Laudan, L.: (1997), 'How About Bust? Factoring Explanatory Power Back into Theory Evaluation', *Philosophy of Science* **64**, pp. 303–316.

Lobo, J.: (1996), 'Sources of Gravitational Waves', in G.S. Hall, J.R. Pulham (1996), pp. 203–222.

Mayo, D.: (1996), *Error and the Growth of Experimental Knowledge*, The University of Chicago Press, Chicago.

Mayo, D.: (1997a), 'Duhem's Problem, the Bayesian Way, and Error Statistics, or 'What's Belief Got to Do With It?' ' and 'Response to Howson and Laudan', *Philosophy of Science* **64**, pp. 222–244, 323–333.

Mayo, D.: (1997b), 'Severe Tests, Arguing from Error, and Methodological Underdetermination', *Philosophical Studies* **86**, pp. 243–266.

Mayo, D.: (2000), 'Experimental Practice and an Error-Statistical Account of Evidence', *Philosophy of Science* **67** (Part II Symposia Papers, ed. by D. Howard), pp. S193–S207.

Mayo, D., Kruse, M.: (2001), 'Principles of Inference and Their Consequences', in D. Corfield and J. Williamson (eds.) *Foundations of Bayesianism*, Kluwer Academic Publishers, Dordrecht, pp. 381–403.

Mayo, D., Spanos, A.: forthcoming, 'A Post-data Interpetation of Neyman-Pearson Methods Based on a Conception of Severe Testing'.

Suppes, P.: (1969), 'Models of Data', in P. Suppes (1969), *Studies in the Methodology and Foundations of Science*, D. Reidel, Dordrecht, pp. 24–35.

Taylor, J., Wolszczan, A., Damour, T. and Weisberg, J.: (1992), 'Experimental constraints on strong-field relativistic gravity', *Nature* **355**, pp. 132–136.

Will, W.C.: (1980), 'The Confrontation Between General Relativity and Experiment, in J. Ehlers, J. Perry, and W. Walker (1980), pp. 307–321.

Will, W.C.: (1986), *Was Einstein Right?*, Basic Books, New York.

Will, C.W.: (1993), *Theory and Experiment in Gravitational Physics*, Cambridge University Press, Cambridge.

Will, C.W.: (1996), 'The Confrontation Between General Relativity and Experiment. A 1995 Update', in G.S. Hall, J.R. Pulham (1996), pp. 239–281.

NEW EVIDENCE FOR OLD

John Worrall

London School of Economics

Abstract If the development of science is, *pace* Kuhn and Co, to be explained as a rational process on essentially empiricist criteria, then there must be some sort of justified epistemic confirmatory premium on predictions of certain kinds, as opposed to "mere accommodations". Only then could the history of mature sciences be explained as consisting of the replacement of well-confirmed theories by even better confirmed theories. There is, of course, a long-running dispute in philosophy of science over exactly this issue. This dispute formed an important part of the debate between Mill, who believed that only "the ignorant vulgar" could think that predictions "count more", and Whewell who claimed that predictive success gave the theory that enjoyed it "the stamp of truth beyond the power of ingenuity to counterfeit". And the issue has surfaced in various forms since then – most recently in the much-discussed "old evidence problem" that allegedly afflicts the Bayesian view of confirmation.

I have argued in earlier papers that there is, in fact, no confirmatory premium for new evidence *as such*: It is true that a crucial distinction needs to be made in confirmation theory between two types of evidence; and true that the distinction coincides *to some extent* with the old/new distinction; however the important difference is not that between new and old evidence but between evidence that has not, and evidence that has, been "used in the construction" of the theory whose confirmation is at issue.

This claim has itself been criticised on various grounds, by Bayesians (like Colin Howson) and non-Bayesians (like Deborah Mayo) alike. Mayo proposes a different view aimed at capturing the same intuitions that underlie my earlier account; while Howson dismisses as "entirely bogus" the whole claim that using evidence in the construction of a theory diminishes that evidence's power to confirm that theory.

In this paper, I shall, show that both Howson's and Mayo's analyses are flawed and hence defend and clarify my original claim. This defence puts the rationale for the claim on a much clearer and firmer footing than hitherto; and thereby, I believe, significantly clarifies this whole important and longstanding debate.

P. Gärdenfors, J. Woleński and K. Kijania-Placek (eds.), In the Scope of Logic, Methodology and Philosophy of Science, Vol. I, 191-209.

1.

The fact that there continues to be a lively debate in the philosophy of science literature about "novel facts" would surely strike the disinterested observer as astounding. After all, the idea that a theory is in general better supported by hitherto unknown evidence that it correctly predicts than it is by already known evidence that it entails seems beyond the pale of credibility. Why should anyone ever have defended the outlandish idea that the time-order of theory and evidence is of epistemological-methodological importance? Let me start then by reminding you why some perfectly sensible people defended this apparently senseless idea.

The clue lies in the surprising complexity of the deductive structure of empirical tests of scientific theories. As many commentators had clearly seen intuitively, but as Duhem was the first to spell out in some detail, the sort of claim that we tend to think of as a "single" scientific theory – Newton's theory, the classical wave theory of light, Maxwell's theory of electromagnetism – has no empirical consequences when considered "in isolation". In order to have a full deduction of a directly empirically testable statement from such a theory, other theoretical assumptions and hypotheses are invariably necessary as extra premises. In order to test Newton's theory through observations of apparent planetary positions, for example, particular – though clearly theoretical – claims about the positions and masses of the main objects in the solar system are required, along with a set of optical assumptions underpinning the working of the telescopes involved and allowing for such hypothesised phenomena as atmospheric refraction. Call these extra assumptions "auxiliaries".

