Miracles and Models: Why reports of the death of Structural Realism may be exaggerated

JOHN WORRALL

Introduction

What is it reasonable to believe about our most successful scientific theories such as the general theory of relativity or quantum mechanics? That they are true, or at any rate approximately true? Or only that they successfully 'save the phenomena', by being 'empirically adequate'? In earlier work[1] I explored the attractions of a view called Structural Scientific Realism (hereafter: SSR). This holds that it is reasonable to believe that our successful theories are (approximately) *structurally correct* (and also that this is the *strongest* epistemic claim about them that it is reasonable to make). In the first part of this paper I shall explain in some detail what this thesis means and outline the reasons why it seems attractive. The second section outlines a number of criticisms that have none the less been brought against SSR in the recent (and as we shall see, in some cases, not so recent) literature; and the third and final section argues that, despite the fact that these criticisms might seem initially deeply troubling (or worse), the position remains viable.

1. The Attractions of SSR

Quantum Electrodynamics predicts the magnetic moment of the electron to a level of precision better than 1 part in a billion. How, it seems natural to ask, could a theory make a prediction about what can be observed that turns out to be correct to such an amazing degree of accuracy, if what it claims is going on 'behind' the phenomena, at the level of the universe's 'deep structure', is not itself at least approximately correct? This may be logically possible, but it seems none the less monumentally implausible.

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

To cite another well-worn example: Fresnel's wave theory of light—that light consists of periodic disturbances transmitted through an all-pervading elastic medium, the 'luminiferous ether'—turned out to predict, completely surprisingly even to Fresnel himself, that if a small opaque disc is held in the light diverging from a point source, then the very centre of what would be the shadow of the disc if geometrical optics were true must in fact be illuminated (indeed just as strongly illuminated at that centre point as if no opaque obstacle had been held in the light beam). This consequence of the theory was, according to an often-told story,[2] regarded by Fresnel's peers as so absurd that his theory was in danger of being laughed out of court (or at least out of the French Academy's prize-competition on the diffraction of light). But Fresnel and Arago performed the experiment with the opaque disc and lo and behold the white spot exists! How, it seems natural to ask, could Fresnel's theory correctly make a prediction that is so at odds with what 'background knowledge' would lead us to expect, *unless* it had somehow or other latched on to the way that light really is? The theory, it seems natural to conclude, must be at least approximately correct if it can get such a striking prediction right.

This is the consideration that makes most of us (and this includes the great majority of scientists) incline toward some version of scientific realism. It has often been dressed up (following Hilary Putnam) as 'the No Miracles Argument' (NMA). If Fresnel's theory, say, were substantially off-beam in what it asserts is going on 'behind' the phenomena in order to produce them, then we would, it seems, be forced to believe that the theory 'just happens' to be correct in predicting effects like that of the 'white spot'—to believe that, despite being quite false, the theory 'just happens' to have consequences about these observable situations that seem so unlikely to be correct but in fact turn out to be so. We would be forced, that is, to accept that the theory's success with this and other predictions was a mere coincidence or 'miracle'. But, so the NMA goes, we should not accept that miracles have happened, at any rate not if there is an alternative non-miraculous explanation. And in this case the assumption that Fresnel's theory

[2] For the real story see my 'Fresnel, Poisson and the White Spot: The Role of Successful Prediction in the Acceptance of Scientific Theories' in G.Gooding *et al.* (eds.) *The Uses of Experiment* (Cambridge: Cambridge University Press, 1989)—but the facts about the history do not affect the issues tackled here.

itself is correct or at any rate approximately correct is exactly such a non-miraculous alternative explanation of its striking predictive success. If the theory were, in particular, outright true then it would of course be no coincidence at all that what it entails about the 'white spot' is correct—all deductive consequences of a true assertion are bound to be true.[3] Hence, the NMA concludes, the reasonable assumption is that Fresnel's theory is indeed (at least approximately) correct. And the same goes for any other theory that has enjoyed comparable striking predictive successes (as all accepted theories in 'mature' science have, since this is a precondition of acceptance).

A lot can be said about the NMA (or, rather, 'the' NMA—since on more detailed analysis it splits into a number of alternative arguments). I shall say some of this later in this paper and my 2007 book[4] goes into greater detail both about its formulation and the role of 'approximate' rather than outright truth within it. However, there is no denying that the intuitions underlying the NMA are powerful and make scientific realism (in some version or other) very attractive.

The (very substantial) fly in the ointment, however, soon becomes apparent when we think some more about, for example, the 'white spot' case. Fresnel's theory, from which this startling and startlingly correct prediction was made, states that light consists of periodic vibrations transmitted through an all-pervading mechanical medium—in Fresnel's final version of the theory this 'luminiferous ether' is held to be an elastic solid. Yet Fresnel's theory was later replaced by Maxwell's electromagnetic theory of light. Maxwell's theory states that light consists of periodic changes of the electric and magnetic field strengths. In what might be called its mature form (which became definitive), this electromagnetic field is *sui generis*: it is just a basic irreducible fact about space that at each point of it and at each instant of time there are well-defined values of the electric and magnetic field strengths; the 'mature' theory explicitly denies that these field strengths can in turn be

[3] The situation is, in fact, not so clear once it is accepted that we can (at best) claim only 'approximate' truth for even our best theories: clearly an approximately true theory is strictly speaking false and hence will have infinitely many false consequences. Here I shall avoid these complications and simply assume that if a theory is approximately true in the appropriate sense then it is no miracle that it gets some prediction correct to within observational accuracy.

[4] J. Worrall, *Reason in 'Revolution: A Study of Theory-Change in Science* (Oxford: Oxford University Press, 2007).

explained via the contortions of some underlying mechanical medium.[5] Hence this later theory, it seems, straightforwardly denies the existence of the most central theoretical (alleged) 'entity' in Fresnel's theory.

This is why Fresnel's theory lies on Larry Laudan's celebrated list of theories that were predictively successful, but which we now 'know' to be radically false.[6] (Laudan plausibly argues that being inconsistent with a theory that science eventually comes to prefer is a sure sign of the falsity of the earlier theory; and that, although the notion of 'approximate truth' remains notoriously vague, if the later theory denies the existence of any 'entity' whose existence is central to the earlier theory, then no sensible account could make that earlier theory count as even approximately true in the light of the later one.) Maxwell's theory in turn was eventually replaced by a theory that makes light consist of photons—weird 'particles' lacking rest mass (and, most of the time, any definite spatial position) that obey an entirely new and probabilistic quantum mechanics. Yet both Maxwell's and the photon theory equally well entail the existence of the white spot and indeed go on to make further impressive predictions of a kind impossible to conceive within Fresnel's theory (in the case of Maxwell, about, for example, the effects of passing a beam of polarised light through an intense magnetic field).

Similarly, many people in the 18th and 19th centuries believed that Newton's theory of mechanics plus gravitation had revealed the truth about the universe (scientists were wont to lament that there was only one truth about the universe and Newton had deprived them of the opportunity to discover it)—this was in large part because of that theory's own impressive predictive successes (with, for example, the precession of the equinoxes, the 'perturbations' from Keplerian ellipses and later, of course, the discovery of Neptune). And yet Newton's theory is based on the assumptions

[5] As is well-known, Maxwell himself continued throughout his life to hold that the field must in the end be the product of an underlying material medium. However, in what might be called the mature version of Maxwell's theory, the field is indeed *sui generis*.

