JOHN WORRALL

RESEARCH PROGRAMMES, EMPIRICAL SUPPORT, AND THE DUHEM PROBLEM: REPLIES TO CRITICISM*

Many of the papers delivered at the Kronberg conference did not directly criticize the LSE position. Rather they suggested alternatives. I shall not attempt to criticize these alternatives here, but shall simply try to show that the various specific criticisms which *were* directed at the methodology of scientific research programmes miss their mark. This will involve me in replying directly only to the papers by Professors Koertge, Musgrave and Post.[1] (I shall also make a few remarks on the paper by Professor Feyerabend.)

In my part of the position paper I argued that MSRP provides novel solutions of two connected problems: the problem of empirical support and the problem of appraising rival scientific theories. I shall consider criticisms of these two proposed solutions in Sections 1 and 2 respectively.

1. AD HOC EXPLANATIONS AND EMPIRICAL SUPPORT

There is a long standing debate in philosophy of science, sometimes called the 'weight of evidence' debate. The point at issue was recently given a sharp formulation by Lakatos [1968], and then by Musgrave [1974]. A large part of my position paper is also devoted to this debate. According to one side (the side usually associated with Mill[2] and with Keynes), whether or not some theory T is supported by a piece of evidence described by the statement e depends solely on the logical relation between T and e. I shall, following Musgrave, call this attractive position the 'purely logical' position. According to it, if two theories, T_1 and T_2, both imply a correct description, e, of some set of facts, then these facts support both T_1 and T_2.

The main argument I urged against this view in Chapter 3 is that it is generally possible to modify a theory in a trivial '*ad hoc*' way so as to entail a correct description of any fact successfully predicted by a rival theory. If, originally, T_1 entails e, but T_2 does not, then it is always possible to produce a T_2', sufficiently similar to T_2 to be

A unified bibliography can be found on pp. 379–383.

G. Radnitzky and G. Andersson (eds.), Progress and Rationality in Science, 321–338.

'essentially the same' theory, which *does* entail e. On the 'purely logical' view e could then in general provide no basis for discriminating between T_1 and T_2'. I argued that on the contrary we should not regard e as supporting such a theory as T_2' if that theory was indeed modified in an *ad hoc* way precisely so as to entail e.

Clearly both Koertge and Post have a good deal of sympathy for the purely logical position on empirical support, because both regard the problem of *ad hoc* explanations as, in fact, a non-problem. Koertge claims (above, p. 269) that "there is no problem of *ad hoc*ness in the MSRP sense of the term and ... the whole business is really very simple". Similarly Post asserts (above, p. 314) that, while a "good deal of the LSE programme can be viewed as an occasionally desperate attempt to eliminate *ad hoc*ness", this attempt is misdirected, for, "from the objective point of view it is quite irrelevant whether a theory has been created pragmatically '*ad hoc*' ".

But neither Post nor Koertge counter my argument that the problem of *ad hoc*ness is fundamental since to ignore it is to give up the hope of discriminating between any two rival theories on the basis of the support they receive from the facts.

Let me state my argument more fully. As Duhem so forcefully demonstrated, those statements (or conjunctions of statements) which we usually regard as amounting to single scientific theories (such as Newton's theory or Maxwell's theory) do not on their own have consequences which are unequivocally testable against experience. In order for such observation statements to be derivable auxiliary assumptions must be added to these scientific theories. I shall therefore regard a testable scientific theory as consisting of two parts: the 'basic' theory (or 'hard core') T, and the auxiliary assumptions, A. Let one theory be $T_1 \wedge A_1$ and another be $T_2 \wedge A_2$, and suppose that $T_1 \wedge A_1$ does, but $T_2 \wedge A_2$ does not, entail e. Then, unless the 'basic' theory T_2 alone entails the negation of e, it is always possible to modify the auxiliary assumptions to A_2', so that $T_2 \wedge A_2'$ is both consistent and entails e. We should certainly speak of this new theory as a modified version of the old one since its 'basic' part remains unchanged. But it follows that we shall not, in general, be able to distinguish between two rival 'basic' theories (say the wave theory of light and the corpuscular theory of light, or classical mechanics and relativistic mechanics) on the basis of the strengths of their empirical support unless we distinguish between genuine and *ad hoc* explanations. A *particular* wave theory of light, say, may, of course, entail correct descriptions of facts not entailed by a particular corpuscular theory. The particular theories would then be distinguishable on the 'purely logical' view. However, it would then be open to the corpuscularist to modify his particular theory so that, as thus modified, it did entail correct descriptions of these facts. Post and Koertge must then say that the two theories receive equal support from the facts.

