkably pervasive time asymmetry of nature.

**1. Introduction.** Experimental research is commonly held up as the paradigm of successful (a.k.a. good) science. The role classically attributed to experiment is that of testing hypotheses in controlled laboratory settings. Not all scientific hypotheses can be tested in this manner, however. Historical hypotheses about the remote past provide good examples. Al-

---

*Received September 2001; revised March 2002.

†Send requests for reprints to the author, Philosophy Department and Center for Astrobiology, 169 Hellems, Campus Box 232, University of Colorado, Boulder, CO 80309; e-mail: cleland@spot.colorado.edu.

‡I am indebted to Alan Franklin, Marc Lang, and Stephen Leeds for comments and helpful suggestions on an earlier draft of this paper. This research was supported in part by a grant from the National Aeronautics and Space Administration to the University of Colorado's Astrobiology Institute.

Philosophy of Science, 69 (September 2002) pp. 474–496. 0031-8248/2002/6903-0000$10.00

though fields such as paleontology and archaeology provide the familiar examples, historical hypotheses are also common in geology, biology, planetary science, astronomy, and astrophysics. The focus of historical research is on explaining existing natural phenomena in terms of long past causes. Two salient examples are the asteroid-impact hypothesis for the extinction of the dinosaurs, which explains the fossil record of the dinosaurs in terms of the impact of a large asteroid, and the "big-bang" theory of the origin of the universe, which explains the puzzling isotropic three-degree background radiation in terms of a primordial explosion. Such work is significantly different from making a prediction and then artificially creating a phenomenon in a laboratory.

Scientists are well aware of the differences between experimental and historical science vis-à-vis the testing of hypotheses. Indeed, it is sometimes a source of friction. Experimentalists have a tendency to disparage the claims of their historical colleagues, contending that the support offered by their evidence is too weak to count as "good" science. A telling example is the startling number of physicists and chemists who attack neo-Darwinian evolution on the grounds that it hasn't been adequately "tested."[1] The most sweeping condemnation of historical science, however, comes from Henry Gee, an editor of the prestigious science journal *Nature.* In Gee's words "they [historical hypotheses] can never be tested by experiment, and so they are unscientific. . . . No science can ever be historical" (2000, 5–8). In other words, for Gee, a genuine test of a hypothesis requires classical experimentation.

Philosophers of science have pretty much ignored the simmering controversy among scientists over the epistemic status of historical claims. Aware that experiment plays different roles in science, skeptical about the existence of a single method for all of science, and unable to provide an epistemically satisfying account of the rationality and objectivity of any scientific practice, philosophers have been reluctant to generalize, let alone make normative judgments, across different disciplines. Moreover, those philosophers who have cast an eye on historical research have spent most of their time on evolutionary biology and archaeology, where issues about teleology or human agency predominate. As a consequence, historical methodology is often characterized (e.g., Goode 1977; Kitcher 1993, 18–34) in terms of narrative histories. Analysis in terms of narrative histories does not, however, do justice to historical work in disciplines such as astronomy and geology that, like experimental physical science, do not involve even a *prima facie* appeal to purposes or ends. A more general understanding of the methodology of historical science and its differences from classical experimental science is badly needed.

---

[1] See, for instance, "A Scientific Dissent from Darwinism" (2001, 23), a statement that recently appeared in the *New York Review of Books* and was signed by approximately 100 (mostly physical) scientists; they listed their fields after their names.

One of the purposes of this paper is to sketch such an account. As we shall see, scientists engage in two very different patterns of evidential reasoning, and one of these patterns predominates in historical research and the other in classical experimental research. These differences in evidential reasoning lie at the heart of the charge that historical science is inferior to experimental science vis-à-vis the testing of hypotheses. But, as I shall also show, it is not an accident that historical research emphasizes one pattern and experimental research the other, nor is it an accident that some investigations utilize both. Using examples from a wide variety of scientific disciplines, I show that these differences in evidential reasoning are underwritten by an objective and pervasive feature of nature, namely, a time asymmetry of causation between present and past events, on the one hand, and present and future events, on the other. Because each practice is tailored to exploit the information that nature puts at its disposal, and the character of that information differs, neither practice may be held up as more objective or rational than the other.

## 2. "Classical" Experimental Science.

As Ian Hacking (1983, 149–166) has emphasized, experiment plays a variety of roles in science besides the testing of hypotheses. Nevertheless, there is little doubt that a significant portion of experimental work is devoted to testing hypotheses in controlled laboratory settings. My central concern in this paper is with this (what I call the "classical") role of experiment since it provides the paradigm to which historical research is often compared unfavorably. Due to space limitations, I shall focus on simple, idealized experiments, ignoring the complications (see Franklin 1999, 13–38) that occur when one is working with fancy equipment, dealing with experiments that cannot (for practical or theoretical reasons) be "repeated" under controlled conditions, etc.

The hypotheses investigated in classical experimental science postulate regularities among event-types.[2] A test condition C is inferred from the hypothesis and a prediction is made about what should happen if C is realized (and the hypothesis is true). This forms the basis for a series of controlled experiments.

In many of these experiments, C is held constant (repeated) while other experimental conditions are varied. When this activity is preceded by a failed prediction in an earlier experiment, it resembles the activity famously condemned by Popper (1963), namely, an *ad hoc* attempt to save a hypothesis from refutation by denying an auxiliary assumption.[3] But

---

[2] These regularities may be statistical, as in quantum mechanics, and they may or may not be causal; experimental scientists are interested in testing functional regularities (e.g., the Boyle-Charles' law for ideal gases), as well as causal regularities.

[3] I am using the term "auxiliary assumption" very broadly to include any assumption whose falsity could be used to salvage a hypothesis in the face of a failed prediction. Thus on my usage, auxiliary assumptions include assumptions about the particulars of

there is an alternative interpretation: It may be viewed as an attempt to protect the target hypothesis from false negatives. Now, admittedly, the distinction between protecting a hypothesis from falsification come-what-may and protecting it from misleading disconfirmations seems subtle. But it is not trivial. A *false* negative does not constitute a counter-example to a hypothesis. Moreover, insofar as an experimental test of a hypothesis involves an enormous number of auxiliary assumptions, the possibility of false negatives is very real. Hence it makes good sense to control for them. Each "control" represents a methodological rejection of an auxiliary assumption about a previous experimental situation. These methodological rejections of auxiliary assumptions amount to more than idle *a priori* speculations. They are incorporated into actual experiments. In other words, they aren't just *ad hoc* attempts to salvage a hypothesis.

Significantly, the same process of holding C constant while varying extraneous conditions often occurs upon a *successful* test of a hypothesis. In this case, however, the activity resembles the sort of activity acclaimed by Popper, viz., an attempt to falsify the target hypothesis. A little reflection, however, reveals that this can't be what is going on. In the first place, such tests are not "risky" in Popper's demanding sense. The hypothesis has survived similar tests and no one expects it to fail under slightly different conditions. Besides, even if it does, the hypothesis won't automatically be rejected. Viewed from this perspective, it seems more plausible to interpret the activity as an attempt to protect the target hypothesis from false positives. This interpretation is reinforced when one considers that C itself may be removed in order to evaluate whether extraneous conditions might have been responsible for the successful results of earlier experiments.

