Progress and Procedures in Scientific Epistemology

Peter Godfrey-Smith

Philosophy Department
Harvard University
Cambridge MA 02138, USA.

April 2007

The 2007 Reichenbach Lecture at UCLA

1. Approaches to Scientific Epistemology

My title is intended to echo Hans Reichenbach's *The Rise of Scientific Philosoph*y (1951), and the phrase "scientific epistemology" is intended in two Reichenbachian senses. One involves the epistemology *of* science; the other involves epistemology undertaken with a scientific orientation. Talk of "progress and procedures" is intended in a similar dual sense.

I start by looking back over the last century, at how a family of problems was tackled by scientifically oriented philosophers. These are problems with the nature of evidence and testing – with how, given our limited access to the world and the ambitious reach of our theories, we can have good reason to believe one such theory over another. These discussions were informed especially by skeptical treatments of the problem in Hume.

We see in this period a number of different theoretical strands. These are characterized by different raw materials, and by different organizing or paradigm cases – different bets regarding the parts of scientific practice that should function as exemplars.

Grouping the work in a rough way, we see first a central tradition which takes deductive logic as starting point and raw material. Deductive logic is seen as the heart of the account of rationality we have developed so far, and as providing the structure with which to build further. When probability theory is used, it is made as logic-like as possible. The scientific examplars in this tradition are the investigation of universal physical laws. The central figures are Hempel and Carnap, with their formal treatments of the inductive confirmation of generalizations. Here I also include hypothetico-deductivism, as it is logic-based although not inductivist in the narrower sense.

A second tradition takes the emerging toolkit of statistical methodology as its raw material. It is inspired by the ways in which parts of 19th and 20th century science *befriended error*, via theories of measurement, the distributions of traits in populations, and inference from samples. In this second tradition the concept of probability is made central, with something like the frequency interpretation of practicing statisticians. Here we see Reichenbach, with C.S. Peirce as a key precursor.

There is also a third tradition, in which the focus is on giving a description of an *idealized rational agent*, and how such an agent handles evidence and decisions. In this tradition there are not clear scientific exemplars, but there is a general epistemic examplar provided by the rational gambler; life is conceived as a gamble with a capricious nature. This is the tradition of Ramsey, de Finetti, and subjectivist probabilism, developing into mainstream Bayesianism in the last quarter of the 20th century. But in this category I also include non-probabilistic theories of ideal agency in Quine and the Jamesian side of pragmatism.

We might also recognize other strands; a fourth is the tradition that focuses on eliminative inference, ruling options *out*. Here we find Popper, and also some defenses of the primacy of inference by elimination outside of philosophy.[1]

These four traditions overlap. The second, statistical strand can often be seen discussing belief management by an ideal agent, for example – though in a reliabilist way that contrasts with the internalism or coherentism of the subjectivists. And as Bayesianism developed in the last part of the century, it drew more and more on material from other strands. But let us think of them as distinct for now.

How did these writers think about the nature of their project, and the status of the conclusions reached? The meta-theory seen in most of this work involves a belief in what we might call "uniform deep structure." The idea is that science and everyday inquiry rely on there being some *basis* for inferences from the observed to the unobserved. This basis can be very unobvious, and the philosopher's task is to reveal it.

Thus we see various forms of the claim that one kind of non-deductive inference is really another kind in disguise. Perhaps all non-deductive inferences can be shown to depend on some one crucial non-deductive bridge between facts or propositions. In Hempel and Carnap, for example, what lies at the bottom of non-deductive inference is the confirmation of generalizations by positive instances. Each observation of an individual that is both *F* and *G* confirms the hypothesis that all *F*'s are *G*; this way in which a particular case can logically "point beyond itself" is the ultimate basis of our ability to rationally choose one body of scientific theory over another.