Moreover, as I showed in my paper *above*, all this is far from being merely a problem of armchair philosophy of science. It is a problem which has been faced by working scientists: the possibility of an apparently superseded theory 'catching up' with its rival by *ad hoc* modifications has often been exploited during science's development. Take the example of the wave and corpuscular theories of light. All the effects (such as those of diffraction and of the 'interference' of polarised light) which are generally taken as supporting the wave theory but not its corpuscular rival, *were* in fact given corpuscular theoretic explanations (many of which were provided by Biot). Scientists at the time were aware of this, and most of them were suitably unimpressed. For example, the experimental physicist Humphrey Lloyd wrote:

> An unfruitful theory may ... be fertilised by the addition of new hypotheses. By such subsidiary principles it may be brought up to the level of experimental science, and appear to meet the accumulating weight of evidence furnished by new phenomena. But a theory thus overloaded does not merit the name. It is a union of unconnected principles ... The theory of emission [of light] in its present state exhibits all these symptoms of unsoundness ... (Lloyd [1833], p. 296).

Translated into MSRP terms, Lloyd is saying that such 'overloaded' theories are not given genuine support by the facts they were concocted to explain.

William Whewell too recognized the flexibility of the emission theory of light. Indeed this recognition may well have been one of the starting points of his philosophical position that a piece of evidence may be entailed by each of two theories and may yet weigh more heavily in favour of one than the other. Whewell writes ([1837] p. 340):

> When we look at the history of the emission theory of light, we see exactly what we may consider as the natural course of things in the career of a false theory. Such a theory may, to a certain extent, explain the phenomena which it was at first contrived

to meet; but every new class of facts requires a new supposition, – an addition to the machinery; and as observation goes on, these incoherent appendages accumulate, till they overwhelm and upset the original framework. Such was the history of the hypothesis of solid epicycles; such has been the history of the hypothesis of the material emission of light. In its simple form it explained reflection and refraction; but the colours of thin plates added to it the hypothesis of fits of easy transmission and reflection; the phenomena of diffraction further invested the particles with complex hypothetical laws of attraction and repulsion; polarization gave them sides; double refraction subjected them to peculiar forces emanating from the axes of crystals; finally, dipolarization loaded them with the complex and unconnected contrivance of moveable polarization ... There is no unexpected success, no happy coincidence, no convergence of principles from remote quarters: the philosopher builds the machine, but its parts do not fit; they hold together only while he presses them: this is not the character of truth.

Of course, Post and Koertge could give up their claim that the problem of *ad hoc*ness is a pseudoproblem without embracing the MSRP account of empirical support. They could, for example, say that if two theories entail correct descriptions of the same set of facts then they are indeed equally supported by the facts, but the two theories may nevertheless be distinguishable on other grounds, for example, on the ground of 'simplicity'. Indeed there are some clear indications that this is the position both Post and Koertge *would* adopt.[3] But so long as one aims to distinguish between empirically equivalent theories on *any* grounds, then one is taking the problem of *ad hoc* explanations seriously. Indeed, those theories which are intuitively regarded as complex (such as Ptolemy's planetary theory or Biot's corpuscular theory of light) are precisely those that have been modified *ad hoc* to explain certain facts which, in their original form, they had been unable to explain.

(It may seem that the heuristic approach I advocate differs only in name from an approach which, while making empirical support a purely logical affair, allows simplicity as an extra ground for rational preference of theories. These two approaches are indeed aimed at capturing the same intuitions about theory appraisal, but it seems to me that the heuristic approach is superior on two counts. First the notion of simplicity is notoriously difficult adequately to characterise in precise terms. The heuristic account of empirical support is certainly not free from sin in this respect, but it does seem to be sharper than any existing account of simplicity. Secondly, in regarding a theory's empirical support and its simplicity as two entirely separate questions the simplicity approach underplays the role of facts. For it is precisely through modifications of the original theory aimed at capturing certain hitherto unexplained facts that those theories intuitively regarded as complex or 'overloaded' have become so. It is originally recalcitrant facts which induce the complexity. In the heuristic approach, on the other hand, the role of facts is given full force: if a theory has been able to 'explain' a certain fact only through *ad hoc* modifications which have made the theory intuitively complex, then it receives no support from that fact.)

According to both Post and Koertge, the heuristic approach to empirical support is not only unnecessary (since directed at a non-problem) but also faulty. Post's criticism of it is that it is 'sociologistic'. (Indeed Post charges MSRP as a whole with 'sociologism'. No approach is acceptable unless it is 'objective'. Post includes as 'sociological' any consideration of the " 'historical' conditions such as the chronological ordering of tests and theories".[4] Now although the heuristic characterisation of empirical support is definitely not 'objective' in Post's sense, it is certainly objective.

