REALISM, HISTORY AND THE QUANTUM THEORY:

PHILOSOPHICAL AND HISTORICAL ARGUMENTS FOR REALISM AS A METHODOLOGICAL THESIS

ROBIN FINDLAY HENDRY

LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

OF THE UNIVERSITY OF LONDON

NOVEMBER 1995

ABSTRACT:

REALISM, HISTORY AND THE QUANTUM THEORY:
PHILOSOPHICAL AND HISTORICAL ARGUMENTS FOR REALISM AS A METHODOLOGICAL THESIS

Scientific realists and non-realists disagree over the reach of scientific knowledge: does it extend beyond the observational realm? Intuitions about abductive inferences are at the heart of many realist positions, but are brought into question by the non-realists' contention that theories are underdetermined by data, and the alleged circularity of realist attempts to show that such inferences *are* reliable. Some realists have tried to *circumvent* this problem by constructing *methodological* arguments for realism: if realism is embedded in scientific practice, the realist's picture of science might provide the best explanation of scientific success. Some non-realists reply by again pointing to the circularity of this strategy, which relies, again, on an abductive inference. Others deny that scientists *do* adopt realist stances. A *methodological* realist position is constructed: realist constraints on the acceptance and pursuit of theories—for instance requirements of intertheoretic coherence, and the avoidance of *ad hoc* explanation—*have* often contributed to progress in science. The position is immune to non-realist worries about the circularity of realist arguments, for it is a thesis about how science is *practised*, not the kind of knowledge it provides.

The argument is pursued within a diachronic account of theory appraisal: Imre Lakatos' *methodology of scientific research programmes* (MSRP) examines the principles that govern the *construction* of theories, and provides criteria—achievement of progress—for the appraisal of research programmes. Although Lakatos may have seen these selection criteria, when fulfilled, as symptoms of something *else*—the fulfilment in the theory's development of some ideal of scientific honesty—achievement of Lakatosian progress can serve as an end in itself. The realist methods mentioned in the last paragraph are then appraised as *means* to this end.

Since the position has a methodological formulation and background, it is applied as a *historical* thesis to case studies in line with Lakatos' *meta*methodology. These comprise two explanatory forays into history: the consistency of Bohr's 1913 model of the atom, and the construction by Heisenberg and Schrödinger of the two original formulations of quantum mechanics. There follows one contemporary application: the construction of explanations in quantum chemistry using approximate models of molecules.

CONTENTS

INTRODUCTION

Among the arguments constructed, criticised and disagreed over by scientific realists and anti-realists are the methodological arguments for scientific realism. The common premise of these arguments is that scientific procedures presuppose the reality of the entities and processes invoked by theories. If science is a successful activity, scientific realism can be inferred as the best explanation, or inferred as a thesis that has been *vindicated* by the success that its assumption delivered. Although methodological theses may not be suitable as the premises of arguments that will convince the anti-realist of scientific realism, the central aim in what follows is to explore the common theme of these premises—methodological realism—and assess it as an historical claim about the construction of successful theories. Arguing in its favour, I propose an account of theory construction central to which is the *intended interpretation* of the equations that express a theory: the intended interpretation provides the theoretical background from which *understanding* (rather than calculating ability) is drawn. The realist's approach to theories might be methodologically important in two ways. Firstly it might have *heuristic* value: for instance, where two theories are consistent according to their intended interpretations, realism provides the rationale for directing attention to models of the conjoined theories. On the other hand, there might be a *regulative* advantage to realism: in pursuit of *realist* aims (unification, explanation), scientists might expect more of their theories than if they regarded their theories as mere tools for prediction. Thus, for instance, if two theories are *in*consistent when interpreted realistically, efforts to achieve consistency must be of prime importance to realists. Only *observational* inconsistency should concern the instrumentalist scientist. Methodological realism will be supported as an historical claim if realist aims and methods have been pursued and applied in successful episodes in the history of science.

The text divides into two sections: the aims of the first—comprising chapters 1 and 2—are philosophical. In chapter 1, Lakatos' methodology of scientific research programmes is defended as a historiographically illuminating account of theory construction. It is then set in a metamethodology that will ground a methodological conditional with a pragmatic consequent: the growth of knowledge. The *antecedent* of this conditional—the realistic attitude to theories—is set up in chapter 2. Put together, antecedent and consequent are linked either by the heuristic and regulative mechanisms outlined in the last paragraph. The resultant conditional grounds for a pragmatic rationale for realism: realistic interpretation might be a good *means* to pragmatic progress *as an end*. This is methodological realism, or realism as a methodological thesis.

The second (shorter) section sets out to support methodological realism via specific historical counterfactuals, taking examples from quantum physics and chemistry. Chapter

3, which centres on Bohr's proposal of his 1913 atomic model, makes an existence claim for the *intended interpretation* in the case of the Bohr atom, and illustrates a regulative version of MR. The existence claim is based on Bohr's re-interpretation of his equations, which allowed him to explain new sets of facts without altering those equations. What must have changed, it is argued, is his intended physical interpretation of his equations. The proposal by Schrödinger and Heisenberg of two curiously similar theories in 1925 and 1926 occupies chapter 4. It is argued that the continued presence of intended interpretations explains three sets of historical facts: (i) the authors' heuristic routes to their respective formalisms; (ii) their taking positions in the debates over the correct interpretation of the joint formalism (and those positions themselves) even *after* the famous equivalence proofs; (iii) the continuity between (i) and (ii). The final chapter illustrates both heuristic and regulative versions of MR for molecular quantum chemistry. Quantum-chemical calculations start from models that are strictly *false* as descriptions of real molecules, for two sorts of reason: it is impossible in practice to enumerate (for instance) all the forces that act in and on a molecule; and even when one writes only as many of the initial conditions as *are* important, it is usually found that the resultant equations are insoluble. Instead, explanations are constructed with the help of *idealised models*. There are two realist lessons here. Firstly, the idealised models come from background classical theories, but the only rationale for their *introduction* into the quantum descriptions is as *approximately true* descriptions. Secondly, worries concerning those models that have been expressed in the quantum chemistry literature can be read as aspirations to apply MR in its *regulative* form.

Acknowledgements

I would like to thank my supervisor, Elie Zahar, for reading and criticising many earlier versions of this thesis. Daniela Slanickova has helped me in an uncountable number of ways all the way through its writing, and crucially aided its final production. She also explained to me the functionalist explanations that enter the discussion in chapter 1. Nancy Cartwright read and criticised an earlier version of chapter 3, and has heard and discussed proto-versions of other parts, especially chapters 2 and 5. Lastly, there is the Philosophy of Science Research Students' Group, whose seminars provided the forum in which the theses defended here were first conceived and argued over.

Research towards this thesis was funded by a British Academy Major State Award.