Let us pull this all together. Much of the activity that goes on in classical experimental science may be interpreted as attempts to protect the target hypothesis from misleading confirmations and disconfirmations. This is subtly but importantly different from attempting to falsify it, or to save it from falsification by *ad hoc* means. For the falsification attempts are aimed at auxiliary assumptions instead of the hypothesis under investigation. It is important to keep in mind that this characterization of classical experimental science is meant to capture an ideal, something to which experimental scientists aspire, even though they may fail to realize it; as noted earlier, it is sometimes difficult to repeat an experiment or to control for suspicious auxiliary assumptions.

My account of classical experimental science bears a superficial resem-

---

an experimental situation (e.g., instruments are working properly, there are no interfering conditions), as well as the more traditional theoretical assumptions about instrumentation, etc.

blance to Lakatos' (1970, 115–138) "sophisticated falsificationism." What I have in mind is nonetheless different. Lakatos was concerned with understanding how scientists protect their theories from refutations.[4] He paid little attention to the fact that scientists also protect their hypotheses from misleading confirmations. Experimentation does not grind to a halt with a successful experiment. Scientists continue to fiddle with extraneous conditions while repeating C. They also try removing C while holding extraneous conditions constant. These two activities play just as prominent a role in actual experimental practice as attempts to protect hypotheses from false negatives. When one considers the vast number of additional conditions (known and unknown) that might affect the outcome of an experiment independently of the truth of the hypothesis, all three of these activities make good sense. It is thus hardly surprising that experimental scientists engage in them when they can.

My sketch of classical experimental science has the advantage of reconciling the almost passionate devotion still expressed by many scientists for falsificationism with the puzzling fact that they rarely reject their target hypotheses in the face of failed predictions. They are engaging in systematic, extended experimentation that sometimes resembles an attempt to falsify a hypothesis and sometimes resembles an attempt to protect a hypothesis from falsification, but is really aimed at something quite different, namely, minimizing the very real possibility of misleading confirmations *and* disconfirmations in concrete laboratory settings. The historical tendency of philosophers to take the isolated experiment as the descriptive unit of experimental research has obscured this character of classical experimental science. For if one focuses on the construction and evaluation of a solitary experimental test of a target hypothesis, it seems as if researchers are either trying to falsify the hypothesis or save it from falsification. Viewed from a temporally extended perspective, however, things look quite different. The experimental evaluation of a hypothesis involves a series of experiments, each one designed in light of the results of previous experiments. In the face of an ostensibly disconfirming result, auxiliary assumptions are modified. Similarly, auxiliary assumptions are also modified in the face of an ostensibly confirming result, and the test condition itself is eventually removed. In the absence of a series of experiments of this character, most researchers are reluctant to submit their work to a peer reviewed science journal.

---

[4] Lakatos suggested that we think of scientific theories as divided into three components: a "hard core," a "protective belt," and a "positive heuristic." The positive heuristic tells scientists how to respond to failed predictions ("anomalies") by revising the protective belt of auxiliary assumptions while keeping the hard core (which often includes the hypothesis under examination) intact.

It is important to be clear about the nature of the interdependency among the experiments in an experimental program. The results of earlier experiments guide the design of later experiments. The 1976 Viking Lander missions to Mars provide a particularly salient illustration of what happens when scientists try to design an experimental program wholly in advance of any data. The Viking Landers were designed to perform robotic experiments testing the hypothesis that Mars had microbes living in its soil. Of particular interest are the "labeled release" (LR) experiments, which generated the most consistently positive results of the three classes of metabolic experiments; for more detail, see Klein et al. 1992, 1221–1233. The initial results of the LR experiment were positive. When the Martian soil sample in the test chamber was injected with a radioactively labeled bacterial nutrient solution it started evolving radioactive $^{14}CO_2$, just what one would expect from terrestrial soils containing microbes. Moreover, when the controlled experiment (determined in advance of any data) was performed and the Martian soil sample was "sterilized" by heating it to 160° C for 3 hours, the reaction stopped, strongly suggesting that the initial result really had been biological; this experiment amounted to a test for a false positive. But the data from the LR experiment were not yet all in. When the same soil sample was subjected to another experiment and given a "second helping" of radioactive nutrients several days later, after the initial reaction had leveled off, the anticipated burst of new activity (from hungry Martian microbes) failed to occur. Even more mysteriously, $^{14}CO_2$ left over from the earlier reaction began disappearing. Scientists were baffled. No one had anticipated results like these.

As the Viking Landers fell permanently silent, disagreement over the interpretation of the LR experiments became acrimonious, with alternative theoretical explanations (Oyama et al. 1976; Levin and Straat 1976) being advanced for the puzzling results. To this day, two of the original investigators (Levin and Straat) still insist that they found life. A majority of the scientific community, however, disagrees, contending that the Martian surface contains a strong oxidant. Various candidates, e.g., hydrogen peroxide (Oyama et. al. 1976), high oxidation states of iron (Tsapin et al. 2000), have been proposed for the oxidant. Nevertheless, there is no empirical evidence that the Martian surface is strongly oxidizing.[5] The basic problem remains. The LR experiments and their controls produced results that no one had anticipated and there was no possibility of performing suitable additional experiments to adjudicate among the many auxiliary assumptions that became the focus of heated speculation. In retrospect, it

---

[5] This is discussed in DiGregorio (1997, 167–174). A well known planetary scientist who does not wish to be in any way associated with Levin's position has also confirmed this in a private conversation.

was naïve of designers of the experiment to think that they could determine in advance of any experimental data what controls would be adequate to rule out false positives and false negatives.

Although my focus has been on the (idealized) practice of an individual researcher (and her co-workers) testing a hypothesis, I am not claiming that her decision to accept or reject a hypothesis is based solely upon her own experimental results, nor am I denying the important Kuhnian insight (1970) that many of the factors involved in a scientific community's decision to accept or reject a hypothesis are based upon sociological and psychological considerations. I am merely trying to characterize what the scientific community expects of a "good" experimental researcher when she goes about testing a hypothesis in her lab, and I am claiming that it is best understood in terms of an extended series of interdependent experiments, as opposed to the solitary experiment.

**3. Historical Science.** In light of the above discussion, let us turn to the practice of historical science. In the prototypical scenario, an investigator observes puzzling traces (effects) of long-past events. Hypotheses are formulated to explain them. The hypotheses explain the traces by postulating a common cause for them. Thus the hypotheses of prototypical historical science differ from those of classical experimental science insofar as they are concerned with event-tokens instead of regularities among event-types. This helps to explain the narrative character of many historical explanations. The complexity of the causal conditions and the length of the causal chain (connecting the cause to its current traces) bury the regularities in a welter of contingencies. Accordingly, it is hardly surprising that historical explanations often have the character of stories that, lacking reference to specific generalizations, seem inherently untestable. Nonetheless, it would be a mistake to conclude that hypotheses about the remote past can't be "tested."