I shall show this by restating the heuristic characterisation in rather more precise terms. According to this approach, facts support not simply a theory but rather a research programme, or a theory *together with* a heuristic. (These are equivalent formulations. A research programme at a particular stage of its development is characterised by a pair of entities: the latest theory it has produced, and its heuristic.) A fact described by the statement e supports (or corroborates) a research programme R at a particular stage of its development if (i) e is implied by T (the latest theory produced by R), and (ii) the programme's heuristic guided the construction of T, *independently of the fact described by e.*

There are certain standard cases in which clause (ii) is not satisfied:

(a) Some constant appears in T in a place where a free parameter had occurred in T's predecessor, and the value of this constant was not dictated by the heuristic of the programme but rather had to be 'read off' from the facts described by e;

(b) (i) e says that in conditions X, Y does not occur; (ii) T says, as did its predecessor, that Y occurs if, *and only if*, conditions Z hold, but (iii) T's characterisation of those cases in which conditions Z hold differs from its predecessor's precisely in excluding the condition X. (This is the general schema which underlies for instance, the Thomas Young example I discussed in my [1976], pp. 140–141.

This characterisation of empirical support remains less precise than one would like, but imprecision does not amount to the abandonment of objectivism. This characterisation makes no reference to, and is quite independent of, any subjective or 'sociological' consideration. No psyches or social structures need be inspected in determining whether the relation holds: one needs to look only at theories, facts and heuristics.

A more serious charge of 'sociologism' is hinted at by Noretta Koertge[5] and explicitly levelled against MSRP by Paul Feyerabend. Feyerabend would, no doubt, admit that MSRP's *criteria* are objective; what he denies is that they have been given an objective epistemological rationale. Feyerabend denies that anyone has shown that preferring that theory which is the better according to MSRP criteria is the most rational course of action, or the one most likely to lead to the truth, *etc.* He claims that MSRP advocates are nothing more than anthropologists with a special interest in the tribe of scientists. But Feyerabend's claim is, I take it, based not on any charge that MSRP's criteria are sociological but rather on the charge that the only *rationale* we have succeeded in giving these criteria is a sociological one: namely that they do seem to capture better than others the inutitive appraisals of competing theories which scientists have as a matter of fact made. But if MSRP is to be *more than* a simple descriptive generalisation of scientists' past preferences, it must give its methodological rules an at least tentative and conjectural underpinning of a *general epistemological kind.*

I think Feyerabend's claim is correct: such an underpinning has not so far been given. (Although I should stress that if MSRP's rules provide an accurate general characterisation of scientist's specific appraisals then this is already quite an achievement. And the tribe of scientists is, after all, a rather exceptional one.) This deficiency can certainly be made good by adopting some suitably vague metaphysical principle which states that God's universal blueprint was 'simple' or 'organically unified'. In that case only those explanations would be acceptable which stemmed from some unifying principle, since patched up *ad hoc* theories could not, given this metaphysical principle, possibly be true of the world. This 'solution' of the problem is analogous to the Popperian 'solution' of the problem of induction advocated by Lakatos.[6] Like the latter, it certainly has all 'the advantages of theft over honest toil'. But perhaps, in this area, theft is the only option.

However, I offer the following tentative proposal for an epistemological underpinning of the MSRP criterion which avoids the simple postulation of a metaphysical principle and hence, whatever its shortcomings, at least involves some toil.

Assume that scientists have in fact preferred those theories which are better supported by the facts in the sense which I have specified. Is this just a reflection of their quirks? Or are there good grounds of an objective kind for this preference?

Presumably if we can agree that the scientific enterprise has certain aims then we can also agree that there are objective grounds for awarding higher marks to those theories which seem to have contributed to the achievement of one or more of those aims than to those theories which have not so contributed. Presumably one of the aims of science is to extend our factual knowledge. It follows that if our view of empirical support were the 'strictly temporal' one (which of course it isn't) we should have an immediate answer to the question posed in the previous paragraph. Those theories which receive genuine empirical support on this view have contributed to one of the aims of science and are, therefore, on this ground objectively better than those which have no empirical support (and so have not contributed to the achievement of this aim). (Of course whether these theories are to be preferred as closer to the truth than others or as more reliable for technological application or whatever are different matters. My aim here remember is the rather modest one of establishing that there are *some* objective grounds for awarding high marks to those theories pronounced best by MSRP.)

But the argument I have just given for the case of 'strictly temporal' empirical support will not work when empirical support is construed, as I have argued it should be, in the heuristic sense. For a theory may be well supported in this latter sense without having taught us any facts of which we were unaware prior to the theory's proposal. Must we then fall back on simply reporting that scientists have as a matter of fact generally preferred those theories which are better supported in this sense?

