JOHN D. NORTON

SCIENCE AND CERTAINTY[1]

In common scientific practice, near certainty is accorded to the basic principles of a mature science and this certainty is said to be based on experimental evidence. I show how two related forms of inference, demonstrative induction and eliminative induction, can be used to support judgments of this type, briefly illustrating their use with the case of quantum discontinuity in the early part of this century.

1. THE UNDERDETERMINATION THESIS

What role can experimental evidence play in the acceptance of new scientific theories and in the reaffirmation of old theories? Two related and currently popular theses severely restrict this role.

The *underdetermination thesis* asserts that a given body of evidence must fail to determine uniquely a single theory. The thesis has a venerable history, with roots extending as far back as Hume's skepticism over the possibility of justified inductive inference.

The related *Duhem–Quine thesis* asserts that theories can only confront evidence as whole. Evidence cannot support or refute individual laws of a theory; evidence inconsistent with a theory tells us at best that one of the laws is incorrect, but not which.[2]

A strengthened variant of the Duhem–Quine thesis asserts that any given law in a theory can be protected from unfavorable evidence by the adjustment of other hypotheses in the theory so that the theory as a whole can accommodate the evidence. Such is the key idea of Lakatos's (1970) "protective belt" – a set of hypotheses of a theory that one decides can be modified in order to protect the "hard core" of a research program from falsification by unfavorable evidence.

This protective stratagem suggests a way of generating variants of a given theory, all adequate to the same body of evidence. If one can modify one hypothesis to protect another from falsification, then surely one could modify two hypotheses in a way that cancels to produce a variant theory bearing the same relation to the evidence as the original theory.

Synthese **99**: 3–22, 1994.

Thus, under the accumulated weight of these theses and their variants, it would seem that the life of the scientist who chooses theories on the basis of available evidence is a precarious one indeed. His evidence cannot pick out a unique theory, for a theory is always underdetermined by evidence. Worse, alternative theories, all equally adequate to the evidence, are readily accessible to him. We might well expect that our scientist will become little more than a vagabond capriciously wandering from theory to theory or perhaps, like Buridan's ass, will be frozen into inactivity by the inability to choose among a plethora of equally viable theories.[3]

2. SCIENTIFIC PRACTICE AND THE PAUCITY OF CHOICE

This portrait of the life of the scientist is, of course, very far from what one encounters in actual scientific practice. There, one rarely finds that the scientist has such an embarrassment of riches. Typically, a scientist is pleased to find even one theory that is acceptable for a given body of evidence. In the case of a mature science, there is most commonly a single favored theory to which near certain belief is accorded and which is felt to be picked out uniquely by the evidence. Challenges to the theory from aberrant hypotheses or experiments are rarely considered seriously.

Examples of the dominance of the single favored theory abound. That there was some kind of conservation of force or energy had become such an entrenched idea by 1775 that the French *Académie des Sciences* decided not to review any more proposals for perpetual motion machines, announcing at least three quarters of a century before the conservation of energy was canonized as the first law of thermodynamics:[4]

The construction of a perpetual motion machine is absolutely impossible.

Similarly, little serious credence is now given to alternative accounts to special relativity of the null result of nineteenth-century experiments aimed at measuring the velocity of the earth through the luminiferous aether, although authors of alternative proposals sometimes find novel vehicles for publication.[5]

Another quite recent example concerns a group of researchers, led by Jacques Benveniste at the University of Paris, that was able to publish a report on research in which highly diluted aqueous solutions of an antibody were still found to have biological activity even at dilutions of 1 in 10^{120} (Davenas et al., 1988). At this high dilution it is improbable that the diluted solutions contain a single molecule of the original antibody. The publication was attended by considerable controversy. A brief note of 'Editorial Reservation' by *Nature*'s editors was appended to the paper. It noted the incredulity of the paper's referees concerning the result, and announced that an independent team of investigators would observe repetitions of the experiment. The journal's deputy editor, Peter Newmark, was reported as saying in a newspaper article:[6]

> We are certain that these results must be wrong, but we have been unable to disprove them.

The depth of this conviction was flagged by the selection as one of the three investigators of James ('The Amazing') Randi, a professional conjurer specializing in the debunking of fraudulent parapsychological experiments. The investigators' report cited significant deficiencies in the experimental methods of the researchers, and concluded that there was no substantial basis for their claim (Maddox et al., 1988).

It is important to see that the widespread reaction of disbelief does *not* derive from the mere fact that the experimental results are unorthodox. Rather, it stems from the fact that the correctness of these experimental results would require the falsity of quite fundamental results of modern biochemistry. These results are that biological activity of this type is chemical in nature and that chemical activity is atomic, with the atomic scale set by Avogadro's number whose value is of the order of 10^{23}.