1

Pragmatic Progress: Scientific Realism and the Methodology of Scientific Research Programmes

The history of science alone can keep the physicist from the mad ambitions of dogmatism as well as the despair of Pyrrhonian scepticism. (Duhem [1914], p.270)

Introduction

Lakatos' philosophy of science is often presented as a corrective development of Popper's brand of fallibilism. While it can be misleading, this approach can provide a fruitful and historically accurate entry into the methodology of scientific research programmes (MSRP). It is misleading because Lakatos' *meta*methodology bore little resemblance to the *a priorist* view of methodology that Popper set out in his [1959], and arguably represented a significant advance. However, the approach is historically accurate because wherever Lakatos concerned himself with the empirical sciences, he actually *did* correct some aspect of falsificationism: hence the proliferating subscripted Poppers of Lakatos [1968c] and [1974]. Furthermore it is *fruitful* because one of Lakatos' chief innovations—the role of *heuristic*—was introduced in response to a central difficulty for Popperian falsificationism: the Duhem problem. This innovation and its consequences will be the central theme of this chapter.

The first two sections of this chapter will provide a detailed exposition of Lakatos' notion of heuristic, and a review of some criticisms of Lakatos' formulation. Suggested amendments will be noted along the way. In subsequent sections (1.3 to 1.7), Lakatos' metamethodology will take centre stage. Exposition will be mixed with some recent criticisms, and the nuances emphasised and the position modified where necessary. Thus in the context of a reply to Newton-Smith [1981], it will be argued that Lakatos' metamethodology appraises methodologies chiefly as a particular type of social theory (although Lakatos might have been less than happy with such a formulation). In 1.6, a naturalistic grounding for this metamethodology will be proposed, in response to recent critiques from Papineau ([1988] and [1989]). The key to that response is the reading of

Lakatos' notion of the "growth of knowledge". Since Hacking ([1979] and [1983]) has identified a philosophical background to Lakatos' work that would *block* the reading that the naturalistic grounding requires, an examination of Hacking's claims occupies 1.7.

1.1. DUHEM, QUINE AND RESEARCH PROGRAMMES

It would be useful to begin this section with a brief review of the Duhem thesis and its background, because it provided the problem situation that prompted Lakatos' switch to diachronic appraisal. The structure of that appraisal—and the units of scientific endeavour to which it was applied—will then be presented.

The Duhem Thesis

The Duhem problem arises because, as Quine famously put it, 'our statements about the external world face the tribunal of sense experience not individually but only as a corporate body.' ([1953], p.41) Quine concludes that 'it is misleading to speak of the empirical content of an individual statement' (p.43). If the statements of science are to be *appraised* through their empirical content via *modus tollens*, this semantico-logical thesis will enjoy *epistemological* import. In the face of empirical anomaly,

> Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. ... Conversely, by the same token, no statement is immune to revision.
> (Quine [1953], p.43)

Of course Quine admitted that the 'system' in question need not include the whole of our knowledge (or 'conceptual system'):

> The holism in "Two Dogmas" has put many readers off, but I think its fault is one of emphasis. All we really need in the way of holism, for the purposes to which it is put in that essay, is to appreciate that empirical content is shared by science in clusters and cannot for the most part be sorted out among them. Practically the relevant cluster is indeed never the whole of science; there is a grading off.
> ([1980], p.viii).

However, given a (possibly vast) negated conjunction, further guidance is needed in selecting one of the conjuncts for rejection.

Lakatos' longest discussion of the 'Duhem-Quine thesis' is in an appendix to his [1970] (pp.184-9), where he attributes *weak* and *strong* readings to Duhem and Quine respectively. On the weak reading, one must find another base for rational choice to supplement the empirical (Lakatos cites Duhem's 'sagacity'). On the strong reading, we must recognise (with Quine, in Lakatos' words) that: 'the Duhem-Quine thesis excludes any *rational* selection rule among the alternatives' ([1970], p.184). In reading Quine thus, Lakatos takes Quine's proposed supplementary criteria (simplicity and pragmatic conservation of ontology) to encode *non*-rational or psychological virtues. Now Quine often *does* use psychological language in his discussions of scientific method (see for instance Quine [1960], pp.19-25):

> Scientific method [is] a matter of being guided by sensory stimuli, a taste for simplicity in some sense, and a taste for old things. ([1960], p.23)

Although it is difficult to see quite as much space between Duhem and Quine on this matter as is discerned by Lakatos, there are crucial differences in presentation. Quine's concerns are primarily logical, and his suggested criteria *substantive*, *forward looking* and essentially *ahistorical.* Duhem, in contrast, expressed criteria that are extensionally similar as *methodological* judgements handed down by the 'good sense' whose application he thought to be so characteristic of the history of physics:

> Now, it may be good sense that permits us to decide between two physicists. It may be that we do not approve of the haste with which the second one upsets the principles of a vast and harmoniously constructed theory whereas a modification of detail, a slight correction, would have sufficed to put those theories in accord with the facts. On the other hand, it may be that we may find it childish and unreasonable for the first physicist to maintain obstinately at any cost, at the price of continual repairs and many tangled-up stays, the worm-eaten columns of a building tottering in every part, while by razing these columns it would be possible to construct a simple, elegant, and solid system.
> ([1914], p.217)

To the extent that good sense is 'vague and uncertain' (p.217), Duhem is aligned with Quine: the added criteria governing theory-choice fail to be categorical, and such choices are *still* underdetermined. However, Duhem peppers his discussion with examples from the history of science, and is therefore aware that although *logic* leaves us free to adopt any of an infinite number of responses to anomaly, hindsight may judge unequivocally:

> The day arrives when good sense comes out so clearly in favor of one of the two sides that the other side gives up the struggle even though pure logic would not forbid its continuation. ([1914], p.218)

At the risk of reading too much into such brief comments, we might conclude that Duhem implicitly separates two questions: (i) how we decide at some point in time which of the jointly-refuted conjuncts to reject; and (ii) whether that decision will turn out to have been 'correct' at a (much) later time. Considering only the first question, Quine concludes that underdetermination will *always* be with us. In adding the second, Duhem realises that choices which are underdetermined *anno* 1800 might not *practically* be so by 1900.

Research Programmes

In MSRP, the Duhem problem is answered by changing the unit of appraisal: instead of theories, *series* of theories are to be assessed. Imagine a historical sequence of theories that speak of (roughly) the same phenomenal domain: *as historical artefacts*, such theories appear in journals and books as sets of statements, often with associated mathematical formulae. Now consider the logical and conceptual relations that hold between these statements: are they logically consistent? Is there some subset of the content of each that is shared by every theory in the sequence? If we follow their authors when interpreting these statements, are they amenable to the same realistic interpretation? Is there a rationale that we can reasonably attribute to the creators of these theories that would *motivate* this sequence of theories? For some such historical sequences, positive answers indicate that something unites the series: a plan or *heuristic*.