[6] L. Laudan, "A Confutation of Convergent Realism" repr. in *The Philosophy of Science*, D. Papineau (ed.) (Oxford: Oxford University Press, 1996), 107–138. Of course what the claim that we now 'know' those earlier theories to be false means is that our current theories (which are objectively better supported than their predecessors) imply that they are false. (Just how 'radically' false they imply them to be will be an issue that looms large in what follows.)

that space is 'flat' and infinite, that two events are either simultaneous or they are not and that there is action-at-a-distance: all assumptions that are outright denied by the theory—of general relativity—that we now accept.

It seems, then, that the facts about theory-change in science show that if it counts as a miracle for a false theory to enjoy striking predictive success, then such 'miracles' occur, if not exactly all the time, then nonetheless with some regularity in the history of science. It is surely true that if, in the light of apparently radical theory-changes like these (so-called scientific revolutions), we are forced to admit that there is no element of continuity between theories accepted at different stages in science, then the NMA is rendered impotent. The history of science in that case would display, just as Laudan argued it does, a whole series of theories that can only be counted as 'radically false' in the light of theories now accepted and yet which enjoyed unambiguous and striking predictive success of the kind pointed to in the NMA. It seems difficult to resist the suggestion that it is not just logically possible, but a possibility that we need to take seriously, that our currently accepted theories will themselves eventually be replaced by theories that stand in the same relation to them as those currently accepted theories stand to the previously accepted ones and therefore, on the supposition we are now making, will look radically false.

This is clearly not a deductively compelling inference from the historical facts concerning theory-replacement—that is why it is usually referred to as the 'Pessimistic *Induction*'. It is of course logically possible that although all previous theories were false, our current theories happen to be true. But to believe that we have good grounds to think that this possibility may be actualised is surely an act of desperation—it seems difficult indeed to supply any halfway convincing reason to hold that we can legitimately ignore the possibility that the future history of science will be similar to the past history of science and therefore to ignore the possibility that our current theories will eventually be replaced in the way that they themselves replaced their predecessors.[7] If these theory-changes

[7] Of course everyone believes that our theories are improving—in the sense at least that later theories are better empirically supported than their predecessors (the so-called phenomenon of 'Kuhn loss' of empirical content being a myth). But this is clearly a question of degree, while in order to justify rejecting the conclusion of the 'pessimistic induction' we would surely need some reason to think that there was a difference in *kind* between earlier theories and the present ones. As it stands, rejecting the

can indeed only count as 'radical'—that is, there is no substantial carry-over, no substantial 'continuity' from one theory to the next so that the earlier theory can only be counted as plain false in the light of its predecessor—then any form of realism seems patently untenable. Only the most heroic head-in-the-sander could then really hold that our current theories can reasonably be thought of as true. If Fresnel's theory can only count as radically false in the light of current theories of light and there is no sense in which that theory is retained (or 'quasi-retained') within those current theories, then to hold *either* that our current theories are true and will never be replaced in the future *or even* that they are approximately true and will be substantially retained within any successor theories that may come along would be a matter of pure, a-rational faith.

Various responses have been developed to the 'Pessimistic Induction'. One general line is to restrict the scope of realism to the level of theories that can be argued to have been entirely unaffected by 'scientific revolutions'. Science may now have radically different views about the fundamental constitution of matter than it did at the time when the chemical elements were thought of as consisting of billiard ball atoms equipped with a number of hooks, but science continues to tell us that one molecule of water consists of two atoms of hydrogen and one atom of oxygen. So, the suggestion goes, we should be realist about theories 'lower down' the theoretical hierarchy, but not about the fundamental theories at the top. Any such position might be called a version of 'partial realism'.

An *apparently* different, and currently widely supported, view is 'entity realism'[8]. Its proponents seem to regard this as an entirely different animal since it claims to eschew realism about theories altogether in favour of realism about entities. But how do we know (or think we know) that some (alleged) entity really is an entity—that is, how do we know (or take ourselves to know) that

---

idea that our current theories are likely to be replaced because earlier ones have on the grounds that our current theories are better supported than the earlier ones (see e.g. Peter Lipton, 'Tracking Track Records', *Proceedings of the Aristotelian Society*, Supplementary Volume LXXIV (2000), 179–205) would be rather like justifying rejecting the idea that it is likely that the current 100m sprint record will eventually be broken by pointing to the fact that the current record is better than the earlier ones.

[8] I. Hacking, *Representing and Intervening*, (Cambridge: Cambridge University Press, 1983) and N. Cartwright, *How the Laws of Physics Lie*, (Oxford: Clarendon Press, 1983).

there is something in reality corresponding to some term involved in our theoretical framework? The answer given by entity realists is that we know this if we can *manipulate* the 'entity' in question. Hacking, for example, discussed some experiments that (*are taken to!*) involve spraying electrons at a particular kind of target and famously remarked 'If you can spray them, they are real!' It is surely patent, however, that entity realism is not a distinctive position at all but simply a (rather ill-defined) version of partial realism. One need only ask *why* we believe that we are spraying electrons at a certain target in certain circumstances or more generally 'manipulating' an electron in certain circumstances. We certainly don't ever directly apprehend the electrons, let alone the manipulation of them. The answer to this question is surely that we believe we are manipulating electrons because we accept certain *theories* that tell us that this is what we are doing and in the light of which we interpret certain observable signs (tracks in a cloud-chamber or whatever) as produced by (alleged!) electrons. Theories are inevitably involved. Entity realists are simply telling us that we should be realists about certain types of theory (ones that are sufficiently low-level and well-entrenched) and not about others (ones that are more fundamental).

Entity realism is, then, just a version of partial realism and hence it shares the main defect of that general view—namely that it surely gives up too easily on the attempt to underwrite at least some sort of realist attitude towards *our most fundamental theories*. It is these fundamental theories, after all, that are the ones that most strikingly elicit the 'no miracles' intuition. It is fundamental theories like Newton's theory of space, motion and gravitation with its prediction of the hitherto-unsuspected existence of Neptune, or Einstein's account of space-time with its prediction of the bending of the light rays by massive objects like the sun, or Fresnel's account of the basic constitution of light with its prediction of the 'white spot' or Quantum Field Theory with its prediction of the magnetic moment of the electron that provide the most striking predictive successes and, hence, the best reason that I can see for being a realist. No one should, of course, even independently of the facts about theory-change, be a fully gung-ho realist about our fundamental theories. There is, for example, a genuine current issue about whether a fully coherent version of Quantum Field Theory can even be formulated; and Quantum Mechanics and General Relativity are, to say the least, uneasy bedfellows. Hence all informed commentators expect one or, more likely, both to be 'corrected' in some not-yet-discovered 'synthesis'. No one, therefore, should claim that it is reasonable to believe that our current fundamental theories are outright true (again: quite independently of the facts about theory-change); but surely one should not give up so easily on the view that it is reasonable to believe they are in some sense *approximately* true.

The only remotely plausible position that does not give up seems to me the 'Poincaré synthesis' (i.e. SSR). Henri Poincaré developed a classic account of the No Miracles Argument, but also fully recognised—long of course before Larry Laudan—the threat to any realist view that seems to be posed by the facts about theory-change in science. Poincaré wrote:

> 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[9].

