

Draft

Chapter from a book, *The Material Theory of Induction*, now in preparation.

Inference to the Best Explanation: Examples

John D. Norton

Department of History and Philosophy of Science

University of Pittsburgh

<http://www.pitt.edu/~jdnorton>

1. Introduction

According to the material theory of induction, there can be no universally applicable schema that fully characterizes inference to the best explanation. At best we can find loose similarities that the canonical examples of inference to best explanation share. These loose similarities were codified in the last chapter (Section 7) in a two-step characterization. First is a comparative step in which one hypothesis or theory is favored over one or more foils. In the second step, this favoring is rendered absolute: we are authorized to infer to the favored hypothesis or theory. It was noted in the last chapter that this characterization was derived from a compendium of canonical examples of inference to the best explanation. That compendium is given in this chapter. The seven examples are developed in Sections 4 to 10. For ease of overview, their fit to the general characterization is summarized in a table in Section 3. Section 11 has a general conclusion. First, however, the next section offers reflections on the importance of the examples.

2. Examples Matter

What this chapter shows is that taking examples seriously is important. I have already lamented in various places in the last chapter how the present literature has often treated its

examples in too great haste. There are two ways that this hastiness has obscured the evidential relations in the examples in science.

First, it is common to employ examples in which human action plays an essential role. This is appealing for, in these are examples, the analysis is easiest and most compelling. However that ease comes just because these examples are poor surrogates for the real examples in science, where the evidential relations are commonly less clear. Specifically, these examples mislead us since, unlike the examples in science, the role of the comparative foil is minimal. Lipton (2004, p. 6) gives the time-worn example:

Faced with tracks in the snow of a certain peculiar shape, I infer that a person on snowshoes has recently passed this way.

Once one has seen the distinctive imprints left by snowshoes, there is really only one account to be given of their origin. We might invent fanciful scenarios just to drive home that there is no real choice. Lipton (2004, p. 56) shows how it is done:

Of course, there is always more than one possible explanation for any phenomenon—the tracks might have instead been caused by a trained monkey on snowshoes, or by the elaborate etchings of an environmental artist—so we cannot infer something simply because it is a possible explanation. It must somehow be the best of competing explanations.

However, entertaining these alternatives rapidly becomes a perfunctory exercise in eliminating the fanciful. We might well dismiss them as comic relief.

Here these human examples are quite unlike the real scientific examples. The alternative hypotheses or theories in the scientific cases were not jokes. Prevailing over them is, almost everywhere, the greater challenge, as we shall see below. The wave theory of light struggled for centuries both with its own early weaknesses and the fact that the competing emission theory had been delivered by the authority of authorities, Isaac Newton himself. Darwin struggled to account for the eye, where his creationist opponents could readily explain the perfection of its design with their designer. The anomalous motion of Mercury's perihelion could be explained by Seeliger's zodiacal light, if only it could be determined that it held enough matter. It was essential to Einstein's general theory of relativity that this quite prosaic account fail, for nothing in the elegance of Einstein's theory could protect it if Seeliger's hypothesis proved workable. Thomson's particle theory of cathode rays had to overcome Lenard's ether wave theory. It is

only in retrospect that we see how precarious was Thomson's victory, for the soon-to-emerge quantum theory did attribute wavelike properties to electrons.

Second, much of the literature on inference to the best explanation mentions examples in science but does not explore them fully. As a result, they draw on dangerously oversimplified caricatures and miss the real moral of the examples. Superficially, for example, big bang cosmology provides an account of Penzias and Wilson's observation of cosmic background radiation rich in explanatory virtue. As a result the inference to the big bang looks immediate and irresistible and can be drawn without much concern for other accounts. However, if one teases out the history, as is done below, one finds that that the explanatory virtue was initially less clear and less decisive. It took decades before the inference was secure; and only popular simplifications of the history could make the inference seem immediate and irresistible.

More importantly, the essential and delicate part of the analysis was not establishing that big bang cosmology could accommodate the result. Almost any cosmology could deliver background radiation in one form or another. All it needed was to include electrically charged matter; and every viable cosmology must do this, else it cannot harbor stars that shine in the electromagnetic spectrum. Rather the burden was first to establish, with some effort over years, a particular thermal form for the background radiation and then to argue in some detail why competing accounts could not recover it. Then the evidential success looks less like a sudden explanatory coup of one theory than a slowly building and widespread failure of the competitors. This dynamic is repeated in many examples.

3. Synopsis of Examples

In the characterization of inference to the best explanation of the last chapter (Section 7), the principal burden is to establish superiority of the favored hypothesis or theory over a competing foil or foils. In the second step, the status of the favored hypothesis or theory can, but may not, be generalized from better explanation to best. The table below indicates in summary how the examples of this chapter instantiate this characterization.

<i>Abduction</i>	<i>Foil</i>	<i>Foil eliminated</i>	<i>Generalization from better to best</i>
Darwin on the origin of species	Independent creation	Refuted by traits without function	Tacit assumption of exhaustive choice
Lyell's uniformitarian geology	Geologies using presently unknown causes	Novel causes incur an undischarged evidential debt.	Known versus unknown causes is exhaustive
Thomson for cathode rays as charged particles	Cathode rays are processes in the ether.	Contradiction with experiment: Ether waves would not be bent by a uniform field	Tacit assumption of exhaustive choice
Lenard for cathode rays as ether processes	Cathode rays are processes in matter	Contradiction with experiment: cathode rays in evacuated tubes	Choice between matter and ether posed as exhaustive dilemma.
Einstein's explanation of Mercury's anomalous motion	Many. Modifications to Newtonian theory. Unobserved masses.	Contradiction with experience. Undischarged evidential debt.	Step not taken.
Cosmic background radiation from the big bang	Alternative cosmologies, especially steady state cosmology	Empirical failure	Taken tacitly
Lavoisier's oxygen chemistry.	Phlogiston chemistry.	Contradiction. Matter has weight (gravity), but phlogiston has levity.	Fact (matter has weight) is one of many warranting facts.
Wave theory of light.	Newtonian corpuscular theory.	Undischarged evidential debt. Contradiction with experiment.	Complicated.

Table 1. Summary of Examples

4. Darwin and The Origin of Species¹

We saw in the last chapter that one of the earliest statements of what we now call inference to the best explanation appeared in Darwin's *Origin of Species*. My task here is focused narrowly on the argument as it is developed in this particular volume of Darwin's writings. My concern is not how the analysis may be developed in other of Darwin's writings.

¹ I thank Zina Ward for helpful discussion of this section.

My concern is definitely not how we might presently make the case for the theory of evolution. The modern case rests on a much larger evidential base and has a greater reliance on supporting sciences, such as Mendelian genetic theory, unknown to Darwin. It resolves many of the problems troubling Darwin's development.

4.1 Darwin's Argument

The whole volume, Darwin tells us, develops what he calls "one long argument" (1876, p. 404). It is an argument that cannot be reproduced here with any fidelity. It depends on a lengthy, massively impressive recitation of detailed facts in natural history. They are explained by a wonderfully simple process. There is in nature a constant struggle for survival by living beings. They grow at a geometrical rate that outpaces the arithmetic growth of resources. Favorable variations give their bearers an advantage. Nature selects them for survival, just as domestic breeders select commercially desirable variations. Those selected flourish, leaving offspring with similar characteristics. Darwin (1876, pp. 102-103) offered a simple summary:

This principle of preservation, or the survival of the fittest I have called Natural Selection. It leads to the improvement of each creature in relation to its organic and inorganic conditions of life:...

The development of the argument in *Origin* then follows a simple formula. Some feature of living beings is displayed and then an account is given of how it could arise through Natural Selection. A reader cannot but be overwhelmed by the sheer mass of facts in natural history that Natural Selection accommodates. No short selection here can do justice to it. Darwin (1876, p. 414) himself tries to convey its weight in a concluding chapter with a rapid recitation of successes:

Many other facts are, as it seems to me, explicable on this theory. How strange it is that a bird, under the form of a woodpecker, should prey on insects on the ground; that upland geese which rarely or never swim, should possess webbed feet; that a thrushlike bird should dive and feed on sub-aquatic insects; and that a petrel should have the habits and structure fitting it for the life of an auk! and so in endless other cases. But on the view of each species constantly trying to increase in number, with natural selection always ready to adapt the slowly varying descendants of each

to any unoccupied or ill-occupied place in nature, these facts cease to be strange, or might even have been anticipated.

These successes lead up to what is, for our purposes, the key evidential claim (1876, p. 421)²

It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified.

Darwin does not justify this key claim. It is, presumably, offered as self-evident. It is plausible, however, that Darwin was following William Whewell. The latter described the consilience of induction as arising when one theory proves, unexpectedly, to explain more classes of facts; and this, Whewell urged, is a powerful indicator of the truth of the theory. (For elaboration, see Thagard, 1977; 1978, Section II.)

In spite of its many successes, Darwin's theory faced serious difficulties. Darwin sought as well as he could to deal with them. They have the character of the taking on of an evidential debt: a supposition needed for the theory to succeed but for which evidence was then lacking. Here are two examples.

The first was that there was some doubt that the earth was sufficiently old for the extraordinary amount of time Darwin's theory required for natural selection to do its work. Darwin's (1876, p. 409) best hope was merely to keep the problem an open question, still to be decided:

With respect to the lapse of time not having been sufficient since our planet was consolidated for the assumed amount of organic change, and this objection, as urged by Sir William Thompson, is probably one of the gravest as yet advanced, I can only say, firstly, that we do not know at what rate species change as measured by years, and secondly, that many philosophers are not as yet willing to admit that we know enough of the constitution of the universe and of the interior of our globe to speculate with safety on its past duration.

² This confident claim was not present in the first edition and, presumably, was added as part of Darwin's response to his critics.

The second was the absence of intermediates. Darwin's theory required all variation to arise through very slow, small gradations. Yet nature has vast gaps between various forms. The evolution of the eye presented a special problem, since its perfection as an optical instrument was not naturally explained as the handiwork of a creator. Darwin strove for pages to make plausible that a light sensitive nerve in some being might eventually develop into an eye. However he could in the end do little better than to appeal to his reader's indulgence (1876, p. 145):

He who will go thus far, ought not to hesitate to go one step further, if he finds on finishing this volume that large bodies of facts, otherwise inexplicable, can be explained by the theory of modification through natural selection; he ought to admit that a structure even as perfect as an eagle's eye might thus be formed, although in this case he does not know the transitional states.

The foil against which Darwin competed was independent creation: the thesis that "species were immutable productions, and had been separately created." (Darwin, 1876, p. xiii) On the development of eye, Darwin (1876, p. 146) was straining merely to match independent creation:

It is scarcely possible to avoid comparing the eye with a telescope. We know that this instrument has been perfected by the long-continued efforts of the highest human intellects; and we naturally infer that the eye has been formed by a somewhat analogous process. But may not this inference be presumptuous? Have we any right to assume that the Creator works by intellectual powers like those of man?

Elsewhere, repeatedly, in the volume, Darwin sought to do better than the thesis of independent creation. The means depended on assuming just what he had suggested we had no right to do. That is, his arguments depended on assuming that a creator would only endow a being with a trait if that trait had some useful purpose; and that if there are similarities across species there must be some discernible purpose for them. With that assumption lingering behind his remarks, time and again Darwin could point out some feature that had no evident purpose, but arose naturally through the slow developments fostered by natural selection.

In besting the thesis of independent creation, Darwin was fond of the superlative "utterly inexplicable," using it at least four times:

...we can clearly understand these analogies [clustering of species], if species once existed as varieties, and thus originated; whereas, these analogies are utterly inexplicable if species are independent creations. (p. 47)

and similarly

This grand fact of the grouping of all organic beings under what is called the Natural System, is utterly inexplicable on the theory of creation. (p. 413)

Such cases as the presence of peculiar species of bats on oceanic islands and the absence of all other terrestrial mammals, are facts utterly inexplicable on the theory of independent acts of creation. (p. 419)

On the view of each organism with all its separate parts having been specially created, how utterly inexplicable is it that organs bearing the plain stamp of inutility, such as the teeth in the embryonic calf or the shrivelled wings under the soldered wing-covers of many beetles, should so frequently occur. Nature may be said to have taken pains to reveal her scheme of modification, by means of rudimentary organs, of embryological and homologous structures, but we are too blind to understand her meaning. (pp. 420-421)

As the examples of his deprecation of independent creation multiply, Darwin rarely speculates on the details of this competing theory. The tacit assumption everywhere is that independent creation delivers immutable species all of whose traits have a purpose. An exception arises when Darwin (1876, pp. 130-31) recalls similarities between equine species:

He who believes that each equine species was independently created, will, I presume, assert that each species has been created with a tendency to vary, both under nature and under domestication, in this particular manner, so as often to become striped like the other species of the genus; and that each has been created with a strong tendency, when crossed with species inhabiting distant quarters of the world, to produce hybrids resembling in their stripes, not their own parents, but other species of the genus. To admit this view is, as it seems to me, to reject a real for an unreal, or at least for an unknown, cause. It makes the works of God a mere

mockery and deception; I would almost as soon believe with the old and ignorant cosmogonists, that fossil shells had never lived, but had been created in stone so as to mock the shells living on the sea-shore.

This is a less visible line of argument in Darwin's text: that there is something defective as a theory in positing a process of independent creation.