Technically, a research programme consists of a series of theories, each member of which contains a *hard core* of basic theoretical assertions plus a *protective belt* of auxiliary assumptions. In his [1970], Lakatos presented the 'progressive problemshift' to the appraisal of research programmes as an attempt to capture the Kuhnian continuity that 'plays a vital role in the history of science' (p.132). The construction of particular theory-versions within a research programme is governed by its *heuristic*, a sharper replacement for the vague Duhemian notion of 'good sense'. The *negative heuristic* stipulates that every member of the series does indeed contain (assume or imply) the statements in the hard core. Thus where a change of theory is envisaged (perhaps in response to anomaly), reforming attentions are directed *away* from the hard core and *towards* the protective belt. Successive theory versions (or 'refutable variants') therefore differ in the constitution of their protective belts. The *positive heuristic* provides 'partially articulated suggestions or hints' (p.135) on how this armour of auxiliary theories should be built up. Together, the hard core and positive heuristic *constitute* the line of research, the refutable variants being its temporary manifestations. The chief intuition behind appraisal in MSRP is that the best research programmes will make significant contributions to the growth of empirical knowledge *drawing only on internal resources*, the key term being *progression*, as against *degeneration*. MSRP makes two requirements: (i) 'that each step constitute a *consistently progressive theoretical problemshift*' (p.134), in that each new refutable variant displays an increase in empirical content over its predecessor; and (ii) that 'the programme as a whole should also display an *intermittently progressive empirical problemshift*' (p.134), in that the novel content is occasionally corroborated. A research programme that satisfies these requirements during some historical period is said to be progressive at that time. Otherwise it is degenerating.

In an endnote to his [1971a], Lakatos specified three ways that 'degenerating problemshift' might occur, fleshing out the intuitive term '*ad hoc*' that was so beloved of Popperian appraisal. Lakatos' original formulation defined *ad-hoc*ness to be a relationship between a theory and its predecessors in the research programme of which it is a product, but Zahar [1978] redefined it as a three-place relation between a problem situation, a theory, and a heuristic. Lakatos' *specification* of the three types was also amended, in Zahar [1973] and Lakatos and Zahar [1976]. The final version is as follows: (i) a theory is *ad hoc*$_1$ if it has no empirical content beyond the problem situation that prompted its construction; (ii) a theory is *ad hoc*$_2$—with respect to a particular explanandum or problem situation—if it *does* have such excess empirical content, but this content has not been corroborated; (iii) finally, a theory is *ad hoc*$_3$ if the assumptions brought in to deal with the problem situation 'sit uneasily' with the hard core, invoking a different (and perhaps inconsistent) metaphysics, or 'do not form an integral part of the positive heuristic' (Lakatos [1971a], note 36). Now *ad hoc*$_3$ moves will turn out to be of some importance, and therefore deserve an example: Newton famously ascribed wave-like properties to otherwise corpuscular entities—the 'fits' of easy transmission—to accommodate the (theoretical) fact that the speed of light increases, rather than decreases, on entering solid media. Whether or not the amendment was *ad hoc* in either of the previous senses, it surely introduced an *interpretive* tension: if light enjoyed some of the properties usually attributed to wave motions, in what sense did the amended theory fit with the corpuscular foundations and models? Analogously, while earlier falsificationism could perfectly well *accommodate* the Duhem thesis, epistemic holism was a natural feature of Lakatos' methodology.

1.2. Heuristic and Appraisal

In the last section, the difference between an arbitrary series of theories and a research programme proper turned on the *heuristic* motivating the latter. Despite this central position, heuristic itself received only the briefest definition there. Likewise, the appraisal of research programmes was only cursorily covered. These notions will now be considered more fully.

Heuristic

Formulated by Lakatos, a heuristic is a set of directives which contains elements of two sorts: (i) the 'don'ts'; and (ii) the 'dos' of research.

(i) The negative heuristic *bans by convention* the construction of theories that do not contain some subset of the shared content of previous theories in the research programme, this subset being the *hard core*. The ban is *conventional* in at least two senses. Firstly, there is no good empirical reason for it *in advance* of successful research activity, although it may reflect a metaphysical commitment. Secondly, if the hard core is partly constitutive of the research programme, then where it is allowed to be 'falsified', the research programme must consequently have changed, much as marriage entails instant exclusion from the set of bachelors. Despite the 'conventional' tag, the hard core is composed of *factual*—albeit highly theoretical—assertions about the world. However, the label 'conventional' was the *methodological expression of a methodological decision*: to use the hard core as the basis for research. In a broadly sympathetic critique of MSRP, Musgrave [1976] doubts that hard cores *are* rendered irrefutable by methodological fiat. Newtonians, for instance, did not treat the law of gravitation as irrefutable: on a number of occasions during the long hegemony of their programme, eminent Newtonians considered adjustments to the supposedly sacred inverse square law in response to anomalous planetary orbits. From this example, Musgrave concludes that one of Lakatos' favourite examples of research activity as a programmatic affair fails to fit his description.

Musgrave then raises a logical problem: a 'refutable version' of a programme contains many assumptions. Why are only *some* of them treated as 'irrefutable', and how are they chosen? The worry is that, as formulated by Lakatos, MSRP 'gives carte blanche to any group who want to erect some pet notion into a dogma' (Musgrave [1976], p.465); Howson and Urbach ([1989], p.96) have made a similar objection. Now in one sense this objection is misconceived: hard cores are identified by the historical facts of their appearance in each of a *systematic* progression of refutable variants. They are appraised via the success of the associated programme. Where a group of dangerous methodological anarchists *do* seize on a dogma in this way, either their researches will be valueless and sterile, or, if successful, they will surely be *vindicated*. Of course this conflicts with the spirit of Popperian methodology, in which award of the label 'scientific' to a theory was made to depend on the state of mind of its adherents. This was not one of its greater strengths, and I think Lakatos was correct to revise it. Musgrave concedes that the spirit of Lakatos' use of the term 'hard core' was clear: hard cores would consist of 'deep and fertile hypotheses, which have the ability to stimulate important mathematical research and to emerge victorious from empirical trials' ([1976], pp.645-6). Thus assertions about voltmeters would be unlikely to find a motivating role at the heart of a durable research effort, but theories of the *ultimate nature* of electricity are a different matter. This, however, is already implicit in the requirement of fertility, because a (logically) stronger assumption will—*ceteris paribus*—be more fruitful than a weaker substitute.[1]

Although Musgrave's interpretation of MSRP is a rather narrow and literal one, the logical objection indicates that Lakatos' original formulation is surely somewhat idealised, and less

[1] Provided, of course, that the surplus content does not turn out to be misleading.

artificial accounts are available. Papineau [1979] offers a Quinean modification to the discrete structure of research programmes. Lakatos thought that a rigid *qualitative* difference between hard core and periphery was required to capture the relative longevity of core assumptions in Kuhnian normal science. In another context, Quine replaced the sharp distinction between analytic and synthetic with a notion of centrality that is a matter of *degree*. Analogously, the positive heuristic does not conveniently label some theoretical assumptions 'hands off' and others 'change where necessary'; some are simply *more* central than others. Any pair of scientists may agree on some assumptions and disagree on others, however central. To capture this diversity (which is in keeping with Lakatos' advocacy of proliferation), Papineau invokes a tree by way of metaphor:

> The trunk consists of those central assumptions common to all scientists in the field. The first branching of the trunk corresponds to those more basic points of initial disagreement which divide the community into a small number of groups. And so on with further branchings, until we get to the thinnest of twigs, to individual disagreements on the least central assumptions. ([1979], p.106)

Each trunk-branch-twig is a 'line' of research. The decision to replace a more or less basic assumption—to revise a line of research—rests on the scientist's individual judgement based on a complex of theoretical and empirical considerations. The relative stability of 'hard cores' is easily explained without resorting to methodological conventions: For a Newtonian, a change in the law of gravitation from $F_G \propto 1/r^2$ to $F_G \propto 1/r^{2.001}$ will entail recalculating every planetary interaction in the model of the solar system, and explaining the more serious anomalies that result, given that the existing model is fairly accurate. Positing the existence of a new planet will allow less radical changes. Continuity then arises because 'scientists should in the first instance always take the line of least resistance' (Papineau [1979], p.107). When these lines-of-least-resistance are exhausted, it is time to turn the attentions of *modus tollens* to more central assumptions. Of course which line of theory replacement *does* offer the least resistance will depend on individual metaphysical predilections. The fact that action-at-a-distance had always been incomprehensible to a number of Newtonians might then explain why amendments to the inverse square law *were* considered at regular intervals (see Musgrave [1976] for details, especially pp.459-63).

(ii) The positive heuristic provides advice on how to construct the protective belts for future theories. Thus it 'saves the scientist from becoming confused by the ocean of anomalies' by setting out

> a programme which lists a chain of ever more complicated *models* simulating reality: the scientist's attention is riveted on building his models following instructions which are laid down in the positive part of his programme. (Lakatos [1970], p.135)

For Lakatos, then, the positive heuristic is 'there as the strategy for both predicting (producing) and digesting [anomalies]' (p.136). If a physical system is modelled as (say) an idealised harmonic oscillator, a natural response to empirical difficulty is to introduce *an*harmonicity into the equations, although the *specific* form of the anharmonicity will depend on the particular application at hand. Like the negative heuristic, Lakatos gave the positive heuristic a methodological formulation: this is a necessary part of formulating a methodology, but it has misled some commentators (see for instance the objections of Newton-Smith [1981], raised and answered in 1.3 and 1.4 respectively). However, Lakatos' intention that the positive heuristic encode substantive content is clear:

> One may formulate the 'positive heuristic' of a research programme as a 'metaphysical' principle. For instance one may formulate Newton's programme like this: 'the planets are essentially gravitating spinning-tops of roughly spherical shape'. This idea was never *rigidly* maintained: the planets are not *just* gravitational, they have also, for example, electromagnetic characteristics which may influence their motion. Positive heuristic is thus in general more flexible than negative heuristic.
> ([1970], pp.136-7)

Now the degree to which theory-modifications 'fit' with advice derived from the positive heuristic plays an important part in the appraisal of each research programme: how could *metaphysical* statements provide such *methodological* advice? Zahar [1989] argues that behind the hard cores of powerful research programmes are metaphysical propositions of a high level of generality, citing as examples the *principle of sufficient reason* (expressed in mathematical form by symmetry requirements) and the *principle of proportionality of cause to effect*. In the context of theory construction, such principles can be read as *meta-statements*, or *stipulations* that equations should have a certain mathematical form. In similar fashion, the *principle of correspondence* requires that new equations are some transform of the equations of refuted but hitherto successful theories, such that the regions of phase space corresponding to the two sets of equations approach identity in the domains over which empirical adequacy *was* achieved by the older equations. Conformity to this principle ensures that the growth of empirical knowledge is continuous and cumulative, rather than revolutionary and revisionist. In conjunction with the more specific constitutive assumptions comprising the hard core, and any further relevant empirical constraints, these meta-statements may in certain cases *uniquely* specify the form of the equations that express the theory. Zahar ([1989], pp.28-33) has provided a detailed example of this in his derivation—*contra* Popper and Duhem—of Newton's inverse square law from Kepler's laws, plus the metaphysical theses expressed as mathematical constraints. Musgrave [1989] supports this *deductive* reading of the positive heuristic.

Postulating an evolutionary explanation for the emergence of such metaprinciples, Zahar takes them to be '*stable* in the sense that they preceded science proper, and have since remained largely constant' ([1989], p.33). He concludes that they must be 'mostly non-technical and often vague', because they have been used in *many* programmes of research:

> If they were all made precise and then conjoined, they might well contain contradictions. (After all, they may have arisen from a confrontation with very different physical situations.) The heuristic of a research programme is determined by the *coherent choice* it operates among these principles and by the more or less sharp formulation it gives to each of them. ([1989], p.33)

Thus only the *specific* forms of the metaprinciples—invoked by *particular* derivations—are precise. If theories can satisfy metaprinciples to greater or lesser degrees, or in different respects, then during the development of a research programme some theory-versions may be revised for purely *internal* reasons, to bring the programme *further* into line with the guiding metaprinciples. Thus 'empirical refutations, though very important, are not indispensable' (p.33). Now Lakatos agreed that the best (and of course *only* the best) theoretical developments are autonomous in this way, but in making the point with typical polemic, he perhaps overstates the case:

> One of the most important points one learns from studying research programmes is that relatively few experiments are really important. The heuristic guidance the theoretical physicist receives from tests and 'refutations' is usually so trivial that large-scale testing—or even bothering too much with the data already available—may well be a waste of time. In most cases we need no refutations to tell us that the theory is in need of urgent replacement: the positive heuristic of the programme drives us forward anyway. ([1970], p.151)

Although similar in effect, the Lakatos and Zahar visions of heuristic are quite distinct. The metaprinciples behind Zahar's positive heuristic are very general, while Lakatos' favourite example of programmatic research was the succession of increasingly-sophisticated Newtonian models of the solar system (cited throughout his [1968c] and [1970]) which indicates that he had in mind theory-development at a much *lower* level. The first refutable variants of the Newtonian programme were based on point-mass planets each interacting only with a stationary point-mass sun, while in the later versions, interplanetary interactions, spherical planets and a host of other improvements were introduced, all for internal reasons. These are low-level adjustments, specific to the constitution of the planets and their interactions; they are not comparable to the very general requirements of symmetry and correspondence. In fact, their function is *qualitatively* different: they add *further detail* to the models, making them *more accurate.* A positive heuristic that is powerful in this sense will do two things. (i) In its *determinate* or *de-idealising* role, it will indicate the respects in which present models are idealised, and suggest future improvements. (ii) In its *open* or *reactive* mode, it will provide an indeterminate list of possible responses to anomaly (for instance 'posit further planets' or 'recalculate deflection of star-images due to atmospheric refraction'). The "de-idealising" role for the positive heuristic explains why Lakatos called for leniency towards underdeveloped research programmes: 'to give a stern 'refutable interpretation' to a fledgling version of a programme is dangerous methodological cruelty.' ([1970], p.151) Leniency is not only merciful; to act otherwise would be to make a mistake:

> The first versions may even 'apply' only to non-existing 'ideal' cases; it may take decades of theoretical work to arrive at the first novel facts and still more time to arrive at *interestingly testable* versions of the research programmes, at the stage when refutations are no longer foreseeable in the light of the programme itself. ([1970], p.151)

In its determinate aspect, the positive heuristic provides a partially articulated vision of how the exact model should look, but this vision must be revisable, for the following reasons. If the auxiliary assumptions associated with the 'virtual' exact theory-version are *correct*, the development of new refutable variants—where each new refutable variant is a more detailed and complicated version of the last—will correspond to the primarily *mathematical* problem of writing down and solving the correct equations. Empirical 'anomalies' that arise before this task is finished could be accommodated by improvements to the model that would have been made anyway. So the perfect positive heuristic 'produces' and 'digests' anomalies—in advance—in the following sense: the best responses to anomalies are manoeuvres that are *not* mere responses to anomaly, but corrections of idealising assumptions that were (perhaps implicitly) recognised at the outset. However, if the 'virtual' exact model is *incorrect*—as is likely to be the case for any moderately complex system—we might expect *unforeseen* anomalies, for which recourse to reactive moves of type (ii) will be necessary.

Despite their transparently *ontological* nature, it seems implausible that assumptions as specific as those that are encoded in Lakatos' 'low-level' positive heuristic could have the evolutionary origin ascribed by Zahar to more general heuristic principles. Opinions about the constitution of the planets would confer scant survival value—and therefore reproductive fitness—on their holder. However, Musgrave [1976], McMullin [1976] and Redhead [1980] have suggested a more recent and prosaic origin for detail-increasing moves suggested by the positive heuristic: the analogies through which we *interpret* the equations that express theories such as Newton's.[2] Take the oscillator example: equations that are written down to describe frictionless simple harmonic motion are equally applicable to circular motion projected onto a line, motion at a point due to a passing wave and the motion of a pendulum. However, their interpretation as any one of these types of idealised system will suggest a *de*-idealisation that is distinct from those suggested by the other

[2] McMullin actually refers to *metaphor* rather than *analogy* ([1976] p.427).

analogies. Under this reading, the positive heuristic will stipulate *which* interpretive analogies are to enrich the equations, and the particular analogies chosen will direct the de-idealising moves. Pursuing the naïve example, a harmonic oscillator that is subject to friction will experience a damping effect due to its motion through the medium in which it finds itself. Pendulums, however, will experience resistance due to the motion of the arm in addition to that of the bob. Where equations accurately describe the motion of damped pendulums and oscillators, the terms that correspond to friction can therefore be expected to have different mathematical forms. Clearly this is what Lakatos must have had in mind when he argued that

> two specific theories, while being mathematically (and observationally) equivalent, may still be embedded into different rival research programmes, and the power of the positive heuristic of these programmes may well be different. ([1970], p.164 fn. 2)

In his [1976] and [1978], Musgrave takes issue with this strong construal of the positive heuristic as a *producer* of anomalies, and suggests that its profile be drastically reduced. The reasons are two. The first is that Lakatos is mistaken in thinking that it was the *positive heuristic* of the Newtonian research programme that directed the de-idealisation of the planetary models. Instead there was a

> strategy for solving, by a method of successive approximation, the difficult mathematical problem of calculating what Newton's theory asserts about planetary motions. Lakatos is quite right that 'falsifications' play no role in this process; for the various 'models' produced do not represent Newtonian predictions to be compared with the evidence, but are rather steps on the way to such predictions. (Musgrave [1976], p.469)

Furthermore, when no gross idealisations remained to be excised, and anomalies began to be accommodated via *reactive* moves (for instance the discovery of Neptune, and when the Newtonian programme hit the buffers of Mercury's perihelion):

> The heuristic did not just happen to run out of steam; rather, the logico-mathematical problem of deriving empirically testable predictions had been solved. ([1976], p.469)

In his [1978], Musgrave's criticism is more pointed:

> I do not think that a positive heuristic, however powerful, can *predict* refutations. Lakatos is led to make this surprising claim by confusing the logico-mathematical of deriving predictions with the empirical problem of testing them. The successive 'Newtonian models' which Lakatos describes are the result of trying to find out what Newton's theory predicts about the solar system by a method of successive approximation. ([1978], p.189)

On this point, I think that Lakatos can be defended against Musgrave's criticism. Musgrave assumes there to be a neat methodological distinction between 'exact' and 'crude' models: it was only from the exact model that 'real' predictions could be derived, and Newton's theory tested; earlier models were just 'steps on the way to such predictions'. If this distinction were to operate during the development of *real* theories, the constructors of theories would need some sort of *access* to the exact model. However, no Newtonian could have known in advance precisely what the 'exact' model of the solar system would look like: there are a very large number of bodies in orbit around the sun, and neither their number nor their masses have ever been calculated (setting aside the problem of calculating their orbits). Now it might be possible to construct a model that is 'exact enough' for observational purposes (and theory testing): some bodies in the solar system exert only negligible gravitational attractions and might safely be left out of the initial conditions of the planetary problem. However, there is always the possibility that a body that was thought to exert only a minor attraction might in fact significantly perturb a neighbouring planet's orbit (perhaps by having a larger mass than was hitherto allowed for). Also, the specification of the—idealised—initial conditions might just *leave out* an unknown body of significant mass, the discovery of Neptune being a case in point. If Newtonians were unable to access their exact model, and could not tell in advance whether an 'approximate' model is exact enough, then the model that figures in the appraisal of a research programme at a particular point in time will just be the *least* approximate one: the latest product of the positive heuristic. The 'exact' model, in effect, is what is left when the heuristic runs out of steam.

Musgrave's distinction between exact and approximate parts of the application of a theory implies that there are two *separate* lists of improvements—those that *definitely* need to be made in order to correct known idealisations, and those that *could* be made as 'natural' responses to unforeseen anomalies, but members of the latter will often need to be transferred to the former. Now if the serious term 'refutation' is reserved for the observational inadequacy of a prediction that is derived from the *exact* solution to an equation that corresponds to the *exact* boundary and initial conditions, *serious* refutations will not occur until the heuristic runs out of steam. Musgrave wants to strip what I have called the 'de-idealising' role from the positive heuristic, leaving it a residual 'reactive' task, but there is no way that the two can be separated *in advance* of research. Lakatos' construal of the undifferentiated positive heuristic fulfilling both requirements more closely reflects the *open nature* of de-idealisation.