This seems extremely unlikely, at first and at second glance, as an account of what scientific inference is "really all about." All those claims about the structure of atoms, the deaths of the dinosaurs – can they really be handled this way? Despite its enormous strangeness when we step back from it as we can now, proponents of this approach had a lot to draw on to make it seem reasonable. One resource was empiricist anxiety about unobservables. Making sense of science, even before we worry about confirmation, seemed to require drawing the *content* of scientific theories closer to experience. This yields a deflation, either strident or low-key, of what the claims about dinosaurs and neutrons seem to say. Another resource is one central to the overall analytic philosophy project. This is the idea that philosophical investigation can reveal hidden logical and semantic properties in a sentence – the general notion of *logical form* is a key resource here. So via a combination of empiricist pressure and a willingness to believe in rich hidden structure, it becomes (almost) reasonable to think that theoretical science is essentially concerned with networks of generalizations that are each confirmed or disconfirmed by their instances.

The same meta-theoretic assumptions are seen in the more statistically oriented strand. Reichenbach said that we can treat all scientific inference as a combination of deductive argument plus one crucial non-deductive pattern. This non-deductive pattern involves the estimation of limiting frequencies of outcomes within sequences of events or trials. If, after seeing m events of type B within n trials of type A, we estimate the limiting frequency of B as m/n, and keep refining our estimate as n grows, then our procedure can be justified as follows. If there is a limiting frequency of events of type B to be found, then our method will eventually get us arbitrarily close to the truth. If there is no limiting frequency, then no method will work anyway.

The main internal objection to this argument is that lots of other estimates, beside m/n, have the same desirable long-run properties as the "natural" estimate, while giving strange-looking results in the short term. Reichenbach saw this, and hoped that something like risk-aversion (1938) or a simplicity preference (1949) might rule out the other options. Most have not been convinced. But for now, let us focus on the external problem: the majority of science does not look *anything* like this.

Reichenbach was uncompromising on this point. He treated cases that look different as "concatenations of inductions" of his kind, mixed in with deductive arguments.[2] For Reichenbach, the only non-tautological or "overreaching" form of argument that can be justified is induction in his specific sense, and this suffices for making sense of the entire non-deductive side of epistemic life. "The aim [of our inductive practices] is predicting the future – to formulate it as finding the limit of a frequency is but another version of the same aim" (1951 p. 246).[3] So again we see the idea that many disparate forms of inference in science are really another sort in disguise.

In my two other strands, the Popperian and Bayesian strands, the meta-theory is a bit different. It would be hard to say that all science is really conjecture and deductive refutation in disguise... so we claim that all *good* science is like this, and the rest should be reformed. Bayesianism has its own highly developed meta-theory, based on behaviorist or interpretivist ideas which I won't discuss here.[4]

The next thing I will do is offer a different meta-theory for this sort of philosophical work.

I start by taking some features of the work at face value, ignoring the accompanying commentary. What this work then tends to look like is the *modeling* of *key fragments* of a very complex structure. Guided by some combination of paradigm cases and the resources of available tools, the philosopher glimpses and tries to capture some particular significant relation between observation and theory. The method employed is the method of the model-builder. This method goes via massive idealization – we ignore a huge amount, and attempt the exact specification of a simpler analogue of the real target structure. We first seek an understanding of how things work in the idealized case, and then use that to cast light on the more complex target phenomenon, via a resemblance between the two.

Some, though not all, theoretical science works via this method. The origins of my characterization of this kind of philosophy lie in another part of philosophy of science itself, especially Giere's 1988 book *Explaining Science* and subsequent work by various people.[5] As I see Giere, despite his totalizing ambitions, what he did was offer a pretty good account of one style or mode of theoretical science, a mode that has grown in importance and self-consciousness over recent decades.

