

Between History and Logic

Some people think that normative philosophy of science is finished. That is, the project of identifying correct methods for empirical enquiry is simply unfeasible. On the other hand, formal learning theory is a rigorous investigation of the formal constraints on empirical enquiry (both self-conscious enquiry by scientists and spontaneous learning by children). In this paper I shall explore the relationship between formal learning theory and the *fin de méthode* view.

I shall argue that reports of the death of normative philosophy of science are exaggerated. However, what we cannot hope for is a science of science, that is, a wholly general and unified account of enquiry as such. This is partly because our understanding of enquiry must draw on a variety of disciplines that have incompatible standards of rigour and therefore cannot be unified into a single discipline (here I shall discuss logic and history, though we could easily include psychology and sociology). Partly, it is because enquiries are too diverse to constitute the domain of a science.

First, I shall review the arguments that are supposed to have done for normative philosophy of science. Then I shall briefly examine some familiar arguments (made by Wesley Salmon, Larry Laudan and Clark Glymour) about Bayesianism. These show that Bayesianism cannot be the whole story about scientific enquiry. Finally, I shall examine formal learning theory in the light of our prior experience with philosophy of science in general and Bayesianism in particular. I shall argue that much of our accumulated wisdom about enquiry makes essential reference to features that vary in structure from case to case, and therefore cannot be abstracted into a general account. The same points tell against any general logic of enquiry. Consequently, normative philosophy of science has to be a tradition of methodological wisdom rather than a science of enquiry. Success, for formal learning theory, means finding a place in that tradition.

This plan requires quite a long run-up before I get to say anything directly about formal learning theory. However, in inter-disciplinary discussions, a review of the journey to here taken by one of the disciplines can prevent a lot of talk at cross purposes. Since much of what follows will have a rather deflationary tone, let me make clear from the outset that in my view formal learning theory is a Good Thing. Others in this volume explain why it is a Good Thing. My aim is to explore its limits and its relations with some of the other Good Things in the philosophy of science larder. However, this voluntary division of labour does not mean that there are no disagreements in this volume. Clark Glymour ends his contribution with the slogan “Epistemology has but two aspects: religion and mathematics”. He argues this claim by reporting that philosophers, when they hear about formal learning theory, usually want to set it aside and return to ‘religious’ (that is, familiar philosophical) questions.¹ I do not doubt that his report is true—because it sounds like the reaction that philosophers tend to give to *any* alternative to their familiar topics. They may not want to hear about mathematical epistemology, but they are scarcely more ready to hear about the history of science and mathematics, except of the most anecdotal *1066 And All That* variety. Epistemology has *many* aspects—of which mathematics is one.

¹ “Is there an external world, are there other minds, is truth relative to belief, what is the best method of enquiry, what can we know?” [page ref—last page of Glymour].

1) *The End of the Myth of Method*

Very few philosophers of science now believe that there is a single logical shape called ‘the scientific method’ that distinguishes science from other kinds of enquiry and accounts for its successes and failures. The collapse of the myth of method is popularly associated with Kuhn and Feyerabend but it is less widely appreciated that what was, in their hands, a scandalous heresy has become an orthodoxy. The basic claim was made in a very sober form at the 1974 meeting of the Philosophy of Science Association. At that time the point was still widely contested, so the author had to proceed carefully, “It may sound strange, if not heretical to suggest, as I wish to do, that there is no such thing as the rationality of science. At best we can talk about rationalities of science.” (Skolimowski 1974 p. 191). Two decades later the cautious tone had given way (in another writer) to something close to sarcasm:

...there was a view of science that commanded widespread popular and academic assent... I shall call it “Legend”... Champions of Legend acknowledge that there have been mistakes and false steps here and there, but they saw an overall trend toward the accumulation of truth, or, at the very least, of better and better approximations to truth. Moreover, they offered an explanation of both the occasional mistakes and for the dominant progressive trend: scientists have achieved so much through the use of SCIENTIFIC METHOD (Kitcher 1993 p. 3).

Kitcher’s sarcasm is directed at the Myth of Method (‘Legend’), not at science itself. His book is subtitled ‘Science without Legend, Objectivity without Illusions’, and aims to show how we can understand progress in science without having to subscribe to this ‘Legend’.² Examples of similarly dismissive attitudes to ‘Legend’ could be multiplied without difficulty, not least because the story of the fall of the myth of method has become standard material for writers of undergraduate textbooks.³

How did this shift happen, and why did it happen just then, in the second half of the twentieth century, twenty-three centuries after *Posterior Analytics* and three and a half centuries after Bacon and Descartes? The crucial development seems to be the coming to maturity of the historiography of science. This is a twentieth-century development.⁴ There were historians of science in earlier centuries, but they were isolated and did not constitute a discipline. Perhaps as a consequence, their work tended to be synoptic and lacked consistent attention to detail. Moreover, such histories were rarely philosophically disengaged. Think, for example, of Whewell’s *History of the Inductive Sciences* (1857). It is certainly philosophically-motivated (a decade earlier Whewell wrote *The Philosophy of the Inductive Sciences, Founded Upon Their History*). Since his history covered the entire history of empirical science in two volumes it could not hope to pay close attention to the fine detail of every episode. Only in the last

² The motto of Kitcher’s book is Shakespeare’s Sonnet 130 ‘My mistress’ eyes are nothing like the sun’.

³ For example John Losee’s *Historical Introduction to the Philosophy of Science* (1993) ends in a debate over the very possibility of normative philosophy of science. Larry Laudan’s 1987 article reverses the rhetoric of Skolimowski’s 1974 piece; in arguing that philosophers should continue to seek *the* scientific method, Laudan consciously sets himself against the grain of current opinion. It should be noted that ‘philosophy of science’ here means philosophy of science *written in English*. The French, for example, have never really forgotten the lessons taught by Duhem and Bachelard.