But Poincaré immediately went on to argue that this apparent threat to realism, and hence to the appeal of the No Miracles Argument, is unreal:

> [The man of the world's] scepticism is superficial; he does not take account of the object of scientific theories and the part they play, or he would understand that the ruins may still be good for something. No theory seemed established on firmer ground than Fresnel's, which attributed light to the movements of the ether. Then if Maxwell's theory is preferred today, does it mean that Fresnel's work was in vain?
>
> No, for Fresnel's object was not to know whether there really is an ether, if it is or is not formed of atoms, if these atoms really move in this way or that; his object was to predict optical phenomena. This Fresnel's theory enables us to do today as well as it did before Maxwell's time. The differential equations are always true, they may always be integrated by the same methods and the results of this integration still preserve their value.[10]

This might seem to amount to a ringing endorsement of instrumentalism—a view that holds that theories should be thought

[9] H. Poincare, *Science and Hypothesis*, (New York: Dover, 1905), 160.
[10] Op. cit. note 9, 160.

of as merely codifications of empirical results and a view that Poincaré is indeed often, but quite mistakenly, taken to hold. In fact he immediately goes on explicitly to reject that interpretation of his view:

> It cannot be said that this is reducing physical theories to practical recipes; these equations [the ones that are retained in the transition from Fresnel's theory to Maxwell's] express relations, and if the equations remain true, it is because the relations preserve their reality [more properly: we still think of them as real]. They teach us now, as they did then, that there is such and such a relation between this thing and that; only, the something which we then called motion [of the particles of the ether], we now call electric current [really: displacement current]. But these are merely names of the images which we substituted for the real objects which Nature will hide for ever from our eyes. The true relations between these real objects are the only reality we can attain ...[11].

Poincaré is claiming, in other words, that if we were to assume for the moment that Maxwell's theory of light is true, then, although we certainly could not continue to hold that Fresnel's theory was also *true*, we can continue to hold that it has correctly identified that part of the 'deep structure' of the universe that governs optical effects—because of the *retention within* Maxwell of Fresnel's mathematical equations governing those optical effects.

To take a straightforward example, Fresnel's theory entails that, any polarised light beam can always be regarded as the superposition of two such beams polarised in orthogonal planes and that when any light beam in air is incident on, say, a plate of glass at angle i some part of that beam will be reflected back into the air at that same angle, while the rest of it will be refracted into the glass at an angle r. His theory moreover entails the exact relative intensities of the reflected and refracted beams:

Letting I^2, R^2, X^2 be the intensities of the components polarised in the plane of reflection of the incident, reflected and refracted beams respectively and I'^2, R'^2, X'^2 the intensities of the components polarised at right angles to the plane of reflection of the incident, reflected and refracted beams respectively, then Fresnel's equations state that these variables will always be related by

[11] Op. cit. note 9, 161.

$$R/I = \tan(i-r)/\tan(i+r)$$
$$R'/I' = \sin(i-r)/\sin(i+r)$$
$$X/I = (2\sin r.\cos i)/(\sin(i+r)\cos(i-r))$$
$$X'/I' = 2\sin r.\cos i/\sin(i+r)$$

where i remember is the angle at which the light is incident on the glass (and therefore also reflected from it) while r is the angle at which the light is refracted into the glass.

These equations are retained entirely intact within Maxwell's theory. Of course, the latter theory radically 'reinterprets' the variables. In Fresnel's theory, the I, R, X, I', R' and X', which are the square roots of the intensities of the various beams, measure the maximum distance by which a particle of the elastic ether is displaced from its position of equilibrium by the passage of the wave. In Maxwell's theory (in its 'mature' form) there is no such medium and those variables instead measure forced variations in the electromagnetic field strengths. From the vantage point of Maxwell's theory, Fresnel was as wrong as he could be about *what* waves are (particles subject to elastic restoring forces and electromagnetic field strengths really do have *nothing* in common beyond the fact that they oscillate according to the same equations), but the retention of his equations (together of course with the fact that the terms of those equations continue to relate to the phenomena in the same way) shows that, from that vantage point, Fresnel's theory was none the less *structurally correct*: it is correct that optical effects depend on *something or other* that oscillates at right angles to the direction of transmission of the light, where the form of that dependence is given by the above and other equations within the theory.

The vantage point afforded by Maxwell's theory is, however, not—and almost needless to say—the *ultimate* vantage point. As Poincaré was writing, the photon theory of light was becoming generally accepted, again yielding a materially quite different view of the ultimate constitution of light than that given either by Maxwell's or by Fresnel's theory. None the less just as Fresnel's mathematical equations had been retained within Maxwell's theory, so the mathematics of Maxwell's theory was again retained (or quasi-retained, courtesy of the correspondence principle—see *below*, 142–144) in the newer photon theory. But this latest theory of light too will no doubt eventually be replaced in its turn ('pessimistic induction' about content)—though, if the history of science is any guide, the structure of that theory will be retained in later theories ('optimistic induction' about structure). As Poincaré puts it, then, the various things that science, at various stages, might be thought of as telling us light *is* are 'merely names of the images which we substituted for the real objects which Nature will hide for ever from our eyes. The true relations between these real objects are the only [persisting] reality we can attain ...'.

This is why I argued in earlier work[12] that Poincaré's position—SSR—is very attractive: in a nutshell, it retains the realism suggested by the NMA (there must be *something* correct about the theoretical claims made by the theory about the 'noumena'—the theory must surely be more than 'empirically adequate') but does not assert a stronger version of realism than seems reasonable in view of the history of theory-change in science (that is, it responds adequately to the 'pessimistic induction').

2. Criticisms of SSR

SSR has, however, itself been subject to a number of criticisms in the recent literature. The main aim of the current paper is to outline (and in the next and final section respond to) just three such criticisms.

(2a) The Fresnel-Maxwell case is maximally atypical

The case that I used, following Poincaré, to motivate SSR (namely the shift considered above from Fresnel's classical elastic solid theory of light to Maxwell's theory of light as a disturbance in the electromagnetic field) is not representative of theory-shifts in the history of science in general. This has been pointed out by Colin Howson amongst others[13]—though I actually explicitly pre-conceded the point.[14] Indeed the Fresnel-Maxwell shift is so far from being representative as to be unique—or so it seems: certainly I know of, and no one else has ever cited, a 'scientific revolution' in which the mathematical equations of the earlier theory are retained *entirely intact* within the 'revolutionary' new theory, as Fresnel's equations are within Maxwell's theory. It would seem that a view of the epistemic credentials of scientific theories that claims to

[12] Op. cit. note 1.

[13] C. Howson, *Hume's Problem*, (Oxford: Oxford University Press, 2000), 39.

[14] Op. cit. note 1, 160.

respond to the facts about theory-change in science, but in fact responds only to one single such change, is not exactly on solid ground.