The reference to self-consciousness is important. Scientists engaged in model-building before they acknowledged that they did. A disconnect between actual and professed methods is common in science, but in the case of model-based science the disconnect can be especially striking. People do one thing, but describe what they are doing quite differently. This should lead us to consider a psychological hypothesis: there is a faculty of model-building imagination, whose products are sometimes oddly handled once created. The faculty works by imagining simpler, often schematic analogues of real systems, and maintaining an ability to assess resemblance relations between these imaginary creations and the real-world targets.[6]

This may give us a good account of much work in philosophy; philosophy is a domain where the model-building imagination operates in a particularly headlong and unconstrained way. And philosophy often shows an especially marked disconnect between imagination and ideology, between what our model-building faculty actually produces, and what we say about these products.

We see this especially clearly in metaphysics, and I would offer this as a diagnostic description of metaphysical systems from Plato through Leibniz to Lewis and Armstrong.[7] But it is seen in the system-building side of philosophy more generally, including a good deal of thinking about evidence in 20th century philosophy of science. Philosophers glimpsed and modeled some key fragments of a complex whole. These models can be genuinely illuminating, but what is usually presented as a treatment of uniform deep structure that must really be there, somewhere, is really a treatment of an idealized but illuminating relative of the real-world phenomenon.[8]

A better future view will embrace this fact rather than pretending it is not there. That last part seems to be the hard part, especially for the mind of the philosopher. Philosophy is full of modeling, in all its extravagance and elegance, and just as full of denials and forgettings of that fact. It is a domain where the possibilities and also the risks associated with model-building are especially prominent. But I think that, informed by attention to how things have gone in the past, to empirical psychology, and to the operation of model-building in the more constrained domain of empirical science, it should be possible to do better in the future. Not because we will necessarily come up with better fragments – probably we will, but we can't tell that in advance. The area where we can do better in a foreseeable way is in our treatment of the status and integration of the fragments, bringing them back into contact with the real subject-matter.[9] That we do by recognizing and working with our model-building tendencies, rather than half the time working within them and the rest of the time denying them – like a slightly scandalous sexual partner we are delighted to meet in private but will not acknowledge in the street.

2. Synchronic and Diachronic Perspectives

The next thing I will do is make a two-way distinction between orientations to the problems of evidence in science. The distinction is between *synchronic* and *diachronic* approaches, between views that treat the problem atemporally, or via the analysis of snapshots that might be taken at a time, and (on the other side) views that explicitly look at processes of change. Neither approach denies the reality of the factors studied by the other, of course, but each makes different bets about how some key problems are best addressed.

The logic-based strand of work discussed earlier is synchronic in approach; the aim is to characterize a support relation between theory and observation that is atemporal in character. On the diachronic side we have the procedure-based approaches of Peirce, Reichenbach, and also Popper. The subjectivist Bayesian strand has a special status here, as it carefully integrates synchronic questions (via the concept of coherence at a time) and diachronic ones (via the treatment of updating).

One bet I would make about the future is that the diachronic orientation will become more important. This is not the way things always have to go – 20th century deductive logic probably gained a lot from its ignoring of "movement" from premises to conclusion. That fact guided synchronic work on the non-deductive side. But I think the future will look different.

3. Small-Scale Change and the Paradoxes of Confirmation

In the rest of the paper I will offer some possible pieces of an overall story of the type sketched in the previous sections. These will presented as two case studies, organized around landmarks in the literature. They are distinguished by scale – small-scale versus large-scale change. Here I mean "scale" both temporally, and in a sense involving the size of the epistemic problem.

The first case features the same kind of relation between theory and evidence seen in the literature on inductive logic in the Hempel/Carnap sense, and in Reichenbach's model. We have some observations, and they are used to assess a hypothesis or answer a question. But the hypotheses under consideration do not involve the use of elaborate theoretical concepts that far outrun the vocabulary used to describe the observation. These are cases like testing a generalization about a link between observable properties, or fitting a curve to data. What we are not doing is introducing explanatory mechanisms far richer than the data, and choosing between rivals of that kind.