Musgrave's second complaint is that—*contra* Lakatos—empirical refutations must, in fact, play a part in the development of research programmes. The argument for this is as follows. The content of the 'low-level' (detail-increasing) positive heuristic resides in analogies that are drawn between the system to be described (for instance Bohr's model of the atom) and some well-known system (the solar system, in Bohr's case). Since Hesse (for instance [1966], chapter 1) has pointed out that between the analogical relata there will be some 'positive analogy' (structural similarity), and some 'negative analogy' (structural dissimilarity), we can expect that the positive heuristic will lead us astray at times. If the strength of the positive heuristic lies in the degree of positive analogy, its power cannot be estimated in advance:

> For the question with any analogy is: "How far does it obtain?" (Musgrave [1976], p.471)

Musgrave concludes that only empirical testing can answer this question, and that theoretical development cannot therefore enjoy the autonomy claimed for it by Lakatos. Unfortunately, in making the useful point that many research programmes *do* require experimental input, Musgrave makes a *specific* claim that it took an *empirical* refutation to force Bohr to change his original circular electronic orbits into elliptical ones. This is incorrect: Bohr planned such sophistication in advance, before his seminal papers were even *published* (see 3.2 and 3.6 for further details). So complete theoretical *autonomy* can be conferred only on a research programme of the *great* heuristic power, the idea being that such a programme would successfully explain and predict interesting facts in a natural—non-*ad hoc*—way. This is most clearly the case when a theory is constructed on purely *theoretical* lines, prior (in logical rather than temporal terms) to any empirical input. Lakatos' favourite examples—the Newtonian programme for planetary orbits, Bohr's atomic model—did enjoy such power, but nowhere did Lakatos claim such power was a *common* occurrence. Lastly, although Lakatos *retrospectively* identified the early stages of Newton's and Bohr's research programmes as examples of powerful—and therefore autonomous—theoretical development, he was aware that this judgement could only be made with hindsight. Neither Bohr nor Newton could have been known how much—or which parts—of the analogical power of their programmes would turn out to be correct, but with hindsight *we* can see that their success was achieved by models that were constructed in accordance with preconceived plans, and developed in a systematic fashion. This does not, however, preclude useful—but fallible—forward-looking judgements of heuristic promise (see Urbach [1978] and Whitt [1992] for relevant discussions).

Normativity and Appraisal

From the sophisticated techniques of appraisal provided by MSRP, it would seem obvious that some sort of normative advice would be derivable. At any point in time, judgements can be made as to which research programmes are progressing and which degenerating. Where there is some overlap in the subject matter addressed by the latest refutable variants of two incompatible research programmes, they are said to be *competing* in that phenomenal arena. In such circumstances, the proper vehicle for research activity in the disputed area would seem to be the programme that is progressing. In an early presentation of MSRP, Lakatos is unequivocal: despite (a series of) 'crucial' experiments favouring the triumphant—progressive—research programme,

> the resistance may last for a long time, for the defeated programme may hold out with ingenious content-increasing innovations unrewarded with empirical success. It is very difficult to defeat a research-programme supported by talented, imaginative scientists. Alternatively, stubborn protagonists of the defeated programme may offer *ad hoc* explanations of the experiments or a shrewd *ad hoc* 'reduction' of the victorious programme to the defeated one. But such efforts we should reject as unscientific. ([1968c], pp.176-7)

Elsewhere in the same paper, however, Lakatos admits that

> Criticism of a programme is a long and often frustrating process and one must treat budding programmes leniently. ([1968c], p.183)

As with Duhem, only *hindsight* can provide the final judgement. Now there is an obvious tension between the stern and permissive comments. There is no reason not to extend the leniency of the second quote to *revived* versions of *old* research programmes, and to vigorous (but of course *theoretically* progressive) attempts to turn an elderly research programme from the path of degeneracy. On the one hand, one of the crucial mechanisms for progress in science is the replacement of degenerating research programmes by their progressive rivals: if science is to progress as fast as possible, the majority of scientists should be working on the most progressive programme in the relevant field. On the other hand, Lakatos frequently stressed the value of theoretical proliferation: a minority might do fruitful work on alternatives. It should be noted, though, that such endeavours are not 'unscientific' if the theoretical work involves *theoretically progressive* moves (see above).

Musgrave [1976] has disapprovingly observed a transition in Lakatos' comments regarding the dispensation of methodological advice to scientists, from the early aspirations to the later caution. Feyerabend ([1970] and [1976]) has gleefully noted the same positionshift, this time with approval:

> scientific method, as softened up by Lakatos, is but an ornament which makes us forget that a position of 'anything goes' has in fact been adopted. ... Such a development, far from being undesirable, changes science from a stern and demanding mistress into an attractive and yielding courtesan who tries to anticipate every wish of her lover. Of course, it is up to us to choose either a dragon or a pussy cat for our company. I do not think I need to explain my own preferences. ([1970], p.229)

If Feyerabend and Musgrave are correct, the later Lakatos resolved the earlier tension in favour of the permissive reading, apparently renouncing the right to issue any *binding* advice to scientists, and therefore any condemnation of those who ignore it:

> My 'methodological rules' explain the rationale of the acceptance of Einstein's theory over Newton's, but they neither command nor advise the scientist to work in the Einsteinian and not in the Newtonian programme. (Lakatos [1971b], p.174)

Lakatos then adds that the only advice he *does* give is that a public record be kept of the conceptual and empirical track records of research programmes, and that 'superseded methodologies should be ignored' (p.174); falsificationism, for instance, should be avoided. There are no absolute injunctions at the individual level:

> I, of course, do not prescribe to the individual scientist what to try to do in a situation characterised by two rival research programmes: whether to try to elaborate one or the other or whether to withdraw from both and try to supersede them with a Great Dialectical Leap Forward. Whatever they *have* done, I can judge: I can say whether they have made progress or not. But I cannot advise them—and do not wish to advise them—about exactly what to worry about and in which direction they should seek progress. ([1971b], p.178)

Musgrave [1976] provides a perceptive construal of this apparently curious position. Advice is community-directed because

> when you think about it, there is something rather odd about a *general* methodological position issuing in advice about what *particular* scientists should do. For what it is rational for an individual scientist to do will depend on a vast number of idiosyncratic factors: his training, his ambition, his (and other's) estimates of his ability, his colleagues, his equipment, the availability of funds, etc. etc. All that a general methodological position like Lakatos's can say to an individual is: "Whatever you do, be honest, and try to live up to the standards of good science". ([1976], p.488 n.73)

So MSRP might describe the criteria on which individuals fallibly appraise the prospects of competing research programmes, based on their heuristic power and previous track record. Public record-keeping is therefore crucial so that individuals have the relevant meta-empirical information with which to make such judgements. Feyerabend famously found this hilarious, arguing that

> one might comment on the futility of a point of view where a thief can steal as much as he wants, is praised as an honest man by the police and by the common folk alike provided he tells everyone that he is a thief. If *that* is the sense in which the methodology of research programmes differs from anarchism, then I am ready to become a research programmist. ([1976], p.216 fn.25)

Musgrave's reading allows a reply. It is not *futile* for the thief to be required only to 'openly declare his thievery', because the "common folk" may decide that they would rather buy their second-hand goods elsewhere: someone with a track-record of sharp-dealing might sell them a video-recorder lacking innards. Likewise, the community of scientists—*if* they are interested in Lakatosian progress—will turn away *en masse* from programmes with a history of botched-together theories, and work on progressive ones where these are available. On this account, MSRP would describe a two-component science. The majority would pursue something like Kuhnian normal science, except that Lakatos provides a rationale—judgements of Lakatosian progress—for their doing so. Meanwhile, an *avant garde* would throw rivals into the ring at regular intervals. Taken as a whole, this picture would explain how Lakatosian empirical progress would be maximised by science so practised.