Perhaps an epistemological underpinning can be derived from the obvious one just given for the 'novel fact' criterion. Begin with the premise that it is 'objectively' reasonable, other things being equal, to prefer those theories which have taught us some new fact by correctly predicting the observable existence of some hitherto unobserved event. Notice, however, that whether or not a theory predicts 'novel facts' depends not just on the theory itself but also on the state of scientific knowledge at the time. Nevertheless there may be a property of the theory itself which makes it *possible* for the theory to contribute to the extension of our factual knowledge (whether or not a theory which possesses this property actually does contribute to the extension of our factual knowledge will then depend simply on the historical conditions). Focussing attention on such a property would mark out the merits of the theory itself and would remove the unfairness of historical accidents. It is, in fact, easy to see that the property of non *ad hoc*ness highlighted by MSRP is such a property. If a theory entails some factual statement *e* and is non-*ad hoc* relative to *e* then it is at least possible for the theory to have contributed to the extension of our factual knowledge by predicting the facts described by *e*. If, on the other hand, the theory is entirely *ad hoc* then it could not, no matter what the historical circumstances at the time of its proposal, have led to the discovery of any facts: for such facts as the theory does entail had to be already known in order for the theory to be constructed. *Ad hoc* theories of *necessity* lag behind the facts; whether non *ad hoc* theories lag behind or anticipate the facts is a matter of historical accident. It thus seems 'objectively' reasonable for scientists to award higher marks to those theories which are given some empirical support (in my sense) to those which are not: for the former could, while the latter could not, have contributed to the achievement of one of the aims of science – the extension of our factual knowledge.

Noretta Koertge has an interesting specific criticism of the heuristic notion of empirical support. It is concerned with the time dependence of the heuristic criterion.[7] Her criticism (which is similar to an earlier one of Alan Musgrave's[8]) is the following:

> Suppose *e* was, as a matter of historical fact, used to construct *S*, but there existed at the time a positive heuristic *P* which could have been used instead. In this case does *e* support *S*? If [the Lakatosians] say 'No', I claim that their criterion is objectionable on the grounds that the order of events enters into it. If they say 'Yes', then I claim they must give up Lakatos's method of comparing research programmes ... [For] suppose I am working in RP_1 and I construct *S* on the basis of *e* in an *ad hoc* way ... Surely it is the intent of MSRP that *e* should not count as a success for RP_1. But suppose there is a competing research programme RP_2 which contains a positive heuristic which could have been used. If one does not care which problem situation in fact resulted in the production of *S*, one must now say that *e* does support *S*. (See, p. 268.)

Now I do not accept Koertge's (unargued) assertion that introducing consideration of the time order of theory and experiment into the question of empirical support is undesirable. But my main answer is this: as I stressed in my original paper, the question "does e support S?" is not, for me, a complete question if S is simply a theory or conjunction of theories. The reason is that, like Popper and like Zahar, I make the relation of empirical support not a two-place, but rather a *three*-place relation.[9] Thus, just as for Popper e may support S given background knowledge b while e does *not* support S given different background knowledge b', so for me e may support S given one method of construction of S while e does *not* support S given a different method of construction. Thus my answer to Noretta Koertge's question "does e support S?" is "It depends". What *is* true, in the case she envisages, is that e supports RP_2 but not RP_1 and this despite the fact that the latest theories produced by both programmes are empirically equivalent. This sounds like an unacceptable consequence of my approach, but, as I stressed in my position paper, it is not. Indeed one of my starting points was the claim that it is usually possible for the proponents of a degenerating research programme to generate out of it a theory which, though not of course logically equivalent to the latest theory produced by its progressive rival, is *empirically* equivalent to it. This, as the passages from Lloyd and from Whewell quoted above suggest, is what happened, for example, in the case of the wave-corpuscule rivalry in optics. Nevertheless despite the empirical equivalence of the latest theories produced by such programmes, one programme may receive more support from the facts than the other. Or, if you like, the theory T as produced by research programme P is better supported by the facts than the *empirically equivalent* theory T' as produced by research programme P'.