In the synchronic approach to such problems, we assume we have available some observations, or sentences describing them, where the observation is a local event. The aim is to show how these observations can somehow *point beyond themselve*. Our data concern specific instances, but our goal is forming generalizations of indefinite scope. That gives us the gap to be bridged – and shows the difficulty of the task.

A diachronic approach to these questions can naturally take the form of a *procedural* orientation. We see this part of epistemology as giving us *models* of how *procedures* can *reliably answer questions*. Roughly speaking, this is what we see in Reichenbach.[10] On a procedural view, specific observations get their evidential status from their *embedding in procedures*, not as a free-standing consequence of their content, of what they say about a localized case.

The two most famous problems for the logical approach here are Hempels' paradox of the ravens (1945), and Goodman's "new riddle of induction" (1955). Both can function as showcases for the superiority of a procedural approach. Here I will briefly discuss Goodman's problem.[11]

Goodman famously asked why, if seeing a large number of green emeralds (and none of any other color) confirms the hypothesis that all emeralds are green, the same collection of emeralds does not also confirm the hypothesis that all emeralds are *grue*. An object is grue if it has been first observed prior to (say) 2010 and is green, or has never been observed before 2010 and is blue. Our actual observed emeralds are positive instances of both generalizations, but the two hypotheses lead us to have different expectations about the first emerald observed after 2010. If all emeralds are grue, then the first new one we see after 2010 should be blue.

There are many proposed solutions to Goodman's problem. Most take a synchronic approach, in the following sense. What we need to find is some feature of the content of the two emerald hypotheses, or a difference in their logical relations to the evidence, that can be used to deny support in the "grue" case. Usually this has taken the form of a restriction on the predicates used in the hypotheses. "Green" is a *projectible* predicate while "grue" is not, so generalizations expressed in terms of grueness are not confirmed by their instances. The basis for this notion of projectibility can be anything from the metaphysical to conventional, but the form of the restriction is the same.

The procedural approach makes a different form of response possible (see also Godfrey-Smith 2004). Initially, let us forget emeralds, and look at the kind of question that would *actually* be answered using simple extrapolation from a sample. Suppose you want to know how many teenagers smoke. The obvious way to answer this question is to collect a random sample of teenagers, find the rate of smoking in the sample, and then extrapolate to the larger teenage population, in a way guided by statistical measures of likely error. The results are sanctioned not only by "intution," but by a model of why the procedure is in principle a reliable one.

There are various ways this method might fail. Maybe you cannot collect a random sample, as the smokers tend to avoid you. Perhaps teenagers will not tell you the truth. But let us consider a third, more unlikely one. Perhaps being asked the question tends to make some teenagers instantly take up smoking, so they truthfully answer yes, but only because they were asked. The process of data-gathering is interfering with the objects you are observing, in a way that makes them an unreliable guide to the unobserved cases. Some statisticians call this a "Hawthorne effect," after a famous case in the 1930s.[12] It also has a kinship with the notion of a confounding variable, although that term is usually applied in the context of causal inference, not estimation. And the problem in this case has nothing to do with good and bad predicates; it has to do with the process of collecting our sample, and some problematic causal relations holding between properties of those objects.

We can then see a relation between this phenomenon and the grue problem. The case of grue features a non-causal analogue of those same problematic dependence relations. When we observe a grue emerald before 2010, we must also note that had we not observed it, it would not have counted as grue (Jackson 1975). The process of observation is interfering with the properties we are interested in. This makes a sample of grue emeralds an unreliable basis for an inference to the larger emerald population; we have violated an assumption of the underlying model of sampling that was the basis for our inference. To adapt a piece of old metaphysical jargon, if the teenage-smoking problem was like a case of confounding, this is a case of *Cambridge-confounding*.[13] "Grue" as a predicate does not have some general overall badness, but it does have a meaning that leads to a malign non-causal interaction between the properties of objects we observe, in the specific context of making inferences from samples. That is why you cannot use random samples to answer grue-questions with grue-observations in the same way you can use such samples to answer green-questions with green-observations. The problem arises *as a feature of procedures*, not a feature of the contents of observation reports and theories considered alone. *If* you had a "non-interfering" way of sampling the emerald population, you *could* estimate the proportion of grue emeralds (how hard this is will depend on the exact "grue" predicate used). Philosophical intuition may balk, but the underlying statistical model tells you that the resulting estimate will be as reliable as it would be in the case of green.