Duhem also in effect pointed out that for most "single" scientific theories, there is a further level of complexity – the "single" theory itself is naturally thought of as consisting of a *core* claim together with some set of *more specific* theories. Consider, for example, the classic case of the two theories vying for acceptance in early 19th century optics. "The" wave theory was made up of the core claim that light consists of some sort of periodic disturbance transmitted through some sort of elastic material medium, together with a set of more specific assumptions (ones which evolved through time) about the type of periodic disturbance and the exact mechanical constitution of the medium. Similarly "the" emission theory consisted of the core claim that light consists of some sort of particle fired off by luminous bodies, together with a set of more specific assumptions (ones which again evolved through time) about the type of material particles involved (what, for example, differentiates those particles that produce red light from those that produce blue light) and the forces to which they are subjected in various circumstances (when reflected from a plane mirror, refracted into glass, or "diffracted" when passing by the edge of opaque objects).

So the full deductive structure of an empirical test is always

a. CENTRAL THEORY

b. AUXILIARY THEORIES

Hence, empirically checkable consequence.

Here the set b is, despite some misleading remarks by Quine and even by Duhem himself, *of course* finite, though invariably large (larger than you might initially think); while a – at least in many cases – itself breaks down naturally into a core component and a set of more specific claims. Call the conjunction of a and b a "theoretical system based on the core idea in the central theory a."

But now let e be some piece of evidence – the result of Young's two slit experiment, say – that seems intuitively to tell against a particular "isolated" theory, in this case the emission theory of light. The above analysis straightforwardly entails that there must be theoretical systems based on the core emission idea that entail that piece of evidence e. And indeed optical scientists of the early 19th century considered such systems – *either* attributing the pattern of dark and light bands to a physiological effect (thus questioning the original auxiliary assumptions underlying the instruments used – in this case our own eyes) *or* attributing those bands to particular (and peculiar) forces that emanate from the double-slit screen and affect the paths of the light-particles (thus questioning one of the original more specific assumptions involved in the central theory).

Since the above is true for *any* piece of evidence and *any* "isolated" theory, it threatens to rule out any account of the rationality of theory-change in science in terms of the winning theory having support from more evidence than the losing theory. *If* having the right empirical consequences is the only criterion, then since any core idea can be incorporated into a theoretical system that has the right empirical consequences, there is no empirically-based rational preference for any such core idea over any other.

2.

So this was what made the initially implausible idea that predictions "count more" appealing: if true it would solve this problem in a crisp and clear way, and hence neutralise the threat to an empiricist account of rational theory-change. "Novel facts count more" would imply that, since interference effects had been predicted by the wave theory, those effects continued to count more for that theory even once they had been accommodated, after the event, by the rival emission theory.[1]

There are, however, at least two problems. One is that although the novelty claim might capture well-established intuitive judgments of support in certain particular cases from the history of science, it clearly contradicts well-established intuitive judgments of support in other cases. One such much-discussed case concerns Einstein's account of the precession of Mercury's perihelion (compared to his prediction of the star-shift data)[2]; another (as I showed elsewhere)[3] concerns Fresnel's 1819 wave account of the straightedge diffraction results first produced by Grimaldi in 1665; and yet another concerns the account by Copernicus's theory of the long known phenomena of planetary stations and retrogressions and of the bounded elongation of Mercury and Venus.[4] In all these cases, the fact that a new theory entailed correctly some already well-known phenomenon counted strongly in its support – just as strongly as any hitherto unknown phenomenon whose existence the corresponding theory predicted.

The second problem with the "novel facts count more" view is that despite its appeal as a solution of some cases of the Duhem problem (and even if, contrary to fact as we just saw, it agreed with the intuitive solution in all particular cases), the idea just remains outlandish. Why on earth *should* the apparently purely contingent historical issue of whether or not a theory was first developed before some particular piece of evidence became available matter at all in an account of the *rational* support that that evidence lends to theory? (Let me say in parentheses that, while the novelty view has been almost universally abandoned there is an interesting exception: Patrick Maher has defended the temporal view in a novel way.[5] Although I do not have the time to argue it here, I believe that Maher's arguments in so far as they are successful at all in fact underwrite the non-outlandish heuristic view that I continue to defend as explained below.)

3.

Zahar and I developed a different view which

(i) aimed, like the novel facts view, to neutralise the Duhem threat to rational theory-change;

(ii) agreed with *all* well-established intuitive judgements of support in particular cases rather than just some of them; and

(iii) unlike the novel facts view has, we argued, a clear epistemic justification.[6]

The view is sometimes called, following Nickles, Mayo and others,[7] the "Use Novelty" (UN) account – though, since the UN charter is that novelty is not the issue at all, the name is somewhat misleading. The account directs attention to what seems really methodologically suspect in cases like the emission theory of light and interference effects. The problem in these cases is not that the evidence involved was already known when the theory was first

articulated, but rather that the evidence was *used in the construction of the theory* – in the paradigm case, this means that the evidence was used to fix the value of some parameter (or parameters) in the theory under consideration.

In such cases, some general theory T involves, let's say for simplicity, a single parameter left free by any consideration of a theoretical kind. T itself does not entail e. But e can be used to fix the value of the parameter, hence producing the more specific theory T' which *does* entail e (*modulo* auxiliaries and initial conditions).

