Miracles, Pessimism and Scientific Realism*

John Worrall

Abstract: Worrall ([1989]) argued that structural realism provides a 'synthesis' of the main pro-realist argument – the 'No Miracles Argument', and the main anti-realist argument – the 'Pessimistic Induction'. More recently, however, it has been claimed (Howson [2000] and Lewis [2001], respectively) that each of these arguments is an instance of the same probabilistic fallacy – sometimes called the 'base-rate fallacy'. If correct, this clearly seems to undermine structural realism and Magnus and Callender have indeed claimed that both arguments *are* fallacious and 'without [them] we lose the rationale for … structural realism' ([2004], p. 333). I here argue that what have been shown to be fallacious are simply misguided formalisations of 'the' arguments and that when they are properly (and modestly) construed they continue to provide powerful motivation for favouring structural realism.

---

* This is a heavily revised version of a paper I first gave at a Lunchtime Colloquium at the Center for Philosophy of Science in Pittsburgh in October 2005. I thank those who commented – especially John Earman, John Norton and Sandra Mitchell – for helpful and constructive remarks. I am grateful to Colin Howson, Paul Teller and especially to Craig Callender and Peter Lewis for comments that led to a number of changes; and to Lefteris Farmakis for research assistance and some criticisms of a previous draft. As usual, I am indebted to Elie Zahar for numerous helpful conversations on Structural Realism.

1 Introduction

The - often breathtaking - predictive success of *some* theories in contemporary science inclines most of us towards scientific realism: surely those theories must at least approximately relate to the 'unseen world' lying 'behind the phenomena' if they can score such dramatic, empirically checkable, successes? The facts about theory-change in science, on the other hand, seem to speak in favour of an anti-realist view: scientists have in the past held theories that were also dramatically predictively successful and yet which are now 'known to be false' (because they are inconsistent with our latest theories). Given this, what guarantee can there possibly be that our latest theories will not themselves be rejected and replaced by quite different ones at some time in the future? And if so, how can we reasonably hold that our current theories are true? Moreover the view is widely held that those theory-changes have been 'radical' or 'revolutionary', how in that case can we hold that our current theories are likely even to be *approximately* true? My [1989] argued that, although these two much-heralded considerations thus seem to pull sharply in opposite directions, they can in fact be reconciled within a version of realism – namely, *structural realism*.

The first, apparently pro-realist, consideration has often been developed as 'the No Miracles Argument' (hereafter the NMA). Roughly: it would be a miracle if current scientific theories enjoyed the success (especially *predictive* success) that they do if what they claim is going on 'behind' the phenomena is not at least approximately correct; but we clearly should not accept that miracles have occurred if there is some non-miraculous alternative; and here the (approximate) truth of what the theories say about the 'noumenal' world is exactly such a non-miraculous alternative explanation of their empirical success. The second, apparently anti-realist, consideration has often been developed as 'the pessimistic (meta-) induction' (hereafter, the PI). Roughly: theories that were accepted in the past (exactly on the basis of the success called upon by the NMA) have subsequently turned out to be (perhaps radically) false; so, we should infer (inductively/probabilistically) that our current theories are (perhaps radically) false too.

However, Colin Howson has recently argued (see his [2000], chapter 3) that the NMA in fact embodies an elementary probabilistic fallacy often called ‘the base-rate fallacy’. While, ironically enough, Peter Lewis ([2001]) has (quite independently) argued that essentially the same fallacy underlies, and therefore vitiates, its seeming-competitor argument – the PI.

If the NMA and the PI are fallacious, then this would seem to destroy the basic problematic at which structural realism is addressed. And Magnus and Callender ([2004], 321-322) have argued that since they are indeed fallacious, ‘the major considerations for and against realism come to naught’, and hence there is a further reason - additional to those already supplied by Arthur Fine (e.g., [1984] and [1986]), Simon Blackburn ([2002]) and others - for ‘dissolving’ the whole scientific realism debate. Magnus and Callender argue, specifically, that, in view of their fallaciousness, the realism debate must do without both the NMA and the PI and ‘[w]ithout these [arguments] we lose the rationale for ... *structural realism*’ (*op.cit*., p. 333)[1]

The first main section of this paper analyses the NMA and the claim that it rests on an elementary fallacy; while section **2** does the same for the PI – in both cases I argue that what have been shown to be fallacious are simply forms of ‘the’ argument that should never have been taken seriously in the first place. I suggest that, when the two arguments are construed properly (and modestly), they are immune to the criticisms raised in the recent literature and, as I argue in section **3**, remain powerful motivations (though, of course, far from conclusive reasons) for adopting structural realism.[2]

1 How, and how not, to work no miracles

(a) The intuitions

[1] Although I concentrate on structural realism here, Magnus and Callender target *any* view that, by taking both the NMA and the PI on board, seeks to position itself somewhere between realism and instrumentalist-empiricism (including, for example, ‘entity realism’).

[2] There have also been a number of direct criticisms of structural realism in the recent literature, many of them based on the ‘Newman objection’ (see Newman [1928], the revival of that argument in Demopoulos and Friedman [1985], in Psillos [1999] and most recently Ketland [2004]). These criticisms are not dealt with in the present paper but are addressed and rebutted in Worrall and Zahar [forthcoming]. I also reserve for a separate forthcoming paper my response to the general ‘dissolution’ thesis as urged by Fine and Blackburn.

Consider a classic, and by now well-worn, example that elicits a strong intuitive 'no miracles response' (at any rate in yours truly). Fresnel's theory of light states that light consists of waves transmitted through an all-pervading elastic medium. We cannot of course observe that medium, but we can observe what Fresnel's theory tells us are the effects of various unobservable motions through it. One such alleged effect, as Poisson demonstrated but as Fresnel himself had not suspected when developing his theory, is that if a small opaque disc is held in light diverging from a point source and if the 'geometric shadow' of the disc (that is, the area of complete darkness that would exist if the laws of geometric optics were strictly correct) is carefully examined, then the centre of that 'shadow' will in fact be seen to be illuminated, and indeed just as strongly illuminated as if no opaque disc were present. Most of Fresnel's peers thought that this entailment represented a clear-cut *reductio* of the theory, yet when Arago performed the experiment it turned out that the 'white spot' does indeed, and contrary to all prior expectations, exist.[3]

Whatever esoteric philosophical considerations may be raised, it is difficult to resist the feeling that if a theory can make such a striking, seemingly improbable prediction that nonetheless turns out to be empirically correct, then the theory must somehow be 'approximately true' – it must have somehow latched on, no doubt in an approximate (but nonetheless substantial) way, to the 'deep structure' of the universe: to how things really are in the 'noumenal world' behind or beyond the phenomena. Duhem, who was (usually!) a structural realist rather than the instrumentalist he is often considered to be, put it eloquently ([1906], 28):

> The highest test, therefore, of our holding a classification as a natural one is to ask it to indicate in advance things which the future alone will reveal. And when the experiment is made and confirms the predictions obtained from our theory, we feel strengthened in our conviction that the relations established by our reason among abstract notions truly correspond to relations among things.

[3] The real history, as I show in my [1989a], was a good deal more interesting and a great deal less clear-cut. However the real historical details, although they do centrally affect the issue of what counts as a successful prediction (and why predictions carry more confirmatory weight), do *not* affect the philosophical issue about successful prediction and realism.

A theory gives us a 'natural classification', according to Duhem, just in case 'the relations' it posits 'truly correspond to relations among things'. Our 'conviction' that Fresnel's theory represents such a natural classification is 'strengthened' because it would, it seems, be inexplicable if that theory could turn out to have such a striking, and empirically correct consequence (that is, it 'indicated in advance things which the future alone [revealed]'), if it did not represent a 'natural classification'.

Some of those developing the NMA have used the idea of science's 'success' in a broader – sometimes vague - sense, rather than in the precise sense called upon by Duhem. Of course science has (sometimes) been 'successful' in a number of ways: one often alluded to its ability to 'unify', to bring together initially apparently distinct areas, such as optics and electromagnetism, under one single theory. Unification is however only really successfully achieved if associated with independent predictive success. Another, and certainly important, sense of success is the way that science 'builds upon itself' – essentially by requiring, when a hitherto accepted theory T fails, that its replacement theory yield T as a limiting case (where the conditions that characterise the 'limiting case' specify the area where T has been unambiguously empirically successful). But again the important fact is that this 'conservatism' has paid off by producing new replacement theories that themselves score independent empirical successes (that is, empirical successes not shared by the theories they replaced). It does seem, then, that it is predictive success that is always the most significant factor in inducing the 'no miracles intuition'.

Notice that the conclusion of any sensible version of the NMA needs to be that the theory concerned is '*approximately* true' or that it 'latches on - *in some approximate way* - to how things are "beneath the phenomena"'. No one should be a gung-ho scientific realist and hold that it is reasonable to believe that currently accepted fundamental theories, even in 'mature' science, are outright true. This is a message that is underwritten by the PI - later (better) theories tell us that Fresnel's theory, for all its predictive success, is strictly false; but the point is independent of that argument. No one believes (or should believe), even ahead of any further 'scientific revolution', that Quantum Electrodynamics, for example, is true – indeed there are questions about whether a fully coherent version of the theory can currently be articulated. No one even believes (or ought to believe) that Quantum Mechanics itself, for all its stunning success, will survive entirely intact. (Its two basic postulates are clearly mutually incoherent; and it also fails to cohere with the General Theory of Relativity.) So the sensible realist claim must in general be that it would be a miracle, not if the fundamental theory at issue failed to be 'outright' true, but rather if it failed to be somehow *approximately* true.[4]

Whether the NMA can be given some more exact articulation than that sketched above will be a central topic in this paper. But it does seem clear that science centrally embodies the underlying *intuition* and would not be possible if it did not. This is reflected in the fact that appeal to what is at least a very similar intuition is implicit in the justification of the standard empirical generalisations that everyone – including those who are anti-realist about 'observation-transcendent' scientific theories - accepts. This has gone largely unacknowledged in the recent literature, but it *was* recognised, and emphasised, by Poincaré, who wrote ([1905], pp 149-150):

> We have verified a simple law in a considerable number of particular cases. We refuse to admit that this coincidence, so often repeated, is a result of mere chance and we conclude that the law must be true in the general case.
>
> Kepler remarks that the positions of the planets observed by Tycho are all on the same ellipse. Not for one moment does he think that, by a singular freak of chance, Tycho had never looked at the heavens except at the very moment when the path of the planet happened to cut that ellipse … [I]f a simple law has been observed in several particular cases, we may legitimately suppose that it will be true in analogous cases. To refuse to admit this would be to attribute an inadmissible role to chance.

It would, Poincaré is claiming, constitute a miracle – an incredible coincidence - if Kepler's simple (first) law were instantiated by all the planetary positions that had so far been checked but was not true in general (that is, was not also instantiated by all the - past and future - unobserved planetary positions). No instrumentalist or

[4] The reference to 'fundamental' theories is important, since, as I reflect later, it is arguable that it *is* reasonable to think of 'middle-' ' or 'lower-range' theories (such as that pure water consists of molecules each of which contains two atoms of hydrogen and one of oxygen) as true.

constructive empiricist known to me fails to endorse the acceptance as rational of standard empirical generalisations (again generally, as the Kepler case illustrates, this 'acceptance' ought to mean acceptance as *approximately*, rather than outright, true) – on the basis of what is of course bound to be a finite set of actual observations. They are therefore all relying – as Poincaré points out - on something that seems practically indistinguishable from the no miracles consideration that they vigorously deny should be thought of as persuasive when it comes to observation -'transcendent', theoretical claims.

The intuition underlying the no miracles argument also underwrites persuasive arguments in a variety *both* of scientific *and* of more commonplace circumstances. Maxwell's work initially left open the possibility that there might be two different 'media' filling the whole of space: the optical ether and the electromagnetic field. But once he had discovered that waves were transmitted through the field at the velocity of light, he immediately inferred that it would be miraculous if there were two separate media each of which just happened to transmit disturbances at exactly the same rate; and hence he inferred that there is only one medium – the field – and that light is an electromagnetic wave. Einstein refused to admit that the parameter measuring a body's responsiveness to an applied force (its inertial mass) and its gravitational action (its gravitational mass) could be identical by accident. It would be a miracle if these two conceptually distinct quantities just happened invariably to have the same value for a given body – so some non-miraculous account must be sought and was of course found in the form of the General Theory of Relativity. We also often argue in the same way in more commonplace settings. Suppose someone claimed that, whatever the appearances, George W. Bush is a man of astute, independent and reliable judgment with a warm, selfless, humanitarian nature. Each of his individual actions, of course, appears to clash horribly with this account – but in each individual case, there happened to be particular circumstances, different in each case, that led to the outcome at issue despite Bush's underlying true character. We all surely reason that it would be a miracle if circumstances had continued to conspire in this way and hence we all (I speak for the serious and sensible amongst us) reject this account in favour of a less rosy view of Bush's underlying character.

Admittedly it is easy to produce alleged 'miraculous coincidences' pretty well at will and many people have been seduced by cooked-up 'coincidences' into accepting conclusions that are themselves quite staggeringly improbable – I think in particular of arguments concerning the so-called anthropic principle and some arguments for the existence of god. So we certainly need to take care in talking about 'miraculous coincidences'. Nonetheless it seems difficult to resist the idea that there is *something* important in the intuitions underlying the truly persuasive instances of the NMA.

Philosophers – on the whole quite rightly – are, however, suspicious of intuitions and try to capture what, if anything, is valuable in them in more rigorous arguments, whose credentials can be examined more sharply. So how, if at all, can the intuitions elicited by the predictive success of at least some theories be captured in some more precisely articulated argument?