These thoughts develop into a direct assault on the explanatory viability of a creator. Darwin (1876, pp. 383) reports:

On the ordinary view of the independent creation of each being, we can only say that so it is;—that it has pleased the Creator to construct all the animals and plants in each great class on a uniform plan; but this is not a scientific explanation.

and then again (p. 422)

It is so easy to hide our ignorance under such expressions as the "plan of creation," "unity of design," &c., and to think that we give an explanation when we only restate a fact.

To tease out Darwin's objection, imagine that we make up a huge list of species and their traits. To add the remark that the creator planned it so, adds nothing of explanatory value.

To sum up, Darwin's argument for his theory rests on its explanatory prowess with a huge array of facts in natural history. The meaning of the term "explanation" is not given. However the familiar covering law account of explanation fits his usage as well as any: the many facts are explained since they are entailed by his theory. More precisely, the *possibility* of the specific facts is entailed by his theory, for natural selection cannot predict specifically each of the many facts Darwin reports. The strength of the explanation resides in the breadth and variety of the facts covered. Perhaps this is a quiet echo of Whewell's notion of consilience of induction.

The foil against which Darwin rails is the independent creation of each species as immutable productions. His claim of its explanatory defects rests on the tacit assumption that each trait of a living creature must have a purpose; and that this is also the case for similarities among different species. Without some assumption of this type, Darwin has no real basis for discarding independent creation. For without it the thesis is so incompletely defined that no evidential test is possible. What in nature might then count as favorable or unfavorable evidence? Finally, Darwin turns this difficulty against independent creation by suggesting that it is no explanatory theory at all.

Darwin's theory has its difficulties and these require him to take on some undischarged evidential debt, such as the supposition of long times for natural selection to work and there were transitional forms now not in evidence. We are to conclude, however, that the foil of independent creation is so troubled that Darwin's theory prevails.

4.2 What Powers the Inference

The delicate but central question in this analysis is just what powers Darwin's inference. Let us review some possibilities for a general account that employs a formal principle.

At some intuitive level, there is a sense of beauty and elegance in Darwin's theory and wonder that it embraces such a diversity of fact. This gives it the ring of truth. That feeling, however, falls well short of what an inductive logic, formal or even material, requires. Is the principle that the evidential support is strong merely if we genuinely and honestly feel it is so? That is not a sustainable principle of logic. Worse, how are we to deal with the case in which the feeling is not widely shared? That is our case. When Darwin announced his theory, the public debate was spirited. Darwin's critics were not swayed.

Might the inference be powered by the general result that Darwin himself cites: that the theory "explain[s], in so satisfactory a manner ... several large classes of facts..."? I will continue to take the otherwise undefined term "explain" to mean "derive their possibility from a few posits of the theory." As noted above, the situation is more complicated. For that is not quite what Darwin's theory does. The derivation does proceed from a few simple posits. However it also draws upon suppositions that are themselves in great need of further evidential support. The theory requires an extraordinary amount of time for its operations to succeed; and many of them, such as the descent of eyes, are presumed possible while required intermediate states are not found. They are also presumed. These are evidential debts that, in other examples in this chapter, are sufficient to lead to the abandoning of a theory. There are evidential strengths and weaknesses to be balanced here before a final decision can be taken. Darwin delineates no general inductive principle to which his analysis conforms. There is no formal theory provided, even in vague outline, that negotiates the complexity of the balancing. In its absence, a formal analysis is unpromising.

The prospects for a material analysis are more promising. For, even lacking a formal theory, Darwin himself and his sympathetic supporters found powerful support for Darwin's

theory in his evidence. They had only facts to draw upon. Another promising sign for a material analysis is that contemporary commentators disagreed so pointedly. They have the same evidence and arguments before them. If these alone are compelling, then disagreement can only come from ineptitude in the logic. If however background factual assumptions also bear crucially on the cogency of the argument, then matters improve. For, if we allow that different commentators harbored different background assumptions, then the disagreement is intelligible. We need attribute no inductive fallacies to the disagreeing commentators.

We can see how material facts could underwrite Darwin's confidence in his theory if we presume that Darwin found his analysis to establish two facts:

1. It is possible that the variety of species arose from descent with modification through natural selection.

and

2. It is unlikely that any other admissible account can accommodate the origin of species.

These facts combined are sufficient to warrant acceptance of Darwin's theory. His account is possibly right; no others are; therefore his has to be right. No formal principle of induction is needed.

The first fact is demonstrated by the massive weight of Darwin's many examples. The second fact is essential, for, without it, merely establishing possibility is insufficient. Unfortunately establishing this second fact is more difficult. For the only other account given serious analysis in Darwin's volume is independent creation. He does cast significant doubt on independent creation. It is contradicted by many arbitrary facts in natural history for which a creator would have no evident purpose. Darwin even calls into doubt whether independent creation counts as an explanatory theory at all.

What is left open is the question of whether there are still other theories possible that may do as well or better than Darwin's. Of course it is hard for us to imagine what these still better theories might be. But that our imagination fails is poor proof that there are no such theories. Perhaps Darwin is expecting us to proceed from a background assumption that we have no reason to expect that *any* theory could do justice to the wealth of fact in natural history. So merely finding one is so extraordinary that we can cease our searching. Or perhaps we might suppose that Darwin poses a dilemma for us: either species descend from other pre-existing

forms, or they did not. Darwin's theory and independent creation are, we are to suppose, the strongest version of each horn. Perhaps, when Darwin's theory bests independent creation, that is enough to establish that Darwin's theory is not just the better of the two but it is the best of all.

5. Lyell's *Principles of Geology*

Charles Darwin was influenced greatly by the uniformitarian geologist, Charles Lyell. Before Darwin left on his formative voyage on the *Beagle* in 1831, its captain, Robert Fitzroy, gave Darwin a copy of Volume I of Lyell's (1830) *Principles of Geology*. Subsequently in November 1832, Darwin received Volume II (1832) in his mail in Montevideo. The volumes had a profound impact on Darwin, who had been recruited as the voyage's naturalist to work in both geology and zoology.

For our purposes, what is striking in Lyell's *Principles of Geology* is that it provides a near perfect template for the argument that Darwin will later develop in his *Origin of Species*. Lyell's concern is to overturn earlier accounts of the origin of the earth's geological features. These earlier accounts supposed that modern features were formed by presently unknown geological processes typically of far greater violence than those now observed. These were the "catastrophist" theories, as Whewell soon called them. They correspond to Darwin's foil of independent creation, for both presume extraordinary occurrences in the past to explain present features: for Lyell, present geology; for Darwin, the diversity of species. Lyell sought to replace these catastrophist theories with a uniformitarian geology, such as defended by James Hutton before him. In it, present day geological feature are explained by very slow geological processes now in operation, while acting over a long time. Correspondingly, Darwin sought to explain the diversity of species by means of natural selection, which employed slow processes present now, acting over a long time.

The connection to explanation is in the title of Lyell's three volume work. It is *Principles of Geology, Being an Attempt to Explain the Former Changes of the Earth's Surface, by Reference to Causes Now in Operation* and the words "explain" and "explanation" appear throughout the text.³ Lyell's overall argument is an inference to the best explanation. Present causes acting slowly over a long time

³ In Volume I, they appear together nearly 100 times.

are a better explanation of present geological features than past cataclysms and, presumably, the best explanation.

Since inference to the best explanation was not then a recognized mode of argumentation, we would expect that Lyell might provide some defense of it. How do we bridge the gap between an hypothesis that explains well and the truth of the hypothesis? Darwin was sensitive to the need to defend his method of argumentation and repeatedly introduced commentary in its defense. Lyell, however, gave no indication that he saw the need to defend the mode of argumentation. As far as I can see, the first two volumes of *Principles of Geology* contain no methodological analysis, beyond chance remarks and colorful reprimands for the errors of past theorists. It is only in the first chapter of Volume III that Lyell gave a more extended defense of his methods.

There he summarized his approach as (1833, p.6)

In our attempt to unravel these difficult questions, we shall adopt a different course, restricting ourselves to the known or possible operations of existing causes; feeling assured that we have not yet exhausted the resources which the study of the present course of nature may provide, and therefore that we are not authorized, in the infancy of our science, to recur to extraordinary agents. We shall adhere to this plan, not only on the grounds explained in the first volume, but because, as we have above stated, history informs us that this method had always put geologists on the road that leads to truth,—suggesting views which, although imperfect at first, have been found capable of improvement, until at last adopted by universal consent.

It was contrasted with the catastrophist foil (pp. 6-7):⁴

On the other hand, the opposite method, that of speculating on a former distinct state of things, has led invariably to a multitude of contradictory systems, which have been overthrown one after the other,—which have been found quite incapable of modification,—and which are often required to be precisely reversed.

⁴ The idea of employing just processes now acting is appealing in the abstract. However it can quickly run into trouble. The steady state cosmology of the mid-twentieth century was a uniformitarian cosmology that led its proponents to wild speculation, such as the continuous creation of matter. The big bang theory, its catastrophist competitor, won the day.

As with Darwin, the strength of Lyell's case rests ultimately on a massive compilation of illustrations of how presently acting causes could generate the geological features now observed. Conveniently for us, in this chapter Lyell selected three examples to illustrate the differences of the two approaches. The first concerned fossil shells and bones. The former view accounted for them as "fashioned into their present form by a plastic virtue, or some other mysterious agency" (p. 4). Lyell instead sought their origin in biological processes just like those in action today. The second concerned the origin of basalt and similar rocks. The former view attributed it to aqueous processes, while Lyell could point to igneous processes now in action that create such rocks. The third concerned the occurrence of fossil shells in rocks in high mountains. The former view sought some unusual process that might dry up oceans and drop their level. Lyell replaced it with processes that elevate land above an otherwise fixed sea level.

In all this, Lyell treated the uniformitarian view as little removed from providing an explanation of some process by directly observing its cause, as opposed to speculating on a novel cause presently not in evidence. He complained of the catastrophists that "...they felt themselves at liberty to indulge their imaginations, in guessing at what *might* be, rather than in inquiring *what is*;..." (p.2, Lyell's emphasis). And then (p.2):

It appeared to them more philosophical to speculate on the possibilities of the past, than patiently to explore the realities of the present, and having invented theories under the influence of such maxims, they were consistently unwilling to test their validity by the criterion of their accordance with the ordinary operations of nature.

Lyell's text becomes more polemical, heaping scorn on the catastrophists. (pp. 2-3)

Never was there a dogma more calculated to foster indolence, and to blunt the keen edge of curiosity, than this assumption of the discordance between the former and the existing causes of change.

This stands in stark contrast with Darwin's cautious defense of his use of causes presently in operation. While we can see nature selecting favorable variations among living beings in processes now in operation, Darwin first offered a lengthy discussion of selection by domestic breeders to convince us of the potency of selection. Perhaps Lyell's task was less formidable. He needed only to establish that processes now in operation might eventually produce a mountain, not an eye.

How does this bear on the concerns of this chapter: the warranting of abductive inferences? In the formal approach, the fact that some hypothesis or theory explains what is observed is confirmatory in virtue of the special character of explanation. Unlike Darwin, Lyell saw no special explanatory relationship between his theory and the geological facts it accommodates that would require any circumspection. The theory, in Lyell's telling, does little more than instruct us merely to observe the causes directly.

The warrant for Lyell's argument for uniformitarianism is readily found in background facts, that is, materially. In analogy with the material warranting of Darwin's argument in *Origin of Species*, we can assume that Lyell seeks to establish two facts:

1. It is possible that present geological features arose over long time periods from causes now operating.

and

2. It is unlikely that any other admissible account can accommodate their origin.

These two facts are sufficient to warrant acceptance of Lyell's uniformitarianism. His theory is possibly correct; no others are; therefore his has to be correct.

The first fact is established by the wealth of examples in Lyell's account. The second proves a great deal easier to establish than the corresponding fact in Darwin's warrant. For Darwin's foil was specifically the thesis of independent creation; and that left open the possibility of many other theories excluded from explicit analysis. Lyell, however, has two cases that are exhaustive. Either present geological features arose from causes now in operation; or they did not. The first case is Lyell's uniformitarianism. The second is a theory that must speculate on presently unknown causes or known causes but of presently unknown intensity.

Lyell has a direct and telling objection to theories of this second type: they are taking on an undischarged evidential debt. If fossil shells were formed by some plastic virtue or mysterious agency, then we are owed independent evidence that such virtues and agencies exist. If high mountains were thrown up suddenly by cataclysmic forces, we are again owed independent evidence that such forces existed. Lyell's theory takes on no corresponding evidential debt. We are assured of the existence of the causes he employs since they are in operation now. Perhaps his only evidential debt is that enough time has passed for these causes to produce the geological features we see now.

6. Thomson's Electron

J. J. Thomson's (1897) "Cathode Rays" marks a turning point in physics. Thomson identified the rays produced in a cathode ray tube as beams of negatively charged particles of a fixed charge to mass ratio. These particles would soon carry the name "electron" and would be the first fundamental particle of the menagerie of particles that would be discovered in the twentieth century.

Describing the achievement as a discovery makes it sound like a "look-see" event, such as the discovery that one has bats in one's attic. It was less that and more the identification by astute reasoning of the nature of a phenomenon long observed and probed. It was also the resolution of a debate between English and German physicists over the nature of cathode rays. Are these cathode rays beams of matter? Or are they waves in the ether? Thomson identified them as matter: particles charged with negative electricity. Lenard, Hertz and others identified them as waves in the ether.

For our purposes, the interesting point is that both sides employed abductive inferences. It was a duel of abduction, won by Thomson. Below we will look at the abductive inference deployed on both sides. We shall see that they are fully controlled by background assumptions. Key to the arguments of both sides is an assumption of exhaustion: that the two alternatives they considered --matter or waves--were exhaustive. For Thomson, the assumption was tacit. For Lenard it became explicit: finding trouble for both matter and waves posed, for Lenard, a troubling dilemma, resolved only by a new, third option.