⁴ See Kuhn (1986) for an indication of the rate of growth of research in the history of science.

century have we seen the establishment of degree programmes, peer-reviewed journals and the rest of the institutional apparatus that allows pioneering and programmatic work to give way to normal research. Only under these circumstances is it likely that scholars will devote themselves to the intense study of narrow domains. And when present-day historians write synoptic histories, they can root their accounts in the monographs and articles of specialists on *this* science in *that* period. No such specialist literature existed in Whewell's day.

How did the development of the history of science as a mature discipline tell against the Legend of Method? This is a large question since neither the history nor the philosophy of science is a monolith, so the story of their encounter cannot be simple. As the case of Whewell shows, they were tangled up together before the history of science emerged as an independent discipline (and indeed before normative methodology became a specialism). Among the philosophers, there was little understanding that history is a separate discipline with its own characteristic standards of rigour. The most naïve philosophers hoped that a simple appeal to the facts of history would bear out their models of the logic of science. Others were more sophisticated—Lakatos, for example, drew on his earlier life as a Marxist to argue that every historical narrative presupposes an ideology (or in the case of the history of science, a methodology), so the competition between methodologies becomes a competition between their respective associated historical narratives.⁵ On the whole, though, most English-speaking philosophers of science had little to say about the philosophy of history. Those who engaged with the philosophy of history at all usually assumed that historical explanations are like explanations in natural science or, if they are not, they ought to be.⁶ Others seemed to believe that historians do no more than set facts in chronological order, ready for philosophers to use.⁷ This naivety about the nature of historiography may explain the optimistic hope that the history of science could serve as an objective test-bed for philosophical theory. If an account of method fails to capture the practice of the greatest scientists in history (went the thought), then it cannot be correct. After all, the great scientists must have been doing it right—the philosophers' job is to work out the logic of what they were doing. Thus, history was to supply philosophy with an objectivity that it is usually supposed to lack. History would teach philosophy by examples. For most English-speaking philosophers of science, this 'historical turn' became problematic in the shape of Thomas Kuhn.⁸

Kuhn claimed that there is something in the very activity of history-writing that is incompatible with the traditional quest of normative philosophy of science for general methodological precepts. His account of historical method is rather sketchy, and the majority of his critics in philosophy of science had little interest in the philosophy of history. It is hardly surprising, therefore, that Kuhn's attempt to argue the incompatibility

⁵ 'History of Science and its Rational Reconstructions' (in Lakatos 1978 vol. 1)

⁶ See for e.g. Hempel (1963). Of course, Hempel did not *thoughtlessly* assume that historical explanations must appeal to 'covering laws'. Rather, he formed a view about explanation in general that drew its most compelling examples from natural science, and then insisted that historical explanations must be of this same sort. In doing so (I think) he articulated and defended the tacit assumption of most English-speaking philosophers of science at the time.

⁷ E.g. Laudan (1987).

⁸ I argue for my reading of Kuhn in Larvor (2003).

of rigorous history and normative methodology generated more heat than light. Nevertheless, from Kuhn and his sources I extract the following.

First, historians treat events, including scientific developments, in their own terms rather than as preparations for the present. Further, they understand as good humanists⁹ that a word or deed means what it means in virtue of what is said and done around and about a given occasion of speech or action. This entails that we cannot extract antique theories from their contexts for the purpose of comparison with later theories, because when so extracted they cease to be themselves. We cannot, for example, compare what Aristotle had to say about bodies under gravity with what Newton thought because Aristotelian bodies, strictly speaking, do not encounter gravity. It does not follow that meaningful comparisons cannot be made, but they require such a labour of translation and explication that we do not get from them regimented data suitable for testing philosophical theses. All we get is a plausible account of the journey from there to here. So we cannot, for example, do a statistical survey of past science to see what logical approaches have worked best.¹⁰ To do so, we would have to decontextualise episodes in the history of science to such a degree that they would become unrecognisable to the historians who know them best. Such surveys of ‘the historical record’ are not on because the historical record is not and cannot be a body of regular, standardised data like Wisden.

Second, making sense of the science of the past often consists in setting it in its proper time and place. In other words, historians of science typically deny *internalism* (the view that the growth and trajectory of science can be explained without reference to anything except the encounter between argument and evidence). For example, in the most recent edition of *Isis*¹¹ the leading article is “Wonderful Secrets of Nature: Natural Knowledge and Religious Piety in Reformation Germany” by Kathleen Crowther-Heyck. Internalist philosophers of science would insist that the effects of religious piety on German science must have been short-lived, inessential and probably regrettable. Historians would consider it unrigorous to approach the question with such prejudices in hand (perhaps religious piety played an essential, long-lasting and laudable role). To take another leading history of science journal, *History of Science*, the March 2003 edition contained the following major articles: “‘Purifying’ Science: E. C. Slater and Postwar Biochemistry in the Netherlands” (Ton van Helvoort); “Herbert Spencer and the Disunity of the Social Organism” (James Elwick); and “‘Men of Science’: Language, Identity and Professionalization in the Mid-Victorian Scientific Community” (Ruth Barton). The rejection of internalism is less obvious from these titles (though the scare quotes around

⁹ That is to say, academics with the mental habits characteristic of the traditional ‘humane’ disciplines of history, literature and philosophy. Good humanists in *this* sense need not be humanitarians, nor need they be atheists.