(Koertge also asks (on p. 268) for clarification of the following problem. Assume that some theory S entails some factual statement e but e does not support S according to the heuristic criterion. Noretta then asks "Does e fail to support S in the same sense as $\neg e$ 'fails to support' S?" Presumably what she is driving at is that some distinction ought to be drawn between facts which refute some particular version of a theory and facts which have been encompassed by the theory even if only *via ad hoc* adjustments. I agree, and I drew such a distinction in my position paper (p. 66, note 32) – though I ought to have drawn it in the text rather than in a note. But, as I argued *above*, this distinction, except in a few rare cases,[10] will be a mobile one. Unless the refuting fact hits not just the whole theoretical system but the 'basic theory' itself there will always be modified versions of the theory within which the originally refuting fact is apparently explained. But such an 'explanation' is more like a consistency proof than a genuine explanation. Of course it is better for some theory to be consistent with a fact than inconsistent with it, but it is still better for a theory to *pre*dict the fact.)

I shall now comment on some of Alan Musgrave's remarks about empirical support. Musgrave in an earlier paper, advocated against both the 'strictly temporal' view (held, for example, by Popper) and the heuristic view (held, for example, by Zahar and me) a 'theoretical' view of empirical support (what he meant by this I shall explain in a moment). Musgrave now complains that, in my position paper, I compared the heuristic view only with the temporal view, thus ignoring the theoretical one. Now I (largely) ignored the theoretical view because my brief was to compare the MSRP position with Popper's and Popper holds a version of the 'strictly temporal' view. However I am very ready to compare the heuristic account of empirical support with the theoretical one.

The theoretical view of empirical support says that if a theory T_2 is proposed as a rival to some already established theory T_1 then T_2 is not given genuine support by any fact that is already explained by T_1. Thus on this account, for example, Einstein's theory is given genuine support only by those facts which it predicts but which were not already explained by Newton's theory. I shall argue that this account both runs into a general difficulty and also fails to square with our intuitions about empirical support in particular cases.

The general difficulty is the following. Suppose T_1 and T_2 are rival theories in some field and that both are independent of e which, however, describes some fact within the theories' field. Then, on the 'theoretical' view of empirical support, it is possible for the proponents of either T_1 or T_2 to produce slightly modified versions of their theories which *are* supported by *e*, *provided they beat their rivals to it*. If the proponents of T_1 happen to hear of e first and immediately produce a new theory T_1' which is in fact just the conjunction of T_1 and e, then e supports T_1', but cannot then support *any* modification of T_2. This is true even if T_2's proponents eventually provide an impressive and contentful modified theory which was not constructed simply to imply e but nevertheless does imply it. But this 'tackers' race' consequence of the theoretical view is surely absurd.

When it comes to particular cases, the theoretical view is both too narrow and too wide. It is too narrow because it rules, for example, that the result of the Michelson-Morley experiment cannot support the special theory of relativity proposed in 1905 since this result had already been explained on the basis of Lorentz's rival theory in 1904. The theoretical view is too wide because it says that this result *must* support Lorentz's theory if this was the first theory to explain it, even if this theory amounted to no more than an *ad hoc* postulation of just sufficient contraction of the arms of the interferometer. The theoretical view is too narrow because it rules that various polarisation effects which are taken to provide dramatic support for the theories Fresnel proposed to explain them in the period 1818 to 1823, cannot, in fact, genuinely support these theories, because Biot had by then already produced a corpuscular theory from which descriptions of these effects follow. And this view of empirical support is too wide because it rules that these polarisation effects *must* be regarded as supporting Biot's corpuscular theory because no previous theory had had these effects as consequences. (Biot's explanation of these effects was in fact condemned as entirely *ad hoc.*) Any theory no matter how 'cooked up' can, on this theoretical view, gain support from the facts so long as it is the first to explain them. None of these cases which destroy the 'theoretical view' provides any difficulty for the 'heuristic view'. The Michelson-Morley result supports both Einstein's 1905 theory *and* Lorentz's 1904 theory of corresponding states, since the result was not involved in the construction of either theory,[11] and yet follows from both of them. On the heuristic account, Biot's theory was *not* supported by the various polarisation effects despite no other explanation of them being then available, because its explanation of them was *ad hoc*: all the essential features of the explanation were 'read off' from these already known facts. But Fresnel's account of these effects *did* derive support from them, despite the pre-existence of Biot's account, because Fresnel did not need to use these effects in constructing his theories.[12]

Despite his complaint about my ignoring it Musgrave himself no longer claims that the theoretical view is superior to the heuristic view, as he did in his earlier article. Rather he now thinks (above, p. 186) that these two views:

> may be complementary rather than competing ... The heuristic view enables us to determine the evidential support of a *single* theory. But when we wish to decide whether some theory is an improvement over its predecessor, then the theoretical view comes into its own. For then we will only count those facts which do not also support the old theory.