Each side had to establish that their favored account fitted the experimental result and, preferably, did so very well. But that alone did not suffice. Each side also needed to demonstrate that the competing account was untenable. Each claims the other's account refuted by the experiment. The assumption of exhaustion then did the critical work of allowing the step from the adequacy of each side's account to its truth. The course of the debate was controlled by the assumptions of exhaustion,

6.1 Thomson: Cathode Rays are Charged Particles

Let us begin with the much-told story. Thomson's argument⁵ in his (1897) "Cathode Rays" depended on the extensive series of experiments reported in his paper. In brief, Thomson shows that cathode rays are deflected by electric and magnetic fields in perfect agreement with the basic law of electrodynamics that we now know as the Lorentz force law. That the charges are negative is also affirmed by directing the rays at a metal vessel, which then becomes negatively charged. Perhaps the most powerful part of Thomson's argument is that the experiments with magnetic and with electric deflection both yield the same value for the characteristic mass to charge ratio m/e for the particles. This same ratio is recovered whatever the material of the cathode emitting the rays.⁶

There are many details here that could be pursued. Thomson's experiments were delicate and the detailed development of his case sophisticated. For our purposes, what matters is that Thomson's favored charged particle hypothesis fits his experimental results quite wonderfully well. He sums it up in a much-quoted passage (Thomson, 1897, p.302) as:

As the cathode rays carry a charge of negative electricity, are deflected by an electrostatic force as if they were negatively electrified, and are acted on by a magnetic force in just the way in which this force would act on a negatively electrified body moving along the path of these rays, I can see no escape from the

⁵ In his 1906 Nobel lecture, Thomson (1906) does use the words "argument" and "proof" to describe the case he makes (my emphasis):

"The *arguments* in favour of the rays being negatively charged particles are primarily that they are deflected by a magnet in just the same way as moving, negatively charged electrified particles.... The next step in the *proof* that cathode rays are negatively charged particles was to show that when they are caught in a metal vessel they give up to it a charge of negative electricity."

⁶ In Norton (2000, §3.2), I have described this part of Thomson's analysis as employing "overdetermination of constants," an argument strategy that has been employed elsewhere to good effect. I also note (§3.3) the overdetermination of constants by itself is not sufficient to rule out competitors, which is the issue the present text now turns to address.

conclusion that they are charges of negative electricity carried by particles of matter.

Thomson does not use the word “explains” or “explanation” here. Unlike Darwin and Lyell, the words are barely used at all. However we can identify Thomson’s overall argument as an inference to the best explanation.

The difficulty of Thomson’s argument is that his summary establishes only that this particle theory fits wonderfully well. Nothing in his summary establishes that other accounts cannot do as well. One might think it excessive to demand anything more of Thomson, for there seems to be no gap at all in Thomson’s argument. However there is a gap. In a few decades, with the rise of quantum theory, propagating electrons will turn out to be waves after all. They might not be waves in a nineteenth century ether. They are waves of quantized particles, so they have wavelike properties nonetheless. More important, these waves have exactly the properties that Thomson found so compelling: they carry negative charge and are deflected just as Thomson found by electric and magnetic fields.

6.2 Lenard: Cathode Rays are Waves

The explicit burden of establishing that no other account can do as well was carried by Thomson’s arguments against the competing view. That view was that cathode rays are a form of radiation in some way akin to light or Röntgen rays (also called X-rays). The then prevalent theory represented such radiation as a wave propagating in the all-penetrating ether.

This wave account was defended by Philipp Lenard,⁷ student and protégé of Heinrich Hertz, who had just died prematurely in 1894 at the age of 36. Lenard’s (1894) poses the problem as one of deciding whether the rays are “processes in matter or processes in ether.” These two possibilities represent the only two possibilities allowed by late nineteenth century physics. A discharge tube can contain ordinary matter and ether; there is no third possibility. So a process such as a cathode ray must be a process within one or both of these. Processes in matter, we soon learn, are akin to the propagation of sound, which is carried by the material substance of air. It is quite plausible that cathode rays are something comparable. The electric potential might

⁷ And alas soon to be a leading light of the anti-semitic, German science movement of the Nazi era.

ionize the gas molecules that are then driven as a ray through the tube by electrical attraction. Processes in ether are akin to light propagation, which is carried by the ether. If of this form, cathode rays would correspondingly be carried as waves in the ether. Any matter present, such as air, would act only as an interference and impede the wave.

This posing of the problem is critical to the further analysis, since it reduces the analysis to deciding between two cases. It is the key assumption of exhaustion. Lenard proposed to decide between the two by means of an experiment in which cathode rays are propagated in a vacuum. He explained (1894, pp. 226-27):

[it affords] the possibility of carrying out the very same fundamental experiments, that had decided for light and sound whether these latter are processes in matter or processes in ether.

Light can propagate in a fully evacuated space without obstruction since ether remains. Sound propagation is suppressed entirely, since its material carrier has been eliminated.

Lenard then reports the results of the experiment; they favor the ether process (1894, p. 248):

Therefore cathode rays also propagate in spaces whose contained matter is only in that extreme dilution in which all known processes in it disappear. One cannot ascribe the mediation of the intensive processes observed to the remainder of the matter, which is more or less completely distant and without influence, but only to the ether, which we cannot remove from any space. If this is accepted, then our experiment on the nature of cathode rays decides that they are processes in the ether.

At first pass, this argument is an abduction: the best explanation of the propagation of cathode rays in a vacuum is that they are ether waves. A more careful analysis, however, shows that explanation as a primitive notion plays no essential role. It is really an eliminative argument. The rays are either material processes like sound or waves in the ether like light. They cannot be the first since they persist in a vacuum. Therefore, by elimination, they must be the second.

Conveniently for us, Lenard then reports others who share the ether process view, thereby giving a contemporary list of those whom Thomson (1897, p. 293) would later identify merely as “German physicists” in the opening of his celebrated “Cathode Rays” paper. They are Heinrich Hertz, Eilhard Wiedemann and Eugen Goldstein.

What comes in Lenard's next paper of the same year is still more interesting. The celebrated quote from Thomson's "Cathode Rays" paper above purports to show that cathode rays are beams of charged particles because they behave in just that way (e.g. "...are acted on by a magnetic force in just the way in which this force would act on a negatively electrified body..."). It is easy for us to read that now as compelling. We might ask, what explains that the rays behave *as if* they are streams of particles? It is that they *are* streams of particles! This is easy hindsight. In 1894, Lenard could dismiss just this argument. He began his second paper on cathode rays of 1894 by noting that cathode rays are deflected by magnetic fields, just as would beams of charged particles (1894a, p. 23):

Here the behavior of cathode rays agrees with the behavior of a stream of massive, negatively charged particles, projected from the cathode.

Lenard then discounted this agreement as superficial:

This agreement between cathode rays and radiating matter--which one finds again in other phenomena of radiation and which has even been seen by many physicists since Crookes' experiments to hold generally—can nonetheless only be superficial, if the result drawn earlier [footnote citation to Lenard, 1894], that cathode rays are processes in the ether, was justified.

Lenard's dismissal is not casual. The main point of his paper is to present experimental results that establish the dismissal. He proceeded to describe the experiments and their results (pp. 23-24):

That the agreement is in fact only superficial seems now to me to be shown especially well in the following experiments, in which the agreement fails completely, when circumstances, which must be of the greatest influence on the speed of radiating matter, turns out to be completely without influence on the magnitude of the magnetic deflection of cathode rays.

The experiments show that the magnitude of the magnetic deflection is not at all influenced by the medium in which the radiation is observed; rather the deflectability of one and same kind of cathode rays remains always immutably the same, in all gases, with all pressures, with each intensity of radiation and even then, if the latter [rays] have passed through a metal wall pushed in front;...

The inference against the particle account depends on the same analysis as Lenard's earlier paper. If cathode rays are streams of matter, then they cannot persist in a vacuum, where there is no matter, just as there can be no sound waves there. Since they depend so much on the matter present, we would expect changes in the matter present to change the amount of deflection of cathode rays by magnetic fields. Yet no such effect is found.

In short, Lenard had asserted three years before Thomson's celebrated paper, that successful explanation of magnetic deflection by the particle theory is not enough. It is an insufficient basis for inferring to the particle theory that can be overruled by the failure of the theory to fit other experimental facts. Lenard claimed that it has been so overruled by his latest experiments. He then recalled another experiment by Hertz. It also precluded the deflected cathode rays merely being a beam of charged particles acted on directly by a magnet (1894a, p. 32):

The deflection of cathode rays is, according to Hertz' experiments, not an effect of the magnet on the rays themselves, but an effect of the latter on the medium through which they radiate; the rays propagate differently in a magnetized medium than in a non-magnetized medium. For if the forces act between the magnet and the rays themselves, then the magnet must also be deflected by the cathode rays, if the magnet is made movable; this is not the case. [footnote⁸]

The basis of the experiment is elementary electromagnetism. If cathode rays are a beam of charged particles, then they behave electromagnetically much the same as a current in a wire. A current carrying wire creates magnetic effects. Oersted had found a magnetic needle deflected in the vicinity of such a wire. Correspondingly we should find magnetic effects in the vicinity of cathode rays. Yet when Hertz sought them, he found none. His delicately balanced magnet was undeflected.

With the failure of the particle theory now assured, Lenard turned to a brief elaboration of the ether theory. How is it that a magnetic field can deflect a cathode ray? The means, Lenard

⁸ The footnote "Hertz, Wied. Ann. **19**. p. 799 f. and 805 f. 1883" is to those parts of Hertz (1883) where Hertz reports the negative result of this experiment. Hertz (1883, p. 807) also draws the same conclusion as Lenard: "the magnet acts on the medium, but cathode rays propagate differently in magnetized than in an unmagnetized medium."

explained, is indirect. The magnetic field affects the ether and, indirectly through that effect, the cathode rays carried by the ether (1894a, pp. 32-33).

The medium, however, whose magnetic alteration is shown through the curvature of the rays, is, as a result of our experiments, the ether itself. For the curvature is found to be fully independent of the nature and the density of any ponderable medium present; in particular, it was also observed in the highest vacuum. [footnote⁹]

Therefore, through their curvature, cathode rays give an immediate indication that the state of the ether between magnetic poles is mutable, as is required by the theory of mediated action at a distance.

This is the ether-wave theorists' account of the magnetic deflection of cathode rays. It becomes the key target of Thomson's argument against the ether-wave theory.

6.3 Thomson: Cathode Rays are not Waves

Thomson's celebrated "Cathode Rays" paper of 1896 begins by posing the problem as a decision between two theories of the constitution of cathode rays (p.293): they are "some process in the aether" ("according to the almost unanimous opinion of German physicists"); or they are "wholly material...particles of matter charged with negative electricity." He continued:

It would seem at first sight that it ought not to be difficult to discriminate between views so different, yet experience shows that this is not the case, as amongst the physicists who have most deeply studied the subject can be found supporters of either theory.

The electrified-particle theory has for purposes of research a great advantage over the aetherial theory, since it is definite and its consequences can be predicted; with the aetherial theory it is impossible to predict what will happen under any given circumstances, as on this theory we are dealing with hitherto unobserved phenomena in the aether, of whose laws we are ignorant.

Thomson's paper then proceeds to recount the well-known experiments that lead up to the conclusion already quoted earlier. Virtually the entirety of the paper and its argumentation are devoted to showing that the charged particle view fits the experiments. He addresses several

⁹ To Lenard (1894, pp. 244 and 246).

objections from the ether wave theorists to the particle theory that can be answered experimentally.¹⁰ However there is no sustained effort to show that the ether wave theory cannot perform just as well experimentally as the particle theory. His argument to this effect is so tersely stated as to be impossible to follow if read in isolation. He inserts within a sentence in the introductory paragraph (p. 293) of the ether process theory of cathode rays

“...in a uniform magnetic field their course is circular and not rectilinear—no phenomenon hitherto observed is analogous...”

The difficulty is not of an experimental character, but theoretical, and presumably that is why it was not elaborated in the heavily experimental “Cathode Rays” paper. Fortunately Thomson had already elaborated the point in his presidential address the previous year to the Sixty-Sixth Meeting of the British Association for the Advancement of Science. There he has expressed his skepticism (1896, p. 702)

... also I think very difficult to account for the magnetic deflection of the rays. Let us take the case of a uniform magnetic field: the experiments which have been made on the magnetic deflection of these rays seem to make it clear that in a magnetic field which is sensibly uniform, the path of these rays is curved; now if these rays were due to ether waves, the curvature of the path would show that the velocity of propagation of these waves varied from point to point of the path. That is, the velocity of propagation of these waves is not only affected by the magnetic field, it is affected differently at different parts of the field. But in a uniform field what is there to differentiate one part from another; so as to account for the variability of the velocity of wave propagation in such a field? The curvature of the path in a uniform field could not be accounted for by supposing that the velocity of this wave motion depended on the strength of the magnetic field, or that the magnetic field, by distorting the shape of the boundary of the negative dark

¹⁰ He shows experimentally that the electric charge is deflected with the rays. Hence they are not merely a distracting secondary effect--“no more to do with the cathode rays that a rifle-ball has with the flash when the rifle is fired.” (p. 294) He corrects Hertz’s experimental result that cathode rays are undeflected by electric fields by repeating the experiment more carefully. (p. 296).

space,^[11] changed the direction of the wave front, and so produced a deflection of the rays.