(b) How (not) to formalise 'the' 'No Miracles Argument'

(i) The scope of 'the' argument

What exact *scope* should we expect such a more precisely articulated argument to have? The idea – suggested by Hilary Putnam ([1978], p. 19), amongst others – that we should think of scientific realism as itself a sort of 'overarching scientific hypothesis' that (allegedly) provides the 'best explanation' (best *scientific* explanation) of the success of 'science' has always seemed to me entirely without attraction. For a start, is there such a thing as science (in general)? It seems that there is instead a wide variety of science*s*, not all of them 'surprisingly successful', certainly not in any sense that elicits the 'no miracles intuition'. Nothing known to me in the sciences of sociology, parts of psychology, dietetics *etc* provides any reason to make one think that their accepted theories have successfully penetrated to the noumenal world 'beyond' the phenomena.

Putnam would (or would when he seemed to be defending this thesis) no doubt refer to the 'let-out' *maturity* clause here: a realist is realist only about 'mature' science – presumably sociology, parts of psychology, etc are not mature. And Magnus and Callender indeed take the 'wholesale realist view' to state that the success of *mature* science is explained by the fact that all (or perhaps, with possible counterexamples in mind, only *most*) mature scientific theories are true.

An immediate problem with this view – pointed out early by Laudan ([1981] and then [1984]) and others, though not commented on by Magnus and Callender – is the vagueness of the term 'maturity'. It seems to me, as I argued in my [1989], sensible, indeed uniquely sensible, for the realist to characterise this notion on the basis of her main sustaining argument – 'the' NMA. As I have explained, I take this to identify the 'success' of theories principally with their empirical *predictive* success: their success in predicting 'new' types of phenomena. Following my suggestion, then, a scientific field would be regarded as having achieved maturity once its fundamental theory had exhibited genuine predictive success. But this indicates that any 'wholesale' argument will simply amount to the union of a number of individual 'retail' arguments for realism about particular theories that have established the 'maturity' of their field by proving predictively successful.

It was always a mistake, then, so I suggest, to think of any overall version of scientific realism as a sort of general inference to the best explanation of some rather loosely characterised phenomenon known as the 'success of science'. What are successful or not, what elicit the 'no miracles intuition' or not, are *individual* theories – such as Fresnel's wave theory of light or Quantum electrodynamics. In so far as there is any sort of general, or 'wholesale', case to be made for scientific realism it is simply as the union of a whole set of specific cases for individual theories (and hence any sensible 'wholesale' view will not extend to the 'whole of science' – whatever that is). Also it is a mistake to defend scientific realism as a thesis about only 'most' successful scientific theories. If there were even a minority of theories that achieved striking predictive success but which can only now be regarded as radically false then all force surely goes out of the realist's contention that it would be a 'miracle' if this were to occur. Finally, there obviously are lots of theories in lots of sciences that could be called 'current theories' at some stage – many characterisable as 'working hypotheses', 'the best guess that we have so far' and so on. But no case for realism should, or can, be made for these and their longer term fates are therefore irrelevant to the issue. The sensible realist is – of course! – not realist about everything in science, but only about its ‘accepted’ theories, where these have proved predictively successful.

It is not, then, that I disagree with Magnus and Callender’s rejection of the ‘wholesale’ realist view - far from it; but rather that I feel that this view was never a target worthy of attack (which is not to deny that very worthy people including Smart, Putnam and Boyd have held the view). But giving up (or, in my case, never having considered holding) the wholesale view and concentrating on (a collection of) retail arguments does not absolve the realist from defending the intuitions underlying ‘the’ NMA – as we have already seen. From this point of view, then, Magnus and Callender’s paper is incoherent: having dismissed ‘wholesale’ realism, largely on the grounds that the NMA is fallacious, they applaud investigation of ‘retail’ realist arguments – for particular theories or particular (alleged) entities (such as ‘the’ atom). But the problem is, of course, that the ‘retail’ arguments can all be construed (and ultimately can only be construed) as instances of some form or other of the NMA. Belief in ‘atoms’ really translates into beliefs that various theories that use the term are at least ‘on the right lines’ because those theories have had striking empirical successes (e.g. in Perrin’s work on Brownian motion), to an extent that seems entirely implausible if they are not on the right lines. Maybe fancy ways of dressing up ‘retail’ realist arguments sometimes disguise their reliance on the intuitions behind the NMA, but reliant on them they certainly are.[5]

Equally clearly in my view, no version of scientific realism – whether wholesale or retail – can be thought of as itself a scientific view. The chief characteristic of acceptable scientific explanations is independent testability. Scientific realism, let’s say, wholesale scientific realism, is the claim that the success of scientific theories is explained by their approximate truth. Suppose we think of this as having been generated to explain the success of ‘mature’ theories *so far*; we can then think of it as ‘predicting’ that the next scientific theory to be accepted in mature science on account

[5] As Craig Callender pointed out to me, this is only an incoherence from the point of view I recommend. If ‘the’ NMA is identified exclusively as an argument for wholesale realism and moreover as essentially one involving ‘proper’ probabilistic reasoning (see next section) then the intuitive ‘NMA-style’ arguments involved in retail arguments are not ‘the’ NMA. But if, as I do, you see only intuitive considerations underlying either form, then since the same intuitions underlie both (attempted) wholesale arguments, which they attack, and the particular retail ones, which they applaud, the position does appear incoherent.

of its predictive success will be true. But this is patently not a testable prediction. For sure, the next theory accepted in mature science will be successful – predictively successful - it would not be accepted in preference to current theories unless it not only replicated those current theories' successes but added predictive successes of its own. But how can we check whether or not it is (approximately) *true*? The whole point about scientific realism, as Magnus and Callender (along, as we shall see, with Peter Lewis) fail firmly to grasp, is that it attempts to defend a link between the effective, decidable notion of success and the essentially undecidable, if you like transcendental, notion of truth (or approximate truth). Of course, we do make *judgments* about a theory's truth (in particular the judgment that it is false in the light of later, better, theories) but these are inevitably conjectural – based in the case of the judgments of falsity on the fact that the theories concerned conflict with currently accepted ones, which are taken (temporarily and for the sake of argument) as true. But truth itself is of course undecidable and ineffective. Scientific realism thus makes no testable predictions and hence cannot be considered to be itself scientific: scientific realism in any form is a *philosophical*, not a scientific, thesis.

(ii) The Form of 'the' argument (or why one shouldn't expect too much)

Any sensible version of the NMA will therefore be of the 'retail' variety: its conclusion will be that it is reasonable to hold that some *particular* theory - the wave theory of light, the general theory of relativity, quantum field theory … - is approximately true. Moreover the success involved in the premise of any sensible version will not be any vague, generic, 'wholesale' notion of success but the genuine *predictive* success of the particular theory at issue: the theory must make a prediction of a general kind of empirical result, one that pans out when tested against actual experiment.

'Prediction' here, as I have explained elsewhere (see in particular my [2002]), need not involve novel, that is, hitherto unsuspected phenomena. The operative condition is that, in order to count as having been predicted, the general phenomenon must not have been 'used in the construction' of the theory at issue. (Obviously this condition will automatically be satisfied by any piece of 'new' evidence that was unsuspected at the time when that theory was first formulated.)

No one is going to exclaim when confronted, say, with some version of Ptolemaic geocentric theory that correctly entails that the planets exhibit stations and retrogressions ‘Wow! That must mean that there is something about the theory’s fundamental claims that must be at least approximately correct, otherwise it would be a miracle if it succeeded with such a striking prediction’. This is because there is a much more homely explanation of its ‘success’: parameters in the general Ptolemaic theory (relating sizes of epicycles and deferents, and the relative epicyclic and deferential velocities) had been fixed precisely on the basis of the previous observation of planetary stations and retrogressions, so that the particular version of Ptolemaic theory with parameters fixed in this way was *bound* to yield the phenomena at issue, irrespective of whether or not the overall theory of which it forms a part has ‘latched on to reality’.

(This demanding predictivist criterion of success already rules out pretty well every theory in Larry Laudan’s famous ‘plethora’ of ‘successful’ theories that we now (allegedly) take to be radically false – with one exception: the ‘classical’ wave theory of light as a periodic disturbance in an elastic medium.[6] Other theories on the list - such as the gravitational and physiological ethers of Hartley and Lesage or the astronomical theory of the crystalline spheres – are surely classic instances of *ad hoc* theories. They respond to a definite explanatory problem – how, to take the last case, do the sun, planets and stars all move around the earth and why do they all orbit in the same direction ? But they ‘solve’ it (in the geocentric version of crystalline sphere theory by assuming that those astronomical objects are all embedded in concentric spheres that are themselves revolving in the same direction but at different rates about an axis passing through the Earth) without the slightest hint of any independent testability. The fact that a theory was taken seriously even by serious scientists is not something on which any sensible realist would rest any part of her case. Only predictive success really counts.)

[6] Stathis Psillos ([1999], chapter 6), while largely agreeing with my criticism of Laudan, but operating with a looser, more ‘intuitive’ notion of ‘maturity’, is inclined to add the caloric theory of heat as another exception. It is certainly true that in some senses the caloric theory of heat is a much better scientific theory than either, for example, the crystalline sphere theory or (still more clearly)Lesage’s theory; but it is not clear to me that the caloric theory ever made a genuine prediction and therefore should count as mature on my characterisation.

At first blush, the impact of a successful striking prediction for a specific individual theory T can be captured by the following informal argument. T has scored some spectacular predictive success; it would be a miracle if T could get such a phenomenon so exactly right if it were not itself at least approximately correct; but we should not accept that miracles have occurred if there is an alternative explanation; and there *is* exactly such a non-miraculous alternative in such cases – namely that T got this prediction correct because it is at least approximately correct; therefore we have reason to infer that T is indeed approximately correct.[7]

Clearly, however, there can be no question of this being, as it stands, a compelling deductively valid argument. Suppose, for example, that T is a mathematical theory relating two variables X and Y through the equation $y = f(x)$. T therefore entails that when X takes the value x_0, Y takes the value $f(x_0) = y_0$. Suppose moreover that (x_0,y_0) is a real datum and that it is somehow 'surprising' that it *is* a real datum. Can we infer that T must be at least approximately correct? As Howson reminds us, Jeffreys showed that it is easy to construct infinitely many (in fact non-denumerably many) rivals to T – rivals in the strict sense that they entail that T is false – and yet which all equally well entail the 'surprising' datum e. Simply take T′ as the theory $y = f(x) + (x - x_0)g(x)$ for any non-uniformly zero function g. Why then could we not equally infer that T′ is true (or approximately true) because it gets the surprising datum e correct? And clearly since the function g(x) is arbitrary, if some such T′ *were* correct then our original theory T, far from being even approximately true, could be intuitively as 'far away from the truth' as you like for all values of X except for x_0. This problem is not restricted to the special case of mathematical theories of the Jeffreys kind (nor to the gruesome case that is resembles). If you consider the

[7] Although the claim that the approximate truth of T would explain its 'otherwise miraculous' success with some surprising prediction *e* might sound plausible, it is by no means obviously true. Clearly if a theory is *true* then so are all its consequences – so if it entails some unlikely prediction that turns out to be correct, it seems reasonable to regard the theory's truth as an 'explanation' of its success. But it has of course never been shown that all consequences of an 'approximately true' theory must themselves be approximately true. Indeed I do not believe that any such demonstration could be produced for any of the accounts of approximate truth that are currently under scrutiny (see Niiniluoto [1987]). These all specify some way in which a theory must correspond to a reality, considered as independently given, in order for it to count as approximately true to some degree and they do so in terms of some measure of the 'relative sizes' of the sets of true and false consequences of such a theory. We shall see however below that for the very special, and minimalist, account of approximate truth involved in Structural Realism, it is indeed automatically the case that the 'approximate truth' of a theory 'explains' its predictive success.

deductive closure of any theory T, restrict yourself to the set of its empirical consequences (however characterised), and consider any conservative extension of that set back into the theoretical language, you will create indefinitely many 'alternatives' to T that equally well entail not just the surprising datum e but *all* the empirical data that supports T. There is more to be said about these Jeffreys-constructions and about underdetermination issues more generally, but certainly they establish (what surely ought always to have been clear) that it is *logically possible* that a theory may entail some 'surprising' and 'startlingly correct' empirical result and yet not be true (or even approximately true).

Logically possible but nonetheless extremely unlikely? Howson, and following him Magnus and Callender, take it that the natural avenue down which to seek a formal account of the NMA intuitions is the (multi-lane) probabilistic avenue. In investigating the prospects for such a probabilistic reconstruction, let's first temporarily lay to one side the issues about approximation. (And also, in line with what was said above, let's consistently follow Howson in making the argument one about a particular theory T that has been predictively successful with phenomenon e.[8]) Talk about it being a 'miracle' if T had got such a phenomenon as e right if it were not true would seem to translate crisply into the assertion that the probability that e would happen were T false is extremely small: $p(e/\neg T) \approx 0$. And the fact that T (together with accepted auxiliaries) entails e 'translates' of course into the claim that (again assuming the necessary auxiliaries) $p(e/T)=1$. Hence we have:

Premise 1: $p(e/T) =1$ (e is entailed by T, *modulo* accepted auxiliaries)
Premise 2: $p(e/\neg T) \approx 0$ (it would be miracle if e had been the case were T not true)
Conclusion: $p(T/e) \approx 1$ and hence, given that e has occurred, $p(T) \approx 1$.