To anticipate a point I shall need to labour below, the claim never was that there is anything suspicious about this procedure *per se* (indeed it plays a regular and indispensable role in the scientific enterprise). But Zahar and I did claim that it means that e cannot be taken fully to support T' despite the fact that T' entails e. Real support for T' must be sought through *independent* evidence – that is, evidence aside from e that T' entails but which its more general predecessor T fails to entail. So, in the example just discussed, Young's two slit result could indeed be accommodated within the emission theory – essentially by giving oneself an expression for the diffracting force produced by the double-slit apparatus with a large number of free parameters and working out *from the details of the known experimental result* what those parameter-values have to be. Since the details of the experimental outcome follow from the rival wave theory without the involvement of any such adjustable parameters – two coherent series of waves just do alternately constructively and destructively interfere depending on the phase relations at particular points – these results support the wave theory over the emission theory, despite the adaptability of the latter theory to them.

The crucial positive feature of Fresnel's theory in respect of these interference or diffraction results was clearly expressed some years after the invention of the theory by the British scientist George Gabriel Stokes:

> The theoretical distances of the several fringes [...] were a matter of pure prediction; for the only unknown quantity involved in the theoretical expression, the length of the wave, had been determined by Fresnel by independent methods [...] so that *not a single arbitrary constant was left to be determined by some measurement of a fringe in some one particular case, whereby an* [...] *accordance between theory and observation might have been brought about.* (Stokes 1884, p. 63; emphasis supplied)

Although the notion of parameter-fixing is a precise mathematical one, there are clear informal analogues. For example, the central "creation science" claim that god created the universe in 4004 BC essentially as it now is, provides itself with an indefinite number of "free parameters" waiting to be filled in with details of how exactly the universe is. Creationists simply use whatever observations are available to fix the details of their more specific theories. This gives rise to what might be called the "Gosse dodge" – the fossil record (or rather *so-called* fossil record) allegedly provides no evidence for Darwinian evolution over creation, since all aspects of this record can be accounted for by supposing that they are just features of the way god chose to create the world: he just decided to plant bone-*like* structures in tar pits and desert sands and just decided to make various rocks complete with pictures that look awfully like the imprints of skeletons of extinct creatures. The UN view straightforwardly captures the intuitively obvious judgment that no amount of Gosse-dodging equilibrates the evidential scales between Darwinism and "scientific" creationism-Darwinism with plausible geological assumptions (pretty well) predicts that there will be fossils around, creationism only accommodates the "fossil" record after the event. But the important fact from the point of view of what really counts as evidence for the relevant theoretical claim is not simply that science already had a fossil record when Creationists developed their theories, but rather that those particular theories were based on, constructed in the light of, tailored to, that record.

So what then exactly does the UN view claim? One purpose of the present paper is to get much clearer on exactly this issue. However for the moment, let's suppose that it is captured – at any rate approximately – by what Nickles called the "no double use rule". This is the rule that "you can't use the same fact twice": once in the construction of a theory and then again in its support. Or alternatively, and more precisely though also more problematically: the bits of evidence that support a theory are those which (i) follow from the theory (*modulo* initial conditions and perhaps other bits of "background knowledge") but (ii) were not used in the construction of that theory – by for example fixing the value of a parameter within it.[8]

4.

Much of this may seem "old hat", but I thought it important to rehearse the general background before coming to the main aim of the present paper. There have been, in the past 15 years or so, a number of important criticisms of the UN charter which call for a response they have not yet received – allegedly destructive criticisms from Bayesians such as my colleague Colin Howson and Allan Franklin and more sympathetic, constructive but no less powerful criticisms by non-Bayesians such as Thomas Nickles and particularly Deborah Mayo.[9] These turn out to raise not just local issues, but a range of central issues in philosophy of science. I cannot attempt to respond adequately to them all here. Instead I shall concentrate on one – which I nonetheless believe to be the most central. (It is certainly one common to all critics.) I shall outline the clarification (or amendment, you can take your pick) of the position that this criticism forces and must leave for another occasion the question of how this more elaborate position ties in with alternative accounts developed by the Bayesians and by Deborah Mayo.

Beforehand, however, I need to emphasise a couple of details which were explicit, or at any rate clearly implicit (or so I claim) in the original position. I emphasise them, not for the boring reason of claiming priority for what are in fact some fairly obvious points, but because they turn out to be significant for what comes later.

First, there never was any suggestion that using phenomena in the construction of theories, fitting theories to known evidence, is a "bad thing".[10] Adams and Leverrier clearly used the initially anomalous data on Uranus's orbit in the construction of the Newtonian theory that predicted the existence of the hitherto unrecognised planet Neptune; and this development is universally and surely correctly regarded as one of the greatest successes of the Newtonian framework.[11] The important aspect of this development, however, was exactly the predictive success – Neptune was found more or less where and with more or less the properties that the modified Newtonian theory entailed. What happened, then, in this case was that there was a specific theoretical system T based on the core Newtonian tenets that entailed the wrong results about Uranus's orbit; scientists like Adams and Leverrier then *retreated* by freeing up certain aspects of that initial system (basically the commitment to a particular number of other planets), and then used the erstwhile refuting evidence from Uranus to refill the parameter values; in this way they created a new theoretical system T' which *both* got Uranus's orbit right *and* predicted the new planet. Given the predictive success in this case, it seems to matter little what we say about the support of the initially refuting data from Uranus's orbit for the eventual system T' – though good bookkeeping suggests that we say here too that the Uranian data (as opposed to that from Neptune) supplies no real support for T'. But clearly, since it led to a theoretical system that did enjoy impressive new support, the use of that data in the construction of that theoretical system cannot be a "bad thing".