Thomson here issues a quite fundamental challenge to the wave theorists. The widely recognized experimental fact of cathode rays is that they are deflected by magnetic fields. The standard mechanism through which waves are deflected is refraction, as manifested by light. When a light wave moves through a medium in which its speed becomes variable, the wave is bent. The amount of bending is recovered by the Huygens construction of elementary wave optics. The most familiar example is the bending of a light ray striking the surface of lens. The effect results fully from the difference of the speed of light in air and glass. It is faster in less dense air and slower in more dense glass.

Lenses alter the direction of light propagation abruptly. A gradual deflection arises with the phenomenon of mirages. Air closer to a heated desert surface is less dense than air at higher altitudes. So the speed of light is faster closer to the ground. The effect is that light grazing the desert surface is deflected upwards. Someone looking at the deflected light sees the blue of the sky but coming from the direction of the ground. The resulting illusion of water is a mirage.

Figure 1 shows how the bending arises. Light propagates from left to right. The wavefront AA' is vertical. Since the wavefront's speed is faster closer to the ground, the subsequent wavefront BB' has been turned upward.

¹¹ In an incompletely evacuated cathode ray tube, there is a dark space in front of the cathode before the cathode rays strike the gas in the tube and trigger light emission. I have been unable to discern precisely Thomson's argument concerning it.

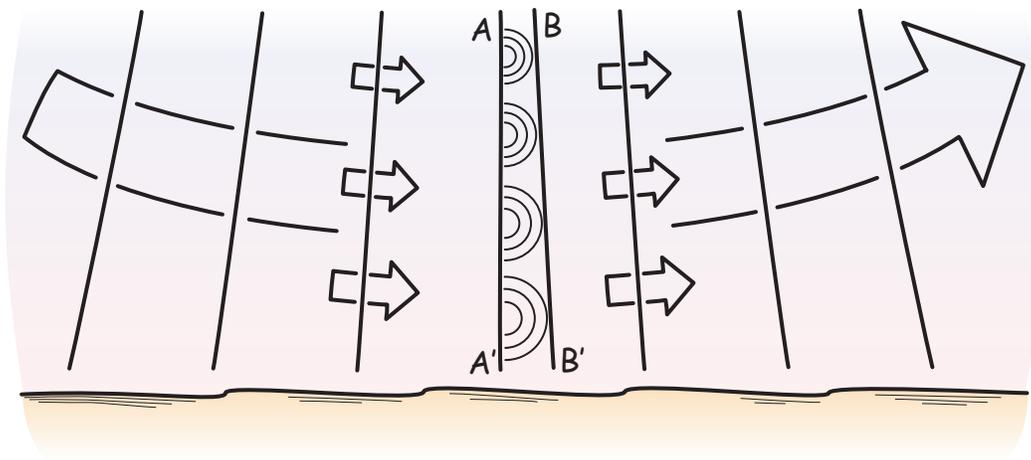


Figure 1. A Mirage: Refraction Bends Light

Thomson's point is that the refractive bending of waves depends essentially on differences in wave speed at different places. A uniform magnetic field, however, is the same everywhere. Hence, Thomson maintains, the effect it has on cathode ray wave propagation must be the same everywhere. There can be no differential alterations in the wave speed and thus no bending of the ray by diffraction. This conclusion would continue to hold even if we allow that the magnetic field might, in some circumstances, induce anisotropic speeds of propagation on the wave; that is, speeds that are different in different directions. Such anisotropy can arise for light propagation in anisotropic media. The corresponding anisotropy cannot arise here, however. The cathode rays are deflected in a plane perpendicular to the uniform magnetic field. The uniform magnetic field is isotropic in this plane.

The charged particle view of cathode rays has no trouble bending the rays. If the charges in the rays have the same initial velocity and start perpendicular to the direction of a uniform magnetic field, then the charges are deflected into a circular orbit in a plane perpendicular to the direction of the field, as shown in Figure 2. This is the "circular course" mentioned by Thomson (1897, p. 293).

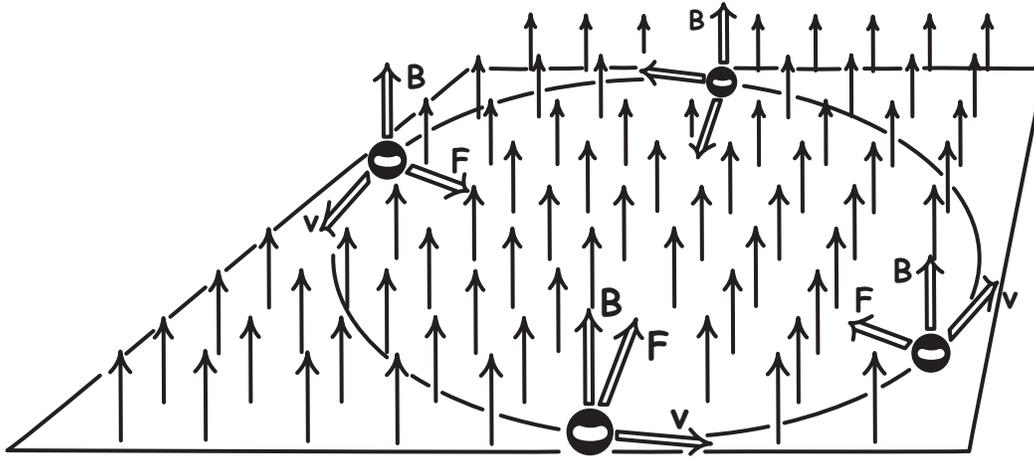


Figure 2. Moving Charge Deflected by a Uniform Magnetic Field

The electrodynamical details, for those who want them, are simple. The force \mathbf{F} on a charge e of mass m moving at velocity \mathbf{v} in a magnetic field \mathbf{B} is

$$\mathbf{F} = m\mathbf{a} = e \mathbf{v} \times \mathbf{B}$$

where the force produces acceleration \mathbf{a} . This acceleration is orthogonal to the direction of the \mathbf{B} field,¹² so the trajectory will remain in a plane perpendicular to the direction of the \mathbf{B} field. The acceleration \mathbf{a} is also orthogonal to the velocity \mathbf{v} ,¹³ which entails that the scalar speed $v = |\mathbf{v}|$ is constant.¹⁴ Since the scalar acceleration a and scalar speed v remain the same in a uniform \mathbf{B} field, the curvature of the trajectory must be the same everywhere; that is, it is a circle.

6.4 “The Dilemma of Accelerated Molecules and Ether Processes” Resolved

In 1906 Thomson was awarded a Nobel prize for “his theoretical and experimental investigations on the conduction of electricity by gases.” The year before, Lenard was awarded a Nobel prize “for his work on cathode rays.”¹⁵ In his Nobel Prize lecture

¹² Since $\mathbf{a} \cdot \mathbf{B} = (e/m) (\mathbf{v} \times \mathbf{B}) \cdot \mathbf{B} = (e/m) \mathbf{v} \cdot (\mathbf{B} \times \mathbf{B}) = (e/m) \mathbf{v} \cdot \mathbf{0} = 0$.

¹³ Since $\mathbf{a} \cdot \mathbf{v} = (e/m) (\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v} = -(e/m) (\mathbf{B} \times \mathbf{v}) \cdot \mathbf{v} = -(e/m) \mathbf{B} \cdot (\mathbf{v} \times \mathbf{v}) = -(e/m) \mathbf{B} \cdot \mathbf{0} = 0$.

¹⁴ Since $(d/dt) v^2 = 2 \mathbf{v} \cdot (d\mathbf{v}/dt) = 2 \mathbf{v} \cdot \mathbf{a} = 0$. To maintain a circular course, we must neglect energy lost by radiation, else v will decrease with energy loss.

¹⁵ Nobel prize citations from

http://www.nobelprize.org/nobel_prizes/physics/laureates/1905/

(1906), Lenard conceded to Thomson, or at least appeared to concede. The lecture is a boisterous history of his work on cathode rays. He describes the apparently irresolvable dilemma posed by cathode rays prior to Thomson's celebrated work of 1897(Lenard, 1906, p.18):

For we knew already that the rays are processes in the ether and not material, so it had to appear as downright amazing, that nonetheless they mimicked accelerated, negatively electrified gas molecules so deceptively. Nothing known had led us out of this dilemma of accelerated molecules and ether processes;...

He then reported Thomson's experiments as decisive and announced the resolution of dilemma. (p. 19, Lenard's emphasis)

The solution of the dilemma therefore was this: The rays are not accelerated, electrically charged molecules, but they are simply accelerated *electricity*. Something we had never believed we had seen: electricity without matter, electric charge without charged bodies. We have found that, therefore, in cathode rays, as already placed in our hands. We have, in some measure, discovered *electricity itself*....

In short, Lenard is defending his long-standing denial that cathode rays are material processes. They are not matter, but pure electricity, an option not considered in the original analysis.

This was not Thomson's view. He did not offer his experiments as finally delivering "electricity itself." Rather the rays were matter, still, but in a new and very finely divided state (Thomson, 1897, p. 312):

Thus on this view we have in the cathode rays matter in a new state, a state in which the subdivision of matter is carried very much further than in the ordinary gaseous state: a state in which all matter--that is, matter derived from different sources such as hydrogen, oxygen, &c.--is of one and the same kind; this matter being the substance from which all the chemical elements are built up.

6.5 Electrons are Waves After All

While the nature of cathode rays seemed secure in the wake of J. J. Thomson's celebrated experiments, the success was short-lived. With the coming of quantum mechanics, electrons were identified as having a dual wave- and particle-like character. The wavelike character of electrons was affirmed experimentally by Davisson and Germer (1927). They found that cathode rays, scattered off a crystal of nickel, produced diffraction patterns. The wavelengths of the associated waves conformed with the quantum formula for de Broglie waves.¹⁶

Thus Thomson's abduction arrived at the wrong conclusion. I state this *not* to impugn Thomson's abduction. It is as good as any. Rather my point is that his inference arrived at the wrong result, because it is dependent completely on background assumptions that proved to be incorrect. This can happen with any inductive inference, for they all depend on background assumptions. The material theory requires that dependence for all inductive inferences. The presence and importance of the background assumptions become quite visible, however, when we try to diagnose where the induction went astray.

In Thomson's case, the fatal intermediate conclusion was that a propagating wave could not *also* be deflected by a magnetic field in just the same way as a beam of charged particles. In quantum theory, neglecting spin, an electron can be represented by the same Hamiltonian as is used for an electron in classical physics. In the quantum case, this Hamiltonian is inserted into the Schrödinger equation to provide an account of an electron as a propagating wave. A standard theorem in quantum theory, Ehrenfest's theorem, assures us that the electron wave is deflected by electromagnetic field just as are classical electrons, as long as the wave packet of the quantum electron is confined to a small region in which the electromagnetic field does not appreciably change. Hence quantum electron waves will also be able to traverse Thomson's "circular course" in a uniform magnetic field.

¹⁶ The elder "J. J." Thomson's son, "G. P." (George Paget), also conducted experiments of this type, affirming the wave character of electrons.

The details of Ehrenfest's theorem for the electromagnetic case are straightforward but tedious. Working through them provides no special illumination.¹⁷ A smaller observation gives a good sense of precisely which assumption ultimately brought grief to Thomson's abduction. He rejected the possibility that cathode ray waves could be deflected by a uniform magnetic field. The key assumption was that a magnetic field could only deflect the waves by the familiar mechanism of refraction, that is, by directly altering the velocity of propagation of the waves and having a different alteration in different parts of space. A uniform magnetic field could not do this, since its effects must be everywhere the same.

What Thomson had overlooked is that the magnetic field might couple to a propagating wave in other ways. Associated with each magnetic field \mathbf{B} is a vector potential \mathbf{A} by the relation $\mathbf{B} = \nabla \times \mathbf{A}$. The Schrödinger equation allows for the effects of magnetic fields on charged quantum particles by coupling the particles to the magnetic field through the vector potential.¹⁸ The \mathbf{A} field associated with a uniform \mathbf{B} field is shown in Figure 3.¹⁹

¹⁷ See Schiff (1968, pp. 177-79) for the derivation. The exact solution for the motion of charge in a uniform magnetic field is given in Landau and Lifshitz (1965, pp. 424-27), but it is unilluminating.

¹⁸ More precisely, using the "minimal coupling" prescription, the momentum operator $\mathbf{p} = -i(\hbar/2\pi)\nabla$ in Schrödinger's equation is replaced by $\mathbf{p} - e\mathbf{A} = -i(\hbar/2\pi)\nabla - e\mathbf{A}$.

¹⁹ If we align the constant \mathbf{B} field with the z -axis of a Cartesian coordinate system, so that $\mathbf{B} = (0, 0, B_z)$, then a compatible vector potential is $\mathbf{A} = -(1/2) \mathbf{r} \times \mathbf{B} = (1/2) B_z(-y, x, 0)$. Since \mathbf{A} is determined up to a gauge transformation only, other representations are possible, such as $\mathbf{A} = B_z(-y, 0, 0)$ in the "Landau gauge." The first however, displayed in Figure 2, preserves the rotational symmetry of the \mathbf{B} field about the z axis and conveys the handedness in the field.