Thus there has been another recent instance of unorthodox experimental result that met with considerable controversy but not with monolithic incredulity because it was not immediately obvious that the results were incompatible with fundamental laws. In fact, part of the allure of the results derived exactly from the possibility that they might well be compatible with these laws. On 23 March 1989, Martin Fleischmann and B. Stanley Pons announced that they had produced a controlled nuclear fission reaction at room temperature by the electrolysis of water with a Palladium electrode in a simple benchtop apparatus. Previously, controlled fusion was thought to require apparatus of massive size and massive cost. Unlike the French researchers, Fleischmann and Pons' results did not call into question fundamental laws such as the atomic constitution of matter or the order of magnitude estimates of atomic sizes. In fact, on the simplest level, they could tell a story that fitted with such laws. During electrolysis in accord with those estimates, the deuterium atoms were supposed to be brought sufficiently closely together within the Palladium electrode to enable a fusion reaction to take place. However, there was considerable exasperation at the unorthodox way in which the results were made public in a press announcement, and that it was with such lack of detail that attempts at replication were difficult. Yet it was not immediately obvious that a fundamental law would be violated if the result were as claimed, so that the criticism voiced in the ensuing controversy was not univocal condemnation. There was a flood of attempts at replication – something not to be expected if the result were dismissed as utterly untenable.[7]

A dominant feature of scientific practice, at least as far as mature theories are concerned, is that the fundamental laws are accorded belief beyond reasonable doubt, at the exclusion of other competitors. I have cited examples of the conservation of energy (and its precursors), special relativity's undetectibility of uniform motion, the chemical nature of biological activity, and the atomic character of chemical activity. There is no sense that these laws are negotiable in the routine practice of science. Anyone who doubts these laws does so at the peril of their reputation. Even claims of experimental disproof of these laws are given no credence, with the appropriate response including the despatching of a professional conjurer adept at the exposure of fraud.

3. EXPLANATIONS OF CERTAINTY

How are we to explain the certainty accorded to the fundamental laws of a mature theory? If we hold to the underdetermination thesis and its variants, then we cannot expect that certainty to be grounded in the evidence. We might seek other explanations of these proclamations of certainty. Scientists are human and science a human endeavor, so we ought to expect that science is prone to human failings as is any human enterprise. Thus we should expect cases in which the proclaimed certainty of some result derives not from evidence but from dogma or wishful thinking, self-interest or peer pressure, fraud or self-deception, or a myriad of other human frailties. Doubtlessly and lamentably, there will be cases of this type. However, are they *all* of this type? To say that they are is to deliver a verdict of the greatest pessimism concerning our abilities to fathom our world. Of all human endeavor, science is the one that has made the most sustained and systematic attempts at controlling human frailty in its inquiries. If science has universally failed in this endeavor, then I see no reason to expect that any other investigative endeavors have fared better. Our understanding of the world – scientific and nonscientific alike – is little more than myth and delusion, and our attempts at rationality are no better than childish games.

I am not such a pessimist. I do believe that there are at least some instances in which the proclaimed certainty of science is based on the proper relationship of the evidence to the theory. I turn to confirmation theory and the study of inductive inference for an account of that relationship.

Before doing so, however, I would like to anticipate a skeptical objection. We must not forget the repeated history of revolutions in science, the objection says. Surely, they show that we cannot expect any warrant for near certainty concerning any scientific theory or law no matter how it is produced. I cannot agree. Almost every example of Section 2 has survived at least one major revolution. We still think now, as in 1775, that a perpetual motion machine is impossible in the domain of common experience, even though the claim has been filtered through a series of revolutions. We do, of course, have a clearer taxonomy of such devices and better accounts of why they might fail.

Moreover, judgments of near certainty ought to be based on a particular body of evidence. Thus they are not irrevocable. In fact, they may be revoked merely through the addition of new evidence without retracting any of the original body of evidence. In this regard, induction is quite unlike deduction. The conclusion of a valid deductive argument cannot be overturned by the addition of new premises logically compatible with the old. That this is not so for induction can be seen in a simple example. Consider a body of evidence that asserts that the speed of light has been measured to be constant in all accessible domains. This evidence inductively warrants the conclusion that the speed of light is a universal constant. New evidence in a hitherto inaccessible domain, that of intense gravitational fields, for example, may reveal a variable speed of light. This would overturn the original conclusion, but the judgment of near certainty of that original conclusion *on the old evidence* would remain valid.