The complaint against creationism, then, that is underwritten by the UN position is not that it uses data about fossils to construct particular theoretical systems based on its core idea, but rather that the *only* data entailed by those systems is the data thus fed into it: unlike the Newtonian case, there is no *independent* evidence for the theoretical system thus constructed. Using data in the construction of a theory is no bad thing, but the data used do not support (or do not fully support) the theory thus constructed, in particular if two theoretical frameworks based on different core ideas are vying for acceptance and if the first framework produces a theory that straightforwardly entails some piece of evidence (already known or not) while the second framework then uses that evidence to construct a particular theory that entails it, then this does not balance the evidential scales – so far as this particular piece of evidence goes at any rate the first theoretical framework remains better supported.

The second important detail that I claim was in the UN account from the start is this. Not only does the UN view not condemn using facts to fix theories in general, it is quite consistent with the fact that sometimes the *best* theoretical explanation of some data is one that has been constructed using those data. For example, the best explanation for the observed lack of stellar parallax in, say, the 18th Century was surely the Copernican/Newtonian one that there is indeed an apparent parallactic motion produced by the orbit of the earth around the sun but, because of the enormous distance to the nearest star compared to the distance to the sun, the extent of that parallactic motion placed it beneath the limits of observability permitted by the then available telescopes. At least at the time this account was first proposed, there were no independent theoretical reasons for adopting any such view about the comparative distances to the sun and nearest (other) star – instead a view about this ratio was "read off" the failure to observe any parallax. The account nonetheless provided the best available explanation – but not because of anything to do with the data about apparent stellar motions, but rather because the account enjoyed empirical success, quite unmatched by any other available account, *in other quite distinct areas*. Given that it was best empirically supported through other data, the framework wins the right to give the best explanation of some phenomenon, even when that explanation is itself entirely *ad hoc* – in the sense that that explanation was constructed from the data without that particular construction being subject to any then available independent test.

5.

After these clarifications of the view, I now consider the direct criticism that has been brought against the UN account. It has been claimed that, just like its temporal novelty predecessor, the UN account is in outright contradiction with well-established intuitive judgements about empirical support in particular cases. Although they make the point in somewhat different ways, Tom Nickles, Colin Howson and Deborah Mayo all in effect claim that the "double use rule" conflicts with scarcely deniable intuitions about particular cases – there are lots of cases in which data used in the construction of a theory also support it (in many cases maximally so). The rule, as my friend Colin Howson puts it, "makes nonsense of quite basic and eminently reasonable scientific appraisals" and is on that and other accounts "entirely bogus".[12]

Colin Howson likes to emphasise standard statistical examples such as the following. We are given that an urn contains only black and white balls though in an unknown proportion; we are prevented from looking inside the urn but can draw balls one at a time from it. Suppose that a sample of size n has been

taken (with replacement) of which k have been found to be white. Standard statistical estimation theory then recommends the hypothesis that the proportion of white balls in the urn is $k/n \pm \varepsilon$, where ε is calculated as a function of n by standard confidence interval techniques. The sample evidence is the basis here of the particular hypothesis constructed and surely also supports it at least to some (good) degree – the evidence for the hypothesis just is that a proportion k/n of the balls drawn were white.

Deborah Mayo cites and analyses in more detail the same case and also cites the following "trivial but instructive example" (p. 271). Suppose one wanted to arrive at what she describes as "a hypothesis *H*" about the average SAT score of the students in her logic class. She points out that the obvious – indeed uniquely sensible – way to arrive at H is by summing all the individual scores of the n students in the class and dividing by n. The "hypothesis" arrived at in this way would clearly be "use-constructed". Suppose it is that the average SAT score for these students is 1121. It would be absurd to suppose that the data used in the construction of the "hypothesis" that the average SAT score is 1121 fails to support that hypothesis. On the contrary, as she writes:

> Surely the data on my students are excellent grounds for my hypothesis about their average SAT scores. It would be absurd to suppose that further tests would give better support.[13]

Exactly so: the data provide not just excellent, but, short of some trivial error, entirely *conclusive* grounds for the "hypothesis" – further tests are entirely irrelevant. This is exactly why it seems odd to talk of a "hypothesis" at all in these circumstances.

Despite this oddness, Mayo's example and the statistical examples cited both by her and by Howson, do point in the direction of an important consideration. Exactly what that consideration is is best seen through the sort of case highlighted by Tom Nickles in his (1987) – that of theories "deduced from the phenomena".

6.

During the heyday of hypothetico-deductivism, Isaac Newton was often regarded as providing classic support for Einstein's dictum that anyone seeking to understand science should concentrate on what scientists do and ignore entirely what they *say* they do. Newton was patently a great scientist, probably the greatest ever, and yet his claims about how to do good science – that hypotheses should be eschewed and properly accredited scientific theories must be "deduced from the phenomena" – seemed to be logical nonsense. Theories transcend any available data – *both* "horizontally" in generalising on necessarily finite data *and* "vertically" in attributing the data to underlying unobserved mechanisms. Even if we allow inductive generalisation, Kepler's and Galileo's "data" tell us how planets and terrestrial bodies move, Newton's theory tells us that they move in the way they do because of the underlying, non-directly observable force of gravity. On the other hand, deductive logic is non-content-increasing. The conclusion of any inference that is deductively valid was already contained in the premises (even though a deduction may of course surprise us psychologically speaking in that we may not have *realised* that the conclusion was indeed implicit in the premises). Since Newton's theory clearly transcends, has content not shared by, Galileo's and Kepler's "data", his – often repeated – claim to have deduced his theory from the data must be logical nonsense.