¹⁰ I have in mind Laudan’s suggestion in his (1987). He proposes treating methodological precepts as hypothetical imperatives, so that the differing cognitive goals of present-day and historical scientists are taken into account.

¹¹ June 2003. George Sarton founded *Isis*, an international journal dedicated to the history of science, in Belgium in 1912. After World War I, Sarton and the journal moved to the United States. Today, *Isis* is edited by Margaret Rossiter at Cornell University, and published and distributed quarterly by the University of Chicago Press.

‘purifying’ and ‘men of science’ are significant¹²). Nevertheless, it is clear that van Helvoort, Elwick and Barton are not simply recording successive encounters between hypotheses and data. These examples were chosen arbitrarily (simply by taking the most recent editions at the time of writing) and could easily be multiplied.

Third, historians are wont to historicise everything, including the logical categories that philosophers need in order to do normative methodology. Here is Kuhn on distinctions such as that between the contexts of discovery and justification¹³:

For many years I took them to be about the nature of knowledge, and I still suppose that, appropriately recast, they have something important to tell us. Yet my attempts to apply [these distinctions] even *grosso modo*, to the actual situations in which knowledge is gained, accepted, and assimilated have made them seem extraordinarily problematic. Rather than being elementary logical or methodological distinctions, which would then be prior to the analysis of scientific knowledge, they now seem to be integral parts of a traditional set of substantive answers to the very questions upon which they have been deployed.¹⁴

Any historian worthy of the name will want to know how the tradition that Kuhn mentions (“a *traditional* set of substantive answers”) arose and developed, in response to what and against what opposition. In this historicist light, the apparently innocent distinctions employed by philosophers suddenly seem question-begging. Kuhn’s importance, in my view, is to have articulated (however obscurely) the stance of professional historians towards the objects of their studies.

Historicism may be natural to historians but it is alien to most philosophers of science, and, more importantly, inimical to their philosophical projects. Some philosophers embraced historicism: history (they insist) teaches that everything is relative to its moment, to its place in the great flux. The search for timeless abstractions called ‘correct methods of enquiry’ is therefore pointless or perhaps even impossible. However, this historicist, relativist, quietist line fails to acknowledge that questions about scientific procedure and propriety are not mere philosophical cobwebs. Sometimes, scientists (and science funding bodies) have decisions to make that involve normative methodological judgments. We cannot be content to say “Do the done thing for your time and place”, because current practice may be mistaken or undetermined. It is precisely in cases of underdetermination or doubt that normative issues become pressing. Then there are those occasions when scientific controversy achieves a wider public significance, such as the struggles over creationism, food safety and global warming. In such disputes, each side inevitably accuses the other of being unscientific and these accusations are not always empty or philosophically naïve. What is more, philosophers have accumulated a rich store of methodological ideas plus some impressive formal theories (such as formal logic, probability and the formal learning theory discussed in this volume). It is trite to dismiss

¹² ‘Men of science’ is the nineteenth-century transitional phrase between ‘natural philosopher’ and ‘scientist’.

¹³ In the ‘context of discovery’ hypotheses are invented; in the ‘context of justification’ they are evaluated. In the Twentieth Century it was widely held (by Popper and most of the Vienna Circle, for example) that there is no logic of hypothesis-invention but there is a logic of hypothesis-evaluation.

¹⁴ Kuhn 1970 p. 9.

all this with a single historicist gesture. The Legend of the One True Method may have been discredited, but the original question remains: *how should we enquire?*

The history of science, then, is not the friendly source of methodological morality tales for which normative methodologists once hoped. On the contrary, historians deride such tales as ‘whiggism’. If methodologists are to address the enduring questions about the nature of enquiry, they must first disengage themselves from the history of science, at least as it is written up by professional historians. This is relatively easy to do. Notice first that the historicist dictum—that historical phenomena owe their identities to their places in the great flux, and must therefore not be abstracted for fear of distortion—is not a conclusion of historical enquiry. It is, rather, a methodological precept. Since philosophers do not share the aims of historians, they need not share all of the historians’ premises and methods. Historians usually want to know why *this* scientist at *this* time and place wrote, spoke and acted as he did. Philosophers have no professional interest in such specifics—their interest is in the argument as such (if such a thing can be identified). In any case, the history of science may not be as relevant as philosophers once thought. Larry Laudan argues that methodological precepts should be read as hypothetical (rather than categorical) imperatives: ‘If you have cognitive goals {A, B, C}, then use method {a, b, c, d...}’. Most of the great scientists of history had different cognitive goals to scientists of the present day (Newton and Boyle, for example, saw the construction of a natural theology as a central task of science¹⁵). Consequently, the successes and failures of their methods are irrelevant to us. This allows Laudan to square our intuition that the great scientists of the past were instrumentally rational (that is, they effectively matched means to ends) with Feyerabend’s observation that they rarely acted in conformity with methodological precepts that seem compelling to us. In any case, there is no *a priori* reason to suppose that the best methods of enquiry lie in the recorded past at all. Perhaps they have not yet been discovered.

Disengagement from the history of science returns methodologists to the problem that reference to the historical record was supposed to solve. How are methodological questions to be resolved? Philosophical argument rarely resolves anything, and philosophical intuitions are hardly objective. The history of science was supposed to provide an objective test-bed for philosophical theories. If the history of science is not suitable for this purpose, then where is normative methodology to find its objectivity? Studies of present-day scientific practice are unsuitable for many of the same reasons that prevent history from grounding philosophical arguments. One attractive option is to abandon the search for empirical-historical tests of methodological theory altogether, in favour of an *a priori* account. Philosophers cannot, nowadays, pretend to offer substantive *a priori* knowledge. But mathematicians can. We come at last (by a lengthy but, I hope, dialectically satisfying route) to the present popularity of formal accounts of enquiry.