(2b) The NMA is invalid and hence the realist ingredient of structural realism is without justification

As indicated in section *1*, SSR sees the NMA (or, at least, the intuitions underlying that argument) as the main basis for being a realist about our successful theories and it then—at least apparently—qualifies that realism in the attempt to pay due regard to the facts about theory-change in science. If, then, it can be shown that the NMA can bear no weight at all—that it is a thoroughly bad argument—SSR would seem to be in obvious trouble. But Colin Howson has argued that the NMA is indeed an entirely worthless argument.[15]

Certainly if 'the' NMA were an attempt to infer deductively the truth (or approximate truth) of a theory T from its predictive success with some surprising piece of evidence *e*, it would be in obvious trouble—since it would amount in effect to a version of the fallacy of affirming the consequent. Clearly it is logically possible that a 'very' false theory could none the less happen to entail some surprising result that turns out to be correct. Indeed, if we are not too demanding about what is involved in 'predicting' a piece of evidence *e* and just take it that it is good enough if T entails *e*, then, as Howson points out, it is trivially easy to produce counterexamples to this deductive version of the argument: grueified constructions will suffice, or, in the case of mathematically formulated theories, Jeffreys-style constructions.

Suppose, to take the latter, sharper case, our theory T links two variables and is of the simple form $y = f(x)$; it predicts that when x takes the value x_0, y will take the value $f(x_0) = y_0$; while when $x = x_1$, it predicts $y = f(x_1) = y_1$; these predictions turn out to be correct when observations are made (and suppose moreover that it is somehow surprising from the point of view of 'background knowledge' that (x_0,y_0) and (x_1,y_1) are genuine data points). Can we then infer that it would be a 'miracle' if T were to get this evidence correct if it were not itself true and hence in turn that T is indeed true? Jeffreys pointed out that there are indefinitely many alternatives T' that share this predictive success (at least in the

[15] Op. cit. note 13, chapter 3.

sense that they equally well entail the data points (x_0,y_0) and (x_1,y_1)): just take T' as $y = f(x) + (x - x_0)(x - x_1)g(x)$ for *any* non-zero function $g(x)$.[16] It clearly would be strange to claim that it would be a 'miracle' if T was successful with (x_0,y_0) and (x_1,y_1) and yet was false, if, as this construction appears to show, there are infinitely many alternatives T', all of which equally entail that data and all of which equally entail the falsity of T.

Even intuitively it seems that what we would want to infer from some predictive success like the 'white spot' is not that it is *impossible* that the theory that enjoyed this success is radically false, but that it seems *extremely implausible* that it would be. A seemingly much more promising line for a formal construal of the NMA is, then, to take it to be a probabilistic argument—leading to the claim, *not* that T *is* (approximately) true, but only that it is *probably* (approximately) true.

In order to investigate this suggestion, let's first lay aside the tricky issues about approximation and operate as if our aimed-for conclusion is that some predictively successful theory T is probably true (as opposed to 'probably approximately true'). The rather nebulous talk about it being a miracle if T had got such a phenomenon as e right if it were not true seems then to translate crisply into the assertion that the probability that e would happen were T false is extremely small: $P(e/\neg T) \approx 0$. While the fact that the truth of T explains *e* can plausibly be seen as 'translating' into the claim that $P(e/T)=1$.[17] Hence the most straightforward 'translation' of the NMA into probabilistic terms seems to be:

Pr 1'. $P(e/T) =1$ (e is entailed by T).
Pr 2'. $P(e/\neg T) \approx 0$ (it would be a 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 these probabilistic formulas mean; but laying these aside too for our purposes, it is easy to show, as Colin Howson again emphasises, that so long as they are indeed probabilities (that is, so long as they satisfy the formal probability calculus), then this reasoning is straightforwardly fallacious.

[16] For more details and references see op. cit. note 13, 40–44.

[17] In fact we would surely want something stronger than this probabilistic condition if we are fully to capture the *explanation* claim—not just that e is entailed by T but that T (and perhaps the 'way' in which it entails e) have some further 'nice' properties. See below 144–147.

Here is a simple, and by now well-known, counter example of the kind cited by Howson. Suppose that we have a diagnostic test for some disease D, and that this test (unfeasibly) has a zero rate of 'false negatives': that is, the probability of testing negative if you do have the disease is equal to 0; and moreover an (again unfeasibly) low 'false positive' rate: of 1 in a 1000, say. Suppose now that some particular person x has tested positive, what is the chance that she actually has the disease? In order to avoid changing terminology later, let T stand for the theory that a given person x has the disease, while e stands for the evidential statement that x has produced a positive result in the diagnostic test at issue. The null false negative rate is then 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—that x has D, given that she has tested positive—is $P(T/e)$.

It is often claimed to be an empirical result about human psychology that most people in these circumstances are inclined, given 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[18]. Such people would seem to be reasoning in perfect agreement with our latest version of the NMA:

Pr 1' holds in the diagnostic case because x is certain to test positive (e) if she has the disease (T) (i.e. $P(e/T)= 1$);
Pr 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 that the probability of x having the disease in view of the positive result—that is, $P(T/e)$—is very high.

Yet, as aficionados are well aware, this inference about the diagnostic test instantiates the famous 'base rate fallacy'. *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 'very high' is arbitrarily close to zero. It all depends, of course, on the *prior probability* of T—the fallacy is to ignore this prior or 'base rate'

In the diagnostic case we can, it seems, reasonably take the prior to be the overall incidence of the disease. If the disease is very rare,

[18] See for example D. Kahnemann and A. Tversky, 'Subjective Probability: A Judgement of Representativeness', *Cognitive Psychology* **3**, No. 3 (July 1972), 430–454.

a lot rarer than the rate of false positives, then the probability that x has the disease may be very low despite her positive test. So, for example, if only 1 in a million people on average have the disease, that is, P(T) = 10^{-6} then the probability that x has the disease, given that she tested positive, is only 10^{-3}.

This is a straightforward consequence of Bayes's 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. Think of drawing balls at random from an urn with 1000000 balls, just one of them red (reflecting the fact that only 1 in 10^6 have disease D) and all the rest white (no disease). Each ball also has either a '+' or a '-' marked on it (corresponding to obtaining either a positive or a negative in the diagnostic test). Given that the test yields 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 a 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. So 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. And yet something has happened, namely the patient 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. Pr 1' and Pr 2' both hold here, then, and yet the conclusion is (very) false. This is a clear-cut counterexample to the probabilistic version of the NMA we are considering and shows in fact that if the initial probability that some theory T is true is sufficiently low, then we can perfectly well have evidence that would be 'miraculous' were T false (probability only 1 in a thousand), and yet the probability that T is indeed false is not only not negligible but is in fact close to 1.

So far, we have taken this probabilistic argument as aiming to establish the truth of some theory T, whereas the sensible realist, already noted, wants to argue only for *approximate* truth. The prospects for producing a non-fallacious version of the NMA along these lines are, however, surely not improved by reintroducing considerations of approximate truth. As we saw, no sensible realist will want to claim anything stronger than that our current theories are approximately true, no matter how 'astounding' their predictive success. But modifying the claim in this way is not likely to help here. 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. We can take it that, the relevant auxiliaries being presumed as given, T logically entails e, while whether or not A(T) entails e is unclear. Nonetheless, since the aim of the NMA is for the success with e to have a major impact on the intuitive credibility of T—here reflected in an increase in its probability—presumably its proponents will need to claim that $P(e/A(T)) \approx 1$. Again, the NMA relies on the idea that the evidence at issue would be very improbable were T not even approximately true, so realists developing this form of the argument would presumably be committed to the premise $P(e/\neg A(T)) \approx 0$. Hence we have a simple modification of the probabilistic argument

Pr 1'' $P(e/A(T)) \approx 1$
Pr 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 then clearly the base rate problem kicks in just as before: depending on the value of the prior probability that T is approximately true, any posterior for T's approximate truth—including a posterior as close to zero as you like—is compatible with the truth of premises Pr 1'' and 2''.