There are, of course, entirely legitimate worries about what exactly the probabilities in these formulas mean, but it is not difficult to show that, so long as they are probabilities at all, then this reasoning is straightforwardly fallacious.

[8] Although Magnus and Callender give the *impression* that they are following Howson's criticism of the NMA, Howson deals entirely with the argument as applied to individual theories and has therefore no notion of the success of a theory itself being a random event.

Here is a simple, and by now well-known, counterexample cited (though using slightly different numbers) by Colin Howson ([2000], pp. 52-54). Suppose there is a diagnostic test for some disease D, and that this test (unfeasibly) has a zero rate of 'false negatives': that is, the probability of someone's testing negative if she has the disease is equal to 0; and moreover the test has an (again unfeasibly) low 'false positive' rate: of 1 in a 1000, say – that is, the probability of someone's testing positive even though they do not in fact have the disease is 1/1000. Suppose now that some particular person x has tested positive, what is the chance that x actually has the disease? In order to avoid changing terminology later, let T stand for the theory that x is suffering from D, while e stands for the evidential statement that x has produced a positive result in the diagnostic test at issue. The zero false negative rate is then just expressed by $p(\neg e/T) = 0$; the low false positive rate by $p(e/\neg T) = 1/1000$; and the probability we are interested in, the probability of x's having the disease given that she has tested positive, is of course $p(T/e)$.

It is often asserted to be an empirical result about human psychology (see, for example, Kahneman and Tversky [1972]) that a large majority of people in these circumstances are inclined, in view of the fact that there is very little chance that x will test positive if she does not have D, to infer from x's positive test result that it is highly probable that she does have the disease – that is, that $p(T/e)$ is high. Such people would be reasoning in perfect accord with the above probabilistic version of the NMA:

Premise 1 holds in the diagnostic case because x is certain to test positive (e) if she has the disease (T) (i.e. $p(e/T)= 1$);
Premise 2 holds because it is extremely unlikely that x would test positive if she did not have the disease ($p(e/\neg T) = 1/1000 \approx 0$);
and the conclusion being drawn is exactly that the probability of x having the disease in view of the positive result - that is, $p(T/e)$ - is high.

Yet, as aficionados are well aware, this inference instantiates the 'base-rate fallacy'. Far from it following that the probability of T given e is very high, *any* non-extreme probability of T, given e, is in fact compatible with the truth of the two premises - even a probability that, far from being high, is arbitrarily close to zero. It all depends, of course, on the prior probability of T. In the diagnostic case we can, arguably, take that prior to reflect the overall incidence of the disease. If the disease is very rare, a lot rarer than the test's false-positive rate, then the probability that x has the disease may be very low, despite her having tested positive. For example, if on average only 1 person in a million has the disease, that is on the 'natural' probabilistic model, p(T) = 10^{-6}, then the probability that the person who tested positive has the disease, p(T/e), is only around 10^{-3}.

This is a straightforward consequence of Bayes' theorem; but the reason the 'posterior' is so low can, as is often pointed out, be more readily seen in an intuitive way using an urn model. Suppose we are drawing balls at random from an urn with 10^6 balls, just one of them red (reflecting the fact that only 1 person in 10^6 has the disease) and all the rest white (no disease). Each ball also has either a '+' or a '-' written on it (corresponding to either a positive or a negative result in the diagnostic test). Given that there are no false negatives, the unique red ball must have a '+' on it. As for the false positive rate of 1/1000, we can't model this exactly with an integral number of balls, of course, since there are 999,999 white balls and we want a probability of one being drawn with a '+' on it to be 1/1000, but clearly the number is close to 1000. So, to a good approximation, there are 1001 balls marked '+' in the urn, all but one of which are white. Hence, if one ball is drawn at random from the urn and it happens to have a '+' on it, then there is to that same good approximation only 1 chance in 1001 that it is red – that is, the chance of any particular patient who has tested positive having the disease is only around 1/1000. And yet something has happened, namely the patient's testing positive, that we know is certain to happen if the patient has the disease and extremely unlikely to happen (only one chance in a thousand) if she does not. Premise 1 and Premise 2 both hold here, then, and yet the (alleged) conclusion is (very) false.

There is no doubt, then, that our initial probabilistic reconstruction of 'the' NMA is fallacious. The prospects for producing a non-fallacious version along these lines are surely not improved by the reintroduction of considerations of approximate, rather than outright, truth. As noted, no sensible realist will want to claim anything stronger than that some theory T is approximately true, no matter how astounding its predictive success might have been. But modifying the claim in this way will not help when it comes to attempting a formal probabilistic reconstruction of the NMA.

Let A(T) be the assertion that T is approximately true. The relationship between A(T) and e is altogether less clear-cut than that between T and e. I am taking it that, the relevant auxiliaries being presupposed, T logically entails e; whereas the relationship between e and A(T) is altogether less clear-cut. Nonetheless, since any version of the NMA requires the evidence e to have large impact on the believability of A(T), the realist who seeks to reconstruct the NMA in this probabilistic way seems committed to the claim that $p(e/A(T)) \approx 1$. And again the fundamental assumption in the argument is that the evidence at issue would be very improbable were T not even approximately true, so the realist seems to want the premise: $p(e/\neg A(T)) \approx 0$. Hence we have a straightforward modification of our initial probabilistic argument:

Premise 1′ $p(e/A(T)) \approx 1$
Premise 2′ $p(e/\neg A(T)) \approx 0$
Conclusion′: $p(A(T)/e) \approx 1$ and hence, given that e has occurred, $p(A(T)) \approx 1$.

But clearly the base-rate problem kicks in just as before: depending on the value of the prior probability of A(T), any posterior for A(T) – including one as close to zero as you like – is in fact compatible with the truth of premises 1′ and 2′.

While there is no disputing the fallaciousness of the ‘base-rate fallacy’, it can certainly be questioned whether these probabilistic arguments accurately capture the intuitive considerations underlying the NMA as elicited by such cases as the wave theory of light and the white spot. If the formal arguments *do* capture those intuitions then the latter will of course have to be surrendered, no matter how appealing they may appear. But do they capture those intuitions? We first need to ask how the relevant ‘base-rates’ could possibly be interpreted in the case of the white spot or any other theoretical success that elicits the ‘no miracles intuition’.

In the diagnostic case, the probabilities at issue could, it seems, readily be interpreted as objective chances, reflecting - or perhaps constituted by - limiting relative frequencies: the test’s false positive rate of 1 in 1000 reflects the assumption that if random selections from the whole population were continually made and the frequency were recorded of those people who tested positive but failed to have the disease amongst all those testing positive, then that relative frequency would converge on 1/1000 as the number of selections increased indefinitely. Similarly the 'natural prior' in that case is the overall incidence of the disease within the population: the population proportion of those suffering from the disease is 1 in every million and hence if a series of selections were made at random from the population and the relative frequency of those having the disease recorded, then that frequency would converge on 10^{-6} as the number of selections increased indefinitely.[9]

But how should we interpret the probabilities involved in our probabilistic reconstructions of 'the' NMA – in particular (a) the probability that evidence e would not occur if theory T were false, and (b) the 'prior' probability that T is true? Any attempt to model these probabilities along the lines of those in the diagnostic case would surely be misguided from the outset. In order to develop such a model, we would have to think of ourselves as drawing a theory at random from some population of theories and noting whether it was true, how probable it made e and so on. But, aside from issues about how we would decide whether a given theory was true (we couldn't), what population of theories would that be? Remember my insistence that the only sensible NMAs are 'retail' arguments for the likely approximate truth of specific theories. It seems, then, that this population should not be thought of as consisting of 'every possible theory' (of what?) – which is just as well, since we surely have no real grasp at all on what that collection would be - , but instead perhaps as the set of all possible alternative, rival theories to the particular theory at issue.

---

[9] Notice, however, that this is hardly the prior that would "naturally" be assumed by the Harvard Medical School Students, upon whom much implicit scorn has been poured (see, e.g. Howson [2000], pp. 52-54). The fame of this particular case is based on the fact that a (small) group of students at Harvard Medical School allegedly systematically got the "wrong" answer when asked what the probability is that x has the disease, given that x tested positive (using similar probabilities to those given above). But one assumption involved in the claim that they got the answer about the posterior "wrong" is that the "true" base rate that they "ignored" is the population incidence of the disease. However, no clinician would intuitively "model" the event of someone's coming through her clinic door as representing a random selection from the population. People tend not to visit clinics for no reason – the very fact that they are there means that the reasonable estimate of the pre-test probability that they have some disease relevant to the clinician's speciality is considerably higher than the population prior. Even in US medicine, where over-testing is rife, the appropriate prior that a patient has some disease ahead of her being subjected to some test, is - thankfully – seldom the overall population prior. (For an antidote to the over-testing venom see Gigerenzer [2002].)

After all, in assessing the impact on, say, Fresnel's theory of light of its success with the white spot, there is no interest at all in the fact that theories from, say, chemistry or biology fail to entail that same experimental result (it is inconceivable that they would) and neither is there any interest in how many theories from those fields are true and/or 'successful' in some generic sense not related to the particular predictive success in question.[10]

But are matters really any clearer if we restrict the population to all possible theories that are rivals to the specific theory for whose likely approximate truth we are arguing? Certainly if we allow in 'gruesome' alternatives, or, in the case of mathematically expressed theories, Jeffreys-style alternatives (as discussed above), we know that that class of alternatives will be infinite, indeed non-denumerable; and well known, surely insuperable, difficulties face any attempt to argue that there is an objectively correct prior probability that a theory drawn from such a set of alternatives has some particular property - say truth or approximate truth. As for the other crucial probability in the probabilistic NMA, namely $p(e/\neg T)$ (or, still worse, $p(e/\neg A(T))$), we might start to think of it as measured by the ratio of all possible alternatives to T (or, still murkier, all possible alternatives to A(T)) in which e holds compared to all such alternatives. But aside from the fact that we again have no real grasp on what the set of alternatives is, the standard Laplacean chances approach here is, as Colin Howson forcefully points out (*op. cit*., 46), crucially dependent on the assumption that all the basic cases are of equal initial weight, and that assumption is surely preposterous here.

Someone still looking for a population from which theory T might sensibly be thought to be drawn will need to restrict the class of alternatives to T (or to A(T)) in some way – but how exactly and with what justification? If we restrict that class to T's *active rivals* at the time of its predictive success, this will normally consist of just one theory T′ (the corpuscular as opposed to the wave theory of light, classical as opposed to relativistic physics, *etc*) and $p(e/\neg T)$ is then readily identified as $p(e/T')$. In the most straightforward case, where we take the theory T′ to come along with all the relevant (then-) accepted auxiliaries, then T′ will standardly deductively entail $\neg e$. The

[10] It might be different if one were to think, as perhaps some reliabilists are inclined to, that there is some single 'scientific way' of producing theories and can legitimately be interested in how reliable that scientific way is. But again this just does not correspond to the reality of theory-production (and theory-*acceptance*) in the sciences.

corpuscular theory of light with natural auxiliaries entails that there will be no ‘white spot’, classical physics, again with natural auxiliaries, entails an incorrect motion for Mercury, *etc*. In that case, it is easy to show that the probabilistic version of the NMA goes through without fallacy, since in fact p(T/e) = 1. The argument is now just the probabilistic version of the deductive rule of disjunctive syllogism (and corresponds in the diagnostic case to there being no false positives, which of course means that any person who in fact tests positive must have the disease, *irrespective of base-rates*).

But the fact that the inference as thus construed is valid can be of no consolation to the realist: the term p(e/¬T) in the probabilistic versions of the NMA cannot simply be identified with p(e/T′) where T′ is T’s main historical rival. The possibility that haunts all versions of the NMA is not that some already available theory, different from T, might share the predictive success e at issue, but that some other, *so far unarticulated*, theory could also predict e, while being radically different from T (and hence entailing that T is radically false despite its predictive success). No one would claim that it was a ‘miracle’ that T would get some prediction correct were it false, if some known (radical) rival T′ (that is, a theory that entails that T is indeed radically false) also made the same prediction. But the worry is that T’s, so far as we know unique, success only *seems* ‘otherwise miraculous’ to us, precisely because we are unaware of some so-far unarticulated possibility T″ that equally well has the predictive success, has other epistemic virtues that make it a still better theory than T, and yet entails that T is way ‘off-beam’ in terms of what it says is going on at the ‘noumenal’ level. Using the NMA to infer that T″ was (approximately) true, it would seem to follow from the fact that T is ‘radically’ false from the viewpoint of T″ that T’s success with e was indeed nothing more than a coincidence or miracle.