Traces provide evidence for past events just as successful predictions provide evidence for the generalizations examined in the lab. Instead of inferring test implications from a target hypothesis and performing a series of experiments, historical scientists focus their attention on formulating mutually exclusive hypotheses and hunting for evidentiary traces to discriminate among them. The goal is to discover a "smoking gun."[6] A smok-

---

[6] I am appropriating this expression from historical scientists. Talk about a "smoking gun" is rampant in informal and popular works by historical researchers. As an example, in a discussion of the asteroid impact hypothesis for the extinction of the dinosaurs, James Powell refers to the "tiresome metaphor of 'smoking gun'" (1998, 115). It should be kept in mind, however, that historical scientists use the term very loosely, and in what follows, I am providing it with a technical meaning that doesn't always coincide with its use among historical scientists.

ing gun is a trace(s) that unambiguously discriminates one hypothesis from among a set of currently available hypotheses as providing "the best explanation" of the traces thus far observed.

Wegener's hypothesis of continental drift provides a salient illustration (for more detail, see Hallam 1973, 135–183). When it was first advanced in 1912, the hypothesis provided a unified account of a large number of puzzling features of the surface of Earth, e.g., the complementary shapes of the Atlantic coasts of Africa and South America, and similarities in geological formations and fossil records on opposite sides of the Atlantic Ocean. The competing hypothesis, the contractionist theory, held that Earth's crust moves only vertically, a result of gradual contraction as its interior cools. The contractionist hypothesis didn't provide as unified an account of the known traces as Wegener's hypothesis. But Wegener's hypothesis had a major defect. There was no known causal mechanism for horizontal continental motion. This was partially remedied in 1960 by Princeton geologist Harry Hess's hypothesis of sea floor spreading, which was formulated in light of new discoveries derived from advances in geological and geophysical oceanography; these discoveries included the global mid-oceanic ridge system, the apparent youth of the sea floor, and the circum-Pacific island arc-trench system with its numerous volcanoes and powerful earthquakes. The subsequent discovery of alternating bands of reversed magnetism, spreading out symmetrically on both sides of the volcanically active Mid-Atlantic Ridge provided the smoking gun for sea floor spreading. The magnetic stripes provided compelling evidence that the Earth's crust moves horizontally, carrying the continents along with it. But widespread scientific acceptance of sea floor spreading had to wait a few more years until a geophysical mechanism for sea floor spreading was worked out in the theory of plate tectonics.

The evolution of Wegener's hypothesis (from continental drift to sea floor spreading to plate tectonics) underscores some crucial characteristics of successful historical work. First, as Philip Kitcher (1989, 430–432) and others have emphasized, the best scientific explanations are unifying. In successful historical science the focus is on the unity that a hypothesis provides for a diverse body of puzzling traces. One of the great strengths of the plate-tectonic hypothesis is the large number of ostensibly independent phenomena that are united under a single explanation. Second, as Salmon (1984, ch. 5) has insisted, good explanations situate observations (in this case, traces) within a causal framework. This was the central problem with Wegener's original hypothesis. From the beginning, it provided more unity for the observed traces than the contractionist hypothesis. But there was no known physical mechanism for producing horizontal continental motion. Widespread acceptance of Wegener's ideas had to wait for the development of the geophysical theory of plate tectonics. In short, successful historical hypotheses explain traces by *unifying* them under a consistent *causal* story.

The greater and more diverse the body of traces that a postulated causal mechanism unifies, the better the explanation. A smoking gun is a trace (or subcollection of traces) that (so-to-speak) cinches the case for a particular causal story. A smoking gun does not, however, uniquely determine a hypothesis outside the context of a set of specific, competing hypotheses; it merely establishes that one of them is superior when it comes to causally explaining the traces thus far observed. Since many of the pertinent traces will be equally well explained by the competing hypotheses, a smoking gun will not include all of them. Furthermore, it is always possible that future observations or theoretical developments will depose a smoking gun and that another hypothesis (new or old) will attain the status of the best explanation. Moreover, some of the factors that go into transforming a trace(s) into a smoking gun for a particular hypothesis may be sociological or psychological, as well as theoretical and empirical; historical research is just as subject to Kuhnian influences as experimental research. In short, a "crucial" test of a hypothesis is no more possible in historical science than in experimental science. But this doesn't mean that historical researchers can't procure supporting evidence for their hypotheses any more than it means that experimentalists can't procure supporting evidence for their hypotheses.

The history of the debate over the extinction of the dinosaurs provides a good illustration of the crucial role played by a smoking gun in adjudicating among rival hypotheses. Prior to 1980 there were many different explanations for the demise of the dinosaurs, including contagion, global climate change, volcanism, and asteroid impact. Luis and Walter Alvarez's discovery (1980) of unusually high concentrations of the element iridium in the K-T (Cretaceous-Tertiary) boundary (ten parts per billion vs. background levels of about three-tenths of a part per billion) focused attention on volcanism or an asteroid impact, since these were the only plausible mechanisms for the presence of so much iridium in a thin layer of Earth's crust; higher concentrations exist in Earth's mantle and in asteroids. The subsequent discovery (Bohor, Foord et al. 1984) of extensive deposits of shocked quartz in the K-T boundary clinched the case for most geologists because, as was soon established (Alexopoulos et al. 1988), shocked quartz of the sort found in the K-T boundary is not associated with volcanism; it is found in only two places on Earth, craters produced by asteroid impacts and the sites of nuclear explosions. Geologists began searching for a crater of the right size and age; they eventually found Chicxulub Crater, which overlaps the Yucatan peninsula and is about 65 million years old and 200–300 km across. It was generally admitted, however, that failure to find the crater wouldn't count heavily against the hypothesis since the active geology of Earth might have obliterated all traces, particularly if the impact had occurred in deep ocean. The combination of iridium and shocked quartz were enough to convince most members of the geological community that a large asteroid had slammed into Earth 65 mya.

While it was widely conceded that the anomalous iridium and shocked quartz provided a "smoking gun" for the impact of a large asteroid, paleontologists (e.g., Clemens et al. 1981) remained unconvinced that the impact explained the extinction. More research was needed to establish a causal link between the impact and the extinction. The extinction event had to be word-wide and geologically instantaneous. The available paleobiological data were very imprecise, unable to distinguish global events occurring within a period of a few years from events that took place at different times throughout intervals of 10,000 to perhaps 500,000 years. Peter Ward's (1990) meticulous studies of the fossil record of the ammonites (which is much more extensive than that of the dinosaurs) were pivotal in marshalling support among paleontologists for the claim that the Cretaceous extinctions were rapid and global. The asteroid-impact hypothesis became the widely accepted explanation for the extinction of the dinosaurs. For of the available hypotheses, it provided the greatest causal unity to the diverse and puzzling body of traces (fossil record of the dinosaurs, fossil record of the ammonites, etc., and iridium anomaly, shocked quartz, Chicxulub Crater, etc.).