Musgrave's new move seems to me a conventionalist strategem of the worst kind. Those holding the various positions on empirical support would, *of course*, all agree that those facts which give one theory extra support over another are those which support the former but not the latter! Thus the proponent of the 'purely logical' position (against which Musgrave argued in his [1974]) would agree completely with Musgrave's new claim: those facts which provide *extra* support for Einstein's theory over Newton's are those which follow from Einstein's theory but not from Newton's. But there is a genuine dispute here which no talk about complementarity can mask. The dispute is over when a *single* theory derives support from a given fact. On the heuristic account a fact may follow from one theory and not from its rival and yet may not provide support for the first theory. But such a fact is bound to support the first theory according to the theoretical view. For example, as I pointed out, Biot's 1816 corpuscular theory was not, on the heuristic view, provided with extra support over its wave theoretical rival by its explanation of polarisation phenomena, despite the fact that no wave theoretical account of these phenomena then existed. Thus the heuristic and theoretical accounts of empirical support are inconsistent and not complementary: Musgrave's policy of 'strategic withdrawal' does not succeed.

2. THE DUHEM PROBLEM

In this section I shall again consider a specific scientific theory as consisting of a basic theory T conjoined with a set of less basic, 'auxiliary' assumptions A. Assume that some such scientific theory is inconsistent with some factual statement. Noretta Koertge hopes to find a methodology which will demarcate those situations in which the better or more promising or more rational solution of the inconsistency is for scientists to look for replacements for T, from those situations in which the better solution is for scientists to look for replacements for A.[13] I wish her luck, but I do not think she will succeed. Certainly if one must produce such a demarcation in order to solve the Duhem problem, then MSRP does not solve this problem. More especially, MSRP does *not* provide the crazy solution of the Duhem problem attributed to it by Noretta Koertge. She says that MSRP implies that, if a theory of the form $T \wedge A$ is inconsistent with the evidence, then "if T is the hard core of a research programme, one should *always* keep T and replace A".[14] But MSRP would be mad to give this advice, since if it were consistently followed it would endow the first research programme in any field with an eternal monopoly. I shall explain what MSRP does say about "holding on to hard cores" in a moment. First, given that it doesn't provide a solution of the kind sought by Koertge, let us see whether MSRP provides *any* sort of solution of the Duhem problem.

As Lakatos often stressed,[15] MSRP consists of a set of criteria for appraising already-articulated programmes. Like Popper, Lakatos explicitly eschews the hope of providing an acceptable methodology of the old Bacon-Descartes kind – *i.e.* one which provides a *method* for finding the truth. MSRP therefore gives only the following solution of the Duhem problem: it places no restrictions on the way a theory may be modified in the event of a clash between theory and evidence; however, once the modified theory has been produced, MSRP's rules will tell whether or not the new theory constitutes progress over the old (and if various modified theories are produced MSRP's rules will be able to say if one constitutes scientific progress over all the others).

But can MSRP say nothing about which part of a theory is *likely* to be modified in the light of clashes with experiment? I think it can say a little. According to MSRP, the development of modern mature science has consisted of the rivalry, not of mere theories, but rather of research programmes. The most important distinguishing mark of a research programme is its heuristic. This will consist in part of mathematical techniques for formulating, and drawing consequences from, the theories produced by the research programme. Another part of the heuristic may consist of techniques for resolving anomalies. When scientists, as Lakatos puts it, declare some of their assumptions "irrefutable by methodological fiat", they do not do so arbitrarily. Restricting themselves in this way would obviously be wrong, unless it brought compensating benefits. It is clear from the nature of a heuristic what these benefits are. If a theory produced by a programme is inconsistent with the result of some experiment, one can modify either the hard core assumptions or the auxiliary assumptions. The former course will involve abandoning the whole programme. The programme's 'hard core' will (in part, at least) underpin its heuristic: abandoning the 'hard core' will also involve abandoning the heuristic. One will then have to build up a whole new programme (unless a new programme is already to hand). This is an enormous undertaking which may involve, amongst other things, the development of entirely new mathematical techniques. By taking this course one puts oneself, at least at first, in a theoretical void. On the other hand, if one sticks to the hard core, and to the programme, and to the programme's heuristic, one's search for modified auxiliary assumptions will be guided in various ways by that heuristic. No wonder then that some theories (those which form integral parts of powerful research programmes) have had such long lives, surviving many clashes with experiment, in which the "arrow of the *modus tollens*" has been directed away from them and towards auxiliary theories.

Thus MSRP explains why it is usually the auxiliary A's and not the 'basic' T's which are modified because of the theory-experiment clashes. The methodology does *not*, however, say that this is what *ought always* to happen, nor even that this is what *ought* usually to happen. The methodology points to the enormity of the step which those who intend to abandon some 'basic' theory T must take, but it cannot advise scientists against taking the step. It is precisely by taking it that certain geniuses like Newton or Fresnel or Einstein have advanced science.