In sum, then, it is no surprise that attempts to reconstruct ‘the’ NMA in probabilistic terms turn the argument into a fallacy. There is, even ahead of consideration of the cogency of the probabilistic logic, no reason to think that the probabilistic rendering of ‘the’ argument (at least when we try to understand those probabilities in any objective way) is at all satisfactory. In particular, as we have seen, if we try to think of the crucial probability p(e/¬T) as expressing a ratio of possible alternatives to T in which e holds to *all such possible alternatives*, then we get into trouble because we have no handle on that class and certainly no reason to think that all possible alternatives have initial equal weight; but if we restrict the possible alternatives to those we know about (which we might plausibly think about as roughly equal in initial weight), then we also get into trouble since then we get trivial answers that have nothing to do with the issue addressed by the NMA.[11]

(c) The correct way to think about the 'No Miracles Argument'

Magnus and Callender (*op. cit.*), implicitly assuming that these probabilistic reconstructions are the only way formally to capture the 'no miracle intuitions', argue that the fact that those reconstructions are fallacies means that we have no option but to reject the intuitions as worthless. This, they suggest, should, on reflection, cause less of a wrench than might be supposed. They point to the work of Kahneman, Tversky and others as (allegedly) showing that (even bright, well-educated) people find the base-rate fallacy intuitively very seductive. No wonder then, they argue, that some of us have found the considerations underlying the NMA intuitively appealing since, as Howson has shown, when properly analysed that argument embodies that same seductive fallacy. They write: 'the no-miracles argument … [is an instance] of the [base-rate] fallacy, … [R]esearch suggests that educated people are apt to commit this fallacy, and thus the intuitive appeal of [the] argument should not be taken as a sign that [it has] any probative force.' (Magnus and Callender [2004], p.330)

But this suggestion is unconvincing. First of all, although Magnus and Callender are careful to say in the passage just quoted that this psychological research only 'suggests' that (even 'educated') people have some inbuilt tendency to commit the base-rate fallacy, they need something much stronger than this if their claim is to carry real weight. And just what the research of Kahneman and Tversky does or does not establish is a matter of dispute. There are many claims in the literature that their results, rather than being produced by any general tendency to commit the base-rate

[11] The situation is clearly not likely to be improved by resort to some intermediate position concerning the relevant 'population' of theories – as do Magnus and Callender ([2004]) in identifying this with the class of 'all candidate theories'. Again this set is ill-defined; again it is hardly likely that each candidate theory will sensibly be modelled as carrying the same weight (or plausibility); and again why should the ratio of successful 'candidate' theories that are true (as if we could ascertain this!) in distant fields such as biology or physiology, say, be at all relevant when assessing the impact of the white spot success on the realist credentials of Fresnel's theory?

and other probabilistic fallacies, are artefacts of the often confusing experimental protocols their subjects were given. And certainly Gigerenzer for one claims that, when people are set the sort of problems used by Kahneman and Tversky in more 'user-friendly' terms, then the great majority are not seduced by any fallacy but instead give the 'correct' answer.[12]

Moreover, even if we were to accept that the base-rate fallacy is seductive, it seems clear that (a) some people are naturally immune (many physicians, for example, without the benefit of any formal acquaintance with Bayes' theorem, have enjoined generations of medical students to remember that 'common complaints are common' – that is, to remember that they are perhaps more likely to be looking at a somewhat unusual presentation of a common complaint than a rather more stereotypical presentation of an extremely rare condition) and (b) immunity, even if not 'natural', is easily achieved via consideration of Bayes' theorem. Since we are talking psychological facts here, it seems legitimate to report that – whether by nature or (philosophical) nurture – I am definitely immune to the base-rate fallacy and yet still keenly feel the intuitive pull towards realism of predictive success. I find it difficult to believe that I am not joined in this by the majority of those in both the scientific and the philosophy of science communities.

This surely lends further weight to the suggestion that, rather than give up on the intuitions, we should recognise that what has failed are the attempts to capture those intuitions in some more formal probabilistic argument. As I already suggested this ought to have been clear in advance: no one should ever have thought of the process of theory construction in science as one of drawing theories from an urn and checking for 'success' and/or (approximate) truth; so no one, even those totally immune from the temptation to commit the fallacy, would seriously think of base-rates (at least in any objective sense) as being a relevant factor for the NMA in the first place.

Is there then some different way toward a more precise formalisation of 'the' NMA? Colin Howson never explicitly mentions what is usually taken to be its obvious

[12] See Gigerenzer ([2002]) for references. As Gigerenzer in effect points out, it is by no means always clear just what the 'correct' answer is that many people are allegedly missing – part of the reason for this, in turn, is the already alluded-to problem of just what the 'natural' prior is in given circumstances.

formalisation (or, as I shall shortly argue, "formalisation") – namely as an 'abduction' or 'inference to the best explanation'. The idea, after all, is that there seem to be only two *explanations* for why some theory T has scored some striking predictive success: (a) that T itself (including those parts of it talking about 'deep structure') is at least approximately correct and (b) that T, although perhaps radically off-beam at the noumenal level, nonetheless just happens to get this particular prediction correct by chance. (a) is much the better explanation (if indeed (b) counts as any sort of explanation at all). Hence we should accept (a).

It is clear that the probabilistic arguments considered above cannot be regarded as anything like full formalisations of any such 'inference to the best explanation'. Obviously the idea that T explains e is not fully captured by the requirement that T (together with accepted auxiliaries) deductively entail e (and hence $p(T/e) = 1$). Jeffreys-style constructions entail the data they were constructed from but do not explain them. (Or to take a real case: versions of Ptolemaic theory with suitable epicycles entail planetary stations and retrogressions but no one would say that they *explain* those phenomena.) In order to explain some datum e, both the way that T entails e and T itself must satisfy further conditions – T must not have been generated precisely to fit e and T must in any event have other 'epistemic virtues' beyond the mere ability to entail e. (Notice also that the 'best explanation' of the success of Jeffreys-style constructions, or of the epicyclic Ptolemaic theory, in entailing the data they were constructed from is not at all that they have likely latched on to some real regularities, but rather that they were indeed explicitly constructed so as to fit, and hence entail, that very data.)

Of course a probabilist might very well retort that the reason her reconstruction does not do full justice to the idea of inference to the best explanation is that no one has any very precise idea of what explanation is, or of what these 'extra epistemic virtues' involve: while probabilistic reasoning is sharp and clear-cut, it is not at all clear just what the abductive account amounts to and hence no surprise that she is unable to give it a precise probabilistic rendering.

I have a lot of sympathy with this reaction: 'unity' is a frequently cited extra epistemic 'explanatory virtue', but although 'unity' is surely a crucial notion, and although it is easy to exhibit theories that clearly possess it and other theories that clearly do not, no one has been able to give anything like a clear-cut general (and non-language-dependent) characterisation of the notion, let alone any independent justification for holding that theories that possess it can reasonably be thought of as likelier to be true than theories that do not.

Calling a 'retail' NMA – that is, remember, an argument for the likely approximate truth of a particular theory – an 'inference to the best explanation' then, while it does no harm, gives the argument an air of precision or formality that it scarcely deserves: we at bottom remain reliant on the intuitions while avoiding owning up to this.

Like the IBE-ers, Bayesians too might like to think that they can elevate the discussion above the level of the merely intuitive. Colin Howson, having argued at length against 'the' NMA as any form of convincing 'objective' probabilistic argument, goes on to point out that it can in fact readily be given a plausible (personalist) Bayesian reconstruction. Indeed it is patent that the fallacy that Howson, and following him Magnus and Callender, exhibit is blocked if, far from ignoring the 'base-rate' or prior, a further premise is incorporated into the formal argument: a premise that asserts that the prior probability of the theory concerned is not low, but in fact reasonably high. This assignment of a prior is, of course, on the personalist Bayesian approach, to be thought of as simply reflecting a personal judgement about the plausibility of the theory, and not as any sort of reflection of some "objective chance" that the theory is true. While this clearly avoids all the problems with urn models noted above, it does mean that it is difficult to see the Bayesian account as doing anything other than re-expressing our intuitions in what *appear* to be more technical terms: it seems implausible that such a 'nice' theory as the wave theory of light should get such a striking phenomenon as the white spot correct and yet not be somehow 'on the right lines' concerning what it says about 'deep structure'. Of course, as will be very clear, I share this intuition, but it is hard to see the Bayesian analysis, which simply translates 'nice' into 'having a reasonable high prior probability', as adding anything to it. As so often, it is difficult to see anything *explanatory* (explanatory at the meta-level) in the Bayesian account.[13]

[13] Magnus and Callender, *op.,cit.,* seem to endorse this point.

We seem to be left, then, essentially with just the intuitions. Should this concern us? I just argued that Magnus and Callender's alleged reason for seeing the intuitions as carrying no probative force is faulty. But the fact that there seems to be no prospect of producing a convincing formalisation (probabilistic or otherwise) of the intuitive considerations involved in striking predictive success might well seem a much more powerful reason for being suspicious of those considerations.

Well, we should not – so it seems to me - expect too much from arguments in philosophy, certainly not at this very fundamental level. There is, of course, no question of a theory's predictive success – no matter how startling and impressive - *proving* that that theory is true (or even 'approximately true') and hence solving the problem of ('vertical') induction (or 'abduction' if you like) at stroke! Perhaps William Whewell believed so. He claimed that the predictive successes enjoyed by the wave theory of light were 'beyond the power of falsity to counterfeit'. But of course they are not *provably* beyond the power of falsity to counterfeit: the truth *may* be something radically different from what any current theory says it is, and it goes without saying that the (complete) true theory will have all the right empirical consequences, including those describing the predicted effect at issue.[14]

Can we expect to show that, although it is of course possible that the truth is very different from what our current theories say it is, this is at least extremely improbable in the light of their predictive success? Well again surely not in any objective sense of probability – the process of theory-production cannot plausibly be modelled as involving the drawing of theories at random from some super-urn of 'all possible theories', or even all possible rivals to some given theory. We have seen why in some detail in the previous section, but I think it ought, on reflection, to have gone without saying.

Proofs and objective probabilities are not what 'the NMA' is about. The impact of predictive success, together with the notion of 'approximate truth', is intuitive – ineliminably so. It is of course possible that our current theories are radically false

[14] Though even Whewell can, I think, plausibly be interpreted as holding only that this is not a "realistic" (as opposed to a merely logical) possibility.

despite their predictive success, but it seems so downright implausible. Implausible enough, I suggest, to set realism as the default position. It is surely not surprising, at least on reflection, that the implausibility here cannot be captured by any sensible analysis in terms of proofs or of objective probabilities. (Some) realists may wish for something stronger from 'the' NMA, but nothing stronger is defensible.

It is, I think, salutary to remember here, Poincaré's surely correct claim that the NMA (or something *very* similar to it) is involved, not just in the case for realism about successful theories, but also in 'ordinary inductive generalisation'. Indeed perhaps Poincaré's clearest statement of the NMA is made in connection with Kepler's 'generalisation' of Tycho's data and was quoted earlier (p.4). Poincaré claims that what would involve 'attributing an inadmissible role to chance' is holding the claim that such a *simple* law as Kepler's first law is not true except in already-observed cases.[15] (Notice, then, in connection with Colin Howson's account, that there is already the arguable-equivalent of a 'premise' about 'reasonably high prior probability' in Poincaré's argument. It would be an amazing coincidence if Kepler's law, *given that it is so simple*, fits all the observed data but is not generally true (where part at least of the meaning of 'simple' for Poincaré is that the law does not contain parameters that are theoretically arbitrary but have been fixed on the basis of some of those planetary observations themselves).) We have learned (some would say 'at last'!) to stop banging our heads against the brick wall of the 'problem of induction'. We have learned to see that there is no solution of 'the problem of induction' (in the original Humean form) either in the form of a convincing deductive argument (by the definition of deductive validity this is *bound* to prejudge the issue!) or in the form of a correct probabilistic argument leading to the conclusion that the generalisation at issue is objectively highly probable, given all the instances. And we (or most of us) have come to see that we should never have expected any such 'solution' to work. Nonetheless we do not doubt that the reasonable, default, view is that the observational generalisations sanctioned in mature science are in fact correct (though notice that, as the case of Kepler cited by Poincaré indicates, this allows for the generalisation turning out to be strictly false but still 'correct within certain limits'). Similarly in the case of the acceptance of observation-transcendent theories,

[15] Poincaré ([1905], pp. 149-150).

which Poincaré correctly regarded as essentially part and parcel of the same process: the fact that we have no proof and no argument for high objective probability does not imply that, again in appropriate circumstances, the reasonable default position is anything other than that those theories are at any rate approximately correct (a position which, as we will see, also allows for those theories turning out to be strictly false but still correct within certain limits).

The intuitions set some form of realism as the default position: there is no more to the "argument" than that. The realist might like to say something stronger, but there is nothing stronger to say that anyone should like.

(d) Can the No Miracles Argument, even when properly construed, be defeated?

All arguments for default positions are of course defeasible. And there are a number of clear-cut ways in which this particular 'realist default' could indeed be defeated. *First* it would be defeated by a demonstration that for any theory T enjoying a striking predictive success one or more rival theories can automatically be created which (i) share T's predictive success and (ii) in the light of which T would have been judged radically false. *Secondly*, it would be defeated by the demonstration that there are indeed lots of theories from the history of science that were genuinely predictively successful but which can, by no plausible stretch of the imagination, still be seen as 'approximately true'. *Finally*, the NMA as I construe it would be defeated if it could be shown that there is another non-chance "explanation" for any successful theory's predictive success that is at least as good as the realist one. Let's consider these three possible defeaters in order.

As mentioned earlier, there certainly are well-known constructions that produce alternatives to any given observational generalisation O (grue-style constructions) or mathematically expressed theory T (Jeffreys-style constructions) that purportedly share the same predictive success as O or T. More generally, we know from Duhem that, if we are ready to accept any *ad hoc* mess as a rival, we can, for any given theory, automatically produce rivals to it that are empirically equivalent (with respect to any given set of empirical results). But does the ready availability of such alternatives defeat the NMA as I am suggesting it should be construed?