So how does prototypical historical research compare to classical experimental research? Historical scientists are just as enamored with falsificationism as experimentalists; as three eminent geologists counsel in a recent textbook discussion of the extinction of the dinosaurs, "a central tenet of the scientific method is that hypotheses cannot be proved, only disproved" (Kump, Kasting, and Crane 1999, 201). Nevertheless, there is little in the practice of historical science that resembles what is prescribed by falsificationism. The search for a smoking gun is a search for *supporting* evidence for a hypothesis. Moreover, a trace that comprises a smoking gun may not (considered just in itself) provide evidence against a competing hypothesis. In the context of the contagion hypothesis, for instance, the presence of shocked quartz in the K-T boundary seems merely irrelevant to the extinction of the dinosaurs. This is not to deny that a search for a smoking gun may turn up evidence against a hypothesis. The point is that it need not. One can establish that a hypothesis provides a more unified causal explanation than another without procuring direct evidence against the latter.

Furthermore, there is little in prototypical historical research that resembles attempts to control for false positives and false negatives. Generating competing hypotheses is not analogous to entertaining auxiliary assumptions. The truth of an auxiliary assumption is independent of the truth of the target hypothesis, which is why a target hypothesis may be saved (in the face of a failed prediction) by rejecting an auxiliary assumption. In contrast, competing historical explanations are incompatible. Moreover, if there is a smoking gun for an historical hypothesis, it already exists. Smoking guns are not produced by systematically varying some conditions while holding others constant. They are uncovered in the uncontrolled world of nature by fieldwork.

This is not to deny that prototypical historical research often involves laboratory work. It is important, however, to be clear about what is actually being investigated in the lab. Most often it is the evidentiary traces, which frequently require sharpening or analysis in order to be identified and properly interpreted. As an example, speculation that life goes back 3.8 billion years rests upon laboratory analysis of carbon isotope ratios in grains of rock as small as 10 μm across weighing only $20x10^{-15}$ g (Mojzsls et al. 1996, 56). Sometimes, however, it is a hypothesis bearing a tenuous logical relation to the target hypothesis. A good example is the 1953 Miller-Urey experiments (Miller 1953), which were touted as an experimental test of the hypothesis that life on Earth originated in a "primordial soup" but really amounted to a test of the supposition that some of the most basic building blocks of life (amino acids) can be produced by electrical discharges on a mixture of methane, hydrogen, ammonia, and water. This is not to deny that the Miller-Urey experiments provide some support for the hypothesis that life on Earth began in an electrified primordial soup. The support is not analogous, however, to that offered by a controlled experiment for a target hypothesis. This is brought out clearly by the fact that scientists now know that amino acids can be produced under a wide variety of different conditions (Chyba and Sagan 1992). Furthermore, most scientists currently believe that the origin of life on Earth is incompatible with the conditions of the Miller-Urey experiment; it is thought that the Earth's early atmosphere did not contain abundant methane or ammonia, and that life probably began on or under the sea floor, near a volcanic vent (Pace 1991). The problem with the Miller-Urey experiment is that the logical relation between the hypothesis actually tested in the lab and the target hypothesis (about the origin of life on Earth) is very convoluted, winding through numerous highly speculative assumptions, ranging from conditions on early Earth to biochemical possibilities for producing amino acids and whole cells. This is fairly typical of experimental work associated with hypotheses about the remote past.

I have been describing two different patterns of evidential reasoning, put succinctly, from causes (test conditions) to effects, with the concomitant worries about ruling out false positives and false negatives, and from effects (traces) to causes, with the concomitant worries about ruling out alternative explanations. As we have seen, although one pattern of reasoning predominates in historical research and the other in experimental work, there is overlap. Historians sometimes reason like experimentalists, and vice versa. This is not an accident. Which pattern of reasoning a scientist employs depends upon her epistemic situation, and although the epistemic situation of an historian typically differs from that of an experimentalist, there are some notable exceptions.

Nature sometimes repeats herself, presenting researchers with multiple examples of a given type of historical phenomenon. In such cases, the body of evidence available to researchers may resemble what an experimental program would provide, and they may exploit this fact by formulating generalizations analogous to those formulated by experimentalists. Stars known as "Cepheid variables" provide a perspicuous example. The length of the period of a Cepheid variable (time interval between two successive states of maximal brightness) is lawfully correlated (the Leavitt-Shapley law) with the magnitude of its intrinsic luminosity. Although one cannot perform controlled experiments on Cepheid variables, one can search for them with powerful telescopes and ascertain whether those that are discovered obey the Leavitt-Shapley law. Insofar as quite a few Cepheid variables have been identified, the body of this work resembles what might be achieved by controlled experimentation.

There are important differences, however. One cannot produce Cepheids with specific periods (the "test condition") at will, let alone control for possibly relevant extraneous conditions. Perhaps the Cepheid variables thus far observed are unusual in some special way. An astronomer may even suspect that a certain stellar condition is pertinent. But she can't remove or introduce it. She can only patiently scan the heavens for its traces. Viewed from this perspective, she is in the same position as a paleontologist hunting for evidence that the great Permian extinction was also caused by the impact of a large asteroid, the main difference being the lack of much evidence—good continuous exposures of rock—dating from the end of the Permian. The point is, without the ability to manipulate suspect conditions, one is at the mercy of what nature just happens to leave in her wake; sometimes she is generous and sometimes she is stingy, but the bottom line is that you can't fool with her. When nature is generous, however, there is little doubt that the body of evidence used to support a hypothesis may more closely resemble that provided by classical experimental research than by prototypical historical research, with its emphasis on finding pivotal evidence—a smoking gun—for discriminating a hypothesis from its competitors.

Just as historians sometimes reason like experimentalists, so experimentalists sometimes reason like historians. In the face of a failed experimental test of a target hypothesis, experimentalists entertain different possibilities for denying auxiliary assumptions, and their reasoning resembles that of a historian choosing from among alternative explanations for puzzling traces. But the subsequent work (removing the suspect condition while holding other conditions constant) bears little resemblance to what goes on in an historical search for a smoking gun. Historical researchers searching for a smoking gun are stuck with what nature has already provided; they must be clever enough to ferret out incriminating evidence that is often well hidden in the messy, uncontrollable world of nature. Similarly, like historians, experimentalists are sometimes faced with unexpected and puzzling effects. Such laboratory phenomena may be produced by accident, during a test of a hypothesis, or deliberately, out of curiosity. For, as Hacking discusses (1983, 155–159), scientists often play around with their equipment to see what it will do, sometimes just for the sake of generating interesting phenomena and other times because they have a vague suspicion that they wish to explore. A well-known example of the latter is Faraday's work on electromagnetism; he suspected that magnetic fields could induce electric effects, and deliberately searched for and found evidence of this. Henri Becquerel's discovery of radiation provides a salient example of the accidental discovery of a new phenomenon; he discovered an intense image on a photographic plate that had been exposed to uranium salts on an overcast day and then haphazardly tucked away in a dark drawer. When faced with a new phenomenon, laboratory scientists explore it, trying to reproduce it and speculating about its cause. That is, like historical scientists, they generate competing hypotheses about its source and search for additional evidence.