How general is this as a solution to the grue problem? It is not fully general. It only applies to the extent that the investigations for which grue-like problems arise can be modeled on inferences to properties of populations from random samples. But that does not mean we should first accept the standard philosophical category of induction, and then note that some cases fall into a subclass that can be handled this way and some cannot. That standard philosophical concept of induction is badly misconfigured, and we can see that from consideration of just these sorts of cases.

What we see in inference from samples is a useable pattern of inference in which: sample size matters, randomness matters, and the philosopher's notion of "law-likeness" or "projectibility" does not matter. Any population that can be randomly sampled can be subjected to this sort of inference, and our model of sampling tells us the reliability properties will be the same.

We can then note a complement to this category. It is possible to develop generalizations, not from random samples, but from a knowledge of the mechanisms operating uniformly in some class of cases. If we have independent reason to believe that a set of objects are uniform in their structure, then we can take *one* case apart, see how it will behave, and make inferences about others. This is category of inference in which sample size *per se* does not matter, randomness does not matter, but the status of the kinds matters enormously. The two strategies of inference distinguished here each involve their own "bridges" between observed and unobserved cases: one goes via the power of random sampling, the other via reliable operation of mechanisms.

Then we see that the philosopher's concept of induction, especially since Goodman, has been a dubious *hybrid* of these; philosophers have supposed that the crucial category of inference is one in which (i) sample size matters, (ii) randomness is not an issue, and (iii) naturalness of kinds does matter. This is a construct that combines elements of two distinct inference strategies in science, in a way that corresponds to nothing real. It is common in philosophical discussion to think that sheer numbers somehow have an epistemic role, independently of randomness of sampling; they are taken to have an epistemic weight that is usually real but can be compromised somehow by failure of naturalness. There is no genuine category here.

So the treatment of grue above is a fragment. It shows in a simple case some phenomena that apply more generally, and also indicates the mis-organization of the landscape by some key 20th century discussions.

This discussion also sheds interesting light on an old encounter between two philosophers: the exchange between Reichenbach and John Dewey in Dewey's "Schilpp Volume" in 1939. Dewey had spurned traditional concepts of induction, especially with respect to the role of weight of numbers. Dewey thought that as generalization really works in science, *one* case is enough for extrapolation, if it is the right case. All the real work goes into showing that the case should be representative. Reichenbach argued that Dewey had not appreciated the crucial role of probability and the significance of convergence results; the real key to projection lies there. I say that both were right in seeing a real phenomenon, a real form of inference and one based on a real bridge between observed and unobserved. But both were too inclined (ironically) to project, treating one case as the key to all.

4. Large-Scale Change and the Underdetermination Problem

My second illustration concerns a larger scale in both temporal terms and in the epistemological problem being addressed. We are no longer concerned with cases where a similar vocabulary is seen in hypothesis and evidence ("these ravens are black; maybe they all are"). Instead the topic is the assessment of explanations for data in terms of hidden mechanisms and structures – the introduction of entire new inventories of causal players and explanatory relations. Here the bridges seen above are not applicable, despite the long history of attempts to show that cases of this kind *are* just cases of the more low-level kind in disguise.[14] And it is hard to see how a uniform story *could* be given.

One interesting fragment has recently been developed – this is how I see the literature on causal learning in "Bayes nets."[15] This literature describes a ground-floor case of this phenomenon: processes by which variants on one specific kind of explanatory structure (a causal network) specified in a vocabulary quite different from the data, can be assessed rigorously – in essence, via Reichenbach's own notion of "screening off."