Instead, as I argued above, these constructions point to an implicit assumption behind the NMA intuitions – namely that the theory at issue is unified and non-*ad hoc*. The intuition is not that it would be a miracle if any old theory happened to have some surprising (and correct) empirical consequence if it were not fundamentally correct, but rather that it would be surprising if this were true of a theory that is unified. This is the reason why the gruefied or Jeffreys-style theories, like *ad hoc* theories in general, would not in fact be judged to share the predictive success of the original theory out of which they are constructed. It is another reflection of this same situation that in the case of theories constructed *ad hoc* in order to yield some phenomenon, we would not be tempted to think that their 'success' in entailing the phenomenon at issue revealed anything about their likely correspondence with reality. Instead – as noted earlier - the explanation for their 'success' is immediately to hand: 'success' in yielding the empirical phenomenon at issue was pre-determined since the theory was constructed precisely so as to yield that phenomenon.

The much more taxing challenge is the *second* one. This is the worry that the realist position as default is defeated, not by any abstract construction of allegedly equally successful potential rivals, but by the real existence of a long list of theories from the history of science that were unified, non- *ad hoc and* predictively successful, but that cannot any longer sensibly be regarded as even approximately true. This is the challenge most often associated with Larry Laudan and, since it is the basis of the 'Pessimistic Induction', it will form the main focus of attention of the next section of the paper. But before that, there is the issue of the *third* possible way to defeat the claim that the NMA sets realism as the default position. The NMA, even in the weak form in which I have endorsed it, certainly relies on the assumption that the only 'explanation' of a theory's predictive success aside from the realist one is that its success is achieved purely 'by chance'. If there are other, better explanations (or even a single rival explanation that is equally good), then the pro-realist force in the argument would evaporate.

The only two such attempted rival explanations that I know of are suggested by van Fraassen's *Scientific Image*. He there explicitly endorses a 'Darwinist' explanation:

> I claim that the success of current scientific theories is no miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive – the ones which *in fact* latched on to actual regularities in nature. ([1980], p.40)

In a prefatory remark, van Fraassen suggests that the Darwinian explanation of why a mouse runs away from a cat is:

> Do not ask why the *mouse* runs from its enemy. Species which did not cope with their natural enemies no longer exist. This is why there are only ones that do (*ibid.*).

Similarly, the suggestion is, there is nothing 'otherwise miraculous' to explain about the success of current scientific theories – they would not *be* current scientific theories had they not been successful.

There is an important ambiguity in van Fraassen's 'explanation'. He talks of successful theories as 'the ones which have *in fact* latched on to actual regularities in nature'. But does this mean that the theories have latched on to the real regularities (a judgment that would clearly involve some strong generalisation) or only the *apparent* regularities – in other words that the phenomena seem to be in accord with these theories *so far*? We can, of course, only *know* the latter: that our theories have so far made the right predictions. The argument from the success of current theories to their general empirical adequacy (which, I take it, is what van Fraassen had in mind) involves, as we already noted in connection with Poincaré, implicit appeal to the NMA; and therefore already concedes at least a good deal of what the realist wants to endorse. Again it seems to me that only confusion can result from failing to distinguish what we can effectively establish from what might be claimed to be reasonably *inferable* (via the NMA or in some other way) about essentially transcendental matters (such as approximate truth or overall empirical adequacy).

Following this line, and adopting my claim that the 'success' we are interested in is empirical predictive success, then van Fraassen's attempted 'Darwinist' explanation of our theories' successes becomes the following. It is no miracle that our current theories are predictively successful, they wouldn't have survived in the competition to become currently accepted if they had not been empirically successful.

But it is surely obvious that this explanation simply involves an equivocation on 'explains the *acceptance* of theory T' and 'explains the *success* of theory T'.

Accepted scientific theories are indeed *pre-selected* for success: at any rate in mature science (where, by my definition of maturity at least, the latest theories have automatically enjoyed predictive success), no new theory would be accepted were it not to score predictive successes (and indeed predictive successes over and above those enjoyed by its predecessor). An entirely reasonable answer, then, to the question 'Why was Fresnel's wave theory of light (for example) accepted by his contemporaries?' is 'because it was successful (predictively successful), it would not have been accepted, in competition with other theories of light, had it not been'. Although the analogy with Darwinian natural selection does not in fact extend at all far, what van Fraassen says on this issue is fair enough.

But this *is* (in fact 'artificial') *selection*: it explains the human decision to accept a theory – it says nothing about the objective epistemic virtues of the theory itself. In particular it is no sort of answer at all to the question 'ok, the theory was accepted by scientists because it scored predictive successes, but *what enabled it to score predictive successes*?' In particular why does a theory which is fundamentally a set of assertions about unobservable entities finish up correctly predicting a range of surprising and striking empirically checkable phenomenon? Nothing that a scientist (or group of scientists) can decide can make this happen – it is a property purely of the theory itself. It is this question, not the one about acceptance, that, I claim, elicits the 'no miracles intuition'. And the only two available answers to this second question so far do seem to be (1) that the theory scores this dramatic success because it has latched on, at least in an approximate way, to the 'deep structure' of the universe and

(2) that its success was a matter of chance. Van Fraassen's 'Darwinist' account does not address this question at all and hence is not the missing third view.[16]

Although this Darwinist one is the non-chance rival explanation to the realist one that van Fraassen explicitly endorses in *The Scientific Image*, another such rival seems more natural given his general position. Again assuming that we are talking about success as an effective notion (so that our best theories have *not* proved successful by proving to be (generally) empirically adequate), then this alternative, non-realist alternative would go as follows. Suppose it is legitimate (it is, though van Fraassen is unconvinced!) to seek an explanation (no matter how weak) for the success of our theories in current 'mature' science in having produced striking empirical predictive successes, then there is no need to invoke anything as strong as the claim that those theories are somehow 'approximately true', their success is explained at least equally well by the weaker supposition that they are 'empirically adequate'. Indeed, as van Fraassen explicitly claims, since all we could ever check – now or in the indefinite future – is the empirical adequacy of a theory, it can be argued that the so-called extra explanatory oomph supplied by the further assumption of approximate truth is just hot

[16] Perhaps the suggestion is that the example of Darwinism shows that there is *nothing to explain* about the success of scientific theories (somewhat along the lines of Arthur Fine). It would be an error to attempt any positive account of, for example, how the mouse gets to be relatively 'fit'; and an analogous error to attempt any positive account of how our accepted theories get to be successful. But this suggestion rests on a misinterpretation of Darwinian theory. Darwinian theory does not *of course* rest content with saying that there is nothing to explain – the mouse would not have been around if it were not evolutionarily fit. It seeks a positive (and preferably independently testable) account of the features that *make* the mouse 'fit' (for example of how it gets to be able to spot predators, to run from them and to gain access to hiding places that the predator cannot reach), and an account of how those features arose and were developed out of some ancestral population through mutation and natural selection (what Philip Kitcher calls a 'Darwinian History'.) Similarly in the realism issue it seems entirely reasonable to seek a positive account of what it is about the theory that has enabled it to achieve its success (and therefore acceptance). What realism provides may not be a very good explanation (there is, as explained above pp. 10-11, sadly no question of independent testability) – but then we are talking philosophy! As weak as it may be, it seems to be the only explanation (or, if you like, 'explanation') around.

An analogous critical reaction to van Fraassen's attempted Darwinist explanation of the success of currently accepted scientific theories to the one presented in the text was developed in Peter Lipton's [1989]. Lipton envisages a club whose membership is restricted to those with red hair. The question 'Why does Ginger (who is a member of the Red Hair Club) have red hair?' is ambiguous as is revealed by the fact that one might give two quite separate answers, corresponding to the two arms of the equivocation I identify in van Fraassen's line. One is 'That's a stupid question, of course he has red hair otherwise he would not be a member of the Red Hair Club'. But this only addresses the selection rules of the club and says nothing about the intrinsic properties of Ginger. These would be invoked by the second style of answer which, if hair colour were a single gene-trait (which it is not, but just for simplicity's sake), would be along the following lines ' Because hair colour is a simple Mendelian trait, Ginger's ancestry in this regard is such and such, and he inherited allele 1 from his mother and allele 2 from his father.'

air. The realist view appears to make an extra, bolder claim but in fact, since this so-called ‘extra’ is not empirically checkable, it is the sort of merely apparent strengthening that ought to be treated with contempt – a case of pretending to stick one’s neck out while in fact remaining in complete safety.

Assume for the moment (as is not in fact the case with the (hopelessly) weak notion of empirical adequacy that van Fraassen explicitly endorsed in his book) that a theory is empirically adequate only if it entails, in conjunction with accepted auxiliaries a large number of empirically correct consequences – idealising: *all* known empirical results in the relevant field. Then it follows from the fact that T is empirically adequate that it will correctly entail some given (general) empirical result e that lies in its field.

In terms of a response to NMA, then, it would appear that ‘It is no miracle that T gets e correct, because T is (generally) empirically adequate’ is at least as satisfactory as that ‘T gets e correct because it is – at least approximately – true’. Indeed if we agreed that empirical adequacy is at least as good an explanation as approximate truth, then it would appear that the former is preferable since it involves a weaker claim.

But again there seems to be an equivocation here – this time in what it is that is being ‘explained’, or equivalently what it is that would allegedly be a ‘miracle’. If we are concerned with ‘observational generalisations’, as Poincaré was when discussing the case of Kepler’s first law (see above p.6), then the ‘miracle’ would be that the ellipse had fitted all observed cases and yet happened not to fit yet unobserved, or never-to-be-observed ones. In such cases, empirical adequacy is the same as truth and the assumption of empirical adequacy/truth avoids any appeal to an incredible coincidence or miracle (though it does, as I already pointed out, appeal to a version of the NMA). But in the case of theories, and particularly fundamental theories, the ‘miracle’ is importantly different. After all, if we take the wave theory of light again as our example, the derivation of particular (general) empirical results (themselves of course empirical generalisations) - such as for example ‘the result’ of the two slit experiment with monochromatic light – begin with assumptions about the form of the vibrations in an unobservable medium. The derivations of a range of other empirical results (the one slit diffraction experiment, the opaque disk’s white spot, the relative intensities of the reflected and refracted beams in partial reflection, etc. etc.) all kick off from the same assumptions or related ones about the same unobservable vibrations. The empirical results at issue of course do not follow within the theory unless such fundamental, essentially theoretical, assumptions are made.[17] The 'miracle' would occur if these theoretical assumptions should lead time and again to empirical predictions that turn out to be verified *while they themselves are not even approximately true*. It is clear that the assertion that the theory is empirically adequate (albeit empirically adequate across the board not just in cases that have already been checked) would do precisely nothing to remove *this* miracle: indeed the relevant 'miracle' would exactly be that the theoretical assumptions appear, and continue to appear, empirically adequate while not being (at least approximately) true.

To claim that it is no miracle that our successful theories have been empirically adequate (so far), they just *are* empirically adequate (in general) is then, in effect, to deny that there is anything to explain about the empirical success of the genuinely theoretical parts of our best theories. And hence to deny that the NMA intuitions have any pull. It is not to provide an attempted explanation of those intuitions to rival the realist one. Neither van Fraassen's explicit Darwinist 'explanation' nor this one that might seem to be implicit in his general view works. *So long as you feel the intuitive pull of the NMA, then some sort of realism remains the only available option*.

The main conclusions of this examination of the considerations underlying 'the' NMA and attempts to formalise it are, then:

1. The considerations apply 'retail' to particular scientific theories; in so far as there is a 'wholesale' view, it is simply the union of particular 'retail' ones.
2. The 'argument' seems to be ineliminably intuitive but it does nonetheless arguably set realism as the default position.
3. The constructability of *ad hoc* alternatives to a given successful theory does not override the default.

[17] Of course should there be any assumption that is 'idle' within the theory, that is we do not need to involve it when deriving the empirical results within the theory, then that idle assumption should have been eliminated in advance. (I am assuming in other words that the distinction that Kitcher makes ([1993], see also Psillos, *op. cit.*) between 'working' and 'idle posits' has been made in advance in deciding just what is the theory that we are inclined to be realist about.)

4. Neither van Fraassen nor, so far as I am aware, anyone else has overridden the default by supplying a defensible rival non-realist non-chance 'explanation' of a theory's predictive successes.

The remaining issue is the central one of whether the (or, as we shall again see, 'the') Pessimistic Induction defeats the claim that the NMA intuition sets realism as the default position.

2 The 'pessimistic induction' re-considered

(a) The intuitions

Poincaré, as we saw earlier, clearly articulated the intuitions underlying the NMA. So too those underlying 'the' Pessimistic Induction. Long before Laudan, Poincaré wrote ([1905], p.160):

> The ephemeral nature of scientific theories takes by surprise the man of the world. Their brief period of prosperity ended, he sees them abandoned one after the other; he sees ruins piled upon ruins; he predicts that the theories in fashion today will in a short time succumb in their turn, and he concludes that they are absolutely in vain. This is what he calls the bankruptcy of science.[18]

Poincaré's starting-point is, then, a claim about the history of science – namely that theories that were once accepted have later been rejected (seemingly left as 'ruins') in favour of theories that are at least logically inconsistent with them (just how ruin-like or '*radically* false' the later theories really make the older ones look will be a central issue in what follows). The best way to think of this is not so much as the premise for any sort of inductive argument, but simply as posing a challenge to those advocating a realist attitude towards at least some currently held theories. However, as is indicated by Poincaré's talk of *predicting* 'that theories in fashion today will … succumb in their turn', it is tempting to think of the fate of earlier theories as constituting some

[18] Notice however that while Poincaré endorses the 'PI intuitions', he immediately goes on to characterise the man of the world's scepticism as 'superficial' (*ibid*. 160.) – for reasons that we shall shortly come to.

sort of inductive grounds for inferring that our current theories will eventually be rejected in favour of new theories inconsistent with them.