2) The Formal Turn: Bayesianism

Insofar as normative methodology is turning to mathematical models of enquiry, it is returning to where it was before its dalliance with history. English-speaking philosophers of science used to look to formal logic and mathematical accounts of

¹⁵ These examples, and the argument here reported, are taken from Laudan (1987).

induction in their attempts to specify the ‘logic of science’.¹⁶ Then, from Kuhn and others, we learned to suspect such models. These logics seemed so abstract as to have lost sight of actual scientific practice altogether, and depended for their intelligibility on a set of arbitrary and question-begging distinctions. Now, having learned that it is hard to sleep with historians without catching their historicist fleas, many philosophers of science seem ready to turn back to mathematics. Dialectical journeys often lead back to a more sophisticated and acutely self-aware version of the starting-point. The trick, of course, is not to forget what was learned *en route*.

The most popular mathematical model of enquiry at the time of writing seems to be Bayesianism.¹⁷ It is worth taking a glance at Bayesianism, since its difficulties are typical of those suffered by formal accounts of enquiry. Moreover, the sort of argument that I wish to make with respect to formal learning theory has already been made in the case of Bayesianism by Wesley Salmon, Larry Laudan and Clark Glymour, among others.¹⁸

In order to use Bayes’ theorem one must have identified a small (or at any rate, a finite) number of ‘serious’ or ‘plausible’ hypotheses. One must already know what counts as a plausible explanation, what counts as relevant evidence and what counts as background belief. These judgments of relevance and plausibility are only possible if one already knows (or at least, believes that one knows) a great deal about the domain under investigation. In case anyone should suppose that relevance and plausibility can be decided by pre-scientific common sense, recall that only a few centuries ago, serious, intelligent people thought that the number of planets is related to the number of holes in the human head. In other words, the infinity of possible worldviews and underlying metaphysical schemes must somehow be cut down to a manageable handful *before* the Bayesian story can start. Therefore, Bayesianism cannot supply a complete account of scientific enquiry.¹⁹ Salmon suggests that the Bayesian algorithm be supplemented with a Kuhnian account of how relevance and plausibility are fixed. Indeed, Salmon hoped for a synthesis between the historical and logical sides of philosophy of science. In the end, though, the Kuhnian supplement comes to dominate the Bayesian element in his account: “The [Bayesian] algorithms are trivial; what is important is the scientific judgment involved in assessing the probabilities that are fed into the equations.”²⁰

¹⁶ And in the case of the Vienna Circle, German-speaking philosophers too.

¹⁷ Bayes’ theorem allows us to evaluate the conditional probability $P(T/E.B)$ that a given hypothesis (T) is true provided both our background beliefs (B) and some new piece of evidence (E) are true:

$$P(T/E.B) = \frac{P(T/B) \times P(E/B.T)}{P(E/B)}$$

To use (even this simple version of) the theorem, we must know i) the conditional probability of the hypothesis being true given only the background beliefs; ii) the conditional probability of the evidence (E) given that both the background beliefs (B) and the hypothesis (T) are true; and iii) the conditional probability of the evidence (E) given that the background beliefs (B) are true.

¹⁸ Glymour (1981); Laudan (2000); Salmon (1990).

¹⁹ Bayesians know this. Most Bayesians regard Bayesianism as no more a set of rational norms on belief formation and modification.

²⁰ Papineau (ed) 1990 p. 287.

Laudan's argument²¹ focuses on just one methodological thought: a good theory ought to address all the phenomena in its field. Other things being equal, it is a shortcoming in a theory T to say nothing about some phenomenon p in its domain of enquiry. In Bayesian terms, T says nothing about p if the conditional probability of p given T equals the prior probability of p . In this case, the likelihood of T in the face of p is exactly the same as its probability prior to p . That is to say, a true Bayesian's confidence in T will be unaffected by T 's failure to address p , even though p is among the phenomena that a good theory in this domain ought to address. A Bayesian might reply that T could still be true—but we do not want scientific theories that are merely true, we want ones with deep explanatory power. Of course, T may offer such significant explanatory benefits on some other front that it effectively re-defines the domain so that p falls outside it. Then T 's failure to address p ceases to be an issue. This, though, is not a judgment that Bayesianism can formalise. What loss of scope is a price worth paying for a gain in explanatory power? Like Salmon's judgments of relevance and plausibility, this part of scientific thinking falls outside of the Bayesian account. It remains to be seen whether *any* formal account can model such judgments.²²

Clark Glymour offers a generalised version of Laudan's argument. Glymour maintains that:

There are a variety of methodological notions that an account of confirmation ought to explicate and methodological truisms involving these notions that a confirmation theory ought to explain: for example, variety of evidence and why we desire it, *ad hoc* hypotheses and why we eschew them, what separates a hypothesis integral to a theory from one 'tacked on' to the theory, simplicity and why it is so often admired, why 'de-Occamized' theories are so often disdained, what determines when a piece of evidence is relevant to a hypothesis...²³

Glymour goes on to explain that, in his view, Bayesianism can explicate some but not all of the items in this list, and that "There are elementary but perfectly common features of the relation between theory and evidence that the Bayesian scheme cannot capture at all without serious—and perhaps not very plausible—revision."²⁴ He goes on to explain that he considers Bayesianism pertinent for statistical reasoning and that it captures some principles of ordinary reasoning. I report his view here but not his arguments for it, since these would put off consideration of formal learning theory even further.