That is, the mere fact that old judgments of near certainty are overturned, as in a scientific revolution, does not mean that the original judgments were improperly made. It only reminds us that the judgment was of *near* certainty and that all induction involves inductive risk. Indeed, the more domains science investigates and the more times it takes an inductive risk, the more likely it becomes that some of its conclusions will be overturned, even if they are all properly drawn.

Let us return to the question of how these judgments can be made. In the following, I shall briefly review the currently most popular accounts of the relationship between theory and evidence, asking how each might accommodate claims of near certainty.

3.1. *Hypothetico-Deductive Confirmation*

This is probably the most popular account of confirmation outside the literature explicitly devoted to confirmation theory. It holds that evidence supports a theory if the evidence is entailed by the theory (with suitable auxiliary hypotheses if needed). In other words, the supported theory must save the evidential phenomena. This simple view of confirmation is unable to establish the certainty of a theory on the basis of a given body of evidence. If one theory entails the body of evidence, it is usually a trivial matter to modify the theory in a way that essentially changes it but leaves the observational consequences of the theory unchanged.[8] This maneuver yields a new theory incompatible with the old which is equally supported by the evidence. Stratagems such as this fit perfectly with the underdetermination thesis of Section 1. A tacit acceptance of this essentially inadequate account of confirmation underpins much of the favorable treatment of the underdetermination thesis.

Some attempts to save this account of confirmation allow that mere entailment of the phenomena by the theory is too crude. The theory must *explain* the phenomena. Such accounts of confirmation are labeled abduction or 'inference to the best explanation'. Their greatest weakness lies in the problem of explicating the notion of explanation. Aside from this problem, this strengthened account still does not provide a truly satisfactory warrant for judgments of certainty. That a theory provides a good explanation of some body of evidence provides no guarantee that there is not another theory that explains it better. Thus the electrodynamic theory of H. A. Lorentz provided a good explanation of the null results of experiments aimed at detecting the motion of the earth through the luminiferous aether, and the physicist of 1905 would be well justified accepting the theory. As it turns out, Einstein's special theory of relativity is judged to give a better explanation.

Another device for rescuing the basic strategy of hypothetico-deductive confirmation is provided by Popper's notions of falsificationism (1959). What we are to reject, according to this account, is a theory entailing observational consequences incompatible with the evidence. What we are to accept tentatively is the theory designed in the wake of such a falsification and that has passed severe experimental test; that is, one that is well "corroborated". The same basic difficulty remains. That one such theory is well corroborated gives us no reason to think that there is not another theory that would be equally well corroborated on the same evidence.

3.2. *Instance Confirmation*

The basic idea of instance confirmation is that a hypothesis is confirmed by its instances. The classic account of this view is due to Hempel (1965). To take the classic example, the hypothesis:

(1) All ravens are black

is supported by evidence that is an instance of the hypothesis:

(2) Raven_1 is black.
Raven_2 is black.
. . .
Raven_n is black.

The results of experiment or observation are usually reported in a language whose vocabulary is a subset of that used in the theory. Thus a simple instance confirmation scheme of this type will be unable to find instances of central hypotheses of a theory entailed by evidence reports if the vocabulary of the hypotheses transcends that of the report. In such cases, instance confirmation cannot be applied. Glymour's (1980) "bootstrap" account of confirmation seeks to remedy this deficiency. The scheme allows that a theoretical hypothesis can be confirmed by an evidence report if that report, in conjunction with other hypotheses, entails an instance of the hypothesis to be confirmed. The hypotheses conjoined with the observation report enable the introduction of the requisite theoretical vocabulary.

In either of its variants, instance confirmation clearly gives us grounds for believing a hypothesis. The problem, however, is this: How many instances are needed to yield near certain belief? (How many swallows make a summer?!) Of course, one has an intuitive feeling of how one's belief ought to change as confirming instances arise. In cases of simple instance confirmation, one is rarely satisfied with just one and expects many cases before one's belief is substantially changed. In bootstrap confirmation, however, there are cases in which just a few positive instances do appreciably increase one's belief. Unfortunately, these intuitive rules are externally provided to the instance confirmation schemes provided by Hempel and Glymour. They are not explicitly codified and may defy such codification.

In any case, even if considerable instance confirmation can be supplied for a given hypothesis (whatever "considerable" may amount to), nothing in the schemes guarantees or even suggests that no other plausible hypothesis might not receive similar or better confirmation on the same evidence.

However, the general form of instance confirmation does provide one notable advance over that of the hypothetico-deductive confirmation schemes. All of these latter schemes suffer from the same basic problem. As long as the entailment relation proceeds *from theory to evidence*, then we have considerable freedom to adjust the theory confirmed by a given body of evidence. We can always strengthen a theory without subtracting from its observational consequences. The only trick is to be sure not to add new, compromising observational consequences. The key feature of instance confirmation is that the entailment proceeds *from evidence to theory*, or to instances of theory. Thus, any logical strengthening of a hypothesis is far more likely to compromise the possibility of deducing an instance of the hypothesis from the evidence.