Finally, I shall remark on some of Alan Musgrave's criticisms of MSRP's solution of the Duhem problem and, more generally, its account of when one scientific theory is better than another. I want simply to take up two points, one of them rather minor.[16]

Musgrave agrees with large parts of the MSRP position, but thinks it is overly anti-falsificationist in a few respects. One of his charges rests on a partial misunderstanding of my position. I claimed that MSRP appraisals, unlike Popper's corroboration appraisals, can distinguish between *some* pairs of theories which are experimentally refuted. This is because one theory may derive genuine support from more facts than the other theory. In other words, the set of facts which genuinely support one system may properly include the set of facts which genuinely support the other.[17] Thus, following Lakatos, I described MSRP as "transferring the methodological spotlight from refutations ... [to] verifications of excess content" (above, p. 52). Musgrave attacks this slogan without caring about the full description for which it was meant to act as shorthand. His attack (on p. 184) consists of considering two theoretical systems S_1 and S_2:

> Suppose S_1 is corroborated in all tests except one, while S_2 is refuted wherever S_1 is corroborated but corroborated in the single test which refutes S_1. All the examined content of S_2 is excess content, and some of it is corroborated.

Now I agree with Musgrave that it would be 'totally unacceptable' to maintain that S_2 has more evidential support than S_1. But I cannot really understand why Musgrave thinks I would maintain this. As he says (*ibid.*) S_2 is not "supported by more facts than its rival". Since support by more facts is the criterion I advocated, I would have thought it obvious that S_2 is *not* better supported than S_1 on my criterion (nor is S_1 better supported than S_2). But, more importantly, Musgrave is wrong to think that this example shows that in assessing the evidential support of S_1 and S_2:

> we must take into account the refutations as well as the corroborations of each (p. 184).

Musgrave may prefer a system in which refutations are taken into account, but then he must argue for it. Certainly it is not true that we 'must' take refutations into account here, for MSRP explains, without considering refutations, why neither is S_1 better supported than S_2, nor *vice versa*. The set of facts which support S_2 is not a subset of the set of facts which support S_1 and nor is the set of facts which support S_1 a subset of the set of those facts which support S_2.

But Musgrave has a much more interesting claim about over-zealous anti-falsificationism within the ranks of MSRP-supporters. This concerns the stress on the large degree of theoretical or heuristic autonomy which some programmes have. No doubt, Lakatos in some ways overreacted to the claim that "scientific theories are often overthrown by experiments, and the overthrow of theories is indeed the vehicle of scientific progress", and overstated the case for heuristic autonomy. In particular, I agree that it is a mistake to think that a 'positive heuristic' can *predict* that certain theories will face anomalies. (Although Musgrave here endows Lakatos with a more extreme position than his writings warrant; Lakatos never says that heuristics can predict the *precise* anomalies a theory will face!) Nevertheless, there is in Lakatos's position, an important kernel of truth, to which Musgrave, to judge from his comments, remains blind. I touched on this kernel when I claimed that a heuristic "will often point to shortcomings in existing theories in the programme ... *quite independently of any empirical* difficulties". Musgrave quotes this disapprovingly (p. 189). Let me try to defend this claim by means of a simple example.

By the time of Fresnel's work in optics, the law of the reflection of a light ray (that the angles which the incident and the reflected rays make with the normal are the same) had been an accepted experimental law for centuries. The law was considered as completely confirmed and entirely free from empirical difficulties. Huygens had, in the 17th Century, produced a wave theory of reflection. Since this theory implied the correct law of reflection, and nothing else, it can scarcely be said to have been in empirical difficulties. Yet Huygens's theory was unacceptable as Fresnel clearly saw. This was because a wave optics *programme* existed whose aim was the explanation of all optical phenomena on the basis of the straightforward mechanics of elastic media. Huygens's theory involved the assumption of a full, particulate ether in which the motion which a disturbed particle passed on to those in contact with it was 'infinitely feeble' in all directions, except one, *viz.*, the direction of the straight line joining the disturbed particle to the original source of the disturbance. This obviously had to be wrong, unless the ether had the most weird and wonderful mechanical properties. Fresnel saw this and produced instead a theory which allowed particles disturbed from equilibrium to pass on this disturbance in 'an infinity of senses'. Fresnel's theory (which also took interference into account) not only explained the success of the usual law of reflection (by yielding it as a 'limiting case'), it also *corrected* it at various points. That is, it yielded entirely different predictions for certain cases, principally where the reflecting surface was very narrow. (These hitherto unsuspected predictions were subsequently confirmed.) Here then it was the heuristic of the programme and not anomalous experimental results which pointed to defects in an existing theory. (*Of course*, once Fresnel had seen that Huygens's theory was unacceptable, and had constructed a new theory more in line with the heuristic of the wave optics programme, the new theory had to be tested against experience.)