It seems, then, that 'the' NMA is in trouble and hence, since the realist element of SSR is based squarely on it, so is SSR.

2(c) The 'Newman argument' 'destroys' SSR

These two arguments seem bad enough news for SSR, but the argument that seems to have convinced most philosophers of science of the untenability of SSR is still a third one. This goes back to a paper of 1928 by the Cambridge mathematician Max Newman, responding to Bertrand Russell's version of structural realism.[19] Newman's argument was brought back to the attention of philosophers of science via a 1985 article by Demopoulos and Friedman[20]. The argument in its crispest form goes as follows:

[19] M.H.A. Newman, 'Mr. Russell's Causal Theory of Perception', *Mind* **37**, No. 146 (April 1928), 137–148.
[20] W. Demopoulos and M. Friedman, 'Critical Notice: Bertrand Russell's *The Analysis of Matter*: Its Historical Context and Contemporary Interest', *Philosophy of Science* **52**, No. 4 (December 1985), 621–639.

1. SSR is committed to the view that the Ramsey sentence of any scientific theory T captures the full 'cognitive content' of that theory.
2. However, as Newman showed, the Ramsey sentence of any theory imposes only a very weak constraint on the universe—it amounts in essence to a mere cardinality constraint, and so if there are sufficiently many objects in the universe then the Ramsey-version of T, for any T, will be true.
3. However it is clear that standard scientific theories impose much more stringent constraints on the universe if they are to be true than merely a constraint on the minimum number of entities the world must include.
4. Hence SSR is committed to an account of the cognitive content of scientific theories that is plainly untenable and is, therefore, itself untenable.

No wonder then that the Routledge Encyclopaedia article on Russell refers to Newman's argument as the 'definitive refutation' of his structural realism:

> Newman's argument is the definitive refutation of the Structural Realism of Russell (1927) ... Russell quickly abandoned SR when Newman showed that any set with the right cardinality could be arranged so as to have the same structure as the world—a result analogous to that claimed in Putnam's model-theoretic argument against realist theories of reference (Demopoulos and Friedman 1989).

Of course this leaves it open that there is something especially faulty with Russell's version of SSR, but in all respects relevant to the current discussion this is not true (or at any rate I do not believe it to be true). If so, then advocates of SSR such as myself and Elie Zahar[21] seem to have shown reprehensible ignorance of the literature in advocating a position that had already been conclusively demolished.

3. Responses to Criticism: why reports of the death of SSR may be exaggerated

Is SSR in straits as dire as the above three criticisms seem to suggest? I consider the criticisms in turn.

[21] E. Zahar *Poincare's Philosophy: From Conventionalism to Phenomenology*, (Chicago: Open Court, 2001).

3(a) The atypical nature of the Fresnel-Maxwell shift

There is no denying that the theory-shift in optics from Fresnel to Maxwell is unrepresentative. Indeed, as indicated, I already emphasised this point in developing my defence of SSR. In all other cases, the best that can be argued is that, once a science has reached maturity,[22] the mathematics of any theory replaced in a 'scientific revolution', while not being retained fully intact, is instead 'quasi-retained' *modulo* the 'correspondence principle'.

The most straightforward cases of the application of this principle are where the equations of the older theory reappear as limiting cases of the equations of the newer theory (and moreover the limiting cases characterise the area in which the older theory had proved entirely empirically successful). A classic case is of course represented by the relationship between the Special Theory of Relativity and Newtonian physics—the Newtonian equations being recovered from the Einsteinian ones as v/c tends to 0 (where v/c tends to 0 as a body's velocity is ever smaller compared to the velocity of light; and where Newtonian and Einsteinian predictions—though always strictly different—are entirely empirically indistinguishable for relatively slowly moving objects).

As Michael Redhead has pointed out[23], not at all applications of the correspondence principle fit this pattern (one example he cites is the transition from geometrical to wave optics, though I am unsure whether geometrical optics can count as a 'mature', that is, in my terms, genuinely predictive theory). Others may well feel that the 'continuity' afforded by the correspondence principle *in general* is hardly worthy of the name. And hence that any 'realism' founded on it is, in turn, hardly worthy of the name.

[22] Larry Laudan complains (op. cit, note 6) that the notion of 'maturity' is introduced by realists as an ad hoc device: whenever it seems like there is no sense in which an earlier theory continues to look 'approximately true' in the light of its successor, the realist can claim that that earlier theory was accepted only when the science that it contributes to was 'immature'. However, as I have explained before (op cit. note 1), it seems that the realist should be ready to 'read off' her notion of maturity from the NMA which is her main support—taking it that a scientific field attains maturity once its accepted theory enjoys genuine predictive success (that is, it predicts some general phenomenon that was either unknown at the time or was not used in the development of the theory concerned).

[23] See, e.g., his 'The Unseen World' in C. Cheyne and J. Worrall (eds.) *Rationality and Reality: Conversations with Alan Musgrave*, (Springer, 2007).

It is surely true that the strength and character of scientific realism, of what the reasonable attitude is toward the purely theoretical claims of current theories, depends on how strong a notion of continuity can be extracted from the history of theory-change in science. And that's the way round it clearly has to be: science, and the history of theory-change within in, strongly constrains the reasonable philosophical view. If the notion of structural continuity via the correspondence principle is not strong enough for your tastes then you will not be happy with calling SSR a 'true' version of realism. But this seems to be a merely semantic issue. The extra ingredient that SSR adds to a van Fraassen-style empirical adequacy view may not be very strong but it *is* an extra ingredient and it comes at no real price. According to the account of theory change that underpins SSR, successive theories in science have not only been successively more empirically adequate, but there has always been a *reason*, when viewed from the vantage point of the later theory, *why* the earlier theory achieved the degree of empirical adequacy that it did—namely that the earlier theory continues to look approximately structurally correct: its mathematical equations are retained *modulo* the correspondence principle.

Given this underpinning, then SSR is just a simple and surely innocuous inductive step away: it seems reasonable to believe that currently accepted theories, if they are replaced at all (as seems highly likely), will be replaced by theories in the light of which they will continue to look approximately structurally correct. If, as seems right, we count as a version of scientific realism any view that asserts that it is reasonable to hold that our successful theories are more than simply highly empirically 'adequate'—they are empirically adequate because they (can reasonably be taken to) 'latch on' in some way to the 'deep structure' of the universe, then SSR counts as realism. It goes on to insist that the way that our theories thus 'latch on' to the 'deep structure' of the universe cannot be further specified—to suppose that it can would be to suppose that we can somehow have access to the universe that is not theory-mediated and thus can directly compare what our theories say with reality. But once articulated this supposition is clearly untenable. This will be disappointing for some. But, as I shall argue in the sub-section *3(c)*, no account of how that relationship might be further specified makes any real sense. If we are talking about coherent, defensible positions, then SSR is as realist as it gets.