More recent years, however, have seen a rehabilitation of Newton's method. Dorling, Glymour, Zahar, Harper and others[14] have all shown that Newton meant what he said – he really did provide deductions of theories from the phenomena –, and that his method is, at any rate on the surface, quite different from anything hypothetico-deductive. Of course not even Newton can defy the laws of logic and what "deduction from the phenomena" really means is deduction from the phenomena *plus some usually very general "background" principles*. Scientific theories are shown to follow deductively not from data alone, nor even from generalised data alone – both claims would indeed be logical nonsense – but from generalised data *plus* general principles that are arguably part of "background knowledge".[15]

Once the real idea of this method is grasped, it is easy to see how ubiquitous its application in science has been. John Norton has shown for example, in some recent papers, how the basic principles of even as "revolutionary" a theory as the quantum theory were argued for by "demonstrative induction" (aka "deduction from the phenomena").[16] Norton argues, I believe fundamentally correctly, that it is the fact that theories are so often deducible in this way that makes the phenomenon of underdetermination of theory by data, so beloved by philosophers, invisible to working scientists. Very often a theory is deducible from, and so uniquely determined by, general principles that scientists already accept plus some – usually new – experimental data.

The details in cases of "revolutionary" science tend to be complex, but "normal" or intra-research programme cases can be very straightforward.

Suppose a scientist already accepts the general wave theory of light – the theory that light from any particular source consists of waves of some wavelength or other transmitted through the luminiferous aether. This general theory does not specify the wavelength of any particular kind of monochromatic light – say light from a sodium arc. The scientist would like a more detailed theory that does specify that wavelength. Does she lay on her couch and make a Popperian bold conjecture? Such an attempt to find a needle in an infinite haystack would be entirely ill-advised. Instead she would deduce the specific theory, involving the specific value of the wavelength, from the phenomena. She would look for some consequence, e, of her general theory T, where e characterises some observable magnitude (fringe separation in some particular experiment, say) as a one-to-one function of the wavelength. She would perform the experiment using light from a sodium arc, measure the magnitude at issue – here, the fringe separation (call the result of this measurement e') – and infer to a more specific theory T'. So for example, subject to a couple of idealisations, one can infer from the general wave theory that, in the case of the famous two-slit experiment, the (observable) distance X from the fringe at the centre of the pattern to the first fringe on either side is related to (theoretical) wavelength λ, via the equation $X/(X^2+D^2)^{1/2} = \lambda/d$ (where d is the distance between the two slits and D the distance from the two-slit screen to the observation screen – both of course observable quantities). It follows analytically, of course, that $\lambda = dX/(X^2+D^2)^{1/2}$. But all the terms on the right hand side of this last equation are measurable. Hence particular observed values will determine the wavelength and so the more specific theory T'.

The relevance of this to the use-novelty issue will be obvious: in cases of theories deduced from the phenomena, the phenomena at issue are clearly used in the construction of the theory (and used in the clearest possible sense, they are essential premises in its derivation) and yet those phenomena not only support the specific theory arrived at, they do so in a near maximal way. When asked why she accepts T' – the particular version of the wave theory involving the particular assignment of a value to the wavelength of sodium light – the answer a 19th century physicist would give is exactly the evidence about the fringe distances from which she derived that particular theory. *Given that she already accepts the general theory*, this evidence is indeed not just support but conclusive support for that particular theory. What better support for a theory from data could there be than that the data deductively entail the theory – given other assumptions that one already accepts?

Deborah Mayo's example cited earlier can be thought of as a special case of deduction from the phenomena: in the SAT score case we just have straight deduction from the data, essentially no background assumptions are necessary. In Colin Howson's statistical case the underlying ideas are, I think, generalisations of those underlying deduction from the phenomena – though obviously the details are different, since in these probabilistic cases the evidence neither deductively implies the constructed hypothesis, nor does the constructed hypothesis deductively imply the evidence. Nonetheless here too the general model – that the process of drawing balls from an urn is a Bernouilli process with unknown parameter k – is part of background knowledge, indeed it is just "given" as an unquestioned assumption. And the evidence (of the relative frequency of white balls in some sample of draws from the urn) is used, not deductively but in accordance with again accepted statistical techniques, to arrive at a specific theory (about the proportion of white balls in the urn as a whole) which specific theory then makes highly probable (and so is intuitively supported by) the evidence that was used in its construction.[17]

7.

So, end of story? Worrall discovers deduction from the phenomena and has to admit that the critics of the UN account of support were right all along – not only are there cases where evidence used in the construction of a theory also provides support for that theory, those cases are exactly the clearest-cut cases of maximal, or near-maximal support. To repeat, what better support could there be than deducibility (or its near relative in Colin Howson's statistical case)?

Well, hold on to that essential phrase that in these cases the data entail (or "quasi-entail") the theory *given other general assumptions one already accepts* and recall the original problem.

Although the argument for support from the interference fringe data for the specific theory T' in our example seems intuitively impeccable, the argument is formally entirely analogous to exactly the sort of most methodologically-suspect cases discussed earlier. Creationists too can deduce their Gossefied theory from the phenomena about the alleged fossils – *given* the general creationist theory that they already accept, then the specific version with the "fossils" as aspects of the creation is deducible from the data.[18]

Or consider another much-discussed example. Immanuel Velikovsky developed a theory about a giant comet that broke away from Jupiter and made a series of orbits around the earth before settling down to a quieter life as the planet Venus. The "close encounters" between the comet and the earth were, according to the theory, responsible for such remarkable (alleged) phenomena as the falling of the walls of Jericho and the parting of the Red Sea. Velikovsky accepted that such cataclysms could hardly have been restricted to selected parts of the Middle East and looked for records of similar natural pyrotechnics in other contemporary record-keeping cultures. The search revealed some embarrassing gaps. But Velikovsky was equal to the task: he postulated that, for the scribes in *some* cultures, the events associated with the close encounter with the comet had proved so traumatic that "collective amnesia" had set in. And he was of course able to deduce the detailed version of his theory that specified which cultures did and which did not suffer from collective amnesia from the phenomena – the phenomena of the existence of (more or less) suitable records from some cultures and their non-existence from others: collective amnesia afflicted precisely those cultures $C_1, \ldots, C_n$ for which no suitable records of cataclysms exist.