(b) How (not) to formalise 'the pessimistic induction'

How might such an inference be formalised? The most straightforward suggestion seems to be as some sort of induction by enumeration. But, although it is reasonably clear what the conclusion of that induction should be – that it is at least likely that currently accepted theories will eventually be replaced by theories inconsistent with them -, it is not clear what the premise/s should be. One suggestion might be to think of the history of science as sliced up into 'stages' and try out the following inference for size (**PI_1**):

Premise: All theories that were accepted at earlier stages in the history of science have eventually turned out to be replaced by theories inconsistent with them.

Conclusion: (inductively/probabilistically) all theories now accepted at the current stage of science will eventually be replaced in their turn by theories inconsistent with them.

Or, taking it for the sake of argument that an older theory's being inconsistent with a theory that we currently accept (accept, after all, on the basis of a *larger* body of evidence than that which supported its predecessor theory) is a sure sign of that older theory's falsity, we could run this version of the argument in a more dramatic-sounding way (**$PI_{1'}$**):

Premise: all theories accepted at earlier stages in the history of science have eventually proved to be false.

Conclusion: All theories accepted at the current stage of science will also eventually prove to be false.

Even before deciding whether there is any force in either 'induction', and thinking about just "how false" replaced theories standardly look in the light of their successors, the first question that arises about the PI as formulated in either of these two ways is surely whether its premise is even remotely true.

What counts as a 'stage' in the history of science is clearly a somewhat conventional matter. Suppose, however, that science is considered to move to a new 'stage' whenever there is a change in fundamental theory in *some* mature science. Several of our current theories, even fundamental theories – the general theory of relativity, for example – are, then, theories that were 'accepted at earlier stages of the history of science' and yet which have *not* - at least so far -'eventually turned out to be false'. There has been at least one major 'revolution', involving the adoption of the new quantum theory, since the articulation of the general theory of relativity after 1915. (Poincaré's talk of our current theories being replaced 'soon' is difficult to reconcile with the history of science – after all, Newton's theory of mechanics and universal gravitation remained accepted for more than 200 years.) Moreover, if we look, not at fundamental theories, but at what might be called 'middle-ranking theories', then there is an enormous number of scientific theories – that pure water consists of molecules each of which in turn contains two atoms of hydrogen and one of oxygen, that light has a finite velocity, *etc, etc* – whose history is one of eventual acceptance followed, it would seem, by continued acceptance at all 'later stages'.

There is no indication whatsoever that such 'middle-ranking' theories are likely to be rejected 'eventually'. So-called scientific revolutions (again in *mature* science) seem directly to affect only the most basic or fundamental theories – theories of the constitution of matter, the structure of space, the fundamental nature of light and of heat and so on. To be sure, everything 'lower down' is, strictly speaking, affected by the changes at the fundamental level: the assertion that water consists of molecules which in turn consist of two atoms of hydrogen and one of oxygen clearly meant something very different at a time when we had a theory of atoms as something like indivisible billiard balls with hooks than it does now that we think of atoms as themselves built out of more fundamental entities, ultimately superstrings or whatever, obeying the probabilistic laws of quantum mechanics. Nonetheless it seems natural to say that 'the' theory of the constitution of water (at the level at which this is identified with 'H_2O whatever the H and the O turn out more precisely to be') once accepted remained accepted; and that we have no reason at all to think that it is likely eventually to prove to be false. Similarly, science's ideas about what exactly it is that has a finite velocity of propagation and produces various specified phenomena – a stream of tiny material particles, an energy flow through an elastic solid medium, a set of 'particles' without rest mass obeying a strange new probabilistic mechanics – have changed in apparently radical ways, but again it seems natural to say that the theory that light has a finite velocity, once accepted as a result of Römer's work on the moons of Jupiter in the 17^{th} Century, has remained accepted ever since. And this is true of many theories – indeed, continuing to concentrate on mature science and laying aside issues of how exactly to individuate theories, it is tempting to say that it is true of the overwhelming majority of them.

It follows that there is no mileage in amending the argument to read (**PI$_2$**)[19]:

Premise: *most* theories accepted at earlier stages in the history of science have eventually proved to be false.

Conclusion: Most theories accepted at the current stage of science will also eventually prove to be false as well.

The premise, although weaker, is not weak enough to escape arguable falsehood *and* **PI$_2$** has two further significant faults. First, it directs attention away from the real issue, which is about *fundamental* theories and their replacement. And secondly, and relatedly, it moves us toward a 'wholesale' version of this argument. As in the case of attempted 'wholesale' versions of 'the' NMA', such an argument would operate with the proportions (or, using some sort of urn model, probabilities) of theories in the past and at present that are various combinations of successful or unsuccessful (in some – alleged - generic sense) and true or false. Such an argument therefore takes the 'convergent realist's' distinctive claim to be that the proportion of true theories amongst all those currently accepted is greater now than at earlier stages in science. This, as I shall soon explain, is quite the wrong way to think about the position of any sensible 'convergent realist'.

The problems with either **PI$_1$** or **PI$_2$** don't end with the falsity of its premise. **PI$_1$**, as it stands, has the form of an induction (by enumeration) to the next instance. Inductions

[19] This is how Peter Lewis interprets Putnam's original argument in his ([2001]).

by enumeration that carry any intuitive cogency are, I believe, always in fact best thought of as more sophisticated enthymematic arguments – to be understood really as eliminative inductions (or, still better, as 'deductions from the phenomena'[20]) but with various extra premises left implicit. Any cogency or persuasiveness carried by the inference from, for example, 'all sampled ravens have been black' to 'the next raven we see will be black', surely stems from the fact that the sample implicitly includes elements from both or all sides of the differences – sex, degree of maturity, geographical location, *etc.* – that more general theories lurking in the background inform us are the usual markers of possible di- or poly-morphism in avian plumage. Knowing what we do about some kinds of birds (for example, peacocks) surely no one would find even surface plausibility in the inference from, say, 'all sampled *male* ravens have been black' to 'the next raven will be black'. For an induction to the next instance to have even surface plausibility, there must at least be no positive reason to think that the next object sampled may be in some significant way different from those already sampled.

However in the case of 'the' PI when construed in this way, the realist who holds that current theories are in general closer to the truth than earlier ones, exactly does assert that there is a difference between the current crop of accepted theories and their predecessors – at least in degree if not in kind. It might be (and sometimes has been[21]) argued, therefore, that not only is the first version inductive, it is a very *bad* induction: arguing for the falsity of current theories on the basis of the falsity of earlier ones, is analogous to Russell's chickens inferring from the fact that he has always done so before, that the farmer will feed them when he comes through the gate on Christmas Eve (or, in my favourite, blacker version of this old chestnut (unless of course you happen to be a chicken), to the guy who jumps from the top of the Empire State Building and says to himself as he passes each storey: 'So far so good, so far so good, …').

The second version, **PI$_2$**, suffers from the same problem – indeed even more overtly. If (mistakenly, but as this version suggests) we think of convergent realism as insisting that the main characteristic of the development of science has been the articulation of theories, an ever greater proportion of which are (approximately) true,

[20] See Worrall ([2002]) and the literature referred to therein.
[21] See for example Lipton ([2001]).

then its invalidity is patent. There is a clear difference between the theories spoken about in the premise (namely they are all from the past) as compared to those mentioned in the conclusion (which are current theories). Moreover this difference is one that the convergent realist, as we are now viewing her, insists is significant. Hence the logical form of $\mathbf{PI_2}$ is:

Most Ts (theories) with property P (accepted in the past) are Q (false).

Therefore, most Ts with property ¬P (accepted at present) are Q.

This is patently an invalid inference form (counterexample: T: Olympic sprinters; P: male; Q: regularly break 10 seconds for the 100m).

Though Peter Lewis begins his analysis with $\mathbf{PI_2}$, he understandably concentrates most of his attention on Larry Laudan's famous 'confutation of convergent realism'. Lewis sees Laudan's argument as amounting to the following *reductio* ($\mathbf{PI_3}$):

1. Assume that the success of a theory is a reliable test for its truth.
2. Most current scientific theories are successful.[22]
3. It follows that most current scientific theories are true.
4. But then it further follows that most past scientific theories are false, since they 'differ from current theories in significant ways'.[23]
5. Many of these past theories were successful.
6. Hence, and contrary to the original assumption, the success of a theory is not in fact a reliable test for its truth.

---

[22] As noted earlier in connection with the NMA, this seems much weaker than is justified and therefore importantly misconstrues the situation: the realist, whose position Laudan, on Lewis's interpretation, is reducing to absurdity certainly claims that *all* current theories (at least in the mature sciences) are successful. In general, Lewis and following him Magnus and Callender misconstrue the debate by treating the success of a theory as if it were a random variable with an associated probability. Success – as Simon Blackburn correctly insists (*op.cit.*) is a *datum* that is neutral between realist and anti-realist. All theories in mature science are 'successful' (especially, I have suggested, predictively successful) and the realism issue is what this effectively decidable notion tells us about the undecidable, 'transcendent' notion of their overall relationship to reality.

[23] Lewis writes 'differ from current theories in significant ways' with a view to warding off the claim that we can continue to regard past theories as approximately true.

Let's not argue about exactly how faithful this is to Laudan's ([1981]) original and evaluate it on its own terms. Lewis points out that issues have been raised about the truth of some of the premises, notably premise 5; but his contention is that this is by-the-by since, whatever might be the case about its premises, the argument is straightforwardly invalid. The fallacious step that Lewis identifies is that from line 5 to the final conclusion at line 6.[24] Lewis himself slips between proportions and probabilities, without drawing any explicit link; but in essence, and employing probabilities, the demonstration that the step from 5 to 6 is invalid proceeds as follows.

A test is reliable, according to the generally accepted standard, if and only if, both the false negative and the false positive rates associated with it are low. Notice again that in making the argument one about reliabilities in this sense, Lewis's analysis turns its back on questions like 'is it reasonable to take the white spot success as some sort of indication that Fresnel's theory has latched on to the deep structure of light?' and turns instead towards 'wholesale' issues about the proportion of theories accepted in science today that are successful in some (alleged) generic sense and the proportion of those that are true. In the case of 'success' as an allegedly reliable indicator of truth, then, writing Tr(T) for T is true and S(T) for T is successful, the claim that both the false negative and false positive rates are low means that both $p(\neg S(T)/Tr(T))$ and $p(S(T)/\neg Tr(T))$ are low. Laudan's 'plethora' of successful but false theories, a plethora that turns into a list of such theories that could be 'extended *ad nauseam*', constitutes his evidence for premise 5. Indeed Laudan backs up this list with the (surely outrageous) claim 'I daresay for every highly successful theory in the past which we now believe to be a genuinely referring theory, one could find half a dozen once successful theories that we now regard as substantially non-referring'.[25] Following which, Lewis takes the assertion that $p(\neg Tr(T)/S(T)) = 6/7$ as a reasonable quantification of premise 5.

---

[24] Depending on how exactly 1 to 6 are formalised there may I think be other fallacies in the argument (for example, interpreting 1 as stating the false positive and false negative rates are low, then the step from it and 2 to 3 looks, as Lefteris Farmakis pointed out to me, itself awfully like the base-rate fallacy) but of course one fallacy is enough to make a piece of reasoning fallacious.

[25] See Laudan ([1981], p. 123). Laudan, seemingly plausibly, takes it that, whatever the ultimately accepted account of approximate truth (if there is one) may be, it will entail that no theory whose central terms are 'non-referring by current lights' can possibly be judged approximately true on the basis of those 'current lights' and must therefore instead count as radically false.

But to pass from premise 5 as thus construed, to the conclusion 6 that 'the success of a theory is not a reliable test for its truth' is to be confused by what Lewis himself calls the 'false positives paradox' and what Magnus and Callender identify on his behalf as another instance of the base-rate fallacy. In order for it to follow from $p(\neg Tr(T)/S(T)) = 6/7$ that success is unreliable as an indicator of truth, it would have to follow that one or other of $p(\neg S(T)/Tr(T))$ and $p(S(T)/\neg Tr(T))$ is not low but reasonably high. But it is easy to see that to infer, for example, that the false positive rate, $p(S(T)/\neg Tr(T))$, is substantial from the fact that $p(\neg Tr(T)/S(T))$ is reasonably high at 6/7 is to commit exactly the inversion fallacy that was exhibited in the diagnostic case and which was therefore taken to invalidate 'the' NMA. The fallacy again results from ignoring the 'base rate': here the probability of a theory's being true ahead of consideration of its 'success'. Lewis readily shows that if, for example, $p(Tr(T)) =1/25$, then both the false positive and false negative rates for success as an indicator of truth can be reasonably low at 1/5 and still false positives would be expected to outnumber true positives by 6 to 1, exactly as in Laudan's claim. (This assignment of probabilities, incidentally, requires implicitly that $p(S) = 28/125$!)

A(n alleged) demonstration that the anti-realist's seemingly most potent argument is fallacious (and there is of course no question but that *Lewis's version* of the argument is indeed fallacious, indeed as we have seen in footnote 24 even "more fallacious" than he himself suggests) might be expected to be music to realist ears. But it is difficult to see the realist, even if she were to accept Lewis's version of the PI, as taking much, if any, consolation from his demonstration of fallaciousness. A rate of true theories of only 1 in 25 amongst what are supposed, presumably, to be theories that have at least been taken seriously in some way both in the history of science and in current science hardly seems a convincing basis for a realist view. And indeed, as we saw in the first part of the paper, the lower the 'prior' that a theory is true, the harder it is for the NMA to supply the sustaining role for her position that the realist seeks.