In sum, to count as a fully fledged version of realism any such view must say something of a realist kind about fundamental, frontier theories, but any version of realism about fundamental, frontier theories is dependent for its plausibility on the production of an account of "continuity through revolution" in the history of science. No one should claim a stronger sense of continuity, and hence a stronger version of realism, than is compatible with the historical record. We should look for the strongest such version and see if it is a continuity worth having. If there is no such notion of continuity worth having, then there is no sustainable version of realism. However, I hold that there is a continuity (admittedly of an approximate kind) at the structural level that is substantial enough to count and hence I hold that SSR is a sustainable version of scientific realism, and indeed, as I shall try to show again later, the only sustainable version.

3(b) What should we expect from 'the' NMA?

Representations of 'the' NMA as either an attempted deduction or as an unadorned probabilistic argument are, as we saw in section *2(b)*, undeniably and straightforwardly invalid. The pro-realist intuitions elicited by particular cases of predictive success remain strong however. This suggests that there may be some other way of developing the argument that is *not* fallacious.

One standard way of running the argument (unacknowledged by Howson) is as an instance of 'inference to the best explanation'. The claim is that the (approximate) truth of the theory T is the 'best explanation' of T's success in predicting e (where, moreover, there is an implicit assumption that T does more than simply entail e, it also 'explains' it). This means that gruesome or Jeffreys-style constructions are doubly barred from counting as counterexamples. First of all, someone construing the argument in this way would allege that, although Jeffreys-style constructions clearly deductively entail the evidence at issue (in our example: the data points (x_0,y_0) and (x_1,y_1)), they do not *explain* those data points. And secondly, and of course relatedly, while it *might* be argued that the 'best explanation' of the original theory T's success with e is that it is (approximately) true (this will depend on the details of the particular case), the best explanation of the Jeffreys-style alternatives 'success' with e is that they were constructed exactly so as to entail it. Hence, although there are undoubtedly infinitely many rivals to T (in the precise sense that they entail that T is false) that equally well entail e, this does not mean that there are infinitely many rivals that equally well 'explain' e and hence that equally well generate the 'no miracles intuition' (nor does it even mean that there is *one* rival that does so).

Although this suggestion seems to me on the right lines, talk of 'inference to the best explanation' gives it an air of precision and formality that the argument scarcely deserves. The fact is that no one has much, if any, idea about how to articulate the extra requirements on a theory, beyond actually entailing some evidence e, for it to *explain* e. A high degree of 'unity' is often cited here, but, although I believe that this is indeed the crucial notion, and although it is easy to point to theories that clearly possess such unity and at others that clearly do not, no one has succeeded in giving a general characterisation of the notion. It would appear easier to characterise what a *dis*unified theory is: essentially one that has been produced from an earlier theory via *ad hoc* modifications designed to remedy defects in that earlier version (standardly empirical anomalies). But this in effect presupposes that the initial theory was unified and again while it is easy to exemplify the idea, it has not proved possible to articulate it in general terms.

What this attempted formulation indicates, however, is that there is indeed an extra implicit assumption in particular applications of the NMA beyond the fact that the theory concerned gets some piece of evidence correct. It would be a 'miracle' if a theory that got some striking piece of evidence correct *and was a theory of the 'right sort'* was nonetheless radically off-beam. Again Poincaré was fully aware of the situation:

> We have verified a simple [sic] 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 [sic] 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.[24]

So the success of Kepler's law in entailing the phenomena is only one feature of the situation for Poincaré: it is also important that

[24] Op. cit. note 9, 149–150.

the law is *simple*. After all, Tycho's data (as Poincaré was surely implicitly aware) fits not only the Keplerian ellipse but also indefinitely many Jeffreys-style versions thereof.

One significant feature to notice is, then, that Poincaré is pointing out that something like the NMA is needed to underwrite what are generally taken to be straightforward inductions to conclusions that are often thought of as 'empirical generalisations'.

As for the NMA in general, he is acknowledging that there is an implicit, and perhaps vague, assumption about simplicity or prior plausibility. It is unlikely that something as simple as the ellipse would fit in all the many cases we have observed (and fail to fit in no observed cases) without being generally true. (It is this that makes van Fraassen's position in *The Scientific Image* so difficult to empathise with. By suggesting that we do indeed accept our best current theories as 'empirically adequate'—where this quite explicitly means empirically adequate *across the board* and not just with regard to empirical results that have already been checked observationally—he is implicitly appealing to the NMA. But why then reject the idea that that same argument has force when it concerns 'theories'?[25])

Poincaré's appeal to simplicity could, of course, be captured within a Bayesian account. And indeed Howson points out (op. cit. note 13) that, although the probabilistic version of the NMA is invalid as it stands, it is easy to make it (probabilistically) valid if, far from ignoring the base rate, one asserts, as an extra premise over and above the claim about the impact on the theory T of evidence e, that the prior probability of T is at least reasonably high.

It is not that this Bayesian rendition seems to me incorrect in any way, but rather (and as usual) it seems to supply no extra 'explanatory force'. The assumption that, in cases where the NMA-intuitions kick in most strongly, the 'prior' of the theory concerned is at least reasonably high seems just a—not especially perspicuous—way of reflecting our intuitive judgment about the unity or simplicity of that theory; and not to add anything to this intuitive judgment. (Indeed the intuitions here seem stronger: they characterise, in a way that (personalist) Bayesianism declares impossible, those cases in which it is, and those in which it is not, *reasonable* to assume a reasonably high prior. It is not a subjective

[25] See B. van Fraassen, *The Scientific Image*, (Oxford: Clarendon Press, 1980).

matter that a theory claiming that planets move in ellipses is simpler than one that claims that their orbits are some Jeffreys-style monstrosity.)

All in all, one should not, it seems to me, expect too much from 'the' NMA. Neither of the ways of running it *apparently* more formally while avoiding patent invalidity—as an 'inference to the best explanation' or as a Bayesian inference with a substantial prior as extra implicit assumption—adds anything of real substance to the underlying intuition. The 'argument' ('consideration' might be more honest) should, so it seems to me, be thought of as doing little more than setting the default position: given the, occasionally staggering, predictive success of our current (and some of our earlier) theories (and given that they have achieved this predictive success without any necessity for *ad hoc* modifications), the reasonable default assumption is that they have latched on in some way to the 'deep structure' of the universe.

There is no question of logical compulsion here—just the suggestion that it seems more reasonable than not that theories that enjoy this sort of success must be 'on the right lines'. Moreover, like all default positions, this one is obviously defeasible; and it would indeed be straightforwardly defeated, if Laudan were right that there is no real continuity (even of a somewhat attenuated kind) between successive theories in 'mature' (i.e. predictively successful) science. The pessimistic induction, if accepted, would trump the NMA. This is because the history of science would then, as already noted, provide a list of alleged 'miracles'—theories that enjoyed striking predictive success but which are not even approximately true by the lights of current science. And, when it comes to 'miracles', familiarity surely does breed contempt. This is why it is important for a defensible realism to establish a way in which successive theories in mature science have indeed been at least quasi-accumulative. And I claim that only the structuralist can successfully establish such an account.

3(c) The 'Newman argument': is SSR really realism?