My point is that contemplation of actual scientific practice, for all it led philosophers to an unwelcome brush with historicism, did leave us with a fund of methodological wisdom. Not the One True Method, but a collection of methodological 'notions and truisms', together with a respect for the sensitivity to the particular situation that is required to use them well. Bayesianism cannot be a comprehensive account of scientific reason unless it can model and account for all these notions and truisms. This point tells only against those who imagine that Bayesianism is the whole story—no-one denies that it has something important to say. Most Bayesians regard Bayesianism as a

²¹ Laudan 2000 p. 173.

²² In this volume, Clark Glymour argues that formal learning theory can articulate some of these trade-offs, even if it cannot calculate them.

²³ Glymour in Papineau (ed) 1990 p. 293.

²⁴ Glymour in Papineau (ed) 1990 p. 294.

set of constraints on the relative strengths with which we hold our beliefs, but insist that how those beliefs are formed and evaluated within these constraints is another question. Bayesianism cannot by itself explain scientific successes (though it may explain those failures in which its formal constraints are violated) nor can it offer advice about the choice of concepts, tools, background metaphysics, problems to address and experiments to attempt.

3) *Formal Learning Theory*

The simple claim that I have just made with respect to Bayesianism seems to come for free in the case of formal learning theory. For, learning theory asks under what conditions a problem is solvable ‘in the limit’, that is, does the sequence of conjectures produced by a given scientist-function stabilise in the long run? This question matters to us even though ‘in the long run we’re all dead’, because “if you can’t know the truth in the long run, you can’t know it in the short run either.”²⁵ So learning theory offers us constraints of the form: learners with *these* computational bounds will never solve problems of *that* logical type. It might seem that we could stop here with the conclusion that learning theory provides theorems about what is possible in the limit that leave a lot of latitude for the exercise of informal expertise (like Salmon’s attempt to yoke Bayes and Kuhn together). However, this will not do, first because learning theory does not entirely respect the short-run/long-run distinction, in the sense that it considers questions of efficiency (which scientist-functions stabilise quickest?). Second, learning theory is able to formalise *some* of our common stock of methodological notions and maxims (such as consistency, conservativeness and decisiveness, for example) in order to assess their effects in the limit. An obvious question arises: can it formalise all of them? If not, what part of our accumulated methodological wisdom escapes formalisation, and why?

To make the question more pointed, consider this programmatic passage from Martin & Osherson’s *Elements of Scientific Inquiry*:

Whatever the motives for studying inquiry, we must begin by appreciating the complexity of the subject matter. As a form of human behavior, science involves a wide range of activities, both in and out of laboratories. Surely such a phenomenon cannot be understood without substantial idealization. By limiting attention to just a few, salient aspects of science we may hope to understand their interaction within the larger scheme, and eventually illuminate further variables that can be added to our model at a later stage.²⁶

We are given no reason to suppose that enquiry (considered as a complex) has a logically modular structure that would allow us to isolate and model some of its ‘elements’ without reference to the whole activity—notice that this atomism is incompatible with the holism proper to historical studies. Nor is it obvious that anything recognisable as human enquiry remains after the ‘substantial idealization’ required for formalisation. Indeed, it may be that a mathematically precise model of enquiry falsifies the phenomenon. Some case-studies seem to show that ambiguity is necessary for progress.²⁷

²⁵ Glymour 1996 p. 282.

²⁶ Martin & Osherson *Elements of Scientific Inquiry* (1998) p. 1. Notice that this title suggests more than a study of formal constraints on inquiry.

²⁷ I have in mind Lakatos (1976), in which the elasticity of concepts is essential to the growth of mathematical knowledge. Also, see Grosholz (2000), who argues that mathematical knowledge sometimes

To return to our question, let us see if there are any methodological notions and precepts that cannot be included in a sequence of progressively more complex formal models as envisaged by Martin & Osherson.

The first thing that leaps out of the literature on formal learning theory is the notion of a ‘paradigm’. This term is clearly owed to yet distinct from the concept introduced by Kuhn.²⁸ One reason to adopt this Kuhnian terminology is this: learning theory has from its inception tacitly accepted the Kuhnian view that enquiry can only get going when there is a consensus on all the hard metaphysical and methodological questions. In Gold (1967), published only five years after *The Structure of Scientific Revolutions*, the class of possible languages is specified in advance. In later expositions, enquiry is modelled as a game played between Nature and a scientist. The players are given a set of possible realities. Nature chooses one of these possibilities as the actuality, and the scientist has to identify this from the data-stream, or ‘environment’ supplied by Nature. Technically, a ‘paradigm’ formal learning theory in has five elements: the set of potential realities; a problem; for each possible world a set of data-streams or ‘environments’; some scientist-functions; and a criterion of success.²⁹ All this has to be fixed in advance, as in Kuhnian normal science (and Bayesianism: recall the seven stars/seven apertures). There is an important difference, in that in Kuhnian normal science the paradigm is socially agreed rather than explicitly stated. Moreover, the social agreement is not a set of rules but rather a collection of paradigmatic examples; and the question of the similarity of new cases to these examples is forever open-ended. Nevertheless, the Kuhnian view that scientists have to fix their methods and metaphysics *before* they can do any science is unusual among formal accounts of enquiry.

However, learning theory says nothing about the central Kuhnian question: how do we get from a failing paradigm to a more promising one? Of course, philosophy of science does not have an agreed, established answer to this question. It does, however, have some methodological notions and truisms. One we have already had from Laudan is that a theory (or paradigm or research programme or tradition of enquiry—the exact unit of analysis is not important) is in trouble if it cannot explain phenomena that clearly lie within its domain.³⁰ Such problematic phenomena need not contradict the theory; they need only embarrass it by showing that its scope is insufficiently broad. This, as we noted in the case of Bayesianism, raises the whole problem of the unity and limits of a domain of enquiry. How do we establish the domain-boundaries that allow us to say that a theory fails to address some of its proper objects? For to do so, we must have some way of identifying the domain that is independent of the theory in hand.