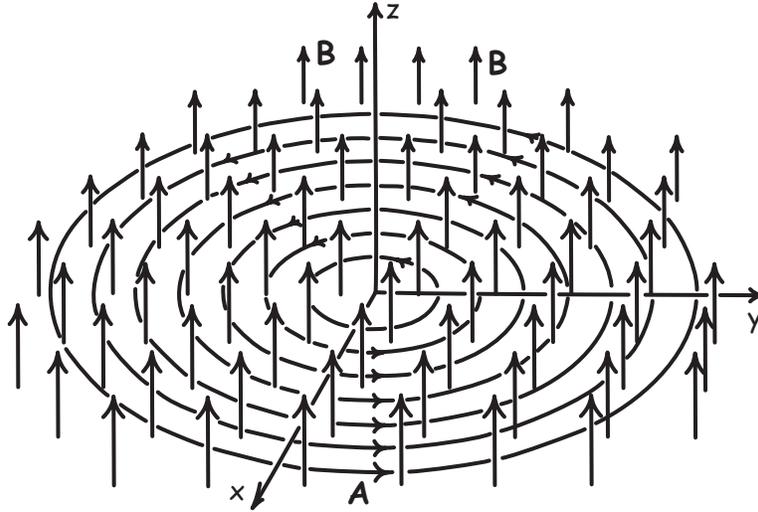


Figure 3. Vector Potential **A** Associated with Constant Magnetic Field **B**.

The integral lines associated with the vector field traces out a circle with the preferred direction of rotation in the figure. When a negatively charged quantum wave packet is coupled to a uniform magnetic field through this vector potential **A**, it will trace out circular trajectories with the same sense of rotation. Its speed, however, is unaltered by the **A** field.

The clue that such coupling is possible is present already in the classical analysis. For one might also ask how a uniform magnetic field can deflect a classical moving charge to the left or the right. If it is uniform, should not both directions be treated alike? They are not treated alike by the magnetic field, once a moving charge is present. A negative charge moving horizontally in an upward pointing magnetic field is deflected to the left, as shown in Figure 2. This derives from the magnetic field vector having what used to be called an “axial” character. That means that it changes sign under mirror reflection of space. (The cross product operator \times and curl operator $\nabla \times$ has similar transformational properties.) The vector potential **A** encodes more clearly how the magnetic field is prepared to deflect charged, moving particles. This preferred sense of rotation of **A** will be replicated by the velocity **v** of the deflected charge: the velocity **v** is linearly related²⁰ to the **A** field by $\mathbf{v} = -2(e/m)\mathbf{A}$. For a negatively charged electron, e is a negative number. Therefore **v** and **A** agree in direction and relative magnitude.

²⁰ For the classical particle, the scalar speed v satisfies $|e|vB_z = mv^2/R$, where R is the radius of curvature of the trajectory. Hence $v = (|e|/m) B_z R$. The two varying components v_x and v_y of the

7. Einstein and the Anomalous Perihelion of Mercury.

In November 1915, an exhausted Einstein was putting the finishing touches onto his general theory of relativity. It was the result of eight years of labor. The final three years had been tense. Einstein had settled upon and published an erroneous version of the theory in 1913. Over the next two years, he had alternated between confidence in the theory and despair over it until he finally found and resolved his errors. In the midst of this resolution, he also found that his theory accounted for a recalcitrant anomaly in planetary astronomy.

According to Newtonian gravitational theory, a planet orbits the sun in an elliptical orbit. The orbit is re-entrant. That means that, in each planetary year, the planet will trace out the same ellipse in space. This familiar results holds exactly only for a two body system of a very massive sun and a single planet. If other planets are present, their gravitational attraction will deflect the original planet's motion away from the re-entrant ellipse. In our solar system, these alterations are very slight and manifest as a very slow rotation of the ellipse of the planet's orbit. In the early twentieth century, careful calculations had accounted for nearly all these motions in the planets. The prominent exception was Mercury. The residual, unaccounted motion of the axis of its orbit was a rotation in the direction of the planet's motion. The planet's perihelion, the point of closest approach to the sun, had an unaccounted advance of roughly 40 seconds of arc per century.²¹

In Einstein (1915), in the passage quoted at the start of this chapter, Einstein reported with delight that his new theory calls for a slight correction to the Newtonian motions that matches exactly this anomalous motion of Mercury. It provides, as the title of the paper asserts, an "Explanation of the Perihelion Motion of Mercury from the General Theory of Relativity."

constant scalar speed v will oscillate harmonically as the charge orbits in a circle. If we locate the origin of the Cartesian coordinates at the center of this circle, we have $\mathbf{v} = (e/m) B_z (y, -x, 0)$, so that \mathbf{v} is function of position in space. Then $\mathbf{v} = -2(e/m)\mathbf{A}$ follows. Different initial positions and velocities for the charge will locate the center of the orbit elsewhere. The appropriately matched vector potential is recovered by a gauge transformation of the original \mathbf{A} field. To relocate the origin to $(x_0, y_0, 0)$, transform \mathbf{A} to $\mathbf{A}' = \mathbf{A} - (1/2) B_z (-y_0, x_0, 0) = -(1/2) B_z (-(y-y_0), (x-x_0), 0)$.

This is a gauge transformation since $\mathbf{B} = \nabla \times \mathbf{A} = \nabla \times \mathbf{A}'$.

²¹ The earlier history of this problem is discussed in the chapter on simplicity.

For our purposes, three aspects of Einstein's claims are important and are developed in the subsections that follow.

7.1 Mere "Confirmation" not "Inference to..."

First, Einstein and subsequent commentators do not carry out a complete inference to the best explanation. They claim only, as Einstein writes (1915, p. 831), "an important confirmation" of the theory. Born's popularization of relativity from the 1920s (1922, p. 254) says: "it [Einstein's theory] is thus already confirmed in advance by Leverrier's calculation [of Mercury's motion]."²² Pauli (1958, pp. 168-69), in his 1921 authoritative review article, is even more cautious. The question of Mercury arises as a "check by experiment" of consequences of Einstein's theory. The agreement of theory and observation constitutes "a great success."

All these affirmations noticeably fall short of an authorization to infer to the theory, as inference to the best explanation allows. The reason is not hard to find. There is no such authorization perceived in this result. The gap between the theory and observation is too great to be closed completely even by as striking a success as this.

Weyl (1921, p. 247) explains the evidential situation quite well in his celebrated *Space-Time-Matter*, after reviewing general relativity's success with Mercury and in two other astronomical tests.

...the actual deviations from the old theory are exceedingly small in our field of observation. Those which are measureable have been confirmed up to now. The chief support of the theory is to be found less in that lent by observation hitherto than in its inherent logical consistency, in which it far transcends that of classical mechanics, and also in the fact that it solves the perplexing problem of gravitation

²² The German is "Genau diesen Betrag aber fordert die Einsteinsche Theorie; sie ist daher durch Leverriers Rechnungen bereits im voraus bestätigt." Unfortunately the later English translation (Born, 1962, p. 348) mangles the German and translates this sentence as "But this is just the amount required by Einstein's theory. The confirmation of this result of Einstein's mechanics was therefore actually anticipated by Leverrier's calculation." That is, the translation mistakenly reports the predicted motion of Mercury confirmed, not the theory predicting it, as in the German.

and of the relativity of motion at one stroke in a manner highly satisfying to our reason.

While we now have more observational and experimental support for general relativity, I believe Weyl's assessment still applies well today. The strongest support for the theory derives from our aesthetic appreciation of the theory.

7.2 Preference for the Better Explanation

While the complete “inference to...” is absent, what is present in this example is a quite thorough implementation of the comparative step: the preference for the better explanation. This is embodied in two facts recognized in the literature. First, all other explanations of Mercury's anomalous motions on offer in the literature had been contradicted by the evidence. Second, other explanations might be possible. However the suggestion was these other explanation would likely take on undischarged evidential debt, by, for example, introducing parameters with arbitrarily set values. Einstein's explanation was distinctive in not requiring any arbitrary parameters.

When Einstein announced his successful explanation of Mercury's anomalous motion, it was very convenient that his colleague, the astronomer Erwin Freundlich, had just published an extensive survey of the problem of Mercury's anomalous motion. Einstein (1915, p. 831) cited Freundlich's account in a footnote to this announcement as support for the failure of Newton's theory to offer an explanation of Mercury's anomalous motion:

E. Freundlich has recently written a noteworthy paper (Astr. Nachr. 4803, Bd. 201 June 1915) on the impossibility of satisfactorily explaining the anomalous motion of Mercury on the basis of the Newtonian theory.

Freundlich's paper listed four ways the astronomers had then tried to explain the anomalous motion. He concluded that none succeeded. That is (1915, p. 51):

...in the explanation of the existing contradiction between theory and experiment, we have progressed no further than since the time of Newcomb.

His final, concluding sentence (p. 56) is:

How the anomalies of these inner 4 planets really come about has unfortunately up to now not been answered thoroughly.

Freundlich cited Simon Newcomb, whose study (1895) of the motion of the four inner planets was then authoritative. Newcomb's (1895, Ch. VI) provided an extensive examination of various hypotheses advanced to explain the anomalous motion of Mercury and for smaller anomalies in the other inner planets. Freundlich then provided an update.

The first candidate was the supposition of as yet unknown planets between the sun and Venus. Freundlich deferred to Newcomb's (1895, pp. 112-115) analysis where he considered the possibility of a single planet or multiple planets in a ring. He was unable to find a suitable configuration that would accommodate the known anomalies. The celebrated but failed supposition of the nineteenth century, of a single new planet, Vulcan does not even bear mention by name. The possibility is dismissed by Newcomb with a casual (p. 115) "But I conceive that a planet of the adequate mass could not have remained so long undiscovered."

The second candidate was of a flattening of the sun, presumably as a result of its rotation. The deviations from sphericity would then lead to gravitational effects that could explain the anomalies. The possibility was ruled out, however, since the flattening would have to be much greater to get the desired effect than is compatible with observations of the sun.

The third candidate was a proposal by Asaph Hall (1894) that the force of gravity might not dilute with distance r as an inverse square $1/r^2$ but very slightly faster as $1/r^{2+\delta}$ where δ is a very small number. Newcomb reports that $\delta = 0.0000001574$ would suffice to create the anomalous advance of Mercury's perihelion. The proposal fails, Freundlich notes, since a value of δ sufficiently large to accommodate Mercury's anomalous motion produces effects in our moon's motions that are incompatible with observation and the then successful theory of Brown for the moon's motions.

The fourth candidate was a proposal by Seeliger. The zodiacal light is a halo of light around the sun. It is presumed due to some diffuse distribution of matter that extends as far as the orbit of Mars. The proposal was that the gravitational action of the matter in this halo might account for the anomalous motion of Mercury. The principal content of Freundlich's (1915) paper was to show that this possibility contradicted other evidence of the zodiacal light. His analysis was complicated. Merely finding the mean

density of the postulated distribution was not enough. Non-uniformities made a difference. Matter within the orbit of Mercury would produce an advance in the planet's motion; and matter outside its orbit would retard it. Freundlich compared the sorts of densities of matter needed and their distribution with other possible properties of the zodiacal light, including how a distribution of massive dust might impede the motion of the planets, including the Earth. His final conclusion was these other properties forced a much smaller density of matter in the zodiacal light than needed to account for the anomalous motion of Mercury.

In sum, at the start of 1915, all concrete proposals for accounting for the anomalous motion of Mercury had been contradicted by further evidence. Freundlich's analysis leaves open the possibility that there might still be some as yet undiscovered account that explains the anomalous motion of Mercury. There proved to be one theory that could do this. Freundlich's paper was written shortly before Einstein perfected his theory and discovered that it accounted for the anomalous motion of Mercury. Might there be still others? Neither Einstein nor responsible commentators at that time asserted flatly that no other theory could accommodate the anomalous motion of Mercury. However they commonly pointed to a single feature of Einstein's explanation that they deemed of great significance.

Other accounts of the motion of Mercury had all required additional suppositions. If extra masses were invoked, their positions and distributions in space needed to be specified. If alterations to Newton's inverse square law of gravity were invoked, then the alterations would add extra parameters, such as Hall's δ above. Einstein's theory, however, required no such additional hypotheses or parameters. Einstein (1915) points to this at the outset with his remark that the explanation succeeds "...without having to posit any special hypotheses." Pauli (1958, p. 169) notes:

Compared with Seeliger's explanation, Einstein's has at least the advantage that no arbitrary parameters are needed.

Born (1922, p. 254; 1962, p.348) also remarks:²³

²³ "Dieses Resultat ist von ausserordentlichem Gewichte; denn in die Einsteinsche Formel gehen keine, neuen, willkürliche Konstanten ein..." This time, the English translation (Born, 1962, p. 348) is accurate.

This result is of extraordinary importance; for no new arbitrary constants enter into Einstein's formula...

Just how does this feature of Einstein's theory come to favor it? They do not say. However, among the ideas developed in this chapter, there is an obvious reading. The introduction of extra, arbitrary parameters or constants is the taking on of an evidential debt. One must eventually provide independent evidence for them, just as one must find independent evidence that there is a planet Vulcan perturbing the motion of Mercury. Until that is done, Einstein's explanation is better supported in the sense that it has no such undischarged evidential debt.²⁴

Hence I take the repeated remark to suggest that any other explanation of the anomalous motion of Mercury is likely to need such extra arbitrary parameters and thus to be weaker than Einstein's. That is, we should not expect a serviceable competitor to Einstein's theory to emerge sooner or perhaps even later. This oblique suggestion is far from a clearly asserted advance from the comparative Step 1. to the absolute Step 2., that is from a preference for the better to the inference to the best. It merely gestures in that direction.