The Newman argument is in effect that no defensible view of scientific theories can be based on the claim that the full cognitive content of such a theory is captured by its Ramsey sentence—because that sentence imposes a mere cardinality constraint on the universe. There is no doubt that SSR *is* committed to the claim that a theory's full cognitive content is captured by its Ramsey sentence.

The issue then clearly is whether Newman indeed establishes that the Ramsey sentence of a theory is as weak as he seems to claim.

The Ramsey sentence of a theory T, as is well-known, is constructed by replacing all the *theoretical* predicates in T by second-order variables and then existentially quantifying over those variables.[26] The claim that this sentence captures the full cognitive content of T simply reflects the fact that, so far as our fundamental theoretical notions are concerned, we know about them only by description—that is, via their role in our theories. If asked what an electron, say, is (or rather what we *think* an electron is), one can do no more than recite our current (full!) theory of electrons—electrons (if they exist at all) are *whatever it is* that satisfy our current relevant theories. But this means that a theoretical term like 'electron' in effect plays the role of a (second-order) 'ambiguous name'; and, as is well-known, in systems of logic that employ ambiguous names α_i, $P(\alpha_i)$ is logically equivalent to there is an x, $P(x)$.[27]

In order to deny that the Ramsey sentence of T captured the full cognitive content of T, one would have to assert that we have some independent way of describing how our theoretical terms 'latch on to' or 'hook up with' reality—independent, that is, of our theories themselves. This can only make sense if something like the causal theory of reference is adopted. But whatever its attractions as an account of how we practically manage to communicate via commonly held assumptions about the reference of ordinary names, the causal theory of reference as an account of the reference of primitive theoretical terms is surely patently hopeless. Think about how one might 'ostend' the electromagnetic field, say, in order to 'baptise' it: clearly we could only know that we are ostending the field (in fact you can point in any direction you like!) via the *theory* of the field. It is just a fantasy (given credence by unthinking reflection on orthodox logical semantics) that we can "stand outside" of our theories and directly compare terms in them with a reality that we can access directly without any theory. No one *really* believes this, though many act (and even sometimes write) as if they do. But once this is recognised as the fantasy it is, then there

26 Of course if you presuppose set theory then the theoretical predicates can be replaced by predicates varying over sets and then the Ramsey sentence is entirely first-order.

27 See, for example, P. Suppes, *Introduction to Logic*, (New York: Van Nostrand, 1957).

just is no question but that the Ramsey sentence of T (in its original, rather than Lewisian form[28]) captures the full cognitive content of T.

It would then be rather surprising if Newman's argument really did establish—as so many commentators now seem to believe it does—that the Ramsey sentence of any scientific theory imposes only a cardinality constraint on the universe. (That is, that all that the Ramsey sentence of any theory requires of the universe in order for that sentence to be true is that the universe includes sufficiently many individuals.) Given that the Ramsey sentence just *has to* capture the full cognitive content of a theory, this would mean that all human theorising is doing no more than imposing a constraint on how many individuals there are in the universe and that is absurd, if anything is.

Yet this is exactly what Demopoulos and Friedman, quoting Newman, seem to assert:

> The difficulty is with the claim that *only* structure [as revealed by the Ramsey sentence] is known. On this view "the world consists of objects, forming an aggregate whose structure with regard to a certain relation R is known, say [it has] structure W; but of R nothing is known ... but its existence; ... [A]ll we can say is *There is* a relation R such that the structure of the external world with reference to R is W." (Newman 1928, p. 144). But "*any* collection of things can be organized so as to have the structure W, provided there are the right number of them" (p. 144 italics added). Thus, on this view, only cardinality questions are open to discovery! Every other claim about the world that can be known at all can be known a priori as a logical consequence of the existence of a set of α- many objects. For, any given set A of cardinality α, can with a minimum of set theory or second-order logic establish the existence of a relation having structure W, provided that W is compatible with the cardinality constraint that /A/ = α. (The relevant theorem from set theory or second-order logic is the proposition that every set A determines a full structure, that is, one which contains every relation [in extension] of every arity on A;...)[29].

[28] D. Lewis, 'How to Define Theoretical Terms', *Journal of Philosophy* **67**, No. 13 (July 1970), 427–446. Lewis's argument for reading the quantifiers as 'there is a *unique* Ø such that' seem to me entirely unconvincing.

[29] Op. cit. note 20, 627–628.

The most obvious mistake that Newman makes (and he seems to be followed, at any rate in this passage, by Demopoulos and Friedman) is to assume that in order to identify the structural claims of a theory T one should replace *all* the predicates occurring in it by predicate variables and existentially quantify over them—that is, that all the predicates in the theory should be treated as theoretical.[30] It is surprising that Russell should have conceded to Newman (as Demopoulos and Friedman correctly report that he did), given that his whole structural realist view was based on a sharp distinction between theoretical and observational notions (in his own terms between things known *by acquaintance* and things known only *by description*).

It is true that if *all* predicates are quantified-over, then the resulting Ramsey sentence is hopelessly weak (though the situation, as we'll see, is rather more complicated than Demopoulos and Friedman, following Newman, assert). But once it is recognised that some predicates are to be taken as observational and therefore as interpreted independently of theory and so emphatically are *not* to be quantified-over, then the charge that the Ramsey sentence of any theory imposes a mere cardinality constraint on the universe is easily refuted. It is in fact well-known that, in that case, the original theory T and its Ramsey sentence will be empirically equally powerful: that is, every sentence that is expressible purely in observational terms and is deducible from T is also deducible from its Ramsey sentence. Thus the Ramsey sentence of T is co-refutable with T; clearly then the Ramsey sentence of any theory that has empirical content imposes much stricter constraints on the universe than any mere cardinality constraint.

(If we suppose, as Newman himself seems to have done, that there is no distinction to be made between observational and theoretical predicates in T, then T's Ramsey sentence will indeed fail to capture its full cognitive content—being in effect a purely mathematical statement. However even then Newman's claim requires modification. He suggested that if structuralism were correct, that is the Ramsey sentence was all that could be known, then this would entail that all that can be known 'theoretically' is

[30] This section of the paper follows the treatment in J. Worrall and E. Zahar, 'Appendix IV: Ramseyfication and Structural Realism' in E. Zahar *Poincare's Philosophy: From Conventionalism to Phenomenology*, (Chicago: Open Court, 2001), 236–251. I am in general greatly indebted to Elie Zahar for many long discussions of structural realism and for invaluable help with some technical issues.

"the number of constituting objects"[31]. But in fact only in exceptional cases does the Ramsey sentence of a theory (formed by quantifying over *all* the theory's theoretical predicates) determine the cardinality of domains that satisfy that theory. For example if T is a contradiction then so is its Ramsey sentence; while if T is $\forall x(S(x) \leftrightarrow \neg(x=x))$ then the Ramsey sentence $\exists\Phi\ (\forall x(\ \Phi(x) \leftrightarrow \neg(x=x)))$ is a logical truth, since $\Phi(x)$ can be chosen to the predicate $\neg(x=x)$. In the first case there are no models and in the second case all interpretations are models. In neither case is there of course any restriction on cardinality. Moreover, by the upwards and downwards Löwenheim-Skølem theorems, if T is any theory with an infinite model of any cardinality then it has models of *all* infinite cardinalities. Hence a logically true T's Ramsey sentence will be true in virtue of a range of structures with domains of all infinite cardinalities. It is only in the case of a theory that has only finite models that its Ramsey sentence determines the size of the domain of individuals. So, for example, if our theory were $\forall x\ Sx\ \&\ \exists x_1,x_2,\ldots x_n\ (\neg(x_1 = x_2))\&\ldots.\&\ \neg(x_{n-1}= x_n)\ \&\forall x(Sx\leftrightarrow(x = x_1)$ v $\ldots$v$(x = x_n))$ then its Ramsey sentence would be equivalent to the assertion that the domain of individuals consists of exactly *n* members.