3.3. *Bayesian Confirmation*

Of the currently popular accounts of confirmation, the most successful treatment of certainty comes from those accounts that represent degree of belief in a hypothesis on evidence by a number and in which the ebb and flow of belief under the impact of new evidence is tracked by the variations in these numbers. The most popular calculus for computing the changes in these numbers is the probability calculus used by Bayesians (Howson and Urbach, 1989). The approach to certainty under the accumulation of favorable evidence is captured in theorems dictating the nature of evidence needed for the probability of a hypothesis to approach unity (certain belief) in the limit.

The basic difficulty of the Bayesian account for our purposes is that the vast majority of scientists do not explicitly offer Bayesian accounts of how their judgments of near certain belief are derived from the evidence. This need not totally dishearten the Bayesian who can still urge that scientists are really secret Bayesians, using inference schemes that are validated by their compatibility with Bayesian principles. This response leaves us, then, with a more immediate question: What are these inference schemes? The next section considers some schemes which are used to establish scientific theories and which the Bayesian might consider analyzing.

4. DEMONSTRATIVE AND ELIMINATIVE INDUCTION

These two closely related forms of induction have been routinely included in older logic treatises. Unfortunately, they have fallen from favor and no longer prominently figure in the mainstream discussions of the bearing of evidence on theory. This is regrettable. These forms of inference, used in the actual practice of science, are of such a form that they give special control of precisely the problem we have been discussing: how judgments of near certainty can be based on evidence.

4.1. *Demonstrative Induction*

Following a generalization of the discussion given in Johnson (1964, p. 210), demonstrative inductions have the form:

Premises of greater generality.
Premises of lesser generality.

Conclusion of intermediate generality.

We should immediately note that demonstrative inductions are nonampliative and, in this regard, violate the modern view of inductive inference, which is now viewed as necessarily ampliative. I shall retain the term "demonstrative induction" simply to avoid unnecessary duplication of terminology.

The nature of demonstrative induction can be shown with the help of a simple example. The classic of instance confirmation allows us to proceed from

Premises of lesser generality:
(1) Raven_1 is black
Raven_2 is black
. . .
Raven_n is black

to the

Conclusion of intermediate generality:
(2) All ravens are black.

We could convert the inference into a demonstrative induction by supplying a premise needed to make the inference from (1) to (2) deductive. That additional premise is

Premise of greater generality:
(3) Ravens are biologically such that they all must have the same color.

At first blush, it seems that we are merely trying to convert an inductive argument into a deductive argument. But the flight to demonstrative induction is far more successful than attempts to convert inductive arguments into deductive arguments by supplying a principle of the uniformity of nature as an extra premise. (This principle is either so vague as to be useless or, when made precise, false.) For the premises of greater generality which are added are typically chosen to be specific to the particular domain and particularly domains of the widest application. In a well-thought-out demonstrative induction, their inclusion will occasion little dissent.

It must be stressed that the flight to demonstrative induction does not and cannot free us of the need to employ ampliative inference. Typically, ampliative inference will be needed to justify the "premises of greater generality". In general, a body of scientific evidence is logically weaker than the theory it supports so that ampliative inferences must be used if we are to infer from the evidence to the theory. Demonstrative induction, however, as I will show shortly, can be especially helpful in assisting our assessment of the strength of that support.

It is to the great credit of Jon Dorling that he has seen the applicability of demonstrative induction in many important cases in the history of physics, including the work of Einstein (see Dorling, 1973, unpublished) for further discussion and numerous examples).

4.2. *Eliminative Induction*

Eliminative inductions have the following form:[9]

> Premises that define a universe of possibilities.
> Premises that enable elimination of members of this universe.
>
> ---
>
> Conclusion: Uneliminated members of this universe.

Eliminative induction is a very familiar form of induction. It is perhaps best known to readers of whodunits, in which the detective begins with a list of suspects, eliminating them one at a time until the culprit is revealed. In Norton (1989), I argue that eliminative induction has more significant application, especially if we allow the universe of possibilities to be infinitely large. For I urge that Einstein's pathway to general relativity was paved with eliminative inductions of this type, so that the argument form not only provided a major component of the theory's support but also the actual carrying out of the eliminative inductions figured in Einstein's discovery process.