7.3 Why Loveliness as an Explanatory Virtue is Overrated

This particular example enables us to mount an interesting test of a core motivation of inference to the best explanation. The idea is that successful explanations gain inductive support because there is something special in the explanatory relation itself. We saw above that Lipton identified explanatory virtues that would underwrite an inference to the *loveliest* explanation. Of all theories in modern physics, general relativity is distinctive in the praise it receives for its immense conceptual simplicity and scope. It is, by any measure, a *lovely* theory. So we might expect that the inductive support it accrues from its account of Mercury's motion would derive from this loveliness.

We can see quite quickly that loveliness has little to do with the support it accrues. That support depends almost entirely on the failure of competing theories to account for Mercury's anomalous motion. A simple thought experiment reveals just how little the loveliness matters. Imagine that, contrary to history, the nineteenth century astronomers did discover a new planet

²⁴ For an account that does not employ the notion of undischarged evidential debt, see Norton (2011).

Vulcan in just the place expected from Mercury's anomalous motion. The discovery would be celebrated as a great triumph of Newtonian physics. It would be a replication of the great success of the discovery of Neptune on the basis of then anomalous motions in the planet Uranus.

In this thought experiment, the tables are turned. The Newtonian theory strains initially to explain the anomaly by taking on the evidential debt of a supposition of a hitherto unseen planet. The Newtonian theory is at a disadvantage. When independent, optical observation finds the planet, however, the evidential debt is discharged and the Newtonian theory prevails. General relativity, however, now finds itself in great difficulty. For the anomaly in Mercury's motion has disappeared, but general relativity still requires an additional advance of the perihelion of 43 seconds of arc per century, beyond what is predicted by the fullest Newtonian account. The observed motion of Mercury, in this fable, now threatens to refute general relativity.

Of course were this fable really to have happened, it is unlikely that this one misadventure would have overturned general relativity. The overall decision would come from a balancing of a greater body of evidence. Mercury's observed motion would weigh against the theory and the loveliness of its treatment of Mercury would have no inductive import at all.

There is a real coda to this fictional tale. In 1918, Hermann Weyl extended Einstein's general theory of relativity in a manner quite in keeping with loveliness of Einstein's original theory. Einstein's theory has incorporated gravity into the metrical structure of spacetime. Weyl now elaborated that structure slightly to allow it to incorporate electromagnetism as well. Einstein was enthusiastic about the theory as piece of theory and praised it strongly to Weyl in correspondence. However Einstein also saw an empirical problem. According to Weyl's theory atomic emission spectra could not retain sharp lines, in contradiction with experience. The loveliness of the theory could not overcome the observational problem and Einstein opposed the theory.²⁵

In sum, loveliness figures prominently in our thought about theories. But its importance is overrated. What matters more is the evidential failures of competitors and, if one's own theory suffers such a failure, loveliness cannot rescue it.

²⁵ For a brief account of this episode, see Norton (2000, pp. 153-54)

8. Cosmic Background Radiation

The discovery in the 1960s that our universe is permeated with a 2.7 degree Kelvin bath of thermal radiation seems tailor made for an abductive inference. For the thermal radiation is readily explained as the residue of the intense heat radiation of the big bang origin of the universe. The fit is so natural that I used it in the opening section of the last chapter to introduce and motivate the idea of inference to the best explanation. It looks like a safe example for philosophy of science textbooks. Hacking (2001, p. 16) gives it in a paragraph, headed *Inference to the Best Explanation*:

Each of the arguments we've just look at is an *inference to a plausible explanation*.

If one explanation is much more plausible than any other, it is an *inference to the best explanation*.

Many pieces of reasoning in science are like that. Some philosophers think that whenever we reach a theoretical conclusion, we are arguing to the best explanation. For example, cosmology was changed radically around 1967, when the Big Bang theory of the universe became widely accepted. The Big Bang theory says that our universe came into existence with a gigantic "explosion" at a definite date in the past. Why did people reach this amazing conclusion? Because two radio astronomers discovered that a certain low "background radiation" seems to be uniformly distributed everywhere in space that can be checked with a radio telescope. The best explanation, then and now, is that this radiation is the result of a "Big Bang."

This compressed account makes the inference look all but instantaneous, much as we infer instantly that the slender cables just glimpsed explain the magician's levitation.

The reality of the example is more complicated in two ways.

8.1 The Thermal Character of the Radiation

First, the discovery of the cosmic background radiation is routinely attributed to work by Arno Penzias and Robert Wilson (1965) and announced in 1965. They found residual radiation with a cosmic source while measuring radio waves bounced off balloon satellites. While this is celebrated as the moment of discovery, merely finding cosmic radiation is not the inductively potent result. For charged matter is posited in all cosmological theories and such matter readily

produces electromagnetic radiation. Without it, the stars cannot shine in the electromagnetic spectrum. To distinguish among the theories, a more distinctive property is needed. That distinctive property is that the radiation has a thermal character with a black body spectrum and, in this case, with the temperature of 2.7 degrees Kelvin, that is 2.7 degrees above absolute zero.

That there should be such thermal radiation was long suspected by cosmologists who worked with the idea of a “big bang” or, as they then preferred to call its radiative part, the “primeval fireball.” They included the physics research group of Dicke, Peebles, Roll and Wilkinson, working at Princeton University, not far from Penzias and Wilson’s Crawford Hill Laboratory in New Jersey. The group had begun its own efforts to detect the thermal radiation, only to find itself scooped by Penzias and Wilson’s chance discovery.

Penzias and Wilson had measured the cosmic radiation at one wavelength only, 7.4 cm. While their results were compatible with black body radiation of a temperature 3.5K +/- 1.0K, it did not establish it. What was needed were measurements taken across a larger range of wavelengths or frequencies to show that the distribution of radiant energy across the range matched the quite precise functional form of the black body curve.

The early history is filled with collections of reports of measurements aiming at establishing this match. Weinberg’s (1972) text includes a table (Table 15.1, p. 512) with reports of 31 measurements of various types. He still finds (pp. 516-517) that the discrimination between black body and gray body radiation rests entirely on one type of mountain top radiometer measurement; and that these are contradicted by rocket and balloon borne measurements.

This difficulty, in addition to the second concern below, allowed only a cautious celebration of the result. Weinberg (1972, p. 506) could give only a begrudging summary report: “It is widely, though not unanimously, believed, that the microwave radiation background discovered in 1965 is just this left-over radiation...”

The evidential difficulties were eventually resolved. The definitive results were delivered by NASA’s COBE satellite. As an index of the completeness of resolution, we can note that Weinberg’s (2008) text leads with the COBE results in the first paragraph of its Preface (p.v):

November 1989 saw the launch of the Cosmic Background Explorer Satellite. Measurements with its spectrophotometer soon established the thermal nature of the cosmic microwave background and determined its temperature to three decimal places, a precision unprecedented in cosmology.

8.2 Competitors²⁶

The second difficulty is that even a thermal spectrum is still not quite distinctive enough to be instantly diagnostic of a primeval fireball. The trouble is that a thermal spectrum arises whenever radiation comes to thermal equilibrium; and there may still be other ways that this spectrum can arise. It is too easily gained.

Once again, this difficulty permitted only measured statements of enthusiasm over the result. Partridge wrote a celebratory survey for the Spring 1969 issue of *American Scientist*. There the cosmic background radiation was offered as something a little less than definitive proof of the big bang, but merely a “new parameter” (1969, p. 39):

The paucity of data in cosmology explains the excitement generated by the discovery of the cosmic microwave background, which we identify with the primeval fireball in which the Universe originated. The expansion of the Universe has now cooled the fireball to a few degrees Kelvin. Measurements of this isotropically distributed microwave radiation have given us a new parameter in cosmology, the temperature of the radiation field, and also one of the most accurate results of observational cosmology, a figure for the isotropy of the radiation field.

Big bang cosmology has the least difficulty in recovering the thermal spectrum. Even there, the recovery is indirect. In the very early universe, matter and radiation come to thermal equilibrium and a thermal spectrum is thus imprinted on the background radiation. However, as the universe expands, matter and radiation eventually decouple. This happens quite early, when the cosmos has cooled to around 3000K. The photons comprising the cosmic background radiation we measure have propagated to us, unimpeded, from this era. Their origins lie in a distant spherical shell surrounding us, the surface of last scattering. The trouble is that these photons have been underway for much of the history of the entire universe. During that time, their frequencies have been greatly reduced by the cosmological redshift that in turn derives from the expansion of space. Will the greatly red shifted distribution still be thermal? A short calculation and some reasonable assumptions, such as given in Weinberg (1972, pp. 506-507),

²⁶ The discussion here barely touches the range of alternative accounts of the cosmic background radiation that arose in the decades following Penzias and Wilson’s measurements. For a survey, see Čirković and Perović (2018).

show that the effect of the redshift is to preserve the thermal character of the radiation while merely reducing its temperature.

That big bang cosmology can eventually accommodate the thermal spectrum of the cosmic background radiation is not decisive. A long section in Partridge's (1969) survey ("B. But Is It the Primeval Fireball?") grapples with the question of whether the cosmic background radiation could arise by other means. Partridge reviews three other mechanisms. In one, short-lived proposal, Kaufman had sought the radiation in emissions from hot intergalactic plasma. Another due to Layzer, posited as the source dust grains heated during galaxy formation.

Partridge's longest analysis was given to proposals generated in the context of steady state cosmology, then the major competitor to big bang cosmology. This alternative cosmology proposed that the universe has maintained its present state on the large scale for all infinity of time. The universe now and the universe any time in the infinite past look much the same. Steady state cosmology was most directly threatened by the discovery of the cosmic background radiation. For background radiation could not be preserved in a steady state within a universe that has been expanding for infinite time. The cooling and diluting effect of the expansion would eradicate it.

Proponents of the steady state theory, Hoyle, Narlikar and Wickramasinghe, rose to the challenge and sought to account for the radiation within their theory in terms of the reradiation of starlight from interstellar grains. Partridge found severe difficulties for the proposal. Nonetheless, his assessment of the overall evidential situation was qualified to the point of awkwardness as "personal bias" (1969, p. 43):²⁷

Also, it is only fair for me to announce my personal bias in advance: I believe the fireball picture to be consistent with all the experimental data, and to be the simplest theoretical explanation of these data. In making this judgment, and in writing this section, I have kept in mind four questions. Can the suggested model for the background radiation explain its intensity? Can it explain the observed spectrum? Can it explain the isotropy of the radiation? And finally, does it survive the cutting edge of Ockham's razor: is it simple, useful, and not *ad hoc*?

²⁷ I have also given this quote at greater length since it is the only place in reading these early sources in which I found notions of explanation entering explicitly with the word "explanation."

For our purposes, the important point is that the assessment is comparative. It is restricted completely to *Step 1. Preference for the Better Explanation*.

In subsequent literature, the assessments remained comparative, but the comparisons quickly reduced to standard big bang cosmology versus just the flagging steady state theory. Peebles' 1971 text, *Physical Cosmology*, retains Partridge's hesitancy. In a list of eight points of evidence for cosmology, the sixth reads (1971, p. 26):

(6) The Universe may contain a Primeval Fireball, blackbody radiation left over from a time when the Universe was dense and hot (Chapter V). If this is substantiated by further measurements it will be direct evidence that the Universe really is expanding and growing less dense, in agreement with the Lemaître cosmology (but not the original Steady State model).

Twenty years later, doubt about the thermal character of the background radiation had gone. However Peebles still made the case for the evidential bearing of the measurements comparatively. The evidence favors big bang cosmology because no other account can accommodate it; and the only other account considered is its old nemesis, steady state cosmology. Peebles (1991, p. 19) wrote:

The thermal form of the spectrum of this radiation is considered to be almost tangible evidence that the universe expanded from a state considerably denser than it is now, because it is exceedingly difficult to see any other way to make the spectrum so close to thermal. Consider for example the classical Steady State theory, in which the mean density of the universe is constant in time...

Peebles proceeded to dismantle the mechanism through which the steady state theorists sought to replicate the measured cosmic background radiation. Radiation, it is supposed, is created cosmically along with baryons in the steady state cosmology. Its spectrum is shifted to a thermal spectrum by absorption and reemission. This absorption corresponds to a certain degree of opacity of space. But the degree required directly contradicts the observed transparency of space.

The same, comparative assessment is repeated at greater detail in Peebles' later 1993 authoritative text, *Principles of Physical Cosmology*, pp. 203-206. He concluded for the absorptive mechanism (1993, p. 204):

The point of this calculation is that if the universe were postulated to be opaque enough at radio wavelengths to have caused the radiation background to relax to the

observed very nearly thermal spectrum of the CBR, space would be predicted to have been too opaque to have allowed the observations of distant radio sources... Peebles then examined the character and density of dust needed for the relaxation mechanism with results once again unfavorable to the proposal.

This comparative assessment seems now to have acquired the status of the standard textbook formulation of the evidential import of the cosmic background radiation. Here it is in a more recent cosmology textbook: (Liddle, 2003, p. 80)

The Hot Big Bang theory therefore gives a simple explanation of this crucial observation. In the Steady State theory, all radiation is supposed to originate in stars and so is at high frequency and is not a perfect black-body; one has to resort to a thermalizing mechanism such as whiskers of iron, which somehow managed to thermalize this into low-energy radiation in the recent past without preventing us from seeing distant objects. It has never been satisfactorily demonstrated that this can be achieved even allowing the *ad hoc* assumptions that the Steady State scenario requires.