8.

Given this formulation of the problem, its solution surely jumps out at you. The judgement of support in the "deduction from the phenomena" cases is ineliminably *conditional*, an ineliminably *intra*-paradigm or *intra*-research programme judgement. *Given* that a general framework, or research programme, is already accepted, then the data give – in the case of a genuine deduction – not just *some* support for the specific theory, but conclusive support. Given that the general form of the wave theory was already accepted, then the fringe data *entail*, and therefore obviously support, the specific form of the wave theory with the particular values for the wavelength-parameters. But equally, if you already accepted the general Velikovskian theory then the cataclysmic record data would entail (or almost entail), and hence give maximal (or near maximal) support for, the specific form of the Velikovskian theory complete with particular values (in that case 0 or 1) for the "collective amnesia parameters".

The correct judgement seems to be, then, that, in all cases of this kind, support is for a particular theory *given* the already accepted general framework or programme, where the conditional remains undischarged. The support, speaking loosely but suggestively, does not spill over to the general framework. The fringe data give you conclusive reason for holding one particular form of the wave theory, if you are going to hold the wave theory at all, but no reason at all for holding the wave theory at all. The "cataclysm" record data (including lack of records) give you (near) conclusive reason for holding one particular form of the Velikovskian theory, if you are going to hold the Velikovskian theory at all; but no reason at all for holding the Velikovskian theory at all.

But is there any other sort of empirical support? Are there cases of support for a specific theory where the support *does* spill over to the general underlying framework idea? The answer is that there are two sorts of case both exemplified in the wave theory example, but not, exactly as one would expect, in the case of Velikovsky's theory.

First in the wave theory there are phenomena that "fall naturally out of the core idea" – this was exactly the case, for example, with straightedge diffraction (remember the passage from Stokes that I quoted earlier). And *secondly*, the more specific version of the wave theory we mentioned, complete with values for the theoretical wavelengths, having been constructed using certain fringe data, then implied *other* independent data – about, for example, fringe separations in different experiments. (Indeed the standard situation in these genuinely scientific cases is that the free parameters are *overdetermined* – there is a range of experimental results $e_1, \ldots, e_n$ *any one of which* would determine the value of the initially free parameters to produce a theory that would then entail, and so obtain independent support from, the other $n-1$ results.)

Neither type of case of the support for the general theory is exemplified in either Velikovsky's or the creationist's theory. Neither of these theories enjoyed any straightforward, non-use-constructed empirical success nor, having produced specific versions in the light of particular evidence, did any of these specific versions go on to make successful independent predictions.

The UN account applies exactly to this kind of support that is not conditional on prior acceptance of the underlying general theory. And it is, moreover these types of case of support – the ones to which the UN rule *does* apply – that carry the more force. The intuition that the fringe data supply powerful support for the fixed-parameter wave theory – because they entail that theory given the general wave theory as premise – relies essentially on the fact that there are other data, not used in its construction that support that general wave theory. The intuition that Kepler's and Galileo's data supply powerful support for Newton's theory, relies exactly on the fact that there is plenty of UN-obeying support for the general principles that underlie Newton's deduction of his theory from that data. The intuition that the lack of suitable records does *not* support Velikovsky's collective amnesia hypothesis and that the alleged fossil record does not support Gossified creationism is based exactly on the following fact. Although, given the general theory, the data in each case entail (or at least come close to entailing) those specific versions of the theory, there is no UN-satisfactory support for either general theory in the first place. The fossil record gives you conclusive reason for holding the Gossified version of Creationism, if you hold any version of Creationism at all, but it gives you no reason at all for holding any version of Creationism. The lack of suitable records gives you (near-)conclusive reason for holding Velikovksy's theory complete with collective amnesia, if you hold Velikovsky's general theory, but it gives no reason at all to hold Velikovsky's general theory.

9.

Some brief concluding remarks:

i. This distinction between two types of support is not itself *ad hoc* within the UN programme. Once you realise that this programme was never committed to the idea that "fitting your theories to the data is a bad thing", the distinction is just an elaboration of the reasons why it need be no bad thing.

ii. Unfortunately – methodological life would be altogether simpler otherwise – the judgment in the UN-satisfying cases that the support "spills over" from the specific to the general underlying theory is *not* a simple reflection of the fact that the general idea entails the phenomena at issue without the need of *any* auxiliaries. The core wave idea on its own does *not* entail the straightedge diffraction phenomena. This would be inconsistent with the Duhem thesis which in fact holds universally. Instead, the straightedge diffraction phenomena were entailed by that core idea together with the "natural" auxiliaries – those motivated, though not entailed, by the general heuristic ideas underlying the wave theory.

iii. This is why the – difficult – notion of naturalness within a framework cannot be avoided. It is also exactly why the UN rule, despite perhaps first appearances, does have an epistemic rationale. Early critics, like Alan Musgrave,[19] were quick to suggest that the UN view was no better off than the temporal novelty view in terms of possession of a cogent epistemic rationale. I have asked why on earth the temporal-order of theory and evidence should in itself have any significance for rational support. But it might seem that there is at least equal justification for asking why on earth a theory's empirical merits, once it is "on the table", should depend at all on how it was constructed – on how it got onto the table in the first place.