It should be admitted that while Demopoulos and Friedman go along with Newman's untenable claim in the passage just cited, they do elsewhere state:

> ... if our theory is consistent and *if all its purely observational consequences are true*, then the truth of the Ramsey-sentence follows as a theorem of set-theory or second-order logic, provided our initial domain has the right cardinality ...[32].

And perhaps this is their real position. It is, of course, not open to the above objection. The claim now seems to be that although the Ramsey sentence of any theory T says everything that T does about observational matters, it is somehow not strong enough to capture the full cognitive content of T.

It is difficult to know what to make of this claim. Notice that the empirical equivalence of T and its Ramsey sentence extends far beyond what we might normally take that phrase to mean. The Copernican and Ptolemaic theories, for example, are (or can be made) empirically equivalent with respect to observed planetary motions (courtesy of delicate adjustment of epicycles); but any

[31] Op. cit. note 19, 144.
[32] Op. cit. note 20, 635, emphasis added.

theory T and its Ramsey sentence are observationally equivalent in the much stronger sense that *any sentence* that is expressed in the observational vocabulary and is entailed by T is also entailed by its Ramsey sentence. Many sentences expressed in purely observational vocabulary should count, in any one's book, as theoretical: one often cited example is the claim that there are unobservables—that is, individuals possessing no observable property.

Moreover, the Ramsey sentence of any theory is itself observational in this sense—trivially so since all the theoretical predicates in the original theory have been removed and quantified-over. This makes the claim just quoted from Demopoulos and Friedman in one sense trivially true since of course (a) T's Ramsey sentence is provably a logical consequence of T and (b) T's Ramsey sentence, like any other sentence, entails itself. (On the other hand, if what Demopoulos and Friedman mean by 'observational consequences' are singular, decidable sentences, then what they say can straightforwardly be proved to be false using the compactness theorem.[33])

Quantifying over the erstwhile theoretical predicates removes them linguistically, but the theoretical terms surely live on within the Ramsey sentence via the structure that they impose on the observational content. (That is, after all, what structural realism is about.) Indeed, as Elie Zahar pointed out to me, if we follow Quine's dictum that 'to be is to be quantified over' (or, better: 'to be asserted to be is to be quantified over in a sentence that you assert') then the Ramsey sentence of some theory T $(S_1,\dots S_n, O_1,\dots O_r)$, where the S_i are theoretical predicates and the O_j are observational, namely $\exists\Phi_1\dots\exists\Phi_n\ (T(\Phi_1,\dots,\Phi_n\ O_1,\dots O_r)$ clearly asserts that the 'natural kinds' $S_1, \dots, S_n$ (the extensions of the theoretical predicates $S_1, \dots S_n$ in the initial theory T) exist in reality just as realists want to say. It is just that—as always—we fool ourselves if we think that we have any independent grip on what the $S_1, \dots, S_n$ are aside from whatever it is that satisfy the Ramsey sentence (assuming that the theory whose Ramsey sentence we are considering is true).

In the end, then, the much-vaunted 'Newman objection' is no objection at all to SSR. In its original form it relies on a false assumption (that SSR regards all predicates as theoretical); and in its modified Demopoulos and Friedman form it simply highlights a consequence of SSR that is essential to it. The fact that it has been found so troubling depends, I think, on nothing more than a

[33] Op.cit. note 30, 235.

residual vague feeling that if this is what SSR amounts to then it doesn't count as *real* realism. So I shall end by confronting that question directly: is SSR 'really' realism?

The answer is 'no' if we take Putnam's characterisation. This, as is well known, has the realist assert that 1) Theoretical terms in mature science refer; and 2) what our currently accepted theories in mature science say using those terms is at any rate approximately true.[34] That is, a realist says that there definitely are electrons (for example) but then is allowed to concede that what current theories say about them may not be exactly true (though will nonetheless be close). This is how 'referential continuity' is supposed to be restored across theory-change—earlier theorists were talking about the same entity when they talked about 'electrons', even though what they said about them was different from (though in some sense nonetheless approximates to) what our current theories say. But why should a realist not be equally as fallibilist and tentative about the mode of reference of the terms in theories as she is about those theories' truth? In any event, no one seriously holds, as I have remarked several times, that we have any theory-independent access to the furniture of the universe that would allow us to compare (even 'in principle') the notions conjectured by our theories with what there really is.

Scientific Realism of any stripe consists of a metaphysical thesis conjoined with an epistemological one. The metaphysical thesis is often taken as the claim that there exists a reality independent of the human mind. But, as Elie Zahar points out, a more accurate rendition in our more enlightened, post-Cartesian-dualism times would be:

> There exists a structured reality of which the mind is a part; and, far from imposing their own order on things, our mental operations are simply governed by the fixed laws which describe the workings of Nature.[35]

I take this metaphysical thesis as a given. Scientific Realism adds the epistemological thesis:

> Not only is this structured reality partially accessible to human discovery, it is reasonable to believe that the successful theories in mature science—the unified theories that explain the phenomena without *ad hoc* assumptions—have indeed latched

[34] H. Putnam, 'What is Mathematical Truth' in *Mathematics, Matter and Method* (Cambridge: Cambridge University Press, 1975), 69–70.
[35] Op. cit. note 21, 86.

> on, in some no doubt partial and approximate way, to that structured reality, that they are, if you like, *approximately true*.

If it is assumed that to be a 'real realist' one must assert that the terms in our current theories refer as part of an acceptance of a correspondence or semantic view of truth as the account of what it means for our theories to have latched on to the real structure of the world, and it is assumed that the realist must develop some sort of weakened version of correspondence as her account of 'approximate correspondence with reality' then SR does not count as 'real realism'.

But there is no reason why the way in which a theory mirrors reality should be the usual term-by-term mapping described by traditional semantics. Indeed, as I have remarked several times already, if we are talking about an epistemically accessible notion then it cannot be! SSR in fact takes it that the mathematical structure of a theory may globally reflect reality without each of its components necessarily referring to a separate item of that reality. And it takes it that the indication that the theory does reflect reality is exactly the sort of predictive success from unified theories that motivates the No Miracles Argument.[36]

SSR may well be more modest than many who have sought to defend some version or other of scientific realism might like. But the modesty involved in SSR is far from undue. No stronger version of scientific realism is either compatible with the facts about theory-change in science or compatible with any truly defensible epistemological view of how our best theories are likely to 'link up with reality'. If SSR isn't realism then nothing defensible is. But SSR is, I suggest, in fact a modest but defensible version of scientific realism—reports of its death have been greatly exaggerated!

[36] My account in this section of the paper is particularly indebted to a number of conversations with Elie Zahar.