Hence Musgrave's 'anti-empiricist' and 'anti-falsificationist' charges are valid only insofar as MSRP does not hold that empirical refutations are the only, or even the principal, driving force of progress in the developed sciences.

London School of Economics

NOTES

* My thanks for critical comments on a previous draft of this paper are due to Peter Clark, Greg Currie, Peter Urbach and John Watkins.

[1] Professor Scheibe also made some interesting points. I entirely agree with his main criticism (which was also expressed by many of the Kronberg symposiasts). This was that the MSRP idea of empirical support has not been made sufficiently precise. Although we have, I think, characterised quite clearly a few specific ways in which facts can be 'used' in the construction of theories ('parameter adjustment' and 'monster adjustment' - see particularly pp. 345–6), we have not yet given a *general* account of what it means for a fact to be involved in the construction of a theory.

[2] Not entirely correctly as Grünbaum persuasively argues above, p. 120 (cf. Watkins's comments below).

[3] Koertge introduces considerations of 'plausibility' (p. 262) and Post refers several times to his earlier attempt to characterise a theory's 'simplicity' (p. 318, note 7). Indeed I find it difficult to understand why Post is so much against our addressing the problem of *ad hoc* explanations when he confronted the *very same* problem in his earlier [1960] paper on simplicity. There (p. 32) he gives the problem the following clear formulation (his *solution* of the problem is, I fear, much less satisfactory):

"The merit of a scientific theory is judged not only by its logical consistency and its correspondence with experience ... (Indeed many crank theories, precisely the ones most difficult to eliminate, would qualify for acceptance under these two criteria!)"

[4] See above, p. 313.

[5] See above, p. 268.

[6] See Lakatos [1974].

[7] Koertge expresses some doubt about whether or not the criterion is, in fact, time-dependent. I had hoped that this was clear. I said, following Zahar, that if a hitherto unknown fact was first discovered as a result of its being predicted by a theory, then this is a sufficient condition for the fact to support the theory, but it is *not* a necessary condition. (See above, p. 49.)

[8] Musgrave [1974].

[9] Some confusion may have been caused by the fact that I sometimes speak of empirical support as a two place relation between a fact and a *research programme*. But as I pointed out above, these are just two equivalent ways of expressing the same thing. For a research programme at a particular stage of its development is characterised by a *pair* of entities: this research programme's heuristic and its latest theory. Thus again *three* entities are really involved in the empirical support relation.

[10] A case which *seemed* for a while to be such an exception is that of the wave theory

of light and rectilinear propagation. It seemed, *e.g.* to Newton, that *no possible* wave theory of light could explain rectilinear propagation. In such cases 'consistency proofs' may be very important.

[11] See Zahar [1973].

[12] See Worrall [1976].

[13] See her contribution above, (especially pp. 261–267). In a much earlier and famous paper, Adolf Grünbaum pursued the same hope. His solution was that the second of the two above courses was the correct one, whenever the (posterior) probability of A was extremely high. This solution seems to me to suffer from two defects. First there is no generally accepted inductive logic to provide us with values for the probability of A in the light of the evidence. Secondly, if there were such an inductive logic it would surely give high probabilities to 'well-entrenched', 'plausible', 'well-tested' theories. Yet many major scientific innovations have involved the overthrow of precisely such theories.

[14] Above, p. 260.

[15] See particularly his [1971], p. 92.

[16] There is one other point which Musgrave makes here, and on which I think he is quite right. I should have been more careful *always* to make explicit the distinction between a 'basic' theory as I called it above and a full theory or theoretical system, consisting of the 'basic' theory together with auxiliary theories. Only the latter can be directly inconsistent with experimental reports. But I did stress this point at least in my Section 3 (p. 52), and I certainly never followed Lakatos in talking about it being permissible to 'ignore' refutations of theories. This locution gives MSRP an unwarranted anti-falsificationist air. All refutations constitute problems. Lakatos did, however have an important point even if it was infelicitously expressed. Lakatos saw that the driving force for science was in large part provided not by refutations of existing theories but by much higher level 'heuristic' considerations. (See p. 336.)

[17] I assume throughout that the facts we are concerned with in each particular case are those 'atomic facts' with which scientists confronted by the two rival theories would be concerned. This blocks the kinds of construction which otherwise would show that the subset relation discussed in the text holds in no interesting case of theory comparison.

The notes refer to the unified bibliography for Parts I and III on pp. 379–383.