This close relationship between demonstrative and eliminative induction is now easy to see. To convert a demonstrative induction into an eliminative induction, one need only read its "premises of greater generality" extensionally, that is, as specifying a universe of possible theories or hypotheses. Thus the extensional reading of premise (3) in the above example is:

Premise that specifies a universe of possible hypotheses:
(3′) All ravens have color_1.
or All ravens have color_2
or All ravens have color_3
. . .
where the set $\{\text{color}_i\}$ exhausts all colors.[10]

Premises (1) and (3′) and conclusion (2) now form an eliminative induction. Conversely, one can convert an eliminative induction into a demonstrative induction by reading the premises that intensionally specify the universe of possibilities.

4.3. *The Virtues of Demonstrative and Eliminative Induction: Rule-Bound versus Assumption-Bound Inductive Risk*

In all inductive inference, one must take an inductive risk. The virtue of demonstrative and eliminative induction is that this risk is relocated to a place where its import, nature, and magnitude can be assessed far more readily. In the other confirmation schemes considered, the inductive risk was located in the inductive rule used.[11] Thus, in the case of instance confirmation from (1) to (2), the inductive risk is associated with the rule of induction used:

> Infer from the instances of a hypothesis to the universally generalized hypothesis.

The assessment of the degree of inductive risk becomes an assessment of the reliability of this rule, a task which is extremely problematic. Once the inference has been recast as a demonstrative or eliminative induction by the addition of premise (3) or (3′), then the task of assessing the inductive risk is transformed. That risk is relocated in the new premise itself. The task of assessing the risk taken in accepting it is now far more tractable. Whether ravens are biologically such that they can admit only one feather coloration is something that can be the subject of further inquiry, and one can reasonably expect results in a way that one would not from an inquiry into the reliability of the above inductive rule.

In the form of a slogan, these inference forms enable the replacement of rule-bound inductive risk by assumption-bound inductive risk.[12]

4.4. *The Approach to Certainty*

How does demonstrative and eliminative induction aid us in understanding how a body of evidence can warrant a law or theory to near certainty and at the exclusion of other candidates? This follows immediately from the eliminative character of the argument form. A well-executed eliminative induction immediately tells us exactly what we want to know. It tells us that on the basis of the evidence incorporated into its premises, there is only one theory or law that can be entertained; the others are eliminated. This solves precisely the difficulty we saw with the hypothetico-deductive and instance confirmation schemes. No matter how much support these schemes provided for a law or theory, we had no insight into the possibility of another law or theory being even better confirmed by the same evidence. An eliminative induction gives us an assurance that no other theory or law in the universe of possibilities is adequate to the evidence.

Of course, such an assurance does not free us of inductive risk. However, the inductive risk can be located in the premises that specify the universe of viable possibilities. Thus, the assurance will be worthless if one has to take a very great inductive risk in adopting these premises. The most satisfactory way of controlling this risk is to seek arguments in which the size of the universe of possibilities is very large and, correspondingly, the "premises of greater generality" of the demonstrative induction are weak. Needless to say, there is no guarantee that demonstrative or eliminative inductions satisfying this requirement can be found to relate known evidence to our theories. As it turns out, however, there are exceptionally good cases of such arguments in the history of science.

5. AN EXAMPLE: QUANTUM DISCONTINUITY

At the turn of the century, one of the puzzles of physics was black body radiation, the electromagnetic heat radiation emitted by a 'black' (totally absorbing) body. The problem was to account for the distribution of energy over the various frequencies of the radiation, and, in 1900, Max Planck had been working for several years on providing an electrodynamic model of black body radiation and its generation by accelerated charges in Hertzian resonators. During that year, the experimentalists Lummer and Pringsheim circulated new experimental measurements that extended deeper into the longer wavelength domain of the spectrum than their earlier results. The new experimental results were incompatible with Planck's earlier theorizing. Planck was able to provide a new distribution law compatible with the data. However, his celebrated derivation of the law depended on an assumption that was to usher in the era of the quantum. Classical electrodynamics allowed Hertzian resonators, as well as the components of a field of electromagnetic radiation, to adopt a continuous range of energy levels. Planck's derivation depended crucially on the assumption that the resonators could adopt only energies[14]

$$0, h\nu, 2h\nu, 3h\nu, 4h\nu, \ldots,$$

although there has been some recent discussion of whether Planck was fully aware of this (see Kuhn, 1978). In any case, within the decade, the nonclassical discontinuity required by the derivation was widely recognized and versions of the derivation, locating the discontinuity in either resonator or spectral component of the radiation itself, were given by Debye, Ehrenfest, Einstein, and Lorentz. It had also become clear from the work of Rayleigh and Jeans that a standard application of classical physics led to a distribution law irreconcilable with the observations and even the finiteness of the total energy of the radiation.