Once experimentalists have determined that a phenomenon is repeatable, however, they begin formulating specific generalizations about it and testing them individually by manipulating test conditions in the manner of classical experimental science. Tom Cech's Nobel Prize winning discovery that RNA can behave as a catalyst (see Zaug and Cech 1986) provides a good example of this transition. While working on an unrelated hypothesis, he was surprised to discover catalytic products in a solution of RNA that supposedly contained no proteins. He formulated competing explanations for this puzzling phenomenon (e.g., the solution was contaminated by proteins after all, RNA can behave as a catalyst), and then proceeded to test each of them in the laboratory in the classical manner by fiddling with suspect auxiliary assumptions (controlling for unsuspected protein contamination) and removing the test condition (the RNA) and seeing whether he still got the catalytic products.

In summary, although there is overlap, there are nonetheless fundamental methodological differences between historical science and experimental science vis-à-vis the testing of hypotheses. These differences in methodology reflect the fact that experimentalists and historians typically find themselves in very different epistemic situations. Experimentalists are primarily concerned with evaluating repeatable generalizations. Their research is focused on generating predictions from a single (sometimes complex) hypothesis, and manipulating repeatable test conditions in a lab while controlling for extraneous factors that might produce false positives and false negatives. Scientists engaged in prototypical historical work, on the other hand, are primarily concerned with evaluating hypotheses about particular past events. They cannot reproduce these events in a lab. They can, however, look for present-day traces of them, and search for a smoking gun that unambiguously sets apart one hypothesis as the best among the currently available explanations for the traces thus far observed. These differences in methodology do not, however, support Gee's charge that prototypical historical research is inferior to classical experimental work. For, as we shall see, the patterns of evidential reasoning upon which they are founded are epistemically underwritten by a corresponding difference in nature, namely, a time asymmetry of causation.

**4. The Asymmetry of Overdetermination.** Many events (broadly construed to include states) are causally connected in time in a strikingly asymmetric manner. Consider, for instance, the shattering of a window by a baseball. It is easy to explain the aftermath (a baseball on the floor of the bedroom and many pieces of glass strewn about on either side of an empty window frame) in terms of a single local event, namely, the impact of the baseball on a windowpane. But now run the process backwards in time. Dozens of widely scattered pieces of glass suddenly twitch and then are rapidly launched into the air, converging on the empty window frame. Meanwhile the baseball also begins moving. It too is ejected into the air, and flies through the empty window frame just before all the pieces of glass reach it. As the baseball clears the frame, each piece of glass simultaneously finds a spot that is precisely its shape and size. The result is an utterly seamless pane of glass. From this time reversed perspective, the events involved in the shattering of the window seem baffling, requiring the concerted but independent action of a baseball and many separate pieces of glass. And, indeed, we never see such sequences in nature: broken windows don't suddenly reconstitute themselves from widely scattered pieces of glass.

That this time asymmetry of nature applies to more than broken windows can be appreciated by considering the difficulty involved in committing the perfect crime. Footprints, fingerprints, particles of skin, disturbed dust, and light waves radiating outward into space must be eliminated. Moreover, it isn't enough to eliminate just a few of these traces. Anything you miss might be discovered by a Sherlock Holmes and used to convict you. Finally, each trace must be independently erased. You cannot eliminate the footprints by removing particles of skin or, for that matter, one footprint by removing a different one. In stark contrast, preventing a crime from occurring is easy: don't hit the baseball; don't fire the gun. In other words, erasing all traces of an event *before* it occurs is much easier than erasing all traces of it *after* it occurs.

David Lewis (1991, 65–67) has dubbed this feature of the world "the asymmetry of overdetermination." The basic idea is that events leave widespread, diverse effects (such as pieces of glass, particles of skin, etc.). Any one of a large number of contemporaneous, disjoint combinations of these traces is sufficient (given the laws of nature) to conclude that the event occurred. One doesn't need every shard of glass in order to infer that a window broke. A surprisingly small number of appropriately dispersed fragments will do. The overdetermination of causes by their effects makes it difficult to fake past events by engineering the appropriate traces since there will typically be many other traces indicating fakery. Similarly, as Lewis notes (1991, 66), there will usually be traces indicating that a questionable trace wasn't faked. Good criminal investigators are very aware of such possibilities: Is the handwritten ransom note left at the Ramsey residence genuine or designed to throw detectives off the trail of the real murderer? This is not to deny that traces may be so small, far flung, or complicated that no human being could ever decode them. All that is required for Lewis' thesis of overdetermination, however, is that they exist.

In contrast, the causal predecessors of an event rarely overdetermine it. A good example is a short circuit that "causes" a house to burn down. Take away the short circuit and the house wouldn't have burned down; the short circuit "triggered" the fire. But the short-circuit isn't sufficient for the occurrence of a fire. Many other factors are also required, e.g., the presence of flammable material, absence of sprinklers. The absence of any one of these additional factors would also have prevented the fire, even supposing that the short circuit had occurred. In other words, the *total* cause of the fire includes more than the short circuit. Local events (such as the short circuit) which are normally identified as the causes of later events (a house burning down) *under*determine them; considered just in themselves, they are not enough (even given the laws of nature) to guarantee that their reputed effects occur. This is the other side of the asymmetry of overdetermination. While erasing all traces of an event requires many separate interventions, a single intervention (and there are usually numerous possibilities) is normally enough to prevent it from occurring in the first place.

The underdetermination of effects by their putative causes is both epistemic and causal; the fire is epistemically *and* causally underdetermined by the short circuit. The overdetermination of causes by their effects, on the other hand, is (strictly speaking) only epistemic. Although one may *infer* that the window broke from any one of a large number of subcollections of shards of glass, the shards are not part of the *cause* of the breaking of the window; they are its effects. But although effects cannot causally overdetermine causes, there is no reason why causes couldn't causally overdetermine effects, and indeed this sometimes happens. The well-worn example is a firing squad, where several bullets simultaneously pierce the victim's heart. Their scarcity underscores the fact that the asymmetry of (epistemic) overdetermination is ultimately founded on a time asymmetry of nature.