But Lewis sees this apparent difficulty as in fact affording the realist a great opportunity – by pointing to 'a natural way for a realist to explain Laudan's historical evidence'. This 'natural way' is to claim that true theories were very rare in the *history* of science but are rather common in *current* science:

… the realist can interpret Laudan's historical cases, not as evidence against the reliability of success as a test for truth, but merely as evidence of the scarcity of true theories in the past. Admittedly, this commits the convergent realist to the empirical claim that successful theories were rare in the past and are common today [because she sees success as a reliable indicator of truth], but the original appeal to success to combat Putnam's argument already entails such a commitment. (Lewis [2001], p. 377)

This is why Magnus and Callender in their development of Lewis's argument differentiate two reference classes: the set of 'candidate theories' *in the past* and the set of 'candidate theories' *now*. Laudan's infamous list only 'establishes' that $p(\neg Tr(T)/S(T)) = 6/7$ relative to the class of historical candidate theories.[26]

I agree, then, that if we were to accept that **PI$_3$** (or indeed either of the attempted formulations we considered earlier) really captures the argument of those who have taken the history of theory-change in science to be a threat to the realist's view, then it would follow that there is no force in what had seemed initially to be the most potent consideration in favour of anti-realism. No need, then, for any serious account of the epistemic status of current scientific theories to take this consideration on board. And in particular no need for structural realism to attempt to produce a 'synthesis' of the NMA and the PI. Both are simply different manifestations of the same fallacy. No wonder that the two sides in the scientific realism dispute talk past each other – they suffer from two different forms of the same delusion - and no wonder that the whole issue produces, in right-minded people, nothing but 'ennui' (Magnus and Callender [2004], p. 322).

But we already saw in section **1** that this attitude is unsustainable in the case of the NMA – the fault there lay, not with the underlying intuitions, but with ill-judged attempts to formalise them. Exactly the same applies to the case of 'the' PI. Indeed, if anything, it applies even more sharply. No one should ever have taken the formalised argument **PI$_3$** seriously. Just like the other two versions I identified above, it is *clearly*

[26] And so – despite the fact that Magnus and Callender continue to call it the base-rate fallacy - the fallacy that Lewis points to is better characterised, on their account, as a reference class fallacy (the fallacy of assuming that a conditional probability p(A/B) remains the same when the reference class shifts and hence A and B, although retaining their *intentional* meaning, in fact become different events from an extensional point of view).

fallacious and misrepresents the real issues. The argument fails to capture the genuine challenge posed to realism by the facts about theory-change in science.

I will rehearse more systematically than above a number of ways in which $\mathbf{PI_3}$ is misguided in the remainder of this sub-section; and then turn in the next subsection to the real underlying challenge (that is, to the issue of how really to think about 'the' PI).

So why is $\mathbf{PI_3}$ misguided? *First,* and as in the case of 'the' NMA, the attempt to model scientists as if they were drawing theories at random from some urn of possible theories and seeing if they turn out to be marked 'successful' and/or 'true' is totally unrealistic. In mature science at least, new replacement theories are developed out of their ailing predecessors in quite systematic, deliberate ways.

Secondly, (and relatedly) it is surely wrong, so far as 'mature' science goes, to take 'success' to be a random variable with some associated probability: as noted when considering the NMA, *all* the *accepted* theories in 'mature science' will be successful – otherwise they would not have been accepted.[27] The realism issue is precisely about whether a particular theory that *has been* successful – and that means, as I have insisted, successful in predicting (often new) general types of phenomena - can reasonably be considered to be at least approximately true: probabilities of success just don't enter in any sensible way. Success, unlike truth, is an effective notion – we know if some theory has been successful, but we never *know* whether a theory is true (or even approximately true 'overall').

Thirdly, the radical change in attitude towards current, as opposed to previously accepted, theories that is required on Lewis's analysis for the convergent realist to counter the argument he sees in Laudan is clearly fanciful. Even if the convergent realist misguidedly traded in terms of proportions of theories from the past that were false compared to the proportion of current theories that are false, she surely could not countenance – assuming that we are speaking of 'mature' science – a shift from

[27] Magnus and Callender's suggestion (*op.cit.*, p.320) that this 'biases' the sample shows how badly they are misconstruing what seems to me at least clearly the best version of the 'convergent realist's' case. Of course it is biased! The realist only makes his case for theories that have already proved their success and therefore been 'accepted' – accepted that is in the sense that Newton's theory was and General Theory of Relativity currently is, not 'accepted' in some weaker sense as an interesting working hypothesis, or the best of an admittedly unappealing bunch so far, etc.

something of the order of 24 out of 25 theories being false in the history of science to something considerably under half currently.

Finally, and most importantly, a sensible convergent realist does *not* interpret her view that theories now are 'less false' than they used to be as meaning that the proportion of accepted theories that are false is less now than it was at earlier 'stages' in science. Instead, the sensible convergent realist accepts that even if all currently accepted (again *fundamental*) theories are false, the improvement is constituted by the fact that *each fundamental theory is 'less false' than its predecessors*.

(c) The correct way to think about the 'pessimistic induction'

'The' PI should, then, certainly not be construed in any of the ways considered by Peter Lewis or by Magnus and Callender. How then *should* it be viewed? As the passage from Poincaré quoted above (p.34) argues, the starting-point is a certain claim about the history of science. I would formulate the fact at issue – and it does indeed seem to be a fact - as follows:

> For every fundamental theory T currently accepted in mature science, there is at least one predecessor theory T′ that used to be accepted and that contradicts T.

Science now accepts the general relativistic account of space-time, which contradicts the Newtonian account of space and of time; quantum mechanics is inconsistent with classical mechanics; the photon theory of light is inconsistent with the waves-in-an-electromagnetic field account and both are inconsistent with the theory of light as a wave in an elastic solid ether; and so on. The ruins (or maybe, as we shall see, *apparent* ruins) are at the 'top' - at the level of our (currently) most fundamental theories. There are few, if any, ruins 'lower down. The lower level stuff (dynamics of relatively slowly moving bodies, finiteness of the velocity of light, etc) is entailed by the new theory no less than the older one (at least 'entailed within observable error'). No claim about 'all' or 'most' erstwhile accepted theories turning out to be false would be justified.

Rather than forming the premise for any inference, this fact about fundamental theories should be looked upon, in the first instance, as posing a challenge for the scientific realist: ‘how can you hold that our current fundamental theories are even approximately true, when previous generations of scientists have felt the same way about theories that are clearly inconsistent with the current ones?’ This is surely a substantial challenge, one that the realist needs to face and answer to if her position is to be at all seriously sustainable. Nothing in either Lewis’ or Magnus and Callender’s analysis touches this challenge at all, let alone makes it any less imposing (or any more ‘ennui-inducing’!).

If we do insist on basing an inference about current theories on this historical fact then the correct one seems simply to be this (it amounts to no more than a different way of expressing the challenge):

> Unless a reason can be given for why our current theories are better supported than their predecessors, not just in degree (everyone accepts this) but in kind, then, although it is of course *possible* that our current theories will never be replaced in the whole future history of science by ones inconsistent with them (and even possible that this is because they are true), it is difficult to see any reason of principle why this should be the case - the assumption that our current fundamental theories will eventually be replaced by ones that stand in much the same relation to them as they stand to their predecessors is, if you like, the *default position*.[28]

It follows from this in turn that no account of scientific realism could possibly be sustainable unless it can somehow accommodate this default position. An anti-realist who holds that ‘the pessimistic induction’ refutes, or at least substantially undermines, any realist account is in effect committed to the view that there is no convincing realist way of accommodating this default assumption – that is, of meeting the above challenge. But two types of response to the challenge are in fact open to the realist.

[28] A similar view was articulated long ago by Mary Hesse ([1976]) as a ‘principle of no privilege’. This states (p.264) that ‘our own theories [should be] held to be as much subject to radical conceptual change as past theories are seen to be’. This principle is surely correct, as Hesse fails to note but as indicated in the text, *unless* some reason can be given why the epistemic situation with respect to our current theories is different in kind from (superior to) the epistemic situations of earlier theories with respect to the theories that they held. Rather than any principle, then, ‘no privilege’ should, as I argue, be regarded as setting the default position that any remotely acceptable version of realism must accommodate.

First she might argue that the epistemic warrants for the theories we hold now are indeed in some principled way different from the warrants for those theories held at earlier stages in the history of science. I assume that no one serious would take the heroic version of the first line and claim that, whatever may have been the case with earlier theories, our current theories are (sensibly regarded as) *true*. As already acknowledged, and as is obvious enough, nothing in the history of science *dictates* that our current theories will eventually be replaced, but that history does surely set this assumption as the default position. The realist taking this first line would clearly owe us an explanation of why our epistemic position relative to current theories is so different – in kind and not just in degree – from that of our predecessors. There is, of course, no doubt that our situation is different *in degree* – there now is *more* evidence in favour of the (currently held) photon theory of light than there ever was for Maxwell's electromagnetic theory of light (in the straightforward sense that every bit of evidence that supports Maxwell's theory also supports the photon theory, while there is extra evidence that supports the photon but not Maxwell's theory). But then the same state of affairs held earlier between Fresnel's and Maxwell's theories – and we now 'know' that a late-19th century scientist would have been wrong if she claimed that we then had the truth about the nature of light. We cannot plausibly claim that we are in a relevantly different epistemic situation unless we can demonstrate a difference in epistemic kind, rather than simply one of (evidential) degree.[29]

There have been certain theory-shifts – that from Aristotelian to Newtonian physics is one – for which it may be reasonable to claim that the epistemic credentials of the second theory are indeed different in kind from those of the first. This is when the adoption of the second theory marks the emergence of the field into *maturity*. As noted earlier, Putnam, when in realist mode, and Boyd, famously did not insist that every theory that has been accepted at some time in the history of science is approximately true - they claimed this is so *only* for theories in fields that have achieved 'maturity'. Laudan ([1981], pp. 122-123) and others have, very reasonably, been suspicious of this appeal to maturity as providing a good deal of room for

[29] Peter Lipton in his ([2000]) argues that, to the contrary, the fact that our current theories are better supported empirically than were earlier theories is itself enough to make it reasonable to suppose that our current theories may not suffer the same fate. But this is surely like arguing that it is reasonable to believe that the current 100m sprint record will never be broken, despite the fact that a whole series of earlier records have been broken, on the grounds that the current record is better than the earlier ones.

adhoccery: whenever a realist cannot see how to make a case that the older theory in some theory-shift continues to look approximately true in the light of the newer theory (this will, of course, be the realist's second line of response that I consider *below*), it can always be claimed that the older theory belonged to an era when that branch of science was immature.

In order to allay these suspicions, as I argued earlier, 'maturity' must be given a clear and defensible characterisation rather than being left as a 'free parameter' that the realist can fill in *ad hoc* in order to avoid possible counterexamples. It seems natural for a realist to do this by 'reading off' the characterisation of maturity from the main consideration that inclines him towards realism – the NMA.[30] The intuition behind the NMA comes forcefully into play only when a theory makes a successful prediction of some general, repeatable phenomenon. It seems natural, then, to define maturity as having been achieved by a field once the accepted theory within it, rather than simply accommodating phenomena *post hoc*, makes a genuine, and confirmed, prediction. So far as I can tell, the corpuscular theory of light, for example, simply took phenomena that were already known and dressed them up *post hoc* in corpuscularist terms (or, more often, *attempted* to thus dress them up). Only with Fresnel's theory of light as a wave motion in an elastic medium did optics produce a theory that made novel predictions – of the white spot diffraction pattern, as we saw, and later, for example, of internal and external conical refraction.

But, however plausible this line of response to the 'Pessimistic Challenge' might appear in the case of some theory-shifts, it certainly will not work for *all*. Suppose it is agreed that the corpuscular theory of light appears 'radically false' from the point of view of the classical wave theory. It seems plausible (and non-*ad hoc*) to argue that – because it was not genuinely predictive – the corpuscular theory had not succeeded in making optics 'mature' and that the realist, therefore, is under no obligation to show that some more nuanced analysis could change that judgment of apparent radical falsity. However, it also seems initially plausible to view the classical wave theory (based, remember, on an elastic solid ether) as 'radically false' from the point of view

[30] See my ([1989], pp. 153-154).

of the electromagnetic theory.[31] If predictive success is the hallmark of maturity then, even if earlier optics was not 'mature', Fresnel certainly made it so – as Laudan remarks concerning the white spot 'if that does not count as empirical success, nothing does!'. (Laudan [1981], p.115)

Fortunately, the realist still has a second line of response to the Pessimistic Challenge open to her. This involves (i) accepting that the only evidential difference between, say, the situation of an early -19th century physicist in respect of the classical wave theory and a 21st century physicist in respect of current theories of light is indeed one simply of degree: we have more evidence now for our photon-based theory than scientists did then for the classical wave theory, but nothing else of qualitative epistemic relevance has changed; (ii) accepting also that the current theory of light and the classical wave theory are of course logically inconsistent and therefore the latter must be false if our current theory is true; *but* (iii) nonetheless arguing that there is, despite their logical inconsistency, a substantial sense in which that earlier theory has been retained in the later theory. If the development of science can be shown to be cumulative (or essentially- or quasi-cumulative) despite the mutual inconsistency of successive theories, then we can retain talk of the earlier theory still appearing 'approximately true' in the light of the later theory (though such talk will amount to no more than a re-expression of that (quasi-) cumulativity).