But the answer is now easy to see. In science we are interested – indeed, I would argue, principally interested – in how much support a specific theory obtains *as representative of the general theory that underlies it.* After all, if we were just interested in codifying, correctly dealing with, some range of empirical phenomena, say fringe phenomena, it would not matter at all, except perhaps pragmatically, which of two rival theories we used if they both in some way accommodated all the phenomena. Scientists in the early 19th Century were principally interested in how much support the facts gave to Fresnel's *specific* theory of diffraction (consisting of the general wave theory plus specific assumptions tieing it to the case of diffraction) *as representative of that general wave theory.* The judgment that principally required justification was that the sensible way forward in *other* areas of optics – such as crystal optics – was to build different specific theories on that same general idea.

But given Duhem's point that the general theory has no empirical consequences of its own, this judgment about "the sensible way forward" involves considerations *both* of how the specific theory based on it stands deductively to relevant facts *and* of how the specific assumptions relate to the general idea. And this latter relationship is exactly what heuristics is about – the complex, but I claim objective, relationships between general motivating idea and particular theories built around it.[20]

iv. Although I have given what I hope is a clear account of the basic idea underlying the heuristic account of support, much remains to be done – especially in terms of detailed comparison of the view with other approaches. It will, for example, no doubt be suggested that the basic idea of the account can be given a much more precise formulation in Bayesian terms. I have no doubt that the heuristic account can indeed be given a Bayesian formulation – but this is exactly the problem with Bayesianism. The fact that every conceivable position in the prediction vs accommodation debate has been defended on the basis of some Bayesian position is a perfect illustration of the fact that "the" position can explain everything and so really explains nothing. However, this clearly requires careful articulation which I am not able to provide here.

A fuller treatment also requires a detailed comparison of the heuristic view I have defended with the accounts of Nickles and Mayo. I have argued that their chief criticism of what they took to be my view misses the mark. But they each present positive accounts that can plausibly be argued to be equivalent to the clarified (or modified?) view I have explained here. Indeed, it might be argued that Deborah Mayo's views about 'severe tests' can be used to develop a deeper account, one that yields my own heuristic view but which operates from an intuitively more compelling starting point. Surely much of what I have said can be reexpressed intuitively using the notion of a severe test – the "fossils" do not provide a severe test of the Gossified version of creationism, planetary stations and retrogressions do provide a severe test of Copernicanism despite having been known for centuries, and so on. However, I would argue that, when it comes to which notion is the deeper, or more explanatory, it is in fact heuristic considerations that explain our intuitions about severe tests rather than the other way round (despite the relative familiarity – due to Popper's influence? – of the latter notion). Again this requires careful argumentation which I cannot provide here.

One final important omission is this. Except for a brief reaction to the criticism of the UN view by Colin Howson, I have been considering throughout the case of deterministic theories where, *modulo* auxiliary assumptions and initial conditions, we do have direct deducibility of the evidence concerned from the theories concerned. In the case of statistical-probabilistic theories, matters may well be more complicated. There, the issue of how well some specific theory fits the data is in general a question of degree. Suppose that we have two separate general classes of statistical theory G_1 and G_2, and specific theories T_1 and T_2 developed from each general class using some data e. It may well be that the extent to which the two specific theories T_1 and T_2 are supported by e, also has consequences for the extent to which the general classes of theory, G_1 and G_2, are supported by e. But this is because, speaking intuitively, the best fit with the data achievable with a specific theory from one of the general classes, G_1, say, is better than the best fit achievable with a specific theory from the other general class G_2. So, for example, the fact that some particular parabola fits the data (of simultaneous values for some variables x and y) better than any particular straightline, gives some reason to think that the general theory that the relationship between x and y is parabolic (plus an "error term") is better supported than the general theory that the relationship is linear (plus an error term). Suppose in contrast, that we have two general deterministic frameworks (say Ptolemaic and Copernican astronomy) and two specific theories P_n and C_m produced within those frameworks on the basis of some evidence e – so both specific theories are "use-constructed" with respect

to e. Here the degree of fit between e and P_n and C_m will be maximal – each deductively entails e (*modulo* accepted auxiliaries). And so it does seem intuitively clear in this case that *e on its own*, while it may give conclusive reason to someone who already accepts the Copernican framework to accept C_m (and similarly give conclusive reason to someone who already accepts the Ptolemaic framework to accept P_n), cannot give any reason to prefer one *general* framework over the other.[21]