There is little agreement about the physical source of the asymmetry of overdetermination. Examples such as the shattering of a window are commonly attributed to the second law of thermodynamics, which (statistically interpreted) says that the natural tendency of physical systems is to move from more to less ordered states. The asymmetry of overdetermination extends, however, to phenomena that do not obviously admit of a thermodynamic explanation. Consider Popper's (1956, 538) well-known example of dropping a stone into a still pool of water. Expanding concentric ripples spread outwards from the point of impact. It is easy to explain these ripples in terms of a stone entering the water at a small region of the surface of the pool. Indeed, one can pinpoint where the stone entered by examining a small segment of the pool's surface. But now consider eliminating all traces of the impact. An enormous number of separate and independent interventions are required all over the surface of the pool. Similarly, try explaining the time reversed process. Ripples, which expanded outwards from the point of impact, now contract inwards to the point of impact. But there is no center of action to explain the simultaneous and coordinated behavior of the individual water molecules involved. From this time reversed perspective, the contracting concentric waves seem to be a miracle; we can understand them causally only by running the process in the other direction (forwards) in time.

Popper's example of the stone and the pool illustrates a phenomenon known as the asymmetry of radiation. Although the asymmetry of radiation is traditionally associated with electromagnetic radiation (light, radio waves, etc.), it characterizes all wave-producing phenomena, including disturbances in water and air. The asymmetry originates in the fact that waves (whether of water, sound, light, etc.) invariably spread outwards rather than inwards. The asymmetry of radiation is thus very widespread in nature.

It is tempting to suppose that the asymmetry of radiation is connected to the asymmetry of thermodynamics. For our purposes, however, it doesn't matter how the asymmetry of radiation is related to that of thermodynamics. What matters is that both asymmetries are objective physical features of the universe. There is widespread agreement about this. Indeed, the only one who comes close to denying it is Price, who contends (1996, 155–161) that the asymmetries characterize only the macroscopic realm of human experience. But even Price admits (e.g., 1996, 76) that it is a physical fact that our universe contains numerous, large, organized concentrations of energy in the form of galaxies, stars, rocks, pools, etc., which give rise to macroscopic asymmetries of radiation and thermodynamics. This is enough for our purposes because this is the level at which scientific investigations (in the field or lab) are ultimately carried out.

This brings us back to the practice of science. Historical researchers investigating particular past events cannot test their hypotheses by performing controlled experiments. But this doesn't mean that they cannot procure empirical evidence for them. Because of the asymmetry of overdetermination, there are usually an enormous number of subcollections of the effects of a past event that are individually sufficient (given the right theoretical assumptions) to infer its occurrence. The trick is finding them. Many of these overdetermining traces (e.g., a splinter of glass, a footprint) occupy small, local regions of space, bringing them within the limited range of human sensory experience. This places scientists investigating the remote past in the position of criminal investigators.[7] Just as there are many different possibilities for catching a criminal, so there are many different possibilities for establishing what caused the demise of the dinosaurs, the origin of the universe, etc. Like criminal investigators, historical scientists collect evidence, consider different suspects, and follow up leads. More precisely, they postulate differing causal etiologies for the traces they observe, and then try to discriminate from among them by searching for a "smoking gun"—a trace(s) that identifies the most plausible culprit among the primary suspects. Unlike stereotypical criminal investigations, however, a smoking gun for a historical hypothesis merely picks out one hypothesis as providing the best explanation currently available; it need not supply direct confirming evidence for a hypothesis independently of its rivals.

Although Lewis characterizes the asymmetry of overdetermination in terms of sufficiency, it could turn out to be a probabilistic affair, with the ostensibly overdetermining subcollections of traces lending strong but, nevertheless, inconclusive support for the occurrence of their cause. Like the determinism in Lewis' original version, the probabilistic support offered by collections of traces for hypotheses would be an objective feature of the world. A body of traces would make its cause highly probable independent of human knowledge, interests, or concerns. Researchers in possession of a decisive body of traces might fail to recognize its significance; they might not have formulated the correct hypothesis or they might lack the theoretical understanding necessary to connect the traces

---

[7] I am indebted to Sheralee Brindell, a self-professed fan of "cheesy" detective novels, for emphasizing the similarities between historical research and a criminal investigation.

with the correct hypothesis. Developing a satisfactory probabilistic interpretation of the asymmetry of overdetermination is beyond the scope of this paper. Needless to say, there are many issues that need to be addressed, e.g., are the probabilities relative to more global background conditions (thus requiring ceterus paribus clauses) or are they absolute? In any case, however, human experience is consistent with either a probabilistic or a deterministic interpretation of the asymmetry of overdetermination. Just as experimental work is irremediably fallible (even in a deterministic world) because of the uneliminable threat of unknown interfering conditions, so the traces uncovered by field work are never enough to conclusively establish the occurrence of a postulated past event, perhaps because we haven't discovered enough of them or perhaps simply because there are no causally sufficient subcollections. But even supposing that it is probabilistic, the asymmetry of (quasi) overdetermination helps to explain the methodology of historical research. It tells us that a strikingly small subcollection of traces is enough to substantially increase the probability that a past event occurred, and that there are likely to be many such subcollections. The existence of so many different reliable possibilities for identifying past events provides the rationale for the historical scientist's emphasis on finding a smoking gun.

The empirical support currently enjoyed by the Alvarez hypothesis provides an excellent illustration of the evidential resources available to historical researchers. In addition to shocked quartz and iridium, the K-T boundary sediments have yielded microspherules (unusual microscopic droplets of glass thought to have been produced by the rapid cooling of molten rock that splashed into the atmosphere during the impact), fullerenes containing extraterrestrial noble gases, and extensive deposits of soot and ash (measuring tens of thousands of times higher than normal levels). Other than a catastrophic extraterrestrial impact, there is no known mechanism for producing these unusual and disparate traces. Had iridium been absent from the K-T boundary, scientists would still have excellent grounds (given currently accepted theoretical considerations) for concluding that a giant asteroid slammed into Earth 65 mya. The increasing number of scientists who believe that the great Permian extinction was also caused by an asteroid impact underscores this point. Sediments at the Permian-Triassic boundary lack an iridium anomaly but they contain a diversity of other traces (most notably, extraterrestrial fullerenes) that are difficult to explain in terms of any other known mechanism (Becker et al. 2001). Similarly, recent high-resolution studies of the fossil record of multiple species, e.g., studies of fossilized plant remains (see Archibald 1996), are in agreement with Ward's ammonite studies about the rapidity and timing of the K-T extinctions. Even in the absence of Ward's work, these studies would provide strong support for a causal connection between the extinctions and the impact. In other words, the body of traces thus far uncovered is more than enough to unambiguously establish the superiority of the Alvarez hypothesis *vis-à-vis its competitors,* and the existence of this large and diverse body of supporting evidence is a consequence of an objective physical feature of the universe, namely, the asymmetry of overdetermination.