This is a "ground-floor" case because the inference is constrained to be inference to some causal network or other (represented with a directed graph, nodes and arrows); there is no openness about the *kind* of explanatory structure to be introduced. Advocates of Bayes nets sometimes find themselves, like just about everyone else discussed in this paper, expressing totalizing commentaries about how this is a fully general account of how theory choice works: theories are *essentially* causal structures and the tools seen in Bayes-net learning are *the way* causal structures are inferred. I see the message of that literature in more restricted terms than that, but this is certainly a useful fragment.

Maybe a rather "particularist" story of large-scale inference will be what emerges in the end. But let us proceed looking for unity, with the mindset of the modeler, willing to idealize in order to capture informative fragments.

I will again organize the discussion via a standard problem, the "underdetermination of theory by evidence." The initial idea is that no matter how much data we might have, there will always be more than one theory that is compatible with the data. So if empirical data is all we have to go on, we can never have reason to accept some particular theory.[16]

There are many versions of the underdetermination thesis. Some are very strong; it is argued that for any theory T_1 we might come to hold, in any domain, there will be an incompatible theory T_2 that we cannot hope to distinguish from T_1 via *any conceivable* evidence. These ultra-strong formulations have various problems (Stanford 2006), so I will consider a formulation that is weaker but still general.[17] This formulation is modified from Psillos (1999, p. 164).

> U: For any particular body of evidence we might have, there will always be more than one scientific theory that can, in principle, accommodate it.

Underdetermination claims are often criticized for a simplistic treatment of the relation between theory and evidence; we only ask whether a theory can *accommodate* a set of data, presumably by implying the data when combined with reasonable auxiliary assumptions. There is no role for probability, explanatory power, and so on. The criticism is fair, but in the interest of simplicity and generality I will work within the framework that is common in these discussions.

So, someone asserts U. How worrying is it? The literature contains much discussion on this point – various brave faces, various gloomy faces... but I say we should not answer the question yet. U gives us only part of the picture. So far at least, U is compatible with another principle that might apply to the situation.

> D: For any particular comparison of two theories we might want to make, there is some possible body of data that will discriminate them.

That is, many of the usual underdetermination anxieties are compatible with a kind of *symmetry*: for any comparison of theories, we can hope to find discriminating data; for any data, there will be rival theories that are not discriminated.

Of course, D might be false. Once we bring in Cartesian skeptical possibilities, it seems that it may well be false. But discussion in the philosophy of science is not supposed to be concerned with those possibilities.[18] And the main problem, as I see it, is the fact that D is not even *raised* in underdetermination discussions; U alone is seen as the crucial point. But what we should be assessing is the consequences of something like a U+D *pair*. If D is false in its simple form, then we still need to be assessing some such pair: U plus whatever D-analogue is defensible.

U seems to acquire its special significance because of the assumption of a particular synchronic point of view. We assume we have some data and a theory T_1 on the table. Principle U then appears as a kind of barrier to theorizing. But this "barrier" is in large part the product of that particular point of view. When we think about a combination of U+D, the natural picture is a diachronic one, in which data are produced to discriminate rival theories; then new theories are introduced that exhaust the discriminative abilities of the data collected so far; and then new data is collected to discriminate among the range of theories now on the table.

And suppose U and D are both true, or have similar standing as approximations. Then we have a "glass half full" and "glass half empty" situation. When we look at U, the glass looks half empty. When we look at D, it seems half full. What must be done more cautiously is the drawing of conclusions solely from the "glass half empty" side.

I imagine that some may say that the "half empty" side still contains the main message. This is because although we might hope to make successful discriminations between theories indefinitely over time, we can never believe, *at* any specific time, that we have found a theory that is *true*.