Notes

1. There is a separate potential argument from within the Bayesian framework. Since, whenever e is already known, it is already part of background knowledge, for any such e, $p(e) = 1$. The view – natural for a Bayesian – that e supports T if and only if $p(T,e) > p(T)$, then entails that no known e ever supports any T. ($p(T,e) = p(T\&e)/p(e) = p(T\&e)$ (if $p(e) = 1$); and $p(T\&e) \leqslant p(e)$.) This would then imply that novelty of evidence is a necessary condition for it to support any theory. However, this was, I think, always regarded as a *problem* for Bayesianism (the "old evidence problem" – see Glymour (1980)) – it seemed forced to say that temporal novelty mattered in this strong sense and yet no one really believed that it could. (Later there were Bayesians – one is Patrick Maher – who at least *appear* to hold that novelty is important.)
2. See, for example, Earman and Glymour (1980).
3. See my (1985).
4. See Lakatos and Zahar (1976).
5. See his (1988).
6. See Lakatos and Zahar (1976) and Worrall (1978) and (1985).
7. See Nickles (1987), and Mayo (1996).
8. I am not, of course, claiming that these are necessary conditions for T to support e – for a start probabilistic theories can be confirmed by evidence which they do not entail. They are however sufficient (whereas the temporal prediction view is neither necessary nor sufficient).
9. Nickles (1987) and particularly Mayo (1996).
10. Contrary to Colin Howson's (1990).
11. Though there have been suggestions that even this case is not as straightforward as has normally been assumed (see Bamford 1996).
12. Nickles (1987), Howson (1990) and Mayo (1996).
13. Mayo (1996).
14. Dorling (1973) and (1974), Glymour (1980), Zahar (1981), Harper (1991) and (1993).
15. For a detailed account of Newton's method, its strengths and the problems it raises see my (2000).
16. Norton (1993) and (1995).
17. But there are special difficulties in extending this account to statistical cases. The present account should be regarded as applying only to the case of deterministic theories.
18. The basis Creation hypothesis is (essentially) that the world was created as it now is observed to be; hence plugging in observations of the way the world is leads deductively to specific versions of the theory.
19. Musgrave (1978).
20. My friend Elie Zahar has a lot to answer for in having made the suggestion (in his (1973)) that heuristics involves consulting private diaries and correspondence.
21. There are no doubt many other complexities in the statistical case that need attention too. In particular the arguments in Foster and Sober (1994). Much of Deborah Mayo's analysis too is, of course, directed at statistical cases.

References

Bamford, G.: (1996), 'Popper and his commentators on the discovery of Neptune: A close shave for the law of gravitation?', *Studies in History and Philosophy of Science* **27**, No.2, pp. 207–232.

Dorling, J.: (1973), 'Demonstrative Induction: Its Significant Role in the History of Physics', *Philosophy of Science* **49**, pp. 360–372.

Dorling, J.: (1974), 'Henry Cavendish's Deduction of the Electrostatic Inverse Square Law from the Result of a Single Experiment', *Studies in the History and Philosophy of Science* **4**, pp. 327–348.

Earman, J., Glymour, C.: (1980), 'Relativity and Eclipses: The British eclipse expeditions of 1919 and their predecessors', *Historical Studies in the Physical Sciences* **11**, pp. 49–85.

Foster, M., Sober, E.: (1994), 'How to tell when Simpler, More Unified, or less Ad-Hoc, Theories will Provide more Accurate Predictions', *British Journal for the Philosophy of Science* **45**, No. 1, pp. 1–35.

Glymour, C.: (1980), *Theory and Evidence*, Princeton University Press, Princeton.

Harper, W.: (1991), 'Newton's Classic Deductions from the Phenomena', *PSA 1990*, vol. 2, pp. 183–96.

Harper, W.: (1993), 'Reasoning from Phenomena: Newton's Argument for Universal Gravitation and the Practice of Science', in P. Theerman, A.F. Seef (eds.), *Action and Reaction*, University of Delaware Press, Delaware City, pp. 30–78.

Howson, C.: (1990), 'Fitting theory to the facts: Probably not such a bad idea after all', in C.W. Savage (ed.), *Scientific Theories*, Minnesota Studies in the Philosophy of Science, vol. 14, University of Minnesota Press, Minneapolis, pp. 224–244.

Lakatos, I., Zahar, E.: (1976), 'Why did Copernicus's Research Programme Succeed Ptolemy's?', in Westman, R. (ed.), *The Copernican Achievement*, University of California Press, Los Angeles.

Maher, P.: (1988), 'Prediction, Accommodation, and the Logic of Discovery', in A. Fine, J. Leplin (eds.), *PSA 1988*, vol. 2, Philosophy of Science Association, East Lansing, pp. 273–285.

Mayo, D.: (1996), *Error and Growth of Experimental Knowledge*, University of Chicago Press, Chicago.

Musgrave, A.E.: (1978), 'Evidential Support, Falsification, Heuristics and Anarchism', in G. Radnitzky, G. Andersson (eds.), *Progress and Rationality in Science*, Reidel, Dordrecht, pp. 118–136.

Nickles, T.: (1987), 'Lakatosian heuristics and epistemic support', *British Journal for the Philosophy of Science* **38**, pp. 181–205.

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

Norton, J.: (1995), 'Eliminative Induction as a Method of Discovery: How Einstein Discovered General Relativity', in J. Leplin (ed.) *The Creation of Ideas in Physics*, Kluwer, Dordrecht, pp. 38–74.

Stokes, G.G.: (1884), *Burnett Lectures. On Light*, Macmillan, London.

Worrall, J.: (1978), 'The Ways in which the Methodology of Scientific Research Programmes improves on Popper's Methodology', in G. Radnitzky, G. Andersson (eds.), *Progress and Rationality in Science*, Reidel, Dordrecht, pp. 18–39.

Worrall, J.: (1985), 'Scientific discovery and theory-confirmation', in J.C. Pitt (ed.), *Chance and progress in modern science: Papers related to and arising from the Fourth International Conference on History and Philosophy of Science*, D. Reidel, Dordrecht, pp. 301–332.

Worrall, J.: (2000), 'The Scope, Limits and Distinctiveness of Newton's Method of 'Deduction from the Phenomena' ', *British Journal for the Philosophy of Science* **51**, March 2000, pp. 45–80.

Zahar, E.: (1973), 'Why did Einstein's program supersede Lorentz?', Parts 1 and 2, *British Journal for the Philosophy of Science* **24**, pp. 95–125, 223–262.

Zahar, E.: (1989), *Einstein's Revolution: A Study in Heuristic*, Open Court, La Salle.