Musgrave, however, complains that Lakatos concedes too much to anarchism. Advice—of an inductive character, based on the track-record of the competing research programmes—*can* be given, although it will be fallible. However, *fallible* advice cannot be *binding*, the danger is that where an Einstein complies with it, the 'Great Dialectical Leap Forward' might be discouraged. Advice that is *not* binding can provide no secure basis for the use of pejorative terms like 'irrationality'. So it appears that Lakatos has abandoned an ambitious attempt to secure wide-ranging—and binding—norms that are to be contravened only on pain of irrationality, and instead offers a vestigial rationality that requires only honest score-keeping and the awareness of the relative chances of success and failure:

> One may rationally stick to a degenerating programme until it is overtaken by a rival *and even after*. What one must *not* do is to deny its poor public record. Both Kuhn and Feyerabend conflate *methodological* appraisal of a programme with firm *heuristic* advice about what to do. It is perfectly rational to play a risky game: what is irrational is to deceive oneself about the risk. ([1971a], p.104)

One point should perhaps be conceded to Feyerabend while rationality is under discussion. Feyerabend begins his [1976] critique of MSRP by observing that a philosophical investigation of science (or any other practice) should address two questions:

> (i) *What is science?* How does it proceed, what are its results, how do its procedures, standards and results differ from the procedures, standards and results of other enterprises?
> (ii) *What's so great about science?* What makes science preferable to other forms of life, using different standards and getting different kinds of results as a consequence? What makes modern science preferable to the science of the Aristotelians, or to the ideology of the Azande? ([1976], p.203)

In addressing (i), MSRP would describe—as a sociological matter—the standards of appraisal *within* science. At times, Feyerabend seems to approve of its performance here, with some reservations (see for instance his [1976], p.202). However, Lakatos is accused of misappropriating the *general* term 'rationality' for ideological purposes, by assimilating it to standards that are relevant to just one community: modern scientists. Defining 'rationality' to be *just* conformity to the intellectual standards encoded in MSRP assumes that question (ii) has been answered, which it has not. Can we just *say* that it is irrational to strive for alternative values? The quick answer is to set aside the tricky question of *transcendent* rationality. Separating (i) and (ii) amounts to the extraction of means from ends, (i) asks what the ends *are*, and the means by which they are achieved; (ii) questions the *desirability* of those ends. Now if MSRP correctly describes scientific standards *as applied*, and explains how their application produces growth-in-knowledge-of-a-certain-sort, it might ground a conditional: *if* you value some commodity, state of affairs or social development that Lakatosian pragmatic progress maximises, *then* MSRP describes how its achievement has been achieved. The preferences of those who eschew pragmatic progress, and *do* prefer to pursue other aims—whatever aims astrological research *does* achieve, for instance—we can put down to taste.

In conclusion then, theoreticians should follow their heuristic noses. For some, a meta-inductive inference based on the previous success of a research programme will recommend its further development. Others will give greater weight to particular conceptual and empirical difficulties, and be spurred to embark on new lines of research that attempt to solve those difficulties. In this way, the load of research effort will be spread between conservative and revisionary approaches. Only subsequent history—in the form of empirical results not yet available—will decide which of the strategies will have been the 'right' one. Feyerabend is surely correct: the labels 'rational' and 'irrational' attached to individual research strategies add nothing to appraisal in this context.

1.3. NEWTON-SMITH ON LAKATOS

In his [1981], Newton-Smith devotes a chapter to a comprehensive critique of Lakatos' views on methodology. First note that Lakatos often took his methodology to be a descriptive and explanatory advance on its perceived predecessor—Popperian fallibilism—with respect to the history of science. Taking Lakatos at his word, Newton-Smith examines the descriptions and explanations that issue from MSRP, and argues that the descriptions are false and the explanations fail to explain. Turning to Lakatos' metamethodology, he finds this too to be inadequate: it will not pick out the methodologies that really explain historical transitions between theories.

MSRP, claims Newton-Smith, *misdescribes* some aspects of real scientific practice. Consider the creation by Newton of a procession of increasingly accurate models of planetary motion. Contrary to Lakatos' claim, the earlier members of this series of theories were not 'theories' in the same sense as their more highly-evolved descendants: it was known that they did not represent exact Newtonian descriptions of planetary motion. In calling the earlier versions 'theories', Lakatos mistakes theory *development* for the serious business of proposing finished theories:

> There is no reason to suppose that Newton seriously posited each model in turn and revised them in the face of observed anomalies. Newton no doubt knew from the start that the initial models would not do. The development of this sequence of models was simply the thought process whereby he arrived at a detailed model worth positing as a theory of planetary motion. (Newton-Smith [1981], p.81)

Thus an episode in the history of science—one of Lakatos' favourite examples—fails to fit the MSRP mould. This, however, is a minor matter, because Newton-Smith moves on to targets—the hard core and two types of heuristic—that are closer to the centre of MSRP's explanatory edifice.

First the negative heuristic and hard core: Lakatos was fond of pointing out that a serious problem for Popperian falsificationism was the unfortunate but fruitful dogmatism displayed by great scientists of the past. This he expressed in methodological form as the negative heuristic's injunction—licensed by Duhem-Quine holism—to deflect the 'arrow of *modus tollens*' away from the hard core. Lakatos even provided a positive rationale for such a rule: *don't kill off budding research programmes before they have a chance to bear fruit*. Disagreeing, Newton-Smith argues that scientists are *not* in fact dogmatically committed to their theories, because alternatives that conflict with widely-accepted theories—but are observationally equivalent—are *often* worked on by mainstream theoretical scientists. Secondly, their commitment cannot be conventional: the success of its associated theories provides reasons to accept—*evidence for*—statements in the hard core. Conventions would not get treated this way. Newton-Smith continues:

> In point of fact what should be recognised is that the scientist's faith is a faith that there is something important in the basic theoretical assumptions and not that those assumptions are exactly right as they stand. ([1981], p.84)