In terms of confirmation theory, the evidence is the experimentally measured black body spectrum and the theory is the discontinuity account. In the work mentioned above, the relation displayed between the theory and evidence is precisely that of the hypothetico-deductive scheme. The theory entails or even explains the evidence. However, the discontinuity theory was by no means a popular theory, and understandably so. It required the falsity of a quite fundamental supposition of classical physics. The mere fact that the discontinuity hypothesis "saved the phenomena" was certainly not sufficient to force its acceptance. Why should one not hope that the phenomena would be saved by some less traumatic variant of the classical theory that preserved continuity?

That there can be no such account, that the observations necessitate quantum discontinuity, is the decisive result established by a number of physicists, most notably Ehrenfest in 1911 and Poincaré in 1912.[15] They were able to construct demonstrative inductions that proceeded *from* the experimental evidence *to* the theory. The "premises of greater generality" were very weak ones indeed. In addition to some essentially uncontroversial auxiliary hypotheses, they merely assumed that a system of black body radiation consisted of a very large number of component subsystems able to exchange energy in a dynamical equilibrium. The observed behavior of the whole system would be the most likely behavior of the combined components. This very general assumption is little more than a statement of the mechanical view of thermodynamical systems that followed from the work in statistical mechanics of Boltzmann and Gibbs, but weakened further so not to require (or exclude) continuity of energy levels.

Their results showed that, if one accepted Planck's distribution for energy over the components of black body radiation, then, with the aid of the above assumptions, one could conclude that the allowed energy levels of the spectral components or resonators were exactly the discontinuous set

$$0,\ h\nu,\ 2h\nu,\ 3h\nu,\ 4h\nu,\ \ldots,$$

However, the Planck distribution law was just one law that fitted the experimental data, and others were known, so that an argument based on specifically accepting that law was perhaps not general enough. To address this worry, Poincaré and, more thoroughly, Ehrenfest presented even more striking demonstrative inductions. As experimental evidence, they merely assumed that the distribution law – not necessarily Planck's – was one that yielded a finite total energy for black body radiation. The result was still the necessity of discontinuity, although now the weakened evidence used allowed only the conclusion that a discontinuity of the type of Planck's theory must occur at the zero energy level. Ehrenfest showed that as one further constrained the form of the law to bring it closer to that of Planck, the form of discontinuity necessitated became more similar to that of the standard theory.

These arguments warrant a very high degree of belief in a quantum discontinuity, for they tell us that any equilibrium statistical account of black body radiation must invoke quantum discontinuity. The ensuing acceptance of quantum discontinuity on the basis of the observed black body spectrum is not, of course, free of all inductive risk. However, the demonstrative inductions described give us some control over that risk. In particular, they tell us that the risk is located in the "premises of greater generality" essential to the inductions. It is not too difficult to make some assessment of the magnitude of the risk buried in them.

The most important premise required the treatment of black body radiation as a system within the compass of a very general statistical mechanics. It was known in 1911 and 1912 that grave difficulties faced anyone who tried to avoid a treatment of a thermodynamic system such as black body radiation in these terms. If the system were some kind of 'pure' thermodynamic system whose thermal properties did not result from the average behavior of very many component subsystems, then the total system would not be expected to exhibit tiny fluctuations in its basic properties such as its energy or pressure. If such a pure system were allowed to interact with another thermal system that did exhibit these fluctuations, such as a kinetic gas, it followed from a 1909 analysis of Einstein that the joint behavior of the two systems would violate the second law of thermodynamics.[16]

Again, one could grant that a statistical treatment of black body radiation was required but seek to escape the arguments of Ehrenfest and Poincaré by assuming that the system was not in equilibrium. Such was the approach taken by Jeans before he abandoned it to become one of the major proponents of the quantum theory. Such an account of black body radiation depends on a detailed analysis of the mechanism of emission of radiation. Jeans found himself unable to provide a classical theory of this type which was compatible both with the observed black body spectrum and with the known behavior of electric charges. In particular, it was difficult to derive a distribution formula whose form would be as independent from the nature of the emitting body as the observed spectrum. Yet exactly this independence followed effortlessly from the simplest thermodynamic arguments available on the supposition of equilibrium.

Finally, it should be noted that the instance of quantum discontinuity is a clear counterexample to the Duhem–Quine thesis (or, at least, the strengthened variant described in Section 1). The observational evidence of the black body spectrum refutes the classical hypothesis of continuity. No amount of adjustment in other parts of the body of classical theory could save the hypothesis, even though a premium was placed on achieving precisely such an adjustment.