This question is especially knotty because domain-boundaries can shift under the influence of a dramatically successful new theory. For example, Galileo at once narrowed an Aristotelian domain, as he separated the science of motion from the general

grows by hybridisation, and that these hybrids “often admit an instability or inconsistency that is however held in place or made tractable by the rational relatedness provided by the abstract structure that holds the domains together” (p. 88). Her case study suggests that instability or inconsistency may be more than regrettable flaws; they may be essential for progress.

²⁸ I am tempted to label this new sense of ‘paradigm’ thus: paradigm₂₃, in honour of the twenty-two senses of the term that Margaret Masterman discerned in Kuhn (1970).

²⁹ Martin & Osherson (1998) pp. 2–3.

³⁰ Laudan calls these ‘non-refuting anomalies’ (2000 p. 166); Lakatos calls them ‘heuristic falsifiers (1976 p. 82).

problem of change; but at the same time he unified the physics of the Earth and the Heavens. Nevertheless, insofar a domain of enquiry is a natural unit, it is so in virtue of commonalities and connections in its subject matter. There is a science of physics in virtue of commonalities among physical processes. There is a science of primatology in virtue of commonalities among primates. It seems to me unlikely that we will ever have a general logical account of the unity of domains because the commonalities that unify a domain are inevitably specific to that domain. The commonalities that unite primates into a domain of enquiry seem to be of a quite different sort from the commonalities among physical processes that render a science of physics possible. If I am right about this, then *no* formal theory of enquiry can ever fully articulate those methodological notions and truisms that refer to the borders between domains of enquiry. Those borders are established and maintained by means and criteria that are local to those domains, and will therefore resist formal modelling. Indeed, for the same reason, no general account (formal or otherwise) can go far beyond the truism that, other things being equal, a theory should address all those phenomena in its domain. The case of Galileo shows why: sometimes, progress requires that domain boundaries be redrawn.

Next, there is the obvious passivity of the ‘scientists’ (that is, scientist-functions) in learning theory. They do not need to carry out experiments, because the data is supplied to them by the ‘environment’. In formal learning theory, the ‘environment’ is nothing but a stream of data. This is a more significant idealisation than may at first appear. An experiment is a complex of physical processes (bombarding *this* with a stream of *those*, mixing some of *that* with a tincture of the *other*, etc.). It takes considerable scene-setting to turn such manipulations of matter into arguments. Doing or seeing something does not, by itself, generate propositions. It is at this point that familiar considerations about the theory-laden-ness of data enter the story. The conduct and interpretation of experiments normally requires expert judgment, tacit knowledge and a trained eye (and if the experiment is to be robust it must be repeated in a variety of different laboratories, requiring a whole community of trained eyes).³¹ All our accumulated sensitivity to the logical and phenomenological subtlety and fallibility of experimentation is abstracted away in formal learning theory. Instead, we have an optimistic sort of empiricism in which the ‘environment’ supplies indubitable facts, ready to use, to a passive observer. This is probably as it should be, as it is not obvious how the experimenter’s skill and tacit understanding could be formally modelled. Here again, if I am right, we have an essential aspect of enquiry that resists formal modelling.³² It is not just that the results of experiments are empirical and therefore less than wholly certain. It is that navigating these uncertainties requires hands-on know-how that is nevertheless part of the logic of empirical enquiry.

The ‘scientists’ of formal learning theory can usually get away with passively contemplating whatever data they are fed because it is normally stipulated that the ‘environment’ drops all the facts into their laps eventually. In the case of language-learning, Gold’s 1967 paper assumed that every string in the target language is presented

³¹ See e.g. Galison (1987).

³² This connects with an issue in artificial intelligence and philosophy of mind. Some philosophers (under various influences including Heidegger, Merleau-Ponty, Kuhn and the later Wittgenstein) deny that know-how can be comprehensively analysed into propositional knowledge-that. If they are correct then experimenters’ skills cannot be entirely coded in a logic-system. See Dreyfus 1992.

to the learner at least once.³³ This is a reasonable assumption if the point is to establish impossibility results, because a language that is unlearnable with this generous assumption is certainly unlearnable without it. However, deciding what data to collect is a crucial part of scientific judgment, not least because it is often the case that data cannot be collected passively. Much of our information about the world cannot be gathered without the use of specialist machines, and will not be gathered unless some scientist believes that it can be gathered and is worth gathering. No-one committed to humour-medicine would bother to go looking for germs. No-one committed to perfect, unchanging Heavens would bother to set up a screen to display sun-spots. Moreover the available technology sets limits to the range of possible experiments and observations. In learning theory the available data is specified when the ‘environment’ is fixed, but (reasonably) the ‘environment’ is not changed by the ‘scientist’s’ learning. The ‘environment’ sends its data in a pre-determined stream regardless of the ‘scientist’s’ state of knowledge. In practice, new knowledge raises new questions, which in turn elicit data that would not otherwise have been forthcoming.