The hard part of historical research is finding traces that are capable of unambiguously discriminating among rival hypotheses. But even when much time has passed and the traces have become very attenuated there is always the possibility of discovering a smoking gun for a long past event. In some cases, one may be able to infer a smoking gun directly from the hypothesis under investigation. A good example is the theoretical work done in the mid-1960s by Robert Dicke and his team of Princeton physicists (see Kaufmann 1977, 267–277). They predicted that if the big-bang theory for the origin of the universe was true the universe should contain an isotropic, microwave background radiation a few degrees above absolute zero. Wilson's and Penzias' subsequent discovery of the mysterious 3-degree background radiation was taken as providing pivotal evidence for the big-bang theory over the steady state theory. Sometimes, however, one just gets lucky and stumbles over a smoking gun as did the Alvarezes in the case of the asteroid-impact hypothesis for the extinction of the dinosaurs; the existence of iridium and shocked quartz in the K-T boundary was not predicted in advance of its discovery. In this context, I want to emphasize that I am not claiming that every true historical hypothesis has a smoking gun. It is unlikely but nonetheless possible for an event to leave no traces; prime candidates are events occurring before the big-bang of cosmology. Moreover, with the passage of time, traces of events become more and more attenuated, and eventually may disappear altogether. Alternatively, they may still be present but very degraded. Finding them may require advances in technology. The discovery of the 3-degree background radiation depended upon the development of very sensitive antennas for communicating with satellites. Similarly, a particle accelerator (cyclotron) was used in the discovery of the iridium in the K-T boundary. Finally, with new evidence and new explanatory hypotheses, the status of a trace as a smoking gun may change. Nevertheless, one can never rule out the possibility of finding a smoking gun, and this is a consequence of an objective fact about nature, namely, most past events are massively overdetermined by localized present phenomena.

The on-going debate over whether meteorite ALH84001 contains fossilized Martian life provides an excellent overview of how historical researchers exploit the asymmetry of overdetermination in their efforts to adjudicate among competing explanations. Discovered in the ice fields of Antarctica in 1984 and subsequently determined to be from Mars, ALH84001 contains a number of unexpected and very puzzling structural and chemical features, including layered carbonate globules, iron sulfides, PAHs, sausage-shaped (a.k.a. bacteria-shaped) structures, and “prismatic” (single domain) magnetite. These features would not be expected to form under a single combination of geochemical conditions. Terrestrial bacteria, however, produce them all. McKay, Gibson, and their team at Johnson Space Center postulated that they had been caused by ancient Martian microbes (see McKay, Gibson et al. 1996). The response of the scientific community was cautious. While conceding that the evidence was intriguing, most scientists nonetheless rejected the claim that it provided strong support for Martian microbial activity. More research was needed.

Samples of ALH84001 (which weighed only 1.9 kg) were carefully doled out to researchers with access to sophisticated analytical techniques and equipment. As of this writing, ALH84001 is one of the most intensely studied rocks on Earth. The results of many of these investigations have been equivocal. As an example, analysis of some of the features (e.g., the magnesium, calcium, and iron compositions found in the carbonate globules) suggests that they were formed at temperatures much too high (> 650°C) for life as we know it (Kerr 1997), whereas analysis of other features (analysis of two adjacent pyroxene grains in the “crushed zone” of the meteorite) suggest otherwise (Kirschvink et al. 1997). The feature in ALH84001 that has attracted the most attention, however, is the prismatic magnetite. Using electron microscopy and energy dispersive spectroscopy (McKay et al. 1996, 926), McKay and his team discovered tiny (20–100 nm) crystals of magnetite in the carbonate globules. Approximately 25% of these crystals have the distinctive feature of being hexagonal in cross section and virtually free of chemical impurities. There are no known nonbiotic processes that produce magnetite of this shape or chemical purity. It is indistinguishable, however, from that produced intracellularly by some magnetotactic terrestrial bacteria (see Thomas-Keprta et al. 2000). It is for this reason that (in an unpublished 1998 “Position Paper” widely distributed among scientists) McKay and his team referred to the prismatic magnetite as a “smoking gun” (p. 5) for Martian life.

Nevertheless, the scientific community remains unconvinced. Many (e.g., Kerr 2001, 1876) feel that more work needs to be done to exclude the possibility of a nonbiotic origin for the prismatic magnetite. Another concern is the lack of new evidence for the biotic hypothesis. If researchers utilizing sophisticated analytical techniques had discovered additional traces that were as difficult to explain nonbiotically as the prismatic magnetite, the case for fossilized Martian life would be greatly enhanced, just as the discovery of shocked quartz (which followed the discovery of iridium) in the K-T boundary was pivotal in convincing skeptics that the Alvarez hypothesis was more plausible than its rivals. Given the overdetermination of the past by the localized present, one would expect such traces to exist if the biotic hypothesis were true. Despite much effort with very sophisticated equipment, however, they haven't been found. Until scientists are convinced that chemically pure, prismatic magnetite cannot be produced by nonbiotic processes or have found additional traces that are as difficult to explain in terms of nonbiotic processes, or, alternatively, a unified, physically consistent, nonbiotic explanation is found for all the perplexing features in ALH84001, the fossilized Martian life hypothesis will remain viable but highly controversial. This is just what one would expect of an historical hypothesis with suggestive but nonetheless ambiguous evidential support.

Just as the overdetermination of past events by localized present events explains the practice of prototypical historical science, so the underdetermination of future events by localized present events explains the practice of classical experimental science. The hypotheses tested in classical experimental science are generalizations (conjectured laws and theories), as opposed to statements about particular events. Nevertheless, like the traces uncovered by historical research, the conditions manipulated in the lab during classical experimental research are local phenomena. But they are only partial causes (triggers) of what subsequently occurs. Accordingly, there is a need to ferret out and control for additional factors that are relevant to the total causal situation; otherwise, the ostensible confirmations and disconfirmations of the target hypothesis may be mistaken. This is why experimental scientists spend so much time systematically rejecting auxiliary assumptions that they accepted in earlier experiments. They are not trying to disprove their hypotheses or to save them from falsification. They are trying to identify false positives and false negatives, which are always a threat since localized events typically underdetermine their effects. In brief, classical experimental research is best interpreted as an attempt to circumvent the inevitable underdetermination of experimental results by test conditions.

**5. Conclusion.** There are fundamental methodological differences between prototypical historical science and classical experimental science vis-à-vis the testing of hypotheses. These differences represent different patterns of evidential reasoning—patterns that are designed to exploit different sides of an asymmetry of epistemic overdetermination. Insofar as they are concerned with identifying particular past causes of current phenomena, historical researchers cannot directly test their hypotheses by means of controlled experiments. They can, however, proliferate alternative explanations for the traces they observe and then search for a smoking gun to discriminate among them. The overdetermination of the past by the localized present provides the rationale for this work, ensuring that the probability of finding such traces is fairly high. In contrast, in their efforts to secure trustworthy laboratory evidence for the generalizations that lie at the heart of their investigations, classical experimentalists face the underdetermination of the future by the localized present. This explains why they spend so much time systematically controlling for assumptions that they accepted in earlier experiments; they are trying to eliminate false positives and false negatives. In other words, although experimentalists and historical researchers both ultimately infer causes from effects, the evidential relations that they exploit are different and this difference reflects the fact that events are causally connected in time in an asymmetric manner. It is important to keep in mind that nothing that I have said contravenes the tentative nature of the support offered by empirical evidence for a hypothesis, nor challenges the claim that many of the factors that play a central role in the acceptance of a hypothesis are psychological and political. But this is just as true of experimental science as it is of historical science. The point is, although there are fundamental methodological differences between historical research and classical experimental research, these differences do not support the contention that prototypical historical science is epistemically inferior to classical experimental science. Indeed, I suspect that the differences that I have identified could be reconstructed as inferences in accordance with Bayes' theorem. Unfortunately, however, it is beyond the scope of this paper to provide a theory of confirmation.