6. COROLLARIES

Once we recognize the importance of demonstrative and eliminative induction in scientific practice, a number of interesting corollaries follow:

6.1. *Textbook Accounts*

There is something very striking about the way that virtually all physics textbooks treat the old quantum theory and, in particular, display the relationship between the experimental evidence of the black body spectrum and the Planck discontinuity theory. They almost invariably show only that the theory can save the evidential phenomena.[17] One might wonder how many students have learned from examples like this that saving the evidential phenomena is all that is needed to establish a theory at the level of textbook certainty. Conversely, one might wonder how many astute students have felt uneasy about the inadequacy of this relation and have seen that the requirement that a theory merely save the phenomena leaves the theory to be accepted radically underdetermined. How much do such textbook treatments contribute to the popularity of the underdetermination thesis?

6.2. *Crucial Experiments*

Whether there are crucial experiments in science is a question that has been debated. These are experiments whose results point decisively at one theory and against others. Whatever the initial reception of an experimental result, we can now see how an eliminative induction might well elevate the experiment to the status of a crucial experiment within the context of later analysis. For within an eliminative induction, it is precisely the report of experimental evidence that is used to pick decisively between the candidate possibilities. Thus the experimental measurements of Lummer and Pringsheim, which seem not to have been seriously in doubt in 1911 and 1912, choose decisively in favor of quantum discontinuity and against even the most general forms of classical continuity.

6.3. *Scientific Revolutions*

If we assume that a body of theory has been well established by demonstrative and eliminative inductions, then we can see why the emergence and acceptance of a new, incompatible theory might well be so traumatic as to warrant the popular label of 'revolution'. This is a term that in the political context is associated with bloodshed, violence, and the complete overthrow of an old order. Demonstrative and eliminative inductions locate inductive risk in very general hypotheses, it is hoped in ones of such generality that they are hard to doubt. The advent of the revolution does not typically involve the rejection of the experimental evidence used to support the older theory. Thus, with the advent of the quantum theory, essentially no correction was made to the plethora of older observations concerning the motion of planets and other massive bodies as well as the accumulated experimental results of the nineteenth century concerning electricity, magnetism, and light. Therefore, if a new theory is to be accepted, the "premises of greater generality" in the demonstrative inductions supporting the older theory must be rejected or at least critically modified – this is what is revolutionary.

In the case of the old quantum theory, both Jeans and Poincaré pointed to the older result that was to be given up. In classical mechanics, it was assumed that very general dynamical equations of a Hamiltonian form were adequate to all fundamental physical phenomena. (Such equations are certainly robust enough to survive the transition to a relativistic physics – both special and general – in which dynamical equations commonly find their most general expressions in equations of a Hamiltonian or the related Lagrangian form.) Hamiltonian dynamical equations lead to a statistical mechanics which was based on a phase space in which time development proceeded according to a probability conserving Liouville flow. Poincaré began his celebrated work (1912) with the promise that he would show that even this extremely general picture was inadequate for the problem of black body radiation.

7. CONCLUSION

There is a great danger that we underestimate what can be established by science on the basis of evidence. Any evidential case for a scientific theory will involve some inductive risk. This risk should not make us skeptics because it can be assessed and controlled, the schemes of inference discussed here being only some of the ways of doing so. Indeed, a great deal of the ingenuity of scientists is devoted to this task, as we saw in the case for quantum discontinuity. Again, we should not be seduced into underestimating what the evidence warrants by inadequate caricatures of its bearing on theory. It is usually easy to show that a particular hypothesis can save the phenomena and this result may well be the only one that survives in the textbooks. If this is the only way that one thinks the evidence can support theory, one readily falls into such skeptical theses as the underdetermination thesis or the Duhem–Quine thesis. The sort of result that is needed to rule them out – the result, for example, that no other minimally acceptable hypothesis is adequate to the evidence – is far more difficult to show and, like the results of Ehrenfest and Poincaré on the necessity of quantum discontinuity, is likely to be forgotten by a science unconscious of its history.