So if we are trying to develop an optimistic or progressivist picture on the basis of a combination of U and D, there must be some rethinking of the *goal* of inquiry, or the virtue that we hope for in a theory that we endorse at any given time. Let us see how that might go. Assume a simple U+D model; we have a sequence of discriminations, each followed by a new discrimination problem. So far we have motion, but not yet motion that is in any sense progressive. Progress requires that there is some epstemically relevant quantity *accumulated* as the process continues, or perhaps an increasingly close approach to a goal. Alternatively, as in Thomas Kuhn's *Structure* (1970), there might be increasing distance *from* "primitive beginnings," but without approach to a goal. We need to be going somewhere worth going or collecting something worth collecting. Formally, the stages we successively reach must be linked by some relation that is both transitive and epistemically or evaluatively relevant.

If the data itself accumulates in a simple way – none is lost as more is added – that will suffice to yield progress on one empiricist measure. Each theory accommodates a larger data set than its predecessor. But my aim is to capture something with a more realist flavor. So here we might make a connection to another literature, long troubled but making recent progress. This is the literature on *approximate truth* (see Oddie 2001 for a good review). In the newer work on this topic, approximation to truth, or truthlikeness, is *strict* truth about a situation *close* to actuality. Jumping over many problems and details, one possibility raised in this discussion is that the best-behaved concept of approximate truth is a *comparative* one. The idea of absolute distance from the truth might be hard to make sense of, but it might make sense to say that *X is closer to the truth than Y*.

If such a notion does make sense, it could give us materials with which to tell a progressivist and realist story within a U+D framework. Closeness to truth would be a transitive relation that may characterize the stages reached along a U+D process.

I emphasize the term "materials" just above. The aim is to describe a way in which a certain kind of progress can in principle, non-accidentally, be achieved. But the process is not in any sense guaranteed to actually get us closer to the truth. This is not an attempt to buttress facile feel-good histories of science, or engage in the misguided task of "rational reconstruction." The actual enterprise of science is affected by much more than this – it has whatever mix order and chaos that it has, chases whatever array of red herrings down blind alleys it chases.

But the model gives us a start. The intended outcome of theory choice is selection of a theory that is closer to the truth than all the others that are, or have been, on the table as rivals. It would hence be closer to the truth than many relevant alternatives, where relevance is shown by the fact that these alternatives were once considered live scientific possibilities. The theory may not be closer to truth than a future theory not yet devised, and indefinite motion towards truth is also compatible with permanently being indefinitely far away from it.

I close with a last set of comments on meta-theory. A U+D model is probably not applicable everywhere. It looks especially applicable in areas like physics, where the level of deductive organization is high and the entities treated are problematic; it looks less applicable in areas like modern cell biology, where the entities are unproblematic and mechanistic knowledge can simply accumulate. And the story, again, used a deductivist treatment of evidence that has many deficiencies.

So I am being pluralistic about evidential relations in science, at least as they appear through the lens of philosophical models. Even if evidence is in some sense one thing, it is a thing so complex that for the foreseeable future we will be understanding it by modeling fragments.

Offering pluralisms can be unsatisfying in philosophy, even when they are true. This is partly for good Occamist reasons; simple theories are not more likely to be true than complex ones, but starting simple is a good "rule of motion." (Simplicity is another area where the diachronic perspective helps.[19]) It is good Occamism to start simple, but bad Occamism to insist on staying simple when a genuine push towards complexity comes.

Further, a structured plurality is more informative than an unstructured one. Here I press once again the underlying psychological picture that puts my treatment of evidence into a larger context. We are philosophers, legitimately asking the most giant and general questions, but much (though not all) of what our minds offer up in response is the product of an idealizing and model-building imagination. We see this in scientific philosophy as much as in *a priori* metaphysics, but it is the vantage point of scientific philosophy that will enable us to recognize how this works, and how our mix of faculties can be combined most effectively.

*   *   *

Notes

[1] See especially Platt's "Strong Inference" (1964).