8.4 Success Through Failure of the Competitors

The measurements of the cosmic background radiation do provide good evidence for big bang cosmology. This review brings into sharper focus how they do it. The accounts above of the success identify no special explanatory relation beyond mere accommodation. Big bang cosmology, with suitable auxiliary assumption, entails the existence of a thermal radiation background. Beyond this accommodation, there is no special explanatory coup through which we can make some philosopher's notion of explanation central to the evidential relation. Thermal radiation is something that can arise easily in any account that hosts energized charged matter and sufficient time for thermal equilibrium to be established. Nothing in the analysis provided by big bang cosmology indicates that it is the only theory that can accommodate the result.

Nonetheless, this exclusivity does turn out to be evidentially decisive. It is not established by examining how big bang cosmology explains the cosmic background radiation. Rather it is established by examining how competitors to big bang cosmology fail to accommodate the result. The decisive fact is not so much about big bang cosmology, but about its competitors. Big

bang cosmology can accommodate the result, where no known competitor can. Big bang cosmology wins the day by default.

The explicit discussion of evidential import is restricted to this comparative result, fully within *Step 1. Preference for the Better Explanation.* of the present account. The second step, acceptance that the evidence supports big bang cosmology specifically and absolutely, is left tacit. That big bang cosmology bests its strongest competitor, the steady state theory, is stressed and, presumably, this victory is intended to lead us to believe that there is no better alternative possible.

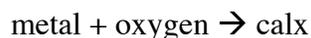
In any case, over half a century after Penzias and Wilson's observation, the origin of the cosmic background radiation is no longer open to serious dispute in the cosmology literature: it is described without apology or qualification as a thermal residue of an early hot universe. Serious consideration is now given to the slight deviations from isotropy in the radiation, for they are now the key to understanding structure formation in cosmology.

9. Oxygen and Phlogiston

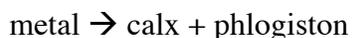
9.1 *The Theories Compete*

The establishment of Lavoisier's oxygen chemistry has been presented as a canonical instance of inference to the best explanation. A closer look will show that an intrinsic explanatory virtue had little to do with the establishment of the theory. Rather the decisive inferences of both Step 1 and Step 2 were warranted by a quite specific fact: that matter has weight.

Oxygen chemistry ascended in the late eighteenth century, when Lavoisier's oxygen theory competed with the phlogiston theory as the correct account of many chemical processes. Combustion illustrates the competition. The oxygen theory portrayed the combustion of a metal as its combination with oxygen from the air to form an oxide, then commonly called a "calx."



The phlogiston theory took all metals to be a compound of a calx and phlogiston; and the combustion of a metal to be the decomposition of this compound into a calx and liberated phlogiston.



There was a close similarity of structure in oxygen and phlogiston chemistry. Just about any reaction accommodated by one was mirrored by a corresponding reaction in the other. To see how the reactions of each theory pair up, you merely need to think of phlogiston as a kind of “anti-oxygen.” Then you can convert a reaction of oxygen chemistry into one of phlogiston chemistry and vice versa. In the phlogiston combustion reaction, for example, substitute anti-oxygen for phlogiston; and then move it from the right-hand product side to the left-hand reactant side, dropping the “anti” prefix. What results is the oxygen combustion reaction. Much of oxygen and phlogiston chemistry were mirror images of each other.

Thagard (1978) presents the triumph of oxygen theory as a canonical case of inference to the best explanation. He quotes his translation of a confident assertion by Lavoisier (pp. 77-78) in support:

I have deduced all the explanations from a simple principle, that pure or vital air is composed of a principle particular to it, which forms its base, and which I have named the *oxygen principle*, combined with the matter of fire and heat. Once this principle was admitted, the main difficulties of chemistry appeared to dissipate and vanish, and all the phenomena were explained with an astonishing simplicity.

While we know that, in the long run, oxygen will win, the situation at the time of the debate was not so clear. Precisely because oxygen and phlogiston chemistry were, to a large measure, intertranslatable, the two theories had considerable overlap in scope. It was not clear that oxygen’s explanatory powers were greater. Thomas Kuhn made this fact a celebrated debating point in the question of the cumulativeness of science, when he used it to illustrate what is now called “Kuhn loss” (1996, pp. 99-100):

The much-maligned phlogiston theory, for example, gave order to a large number of physical and chemical phenomena. It explained why bodies burned--they were rich in phlogiston--and why metals had so many more properties in common than did their ores. The metals were all compounded from different elementary earths combined with phlogiston, and the latter, common to all metals, produced common properties. In addition, the phlogiston theory accounted for a number of reactions in which acids were formed by the combustion of substances like carbon and sulphur. Also, it explained the decrease of volume when combustion occurs in a confined

volume of air the phlogiston released by combustion “spoils” the elasticity of the air that absorbed it, just as fire “spoils” the elasticity of a steel spring.

Whatever its other explanatory virtues, oxygen chemistry could not provide an explanation for the common properties of metals, as could phlogiston chemistry.

9.2 Weight and Levity

What turned the tide in oxygen’s favor and formed the basis of Lavoisier’s case for oxygen was weight. When a metal burned to form a calx, the calx weighed more, while the air above lost 1/6th of its volume; and when the calx was reduced back to metal, in the case of mercury calx, it lost weight and returned just the missing portion of air. These gains and losses of weight could be explained by the phlogiston theory if we assume that phlogiston had negative weight, that is, “levity” as opposed to “gravity.” It now seems a curious assumption, but it saves the phenomena. When a metal forms a calx, it loses the levity of phlogiston. This is a loss of a negative weight. Taking away a negative has the effect of adding a positive. It results in a calx that weighs more than the metal.

Phlogiston chemistry fails if we deny the admissibility of levity and insist on the background fact that matter must have weight. John Herschel summarized the failure, writing a few decades later (1840, p. 301):

So far as weight is concerned, it makes no difference whether a body having weight enters, or one having levity escapes; but there is this plain difference in a philosophical point of view, that oxygen is a real producible substance, and phlogiston is no such thing: the former is a *vera causa*, the latter an hypothetical being, introduced to account for what the other accounts for much better.

More picturesquely, Herschel characterized the question of weight as the crucial factor in deciding between the two: (p. 300; Herschel’s emphasis)

...of two possible roads the wrong was chosen; and a theory obtained universal credence on the credit of great names, ingenious views, and loose experiments, which is negatived, *in every instance*, by an appeal to the balance.

His language is reminiscent of Bacon's "crucial instances," which Bacon had described with an analogy to signposts directing us at branches in a road.²⁸

Herschel's account leaves unsupported his conclusion that levity-bearing phlogiston cannot be a real substance. William Whewell (1847, pp. 409-11) lays out a more elaborate case. He too based the decision in favor of oxygen chemistry in this fact about matter. The levity of phlogiston was "rejected by all the sounder philosophers," he wrote, and "It is assumed, it appears, that all matter must be heavy..." He proceeded to attempt a quite general argument that deduces the heaviness of matter from the very idea of substance. One part of his argument returns to phlogiston:

For if weight is not the criterion of the quantity of one element, phlogiston for instance, why is weight the criterion of the quantity of any other element? We may, by the same right, assume any other real or imaginary element have levity instead of gravity; or to have a peculiar intensity of gravity which makes its weight no index of its quantity.

We can now reassess just how the decision in favor of the oxygen theory was taken. While Lavoisier had boasted of the explanatory prowess of this oxygen theory, at the time of the decision there was little to choose between the explanatory capacities of oxygen and phlogiston chemistries. What was decisive, however, was a fact: matter has weight. That fact was compatible with oxygen chemistry but not with phlogiston chemistry, in so far as phlogiston was supposed to be material.²⁹

Once again we see the two-step structure emerging for the inference. The first step is comparative between oxygen chemistry and the foil of phlogiston chemistry. The decision is not derived from some superior, intrinsic explanatory virtue in the favored oxygen explanation. Rather, the foil is rejected because of a logical incompatibility with a background fact: matter has weight. Oxygen chemistry thereby prevails.

This same fact mediates in the second step: that we should not just prefer oxygen chemistry over phlogiston chemistry, but that we should infer to and accept oxygen chemistry.

²⁸ A portrait of Francis Bacon is on the title page of Herschel's *Preliminary Discourse*.

²⁹ I pass over the rather great awkwardness for Lavoisier that he had also allowed caloric, the matter of heat, into his table of elements, even though no sensible weight for it had been found.

There are many component inferences and further factual assumptions required for the second inference. But the course of each component is unremarkable. The full accounting would need to look at many different chemical changes. Here is how one proceeds. When a metal calx transforms into a lesser weight of metal and a released gas, we read directly that this is a decomposition of the calx into its constituent metal and the gaseous component, oxygen. The further assumptions needed to make this inference from the observation a few instances to the generality would include: that matter is conserved, so that any weight lost must reappear in the matter of the gas; and that the calx is a pure substance all of whose samples have the same properties. Then the behavior of one sample can stand for all.

10. The Wave Theory of Light

Darwin (1876, p. 421) indicated the wave theory of light (“undulatory theory”) as one established by the same abductive methods as he used in *Origin of Species*. Thagard (1978, pp. 77-78) includes it as one of his canonical scientific examples of inference to the best explanation. As a result, one might expect that it would be straightforward to reconstruct the abductive inference. Matters prove otherwise. The wave theory evolved slowly into its modern form, only gradually acquiring evidential support in a temporally extended process of great complexity. While the fuller evidential case cannot even be sketched here, we can see enough of it to know that it conforms to the pattern already seen. The two step character is present. The explanatory prowess of the wave theory was almost invariably compared with the foil of Newton’s corpuscular theory, which gave it real competition. The latter was vanquished eventually either by its need take on undischarged evidential debt or by direct contradiction with experiment. The second step long remained fraught. At any moment, the explanatory achievements of the wave theory were threatened by new, as yet unexplained optical phenomena.

The complexity of the example derives from a pair of coupled circumstances.

First, *the* wave theory of light is a misnomer. There is a long history of theories that attribute wavelike properties to light, extending back to the seventeenth century in the work of Hooke and Huygens. However the theories adopt many forms as they develop, sometimes adapting to then current developments in the surrounding sciences. The earliest theories simply presumed light to be a propagation in some medium, akin to sound propagation in air. Later

theories retracted, for sound waves are longitudinal rarefactions and compressions, whereas light waves proved to be an oscillation that was transverse to the direction of propagation. Ultimately, light was absorbed into electromagnetic theory as the propagation of a wavelike disturbance in the electromagnetic field.

Second, the behavior of light was examined carefully in many experiments. As a result, the range of experimental results to be accommodated by a theory of light was large and growing. They includes results on the speed and direction of light propagation, its decomposition into colors, reflection, refraction in media, colored bands in thin plates (“Newton’s rings”), the polarization of light, stellar aberration, various interference patterns including fringes around shadows, double refraction in crystals, and more. The character of the wave motion attributed to light developed in concert with these developments.

The history of the establishment of the wave theory of light is a history of its competition with the Newtonian corpuscular theory, also known as the emission theory. The competition was quite real. In the seventeenth century, the wave theory was rudimentary. It was based, according to Huygens (1690, p. 11), on the supposition that light is “some motion impressed upon the matter which lies in the intervening space” and that the motion “is propagated, as that of sound, by surfaces and spherical waves.” The explanatory successes of Huygen’s theory are now well known. His constructions enable recovery of familiar processes of reflection and refraction.

10.1 Early Competition of Wave and Emission Theory

Huygen’s theory faced considerable explanatory competition from Newton’s corpuscular theory. The latter theory merely supposed that light consists of very small corpuscles, moving very quickly. The theory was ontologically frugal. Both posited the existence of matter. For the corpuscular theory, the matter posited just was the light seen. For the wave theory, vastly more matter needed to be supposed in the form a space-filling, all-pervading substance in which light would propagate as vibrations.

Newton’s theory could deal quite effectively with these same phenomena as the wave theory. There, Newton’s theory had advantages. Light propagates in straight lines. Wave propagations in media, such as sound, do not propagate linearly but follow tortuous pathways according to alterations in the medium and its motion. This problem, according to Shapiro (2002, p. 232) remained Newton’s principal objection to the wave theory throughout his life. There

were other explanatory advantages for the corpuscular theory. The equal angles of reflection of light matches perfectly with the behavior of bodies undergoing elastic collision. Newton had found that white light decomposes into rays of definite colors and that these rays were quite fixed in their color. It was not altered by reflection, refraction and other like processes. That constancy was easily accommodated into a corpuscular theory by assuming that the different colors correspond to different types of corpuscles with stable characters. It was less clear that mere vibrations in some unseen, all-pervading substance could provide the same stability.

10.2 The Emission Theory Weakens

The tide began to turn against the Newtonian theory with the work of Thomas Young in the early nineteenth century and then its development by Augustin Fresnel. They were able to account for many optical effects as arising from the constructive and destructive interference of light waves. Newton's theory could accommodate such effects to some extent. The most celebrated of these effects was "Newton's rings," that is, rings of light and dark that form in the small, intervening space when a lens sits on a flat sheet of glass. Newton's account was complicated, depending on "fits of easy transmission and reflection."

The details are too complex for recapitulation here. What is relevant, however, is William Whewell's (1858, p. 89) assessment of them in his *History of the Inductive Sciences*, written from the perspective of someone close to the episode. In spite of Newton's status as a national hero, Whewell was quite scornful of Newton's hypotheses (p. 89, Whewell's emphasis):

The colors of thin plates. Now, how does Newton's theory explain these? By a new and special supposition;--that of *fits of easy transmission and reflection*: a supposition, which, though it truly expresses these facts, is not borne out by any other phenomena. But, passing over this, when we come to the peculiar laws of polarization in Iceland spar, how does Newton's meet this? Again by a special and new supposition;--that the rays of light have *sides*. Thus we find no fresh evidence in favor of the emission hypothesis springing out of the fresh demands made upon it.