NOTES

[1] I am grateful to Peter Achinstein, Don Howard, and the other participants at the conference, 'The Role of Experiments in Scientific Change', Virginia Polytechnic Institute and State University, 30 March to 1 April, 1990, for helpful discussion, and especially to Ron Laymon for his discussion comments presented at the conference on an earlier version of this paper.
[2] Duhem (1962, p. 187); Quine (1953, pp. 40–41).
[3] Lakatos (1970) sought to solve the problem by comparing the fecundity of the theories, adopting those that anticipate novel evidence. This does not solve the problem because the above mechanism can produce arbitrarily many variant theories all able to anticipate the new evidence.
[4] Quoted in translation from Elkana (1974, pp. 29–30).
[5] Josiah Alfred Briscoe, 'Gauge for Measuring Absolute or Cosmic Velocity and Direction', Patent Specification No. 15089/58, Published 20 December 1961, The Patent Office, London. The example has special irony given Einstein's occupation in 1905, the year of first publication of this work in special relativity. (I am grateful to Frank Dickson for drawing this patent to my attention.)
[6] M. W. Browne, 'Impossible Idea Published on Purpose', New York Times, 30 June 1988.
[7] For an account of the controversy, see Peat (1989).
[8] See, for example, the strategy of "deoccamization" discussed in Glymour (1980, pp. 31–32).
[9] In Norton (1989), I give the argument its most general form by allowing the inference from premises to conclusion to precede either inductively or deductively. In most cases, one prefers to restrict the inference to deductive inference and I shall consider only that case here.
[10] The "or" is exclusive.
[11] Ron Laymon has pointed out to me that the Bayesian scheme also locates inductive risk in assumptions when it distributes its prior probabilities. The Bayesian scheme does involve rule-bound risk associated with accepting the probability calculus as the calculus of inductive inference.
[12] In the more general case of eliminative induction in which the premises only inductively support the conclusion, this effect will still occur. The introduction of an eliminative induction can still be associated with the relocation of some inductive risk from a rule to a hypothesis. However, the relocation will not be complete.
[13] For a detailed treatment of this example, see Norton (1993).
[14] h is Planck's constant and ν the resonant frequency of the resonator.
[15] It must be stressed that these were not the only consideration in the acceptance of quantum discontinuity. Its success with the problem of specific heats at low temperatures was especially important at this time, to which was soon added the success of quantum discontinuity in the theory of atomic spectra, not to mention Einstein's long-resisted work on the light quantum.
[16] For a simple account of the analysis, see Norton (1991, pp. 132–34).
[17] See, for example, Bohm's well-regarded text (1951, Chap. 1).

REFERENCES

Bohm, D.: 1951, *Quantum Theory*, Prentice-Hall, New York (Dover repr.: 1989).

Davenas, E., et al.: 1988, 'Human Basophil Degranulation Triggered by Very Dilute Antiserum against IgE', *Nature* **333**, 816–818.

Dorling, J.: 1973, 'Demonstrative Induction: Its Significant Role in the History of Science', *Philosophy of Science* **40**, 360–72.

Dorling, J.: unpublished, 'Einstein's Methodology of Discovery was Newtonian Deduction from the Phenomena'.

Duhem, P.: 1962, *The Aim and Structure of Physical Theory*, Atheneum, New York.

Elkana, Y.: 1974, *The Discovery of the Conservation of Energy*, Hutchinson, London.

Glymour, C.: 1980, *Theory and Evidence*, Princeton University Press, Princeton, New Jersey.

Hempel, C.: 1965, 'Studies in the Logic of Confirmation', in *Aspects of Scientific Explanation*, Free Press, New York, pp. 3–51.

Howson, C. and P. Urbach: 1989, *Scientific Reasoning: The Bayesian Approach*, Open Court, La Salle.

Kuhn, T. S.: 1978, *Black-Body Theory and the Quantum Discontinuity*, Clarendon Press, Oxford.

Johnson, W. E.: 1964, *Logic*, Part II, Dover, New York.

Lakatos, I.: 1970, 'Falsification and the Methodology of Research Programs', in I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, Cambridge University Press, Cambridge, pp. 91–195.

Maddox, J., J. Randi and W. W. Steward: 1988, '"High Dilution" Experiments a Delusion', *Nature* **334**, 287–90.

Norton, J.: 1989, 'Eliminative Induction as a Method of Discovery: How Einstein Discovered General Relativity', paper prepared for 'How Theories are Constructed: The Methodology of Scientific Creativity', Greensboro Philosophy Symposium, 24–26 March 1989, Greensboro, North Carolina.

Norton, J.: 1991, 'Thought Experiments in Einstein's Work', in T. Horowitz and G. Massey (eds.), *Thought Experiments in Science and Philosophy*, Rowman and Littlefield, Savage, Maryland, pp. 129–48.

Norton, J.: 1993, 'The Determination of Theory by Evidence: The Case for Quantum Discontinuity, 1900–1915', *Synthese* **97**, 1–31.

Peat, F. D.: 1989, *Cold Fusion: The Making of a Scientific Controversy*, Contemporary Books, Chicago.

Poincaré, H.: 1912, 'Sur la théorie des quanta', *Comptes Rendus* **153**, 1103–08.

Popper, K.: 1959, *The Logic of Scientific Discovery*, Hutchinson, London.

Quine, W. V. O.: 1953, 'Two Dogmas of Empiricism', in *From a Logical Point of View*, Harvard University Press, Cambridge, pp. 20–46.

Department of History and Philosophy of Science
University of Pittsburgh
Pittsburgh, PA 15260
U.S.A.