[2] "The connecting link, within all chains of inferences leading to predictions, is always the inductive inference. This is because of all scientific inferences there is only one of overreaching type: that is the inductive inference. All other inferences are empty, tautological: they do not add anything new to the experiences from which they start." (*Experience and Prediction*, p. 365).

[3] The aim of inductive inference is "to find series of events whose frequency of occurrence converges towards a limit" (p. 350, italics removed). He insists this is broader, not narrower, than Hume's sense of induction; "it conceives the aim of induction in a generalized form" (*Experience and Prediction*, 1938, p. 350). The 1951 quote is from *The Rise of Scientific Philosophy*.

[4] Though perhaps this official meta-theory is becoming less popular now.

[5] See also my "The Strategy of Model-Based Science," Weisberg's "Who is a Modeler," and some strands of Cartwright's work.

[6] See Nancy Nersessian's work. Relevant work in cognitive psychology is also cited in my "The Strategy of Model-Based Science."

[7] This is discussed in more detail in my "Theories and Models in Metaphysics" (2006).

[8] Reichenbach's choice of fragment is interesting here. He did not discuss inference from samples which are surveyed as wholes, but inferences about frequencies where the observations come in one at a time. In a sense, he was drawing simultaneously on both the technical material

he found illuminating, and on a very general picture of our epistemic situation. In a general sense, we are located in time, seeing events one by one, trying to predict the next.

Some writers in the movements I have been discussing do talk of a "model" when they describe their work – the Bayesians, especially. But not in the same sense perhaps, or if so, without realizing the extent of the meta-theoretic shift that this idea makes possible.

[9] For this theme, see Dewey's *Experience and Nature* (1929), Chapter 1.

[10] For more detail, see my *Theory and Reality*, Chapter 14, and Kelly and Glymour "Why Probability Does Not Capture the Logic of Scientific Justification".

[11] For an analogous treatment of the ravens, see *Theory and Reality*, Chapter 14.

[12] The case is sketched in my 2004 paper.

[13] Here I draw on Peter Geach's notion of a "Cambridge change" (1972).

[14] Here again, Reichebach discusses the challenge from large-scale change to his theory of induction, and refuses to make concessions. He imagines this objection: "Your theory of induction as an interpolation, as a method for continual approximation by means of anticipations, may be good enough for the subordinate problems of scientific inquiry, for the completion and consolidation of scientific theories.... [T]he genius follows other ways, unknown to us.... Is not Einstein's discovery of new laws of the motion of planets, of the bending of light by gravitation, of the identity of mass and energy, etc., a construction of ideas which has no relation to diagrams of curves of interpolation, to statistics of relative frequencies, to the slow driving of approximations, step by step?" (p. 381)

He replies that once we distinguish the contexts of discovery and justification, we will see that these phenomena do not "constitute any objection against my theory of induction as the only possible means for the expansion of knowledge." (p. 381) "Einstein saw – as his precursors had not seen –... that an inductive expansion of the known facts leads to the new theory." (p. 382)

[15] See Pearl *Causality*, Spirtes, Glymour, and Scheines, *Causation, Prediction....* Gopnik and Schulz, *Causal Learning.*

[16] This is usually expressed as a problem for "scientific realism." See my *Theory and Reality* for why this is not a good way of setting up the scientific realism issue.

[17] As Stanford (*Exceeding our Grasp*) argues, these versions of the argument tend to rely on extreme skeptical hypotheses (often of a Cartesian kind), or on small manipulations of T_1 that produce a variant that is not scientifically interesting. There are worked-out illustrations of underdetermination for some particular physical theories, usually involving space, time, and motion, that are neither ultra-skeptical nor trivial, but certainly not for all theories.

[18] And perhaps in the case of some specific scientific domains, D is again a vain hope. But that, again, is not the usual focus of discussion.

[19] See my "Popper's Philosophy of Science: Looking Ahead" and various papers by Kevin Kelly.