If this second response can be made to succeed, then the 'Pessimistic Challenge' is indeed met: there is no reason at all to believe that our current theories will not in their turn eventually give way to theories that contradict them, in just the same way that those current theories contradict their predecessors; we cannot of course *prove* that the replacement theories will nonetheless retain the current ones in the same way that those current ones retain their predecessors – they *may* reject them in a much more wholesale way than we have seen exhibited, according to this hoped-for account, in the history of science so far; but it nonetheless seems reasonable (though far from deductively mandatory) to make the *optimistic* induction that this will not be the case, and that our current theories will be retained in the same way as their predecessors have been. In that sense, and in that sense alone, it would be reasonable

[31] This is certainly the case on Laudan's account which takes it that if some earlier theory involved a 'central' theoretical term that is denied real-world reference by a subsequently-accepted theory, then the earlier theory is, in the light of the later theory, not even approximately true.

to hold – as the 'default' pending some objection stronger than anything provided so far - that our current theories in the mature sciences are approximately true.

All this, however, is predicated on a big 'if'. Laudan, as noted, seemingly quite reasonably takes it that if T_2 entails the non-existence of what according to T_1 is a central theoretical entity, then it would be absurd to claim that T_1 continues to look approximately true in the light of T_2. Yet the theory of light as a disturbance in a *sui generis* electromagnetic field outright rejects the idea that there is an elastic solid mechanical medium filling the whole of space; while current theories of light based on particles without rest mass and obeying the probabilistic laws of quantum mechanics not only re-emphasise that rejection but make Fresnel's elastic waves seem even more remote.

If, I suggest, there is any sense to be made of 'quasi-cumulativity' through theory-change, despite the above considerations, then it is of cumulativity at the level of *structure* and the corresponding version of scientific realism that becomes defensible is *structural* scientific realism. Essentially metaphysical ideas about how the mathematical structures involved in our best theories are instantiated in reality may indeed seem to change radically as science progresses, but those mathematical structures themselves are invariably retained (or 'quasi-retained'). Maxwell's theory (in its mature form)[32] may do away with the elastic solid ether on which Fresnel's theory was based, and hence Fresnel was indeed as wrong as he could be about *what* waves to constitute the transmission of light, but his theory continues to look correct from the vantage point of the later Maxwell theory in that it agrees that optical effects fundamentally depend on *something or other* that waves at right angles to the direction of the transmission of light. Hence Fresnel's equations – though not his preferred interpretation of the terms within them (such as amplitudes) – are retained in the later theory; and also, though this time in an approximate way, in the theories, based on quantum mechanics, that eventually replaced Maxwell's theory in its turn.[33]

---

[32] As is well-known, Maxwell himself continued to hope for some sort of 'reduction' of his electromagnetic field to some underlying 'mechanical' medium (just as Newton seems to have continued to hope for some sort of 'reduction' of his universal force of gravity to pressure gradients in some all-pervading mechanical medium). However the 'mature' view (taken by Maxwell's successors) was that the electromagnetic field is *primitive.*

[33] Let me here clear away one misunderstanding of my views: as I had taken to be obvious and implicit, it is not *just* the mathematics, but also the way the mathematical terms are tied to observation that must

As I put it in earlier work, Fresnel was, from the vantage point of the later theory, as wrong as he could be about the *nature* of light (there is no such thing as the elastic solid ether and *a fortiori* no such thing as waves transmitted through such an ether), but he was correct about its *structure* (light really does depend on *something or other* that vibrates at right angles to its direction of transmission).

The Fresnel-Maxwell case, as I acknowledged all along, pays for being such a pure and clear instantiation of the structuralist's case by being entirely unrepresentative: it is , so far as I can tell, the only case in the history of science in which a successor theory in a 'revolution' retains the mathematical equations of its predecessor *completely intact.* Certainly, if this story of cumulativity is to be extended to other cases – for example to those theories of light that came after Maxwell's, or to the theory-shift between classical physics and the general theory of relativity – then liberal use will need to be made of the *correspondence principle*. The reason I often talked above of '*quasi*-cumulativity' is that by far the more usual situation is that the equations of the older theory reappear in 'modified form' as 'special cases' of the newer theory, via the correspondence principle. Often, but by no means invariably, this means that, as some parameter in the newer theory tends to some limiting value, the newer theory's mathematical equations tend to the older theory's equations.

But cumulativity of structure *modulo* the correspondence principle is surely substantial cumulativity. And moreover this degree of continuity underwrites the fundamental pro-realist intuition that the predictive success of theories in mature science cannot plausibly be accidental.

3 Conclusion: structural realism lives! (So far)

If Laudan had been right then no version of scientific realism could have been defended: 'the' PI, if established, clearly trumps 'the' NMA. If it could be shown that theories radically at odds with the ones that we currently hold had themselves been predictively successful, then it could hardly be argued that it would be a 'miracle' if our current theories were radically false and yet enjoyed the predictive success that

remain the same (or essentially the same) in order for a judgment of 'quasi- cumulativity' to be justified.

they do. The predictive successes enjoyed by that list of earlier theories would then be multiple and ungainsayable instances of this alleged miracle. And with 'miracles' familiarity surely does breed contempt: if an event has many instances then it can hardly count as a miracle.

However, if structural realism is correct that there is a significant element of (quasi-) cumulativity in the development of mature science, then the NMA is, on the contrary, strengthened: from the point of view of later science, it was indeed no 'miracle' that Fresnel's theory scored the predictive successes it did, because, although Fresnel entirely misidentified the nature of light, he was correct about the structure of optical phenomena. His predictions, for example, about the relative intensities of light in the refracted and reflected polarised beams in partial reflection were inevitably correct, as judged from the later viewpoint, because they follow just from the mathematical equations that are also sanctioned by later theories – which we are supposing for the moment are true. It then seems unreasonable *not* to suppose that our current theories of light, should they be replaced in their turn, will also appear structurally correct (of course, and by definition, in some currently unspecifiable way) in the light of those future replacements. Similarly (and more representatively), it is, from the vantage point of the special theory of relativity, no 'miracle' that Newton's theory scored predictive successes (with, for example, the precession of the equinoxes), because, although it did not fully grasp the structure of kinematical and dynamical phenomena, it did do so approximately – Newtonian mathematics applies, within observational error, when a body's velocity is small compared to that of light, and Newton's successes had all concerned such bodies.

And this in turn means that structural realism, through its response to the 'Pessimistic Challenge', strengthens in a further way the NMA on which it is partly based. As noted earlier (footnote 7, p. 13), it is initially unclear whether the assumption that a theory is approximately (as opposed to fully) true can legitimately be said to *explain* some predictive success that it has enjoyed. If a theory is true then so, of course, are all of its deductive consequences – no wonder then that if a theory is true, its empirical consequences turn out to be true as well. But in the absence of any agreed account of approximate truth, there is certainly no *guarantee* that because a theory is 'approximately true' so automatically will be its empirical consequences. So there is no automatic underwriting of the claim that it would be no wonder if a theory scored the predictive successes it did if it were approximately true; and hence no automatic underwriting of the claim that the theory's approximate truth 'explains' its empirical success. However, within structural realism, 'T is approximately true' or rather 'approximately true as judged from the vantage point provided by theory T′' just *means* that the structure of T (or a facsimile of it, via the correspondence principle) is preserved within T′. Since T's structure is enough to entail all the observational results T ever did, the 'approximate truth' of T does indeed explain its success with any given empirical phenomenon e. In cases where the correspondence principle is involved we cannot tell ahead of time what limits the area of empirical success of our current empirically successful theory, but we can be sure that, so long as the correspondence principle applies, the successes, such as e, will all lie within the area that the later theory will specify. In the standard case it will turn out that all the empirical successes of the earlier theory are characterised by values of some parameter invoked by later theory that fall close to the limit, whatever it turns out to be, where the older theory's structure re-emerges.

Finally, the intuitive NMA is strengthened still further by the - underemphasized – fact that it is invariably itself no accident that the structure of a successful theory reappears within, or reappears as a 'limiting case' of, its successor. On the contrary it is an explicitly adopted heuristic requirement that this be the case: new theories are usually *deliberately constructed* so as to be linked by the correspondence principle to their empirically successful predecessor.[34] The fact that this heuristic requirement has so often been imposed and has produced new theories that both automatically share the empirical success of their predecessors (because the new theories yield their predecessors as limiting cases) *and score new predictive successes of their own* is surely the basis for a related, but slightly different, version of the NMA. How could insisting on the requirement that the older theory be a limiting case of its successor, have so often produced a successor theory that is itself strikingly and independently predictively successful unless both theories were 'on the right lines' concerning the 'deep structure' of the universe?

[34] See Norton and Bain ([2001]) and Zahar ([1988]).

In summary: Although the facts about the startling predictive success of some of our theories and about theory-change in the development of science can be, and as we have seen *have been*, made the starting-points for arguments that are clearly defective, the facts themselves, and their importance, remain. The problems lie not with the facts themselves and the intuitions they elicit but with the arguments that have inappropriately been developed out of them. The central issue concerning scientific realism is whether some version of it can be developed that reconciles the facts about theory-change in science with the view that the predictive success of (some of) our theories cannot have been a gigantic coincidence. The only such account that seems at all promising at present, so far as I can see, is structural realism. If - as many recent and not so recent critics have suggested - [35] this promise is illusory, because structural realism is subject to fatal objections, then it would seem that no scientific realist account can be defended. Whether or not this is so, is not an issue that should fill any philosopher of science with 'ennui'.

References

Blackburn, S. [2002]: 'Realism: Deconstructing the Debate', *Ratio* **15**, pp. 111-133.

Demopoulos, W. and M. Friedman [1985]:, 'Critical Notice: Bertrand Russell's *The Analysis of Matter*: Its Historical Context and Contemporary Interest', *Philosophy of Science,* **52**, pp. 621-639.

Fine, A. [1984]: 'The Natural Ontological Attitude' in J. Leplin (*ed.*) *Scientific Realism*, Berkeley, University of California Press.

Fine, A. [1986]: *The Shaky Game*, Chicago: University of Chicago Press.

Gigerenzer, G. [2002]: *Reckoning with Risk: Learning to Live with Uncertainty*. London: Penguin.

Hardin, C. and A. Rosenberg [1982]: 'In Defence of Convergent Realism', *Philosophy of Science*, **49**, pp. 604-615.

Hesse, M. [1976]:, 'Truth and Growth of Knowledge' in F.Suppe and P.D.Asquith (*eds*) *PSA 1976*, volume 2, East Lansing, MI: Philosophy of Science Association.

[35] See above footnote 2.

Howson, C. [2000]: *Hume's Problem*, Oxford: Oxford University Press.

Kahneman, D. and A. Tversky [1972]: "Subjective Probability: A Judgement of Representativeness", *Cognitive Psychology*, **3**, pp. 430-454.

Ketland, J. [2004]: 'Empirical Adequacy and Ramsification', *British Journal for the Philosophy of Science*, **55**, pp. 287-300.

Kitcher, P. [1993]: *The Advancement of Science*. Oxford: Oxford University Press.

Laudan, L. [1981]: "A Confutation of Convergent Realism" in D. Papineau (*ed.*) *The Philosophy of Science*, Oxford: Oxford University Press, [1996], 139-165.

Laudan, L. [1984]: *Science and Values*, Berkeley: University of California Press.

Lewis, P. [2001]: 'Why the Pessimistic Induction is a Fallacy', *Synthèse*, **12**, pp. 371-380.

Lipton, P. [2000]: 'Tracking Track Records', *Proceedings of the Aristotelian Society*, Supplementary Volume **LXXIV**, pp. 179-205

Magnus P. D. and G. Callender [2004], 'Realist Ennui and the Base Rate Fallacy', *Philosophy of Science,* **71**, pp. 320-338.

Newman, M.H.A. [1928]: 'Mr. Russell's Causal Theory of Perception', *Mind*, **37**, pp. 137-148.

Niiniluoto, I. [1987]: *Truthlikeness*, Dordrecht: Reidel.

Norton, J. and J. Bain [2001]: 'What Should Philosophers of Science Learn from the History of the Electron' in J. Z. Buchwald and A. Warwick (*eds*), *Histories of the Electron.* Cambridge MA: MIT Press, pp. 451-65.

Poincare, H. [1905]: *Science and Hypothesis*, repr. New York: Dover, 1952.

Putnam, H. [1978]: *Meaning and the Moral Sciences,* Boston: Routledge and Kegan Paul.

Worrall, J. [1989]: 'Structural Realism: The Best of Both Worlds', repr. in D. Papineau (*ed.*) [1996], *The Philosophy of Science*, Oxford: Oxford University Press, pp. 139-165.

Worrall, J. [1989a]: 'Fresnel, Poisson and the white spot: the role of successful predictions in the acceptance of scientific theories' in D. Gooding, T.Pinch and S. Shaffer (*eds*) *The Uses of Experiment*. Cambridge: Cambridge University Press, pp. 135-157.

Worrall, J. [2000]: 'The Scope, Limits and Distinctiveness of the Method of "Deduction from the Phenomena": Some Lessons from Newton's

"Demonstrations" in Optics', *British Journal for the Philosophy of Science*, **51**, pp. 45-80

Worrall, J. [2002]: 'New Evidence for Old' in P. Gardenfors et al. (*eds*) *In the Scope of Logic, Methodology and Philosophy of Science*, Amsterdam: Kluwer.

Worrall, J. and E.G. Zahar [forthcoming]: 'The "Newman Objection" to Structural Realism – *what* objection?'

Zahar, E.G. [1988]: *Einstein's Revolution – a Study in Heuristic*. Illinois: Open Court.