Now, learning theorists know that their ‘scientists’ are unnaturally passive. Therefore they developed the concept of an ‘oracle’ that collects data from the ‘environment’ and feeds it to the ‘scientist’. The oracle decides what data to collect next as a function of the information received so far.³⁴ This, though, merely raises questions of efficiency (that is, which oracles help scientists to reach their success-points quickest?). So far as I know, learning theory does not model the fact that a scientist may never see certain data because his false beliefs prevent him from looking for them. The oracle stands between the ‘scientist’ and the ‘environment’, and determines the order in which the data arrives. The oracle is not modified by what the ‘scientist’ learns (though in advanced versions of learning theory the ‘scientist’ may be modified by what *it* learns). Real scientists change their search-strategies as well as their hypotheses in the light of new information. In learning theory, the real scientist has been split into a ‘scientist’ that forms beliefs and produces hypotheses, and an oracle that decides what question to ask next. The point here is that changes to the ‘scientist’ ought to induce changes in the oracle. This brings us back to the first point: there are no paradigm-shifts (or problem-shifts, or changes of background metaphysics) in learning theory.

A third methodological notion that learning theory may struggle to articulate is that of the *ad hoc* hypothesis. In fact there are two principal senses in which a hypothesis may be *ad hoc*. One is straightforward: a hypothesis is *ad hoc* if it has been cooked up solely to save a theory from one particular counterexample, but offers no explanation or insight into anything beyond that one case (“All swans are white—obviously, this pink one has been dyed.”). The other sense of *ad hoc* requires reference to a research

³³ Gold 1967 p. 448. This presentation can take two forms: ‘text’ and ‘informant’. A ‘text’ for a language L is “a sequence of strings x_1, x_2, \dots from L such that every string of L occurs at least once in the text” (*ibid.* p. 450). An ‘informant’ for L “can tell the learner whether any string is an element of L , and does so at each time t for some string y_t ” (*op. cit.*). In the case of learning from an informant rather than from a text, it is assumed that every finite string of the alphabet common to the possible languages will come up at least once. No other strings are presented, and each string is accompanied by information as to whether the string is correct or incorrect. He briefly considers ‘request informants’ that answer queries about strings chosen by the learner, but quickly proves that these are equivalent to the ‘arbitrary informant’ that rules on all the possible strings in an arbitrary order (*ibid.* p. 467).

³⁴ Martin & Osherson (1998) pp. 87–90.

programme: a hypothesis is *ad hoc* in this sense if it is developed using resources from outside the proper repertoire of the research programme. A hypothesis could be *ad hoc* in the second sense but not in the first. That is, a hypothesis could be an insight and a step in the right direction, and yet be *ad hoc* with respect to a research programme that makes use of it (“The flow of humours through a wound may be blocked by tiny animals, so we should boil our instruments and bandages to kill them.” Here, germ theory is *ad hoc* with respect to humour-medicine, but is in itself an insight.). It may be possible to formalise the first sense of *ad hoc*. However, general accounts of methodology, including formal learning theory, will struggle to articulate the second notion of *ad hoc*-ness because it requires the notion of a sustained and unified enquiry (or paradigm, or research programme, *etc.*) as a heuristic whole. Here, the point is the same as I argued above for domains of enquiry. The principle of unity of a research programme is specific to that programme. If the unity of a research programme cannot be formally modelled, then it must also be impossible to model the case in which a given hypothesis violates that unity (excluding trivial models in which the unity of the programme is an unanalysed black box).

The pattern of argument should be obvious now if it was not before: I have deliberately picked out methodological notions and truisms that require some reference to the specifics of the domain of enquiry or the unity of the enquiry itself. It is, ultimately, the unity of these complex wholes (domains of enquiry and research programmes) that resists formalisation. There is another notion that depends on them, which for the sake of a label we might call the ‘Whewell bonus’. This is the case in which a theory predicts or explains a phenomenon that it was not originally intended to address; the case in which even the theory’s champions are pleasantly surprised by its success. This notion is notoriously difficult to formalise because as it stands it involves the intentions of the theory’s originator. These intentions (being psychological items) ought to be irrelevant if the Whewell bonus is an objective virtue of theories. The challenge is to capture the thought that a theory has reached beyond its original scope without making any psychological reference. Here again, I would argue for particularism: there is no general logical account of the original scope of theories. If in any given case it is possible objectively to determine the original scope of a theory, it is only by paying attention to the details in hand. Apart from anything else, it may be that the proper scope of a theory is in part a function of the history of the discipline in question prior to the formulation of the theory. The same point comes up if we consider unification by consilience, in which a pair of distinct theories in hitherto separate domains of enquiry come to reinforce each other (for example, if studies of population DNA and the evolution of language groups produce the same hypotheses about human kinship and migration).

The Whewell bonus raises another class of methodological notions and truisms, namely, those that exploit the insight that theories are as much tested against each other as they are against nature. We accept our current scientific orthodoxies because they are our least lousy theories so far. Therefore, we cannot fully formalise a set of criteria for accepting or rejecting theories without articulating some principle(s) of comparison between theories. In particular, a Whewell bonus is only a real bonus if the phenomenon has not already been adequately explained by some other theory. I shall not insist on this point, however, as it may be possible to modify formal ‘scientists’ so that they watch each other’s results as well as their own and modify their enquiries accordingly. Indeed,

there are already some results about teams of 'scientists', though these do not answer the present point.³⁵

4) Conclusion

So far I have argued that there are methodological notions and truisms that formal learning theory has not captured, and probably will not ever capture. This is essentially the same claim that Glymour, Laudan and Salmon made about Bayesianism. Since this is a negative claim, let me now allay some natural fears. The point is not to 'attack' formal learning theory, but rather to learn a lesson from the history of our own discipline. When philosophers acquire a new tool, there is a danger that the rising generation of researchers may be so impressed with it that much of what was known before is forgotten. The more impressive the tool, the more intense the danger. Something of the sort seems to have happened to the philosophy of science and mathematics with the development of formal logic from Frege onwards, and it took a rather wrenching 'historical turn' for philosophers to rediscover the sensitivity to history and practice that previous generations took for granted. Let us not make the same mistake again. Let us learn what we can from formal learning theory without losing sight of what we have learned from historical, sociological and phenomenological studies of science. These studies have shown us that much of the rigour of science is bound up with the subject-matters and specialist techniques of individual sciences. Dendrochronology, for example, depends on truths about logic, trees and the weather, plus the judgment of experts and the reliability of their computers. To isolate the logical aspect and forget the rest makes a mystery of the effectiveness of the technique, and this is to invite mystification and scepticism. It is precisely because the rigour of each special science depends on its characteristic objects and techniques that enquiry is not the sort of thing of which there can be a science.