In present terms, the problem was not that Newton's account was incompatible with experiment. Rather it required an undischarged evidential debt in the form of the hypotheses identified by Whewell.

One might imagine that the explanatory advantage of the wave theory was absolute by this time. However that was not so. It still did require a medium of unusual properties. Since light propagates in empty space, that medium—the luminiferous ether—must be all pervasive. It must be entirely unaffected when ordinary matter is evacuated from vessel, where such evacuation would completely suppress sound propagation. Tyndall (1873, pp. 47-48) could report as late as 1873 of the persistence of doubt over this assumption of the medium. He wrote of David Brewster (1781-1868), a celebrated pioneer in optical science:

In one of my latest conversations with Sir David Brewster he said to me that his chief objection to the undulatory theory of light was that he could not think the Creator guilty of so clumsy a contrivance as the filling of space with ether in order to produce light.

10.3 Wave Theory Triumphs

Thus the competition proceeded. It is quite hard to locate simple cases of explanatory competition between the emission theory and a wave theory of light, suitable for a brief exposition here. Lloyd (1873, pp. 11-12) reports one such case. By the time of Lloyd's writing, it had been ascertained experimentally that the speed of light is the same everywhere in empty space, whatever the source of the light. Lloyd found it incredible that all the different processes that accelerate the corpuscles into propagating light should produce exactly the same speed. More puzzling is that they could retain that speed when the gravity of celestial objects would slow them down. He reported Laplace's computation that the gravity of a star 250 times as great as our sun, but of the same density, would stop the motion entirely. There was a desperate rescue possible:

The suggestion of M. Arago seems to offer the only way of escaping the force of this objection. It may be supposed that the molecules of light are originally projected with different velocities, but that among these velocities there is but one which is adapted to our organs of vision, and which produces the sensation of light.

The constancy of the speed of light, however, followed naturally if light is a wave propagating in a medium. The speed depends only on the elasticity and density of the medium, which are assumed constant.

We see in this simple example that the wave theory accommodates the constancy of the speed of light fairly well. The accommodation is dependent on a special hypotheses, the uniformity of the medium. Since the constitution and nature of the medium remained uncertain, the wave theory account is not problem free. The emission theory, however, is in great trouble. Any reasonable mechanics of the era for corpuscles predicts many speeds. The fact that only one is observed is a refutation. The emission theory can be protected, but only by taking on a dubious hypotheses about our vision; that is, by taking on a quite significant evidential debt.

A decisive turning point came with experiments around 1850 that directly measured the speed of light in media. When light propagates from a less dense to a more dense medium, it is refracted towards the denser medium. This familiar effect is the basis of the functioning of optical lenses. It is explained quite differently by the wave and emission theories. The wave theory assumes that the speed of light in the denser medium is reduced and the angle of refraction is recovered by a Huygens construction. The emission theory, however, explains the refraction towards the denser medium by attractive forces that accelerate the light corpuscles into the denser medium. That is, the speed of light increases in a denser medium.

This stark difference of prediction was finally put to the test. The wave theory prediction was borne out. Crew, in 1900, reports the victory (1900, p. xii)

It was in the year 1850 that Fizeau and Foucault measured directly the speed of light in air and water, and found the ratio of these speeds numerically equal to the ratio of their refractive indices. This experiment has sometimes been called the *experimentum crucis* of the wave theory; but with scant justice we venture to think, inasmuch as no great doctrine in physics can be said to rest upon any single fact, though modification may be demanded by a single fact.

Crew's caution was prudent. While this result may have ended the emission theory's prospects, the wave theory of 1850 still had obstacles to overcome. Its dependence on a medium of uncertain properties, the luminiferous ether, would fester and eventually become a focus when Einstein published his special theory of relativity in 1905.³⁰

³⁰ More relevantly, Einstein also published a startling result in 1905 concerning light. His light quantum hypothesis asserted that the energy of high frequency light was spatially localized into points; and that was quite reminiscent of Newton's tiny corpuscles.

By this time, the wave theory of light was no longer an independent theory that would rise and fall according to new experimental results on light alone. Since the 1860s, light had been identified as a wave propagating in an electromagnetic field, so that the success or failure of the wave theory became intimately tied to that of electromagnetic theory. A fuller account of the final victory of the wave theory would have to include an account of the rise of electromagnetic theory upon which it came to depend.

By the turn of the century, the complex, lingering competition between emission and wave theories of light was reducible to a few brief sentences in the opening pages of a textbook. Walker (1904, pp. 1-2) summarizes it as:

...the emission theory is lacking in simplicity, and overcrowded with hypotheses; moreover it contradicts the facts in an important particular, for it leads to the result that the propagational speed of light is greater in a dense medium, such as water, than it is in air, whereas direct experiments show that the reverse is the case.

This summary serves us quite well, for it encapsulates failures of the Newtonian foil in Step 1 of the abductive inference. It is defeated by the undischarged evidential debt of special hypotheses and by contradiction with experiment. The second step, the elevation of the wave theory from the better explanation to the best and the one to which we infer, is too complex to gloss here.

11. Conclusion

The standard philosophical account of this argument form tells us that we may infer to some hypothesis or theory because that hypothesis or theory displays some powerful and distinctive explanatory prowess. This chapter has examined canonical examples in real science of inferences to the best explanation and finds something different. The favored theory or hypothesis does not gain favor because it implements some philosophically distinctive notion of explanation. The evidential successes are more successes of accommodation, albeit at times noteworthy. The real evidential challenge for proponents of the favored hypothesis or theory is to display the evidential failure of competitors. The favored theory or hypothesis does not so much prevail because of its own intrinsic virtue. It prevails by default because of the evidential failure of the competitors. The failures of the competitors are not explanatory failures. They are simpler and consist in two modes. Either the competitor is contradicted by the evidence; or its survival

requires it to take an evidential debt, that is, to make suppositions for which there is insufficient evidential support.

As a result, it was possible in the last chapter (Section 7) to characterize these inferences in loose and general terms as “inference to the best explanation without explanation.” The emphasis in the examples on comparison led the characterization to have two steps. The first and dominant step is comparative: one hypothesis or theory is favored over the competing foil. This step is clearly visible in the examples recounted here. The second step is logically very strong. It dispenses with comparisons: “favoring” is replaced by “inferring to.” However it has little explicit presence in the examples. The step, if taken at all, is made tacitly. The competing foils are defeated and that is enough to let the victor ascend.

References

- Born, Max (1922) *Die Relativitätstheorie Einsteins und ihren physikalischen Grundlagen*. 3rd ed. Berlin: Julius Springer.
- Born, Max (1962) *Einstein's General Theory of Relativity*. Methuen, 1924; revised and enlarged, New York: Dover, 1962.
- Ćirković, Milan M. and Perović, Slobodan (2018) “Alternative Explanations of the Cosmic Microwave Background: A Historical and an Epistemological Perspective,” *Studies in History and Philosophy of Modern Physics*, **62**, pp. 1-18.
- Crew, Henry (1900) *The Wave Theory of Light: Memoirs by Huygens, Young and Fresnel*. New York: American Book Co.
- Darwin, Charles R. (1876). *The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. 6th ed. London: John Murray.
- Davisson, Clinton J. and Germer, L. H. (1927) “Diffraction of Electrons by a Crystal of Nickel,” *Physical Review*, **6**, pp. 705-740.
- Einstein, Albert (1915) “Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie,” *Preussische Akademie der Wissenschaften, Sitzungsberichte*, 1915 (part 2), pp. 831–839.

- Einstein, Albert (2000) "Nature in the Realization of the Simplest Conceivable Mathematical Ideas¹: Einstein and the Canon of Mathematical Simplicity," *Studies in the History and Philosophy of Modern Physics*, **31**, pp.135-170.
- Freundlich, Erwin (1915) "Über die Erklärung der Anomalien im Planeten-System durch die Gravitationswirkung interplanetaren Massen," *Astronomische Nachrichten*, **201**, No. 4803, pp. 51-56.
- Hacking, Ian (2001) *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.
- Hall, Asaph (1894) "A Suggestion in the Theory of Mercury," *The Astronomical Journal*. **14**, pp. 49-51.
- Harper, William (2002) "Newton's Argument for Universal Gravitation," Ch.5 in I. B. Cohen and G. E. Smith, eds., *The Cambridge Companion to Newton*. Cambridge: Cambridge University Press.
- Herschel, John (1840) *Preliminary Discourse on the Study of Natural Philosophy*. New edition, London: Longmans.
- Hertz, Heinrich (1883) "Versuche über die Glimmentladung," *Annalen der Physik*, **19**, pp. 782-816.
- Huygens, Christian (1690) *Treatise on Light* in H. Crew (1900).
- Kuhn, Thomas S. (1996) *The Structure of Scientific Revolutions*. 3rd ed. Chicago: University of Chicago Press.
- Landau, Lev D. and Lifshitz, Evgeny M. (1965) *Quantum Mechanics: Non-Relativistic Theory*. 2nd ed. Trans. J. B. Sykes and J. S. Bell, Oxford: Pergamon.
- Lenard, Philipp (1894) "Über Kathodenstrahlen in Gasen von atmosphärischen Druck; und im äussersten Vacuum," *Annalen der Physik*, **51**, pp. 225-267.
- Lenard, Philipp (1894a) "Über die magnetische Ablenkung der Kathodenstrahlen," *Annalen der Physik*, **52**, pp. 23-33.
- Lenard, Philipp (1906) *Über Kathodenstrahlen: Nobel-Vorlesung*. Leipzig: J. H. Barth.
- Liddle, Andrew (2003) *An Introduction to Modern Cosmology*. 2nd ed. West Sussex: John Wiley & sons.
- Lipton, Peter (2004) *Inference to the Best Explanation*. 2nd ed. London: Routledge.

- Lloyd, Humphrey (1873) *Elementary Treatise on the Wave-Theory of Light*. London: Longmans, Green and Co.
- Lyell, Charles (1830, 1832, 1833) *Principles of Geology*. Vol. 1, 2, 3. London: John Murray.
- Newcomb, Simon (1895) *The Elements of the Four Inner Planets and the Fundamental Constants of Astronomy*. Washington: Government Printing Office.
- Norton, John D. (2000) "How We Know About Electrons," pp. 67- 97 in R. Nola and H. Sankey, eds., *After Popper, Kuhn and Feyerabend; Recent Issues in Theories of Scientific Method*. Dordrecht Kluwer.
- Norton, John D. (2011)"History of Science and the Material Theory of Induction: Einstein's Quanta, Mercury's Perihelion." *European Journal for Philosophy of Science*. **1**, pp. 3-27.
- Pauli, Wolfgang (1958) *Theory of Relativity*. Oxford: Pergamon.
- Partridge R. Bruce (1969) "The Primeval Fireball Today," *American Scientist*, **57**, pp. 37-74.
- Peebles, P. J. E. (1971) *Physical Cosmology*. Princeton: Princeton University Press.
- Peebles, P. J. E. (1991) "The Emergence of Physical Cosmology," pp. 17- 30 in A. Blanchard, et al., eds., *Physical Cosmology*. Gif-sur_Yvette, France: Éditions Frontières, 1991.
- Peebles, P. J. E. (1993) *Principles of Physical Cosmology*. Princeton: Princeton University Press.
- Penzias, Arno and Wilson, Robert (1965). "A Measurement Of Excess Antenna Temperature At 4080 Mc/s," *Astrophysical Journal Letters*. **142**, pp. 419–421.
- Shapiro, Alan E. (2002) "Newton's Optics and Atomism," pp. 227-255 In I. B. Cohen and G. E. Smith, eds., *The Cambridge Companion to Newton*. Cambridge: Cambridge University Press.
- Schiff, Leonard I (1968) *Quantum Mechanics* 3rd ed. Tokyo: McGraw-Hill Kogakusha.
- Thagard, Paul R. (1977) "Discussion: Darwin and Whewell," *Studies in History and Philosophy of Modern Physics*, **8**, pp. 353-56.
- Thagard, Paul R. (1978) "The Best Explanation: Criteria for Theory Choice," *Journal of Philosophy*, **75**, pp. 76-92.
- Thomson, Joseph J. "Presidential Address." pp. 699-706 in Section A. Mathematical and Physical Sciences. *Report of the Sixty-Sixth Meeting of the British Association for the Advancement of Science Held at Liverpool in September 1896*. London: John Murray, Albemarle St.
- Thomson, Joseph J. (1897) "Cathode Rays," *Philosophical Magazine*, **44**, pp. 293-316.

- Thomson, Joseph J. (1906) "Carriers of Negative Electricity," Nobel Prize Lecture, December 11, 1906.
- Tyndall, John (1873) *Six Lectures on Light: Delivered in America in 1872-1873*. London: Longmans, Green, and Co.
- Walker, James (1904) *The Analytic Theory of Light*. Cambridge: Cambridge University Press.
- Whewell, William (1847) *The Philosophy of the Inductive Sciences*. Vol. 1. London: John W. Parker.
- Whewell, William (1858) *History of the Inductive Sciences from the Earliest to the Present Time*. Vol. II. 3rd ed. New York: D. Appleton and Co.
- Weinberg, Steven (1972) *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. New York: John Wiley and Sons.
- Weinberg, Steven (2008) *Cosmology*. Oxford: Oxford University Press.
- Weyl, Hermann (1921) *Space-Time-Matter*. 4th ed. Trans. H. L. Brose. London: Methuen.