## REFERENCES

Alexopoulos, J. S., R. A. F. Grieve et al. (1988), "Microscopic Lamellar Deformation Features in Quartz: Discriminative Characteristics of Shock-Generated Varieties", *Geology* 16: 796–799.

Alvarez, L. W., W. Alvarez et al. (1980), "Extraterrestrial Cause for the Cretaceous-Tertiary Extinction", *Science* 208: 1095–1108.

Archibald, J. David (1996), Dinosaur Extinction and the End of an Era: What the Fossils Say. New York: Columbia Press.

Becker, Luann, Robert J. Poreda et al. (2001), "Impact Event at the Permian-Triassic Boundary: Evidence from Extraterrestrial Noble Gases in Fullerenes", Science 291: 1530–1533.

Bohor, B., E. E. Foord et al. (1984). "Mineralogic Evidence for an Impact Event at the CretaceousTertiary Boundary," Science 224: 867–869.

Chyba, Christopher F. and Carl Sagan (1992), "Endogenous Production, Exogenous Delivery, and Impact-shock Synthesis of Organic Molecules: An Inventory for the Origins of Life." Nature 355: 125–131.

Clemens, W. A., J. David Archibald et al. (1981), "Out with a Whimper Not a Bang," *Paleobiology* 7: 293–298.

DiGregorio, Barry (1997), *Mars: The Living Planet.* Berkeley: Frog, Ltd.

Franklin, Allan (1999), *Can That Be Right?.* Dordrecht: Kluwer Academic Publishers.

Gee, Henry (2000), *In Search of Deep Time.* New York: The Free Press.

Goode, Terry (1977), "Explanation, Expansion, and the Aims of Historians: Towards an Alternative Account of Historical Explanation." *Philosophy of the Social Sciences* 7: 367–384.

Hacking, Ian (1983), *Representing and Intervening.* Cambridge: Cambridge University Press.

Hallam, Anthony (1973), *A Revolution in the Earth Sciences.* Oxford: Oxford University Press.

Kaufmann, William (1977), *The Cosmic Frontiers of General Relativity.* Boston: Little, Brown, and Company.

Kerr, Richard A. (1997), "Martian 'Microbes' Cover their Tracks", *Science* 276: 30–31.

———(2001), "Are Martian 'Pearl Chains' Signs of Life", *Science* 291: 1875–1876.

Kirschvink, Joseph L., Altair T. Maine et al. (1997), Paleomagnetic Evidence of a Low-Temperature Origin of Carbonate in the Martian Meteorite ALH84001", *Science* 275: 1629–1633.

Kitcher, Philip (1989), "Explanatory Unification and Causal Structure," in Philip Kitcher and Nathan Salmon (eds.), *Scientific Explanation.* Minneapolis: University of Minnesota Press, 410–505.

———(1993), *The Advancement of Science.* Oxford: Oxford University Press.

Klein, Harold P., Norman H. Horowitz, and Klaus Biemann (1992), "The Search for Extant Life on Mars", in H. H. Kieffer and B. M. Jakosky et al. (eds.), *Mars.* Tucson: University of Arizona Press, 1221–1233.

Kuhn, Thomas (1970), *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kump, Lee R., James F. Kasting, and Robert G. Crane (1999), *The Earth System.* New Jersey: Prentice Hall.

Lakatos, Imre (1970), "Falsification and the Methodology of Scientific Research Programs", in Imre Lakatos and Allan Musgrave (eds.), *Criticism and the Growth of Knowledge.* Cambridge: Cambridge University Press, 91–196.

Levin, G. V. and P. A. Straat (1976), "Recent Results for the Viking Labeled Release Experiment on Mars", *Journal of Geophysical Research* 82: 4663–4667.

Lewis, David (1991), "Counterfactual Dependence and Time's Arrow", reprinted in Jackson (ed.), *Conditionals.* Oxford: Oxford University Press, 46–75.

McKay, David S., Everett K. Gibson et al. (1996), "Search for Past Life on Mars: Possible Relic Biogenic Activity in Martian Meteorite ALH84001", *Science* 273: 924–930.

McKay, David S., Everett Gibson, and Kathie Thomas-Keptra (1998), "Possible Evidence for Life in ALH84001", Position Paper, NASA.

Mojzsls, S. J, G. Arrhenius et al. (1996), "Evidence for Life on Earth before 3,800 Million Years Ago", *Nature* 384: 55–59.

Miller, Stanley L. (1953), "A Production of Amino Acids under Possible Primitive Earth Conditions", *Science* 117: 528–529.

Oyama, Vance I., and B. J. Berdhal (1976), "The Viking Gas Exchange Experiment Results from Chryse and Utopia Surface Samples", *Journal of Geophysical Research* 82: 4669–4675.

Pace, Norman R. (1991), "Origin of Life—Facing up to the Physical Setting", *Cell* 65: 531–533.

Popper, Karl (1956), (letter), *Nature* 177: 538.

———(1963), "Science: Conjectures and Refutations", in Martin Curd and Jan Cover (eds.), *Philosophy of Science.* New York: Norton, 3–10.

Powell, James L. (1998), *Night Comes to the Cretaceous.* San Diego: Harcourt Brace.

Price, Huw (1996), *Time's Arrow and Archimedies' Point.* Oxford: Oxford University Press.

Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.

Thomas-Keprta, Kathie L., Dennis A. Bazylinski et al. (2000), "Elongated Prismatic Magnetite Crystals in ALH84001 Carbonate Globules: Potential Martian Magnetofossils", *Geochemica et Cosmochimica* 64, 4049–4081.

Tsapin, A. I., M.G. Goldfeld et al. (2000), "Iron (VI): Hypothetical Candidate for the Martian Oxidant", *Icarus* 147: 68–78.

Ward, Peter D. (1990), "The Cretaceous/Tertiary Extinctions in the Marine Realm", in Buck Sharpton and Peter Ward (eds.), *Global Catastrophes in Earth History.* Boulder: Geological Society of America, 425–432.

Zaug, Arthur J. and Thomas R. Cech (1986), "The Intervening Sequence RNA of *Tetrahymena* Is an Enzyme", *Science* 231: 433–524