Philosophy ought not to ignore developments in logical theory. But the alternative to mathematical accounts of enquiry is not mere 'religion', and philosophy ought not to ignore historical and phenomenological studies of scientific practice either. Philosophy cannot (as Salmon hoped) simply combine these approaches because their standards of rigour are incompatible. Abstraction is a methodological sin among historians, while historicism is anathema to logicians. Philosophers, caught in the middle, are at once chided by mathematical logicians for lacking the clarity of the exact sciences, and by historians for concocting abstractions while neglecting concrete reality. The point of this article is to explain to exponents of formal learning theory that, if philosophers do not embrace it with both hands, it may be because one hand is already carrying a rather unwieldy load of historical and phenomenological insight.

References

- Dreyfus, Hubert L. (1992) *What Computers Still Cannot Do: A Critique of Artificial Reason*. Cambridge, Mass.: The MIT Press.
- Feyerabend, Paul (1993) *Against Method (3rd edition)*. London: Verso. First edition 1975.
- Galison, Peter L. (1987) *How Experiments End*. Chicago: University of Chicago Press.

³⁵ Osherson, Stob, & Weinstein (1986).

- Glymour, Clark (1981) 'Why I am not a Bayesian'. In *Theory and Evidence*. Chicago: University of Chicago Press, 63–93. Reprinted in D. Papineau (ed.) *The Philosophy of Science*. Oxford: Oxford University Press, 1996.
- Glymour, Clark (1996) 'Hierarchies of Knowledge and the Mathematics of Discovery' in *Machines and Thought: The Legacy of Alan Turing*. P.J.R. Millican & A. Clark (eds). Oxford: Clarendon Press.
- Gold, E. Mark (1967) 'Language Identification in the Limit'. *Information and Control* 10, 447–474.
- Grosholz (2000) 'The Partial Unification of Domains' in *The Growth of Mathematical Knowledge*. Breger & Grosholz (eds). Kluwer (Synthese Library/ volume 289) 81-91.
- Hempel, Carl G. (1963) 'Reasons and Covering Laws in Historical Explanation' in *Philosophy and History*, ed. Sidney Hook (New York) pp. 143-163.
- Kitcher, Philip (1993) *The Advancement of Science*. Oxford: OUP.
- Kuhn, Thomas S. (1970). *The Structure of Scientific Revolutions* (second edition). Chicago: University of Chicago Press, (first edition 1962).
- Kuhn, Thomas S. (1986) 'The Histories of Science: Diverse Worlds for Diverse Audiences'. *Academe* (July-August).
- Lakatos, Imre (1976) *Proofs and Refutations*. Cambridge University Press, Worrall and Zahar (eds).
- Lakatos, Imre (1978) *Philosophical Papers* (volumes 1 and 2). Worrall and Currie (eds), Cambridge: CUP.
- Larvor, Brendan (2003) 'Why did Kuhn's *Structure of Scientific Revolutions* cause a fuss?' *Studies in History and Philosophy of Science* vol. 34/2 pp 369–390 (June).
- Laudan, Larry (1987) 'Progress or Rationality? The prospects for Normative Naturalism'. *American Philosophical Quarterly* 24 (1): 19–31. Reprinted in D. Papineau (ed.) *The Philosophy of Science*. Oxford: Oxford University Press, 1996.
- Laudan, Larry (2000) 'Is Epistemology Adequate to the Task of Rational Theory Evaluation?' In R. Nola & H. Sankey (eds) *After Popper, Kuhn and Feyerabend* 165–176. Dordrecht: Kluwer.
- Losee, John (1993) *Historical Introduction to the Philosophy of Science*. Oxford: Oxford University Press (3rd edition).
- Martin, Eric & Osherson, Daniel (1998) *Elements of Scientific Inquiry*. Cambridge, Mass.: MIT Press.
- Osherson, Daniel; Strob, Michael & Weinstein, Scott (1986) *Systems that Learn*. Cambridge, Mass.: MIT Press.

- Salmon, Wesley C. (1990) 'Rationality and Objectivity in Science or Tom Kuhn meets Tom Bayes'. In *Scientific Theories*, Minnesota Studies in the Philosophy of Science, vol. 14 (Minnesota Press), 175–204. Reprinted in D. Papineau (ed.) *The Philosophy of Science*. Oxford: Oxford University Press, 1996.
- Skolimowski, Henryk (1974) 'Evolutionary Rationality' in R.S. Cohen *et al.* (eds) PSA 1974 pp. 191-213 (published as *Boston Studies in the Philosophy of Science vol. XXXII*. Dordrecht: D. Reidel Publishing Company, 1976).
- Whewell, William (1847) *The Philosophy of the Inductive Sciences, Founded Upon Their History*, 2nd edition, in two volumes, London.
- Whewell, William (1857/1873) *History of the Inductive Sciences, from the Earliest to the Present Time*, 3rd edition, in two volumes, New York.