In place of the negative heuristic, a weaker constraint on theory-construction is proposed:

> while progress is being made, only those variants on the basic assumptions which preserve the observational successes of the programme should be explored. ([1981], p.84)

Newton-Smith next argues that on Lakatos' characterisation, the positive heuristic must fail in the crucial methodological role allotted to it: driving the pragmatic progress that constitutes success in MSRP. According to Lakatos' *strong* construal, the positive heuristic provides the 'preconceived plan' (Lakatos' words) by which theories are constructed and anomalies turned 'victoriously into examples'. For Newton-Smith, however:

> It is implausible in the extreme to suppose it to be characteristic of successful theories that they come equipped with this sort of advance warning system. ([1981], p.84)

This is because some anomalies *never* get addressed (where is the advance warning in *these* cases?) and because:

> Response to anomalies, empirical or conceptual, comes after the fact of their discovery. And so it should be. For it would be a most inefficient use of our intellectual resources to formulate now what our response would be to entirely hypothetical anomalies. ([1981], pp.84-5)

As evidence of Lakatos' *strong* construal of the positive heuristic, Newton-Smith points to his [1970] remark to the effect that two formally equivalent theories might be embedded in two different research programmes and be developed into theories with very different content. Unfortunately, on this strong reading, there is:

> no reason to think that the success of an *SRP* indicates anything more than the power of the heuristic itself to generate successful new predictions. ([1981], p.87)

Newton-Smith argues that this is fine for instrumentalists, but Lakatos—*qua* realist—needs support to accrue to the metaphysical hard core. It is, however, an inevitable consequence of the curious way that MSRP labels dogmatic acceptance of the hard core *conventional*: one doesn't seek evidence for conventions. Newton-Smith takes the hard core to consist of the 'primary, basic empirical assertions about the world' (p.86), but:

> Lakatos's invocation of conventions in this context has blinded him to the fact that the crucial problem is that of obtaining evidence for the approximate truth of theoretical assumptions. ([1981], p.87)

Mirroring Musgrave's [1976] critique, Newton-Smith detects an alternative understanding on the part of Lakatos' students, citing Worrall [1976] and Zahar [1973]. Identified by extension (*i.e.* examples given by Worrall and Zahar), the *minimally* construed positive heuristic is not strong enough to differentiate rival research programmes, and could not therefore explain their differential rates of progress. They would need to be distinguished instead by their hard cores, and the positive heuristic would therefore have 'no real role to play' (Newton-Smith [1981], p.87) in theory appraisal.

In MSRP, rival research programmes are comparatively evaluated by assessing the extent to which their successive theory-versions are *ad hoc*, which is in turn defined in terms of *novelty*. According to Newton-Smith, Lakatos originally had temporal novelty in mind, but was corrected by Zahar, who recognised that 'the explanation of a known fact can be as important in providing evidence for a theory as the generation of true novel predictions' (Newton-Smith [1981], p.87). The notion of heuristic novelty raises an irrelevant *psychological* question of which perceived problem-situation prompted the theory's construction. This is unnecessary and misleading, claims Newton-Smith: the same assessment can be made if we just weigh up the number and variety of facts explained: an *ad hoc* theory will be unsatisfactory on these grounds because it will fail to provide explanations of any number or quality outside the domain for which it was 'cooked up'. An old objection is then raised to the relative appraisal of theories by their explanatory and predictive powers by this method: where two theories are compared, Newton-Smith claims that MSRP

> is making intuitive use of a notion of the relative size of classes of successful predictions and successful explanations of known facts. ([1981], p.88)

If all such classes are denumerably infinite and therefore equinumerous, MSRP will fail to differentiate good theories from bad ones. Lastly, Lakatos' model is 'too simplistic' because it 'accords no role to conceptual evaluation' (p.89).[3]

Newton-Smith then assesses whether MSRP performs well in the three roles that Lakatos allotted to methodology: (i) providing a demarcation criterion; (ii) evaluating research programmes; and (iii) explaining scientific change. Newton-Smith argues that the first of these roles is 'pointless' (p.90), and figured only as a rhetorical device in Lakatos' denunciations of research activities of which he (following Popper) happened to disapprove. 'Good' versus 'bad' is a more useful dichotomy than 'scientific' versus 'pseudo-scientific', claims Newton-Smith, and turns to the evaluation of MSRP's machinery of evaluation. Interpreting methodology normatively, Lakatos envisaged journal editors and funding bodies refusing to cooperate with those who cling to degenerating research programmes. However, Lakatos had to admit that a programme that has been degenerating for many years may be 'turned around' by vigorous research activity and that a progressive one may run out of heuristic steam. If MSRP provides criteria for the normative assessment of theories, *it also dictates that they can never be applied.*

[3] See Feyerabend's criticism of Laudan's similar claim (Feyerabend [1981a], p.235 fn.12).

The third use for a methodology—explaining scientific change—brings us to Newton-Smith's objections to Lakatos' metamethodology, because for Lakatos, we chiefly appraise a methodology on its explanatory power *vis-à-vis* the history of science via a methodology of *historiographical* research programmes (MHRP), of which more in 1.6. Newton-Smith agrees with Lakatos that history and philosophy of science require 'mutual interaction' (p.92) but disagrees as to the form that the exchange should take. In line with his requirement that methodology be true to the historical facts, Newton-Smith argues that if a methodological principle is to explain an event in the history of science, we must have reason to suppose that the protagonists *actually* believed in it. Explanations provided by MSRP do not satisfy this criterion, because Lakatos assumes that where there was a transition from a theory T_1 to its successor T_2, we have explained the episode merely if we can show that T_2 is superior to T_1 according to *our* methodology.

For Lakatos, the *best* methodology is that which requires the least explanatory recourse to external factors in its reconstructed historical stories, where there is no independent evidence for interference. Claiming that Lakatos provides no arguments for this, Newton-Smith points out that there is some dispute over how important 'external' influences *really* are to good science. In other words, whether the history of science *does* require rational (as opposed to non-rational) reconstruction is *itself* an issue; one that Lakatos settles by *fiat*. Where two incompatible reconstructions are historically equivalent, we cannot decide in favour of the rational one *a priori*. We need independent evidence that it was *consciously* applied. To express these worries more succinctly, Newton-Smith offers two counter-examples to Lakatosian metamethodology as amended by Worrall [1976]:

(1) Suppose that a methodology M_1 to which *we* subscribe privileges a theory B over its rival A. Historically, scientists chose A. We can explain their choice by reference to external factors, but unfortunately we can find no other evidence for their presence. Under MHRP, this disconfirms M_1. Newton-Smith objects that past scientists might have used another methodology M_2: we and they may just *disagree* on "what makes a theory a good one". If we have an *a priori* rationale for M_1 as a normative theory, we certainly should *not* abandon it for flimsy historiographical reasons: MHRP's disconfirmational procedures are therefore unsatisfactory.

(2) Now consider a methodology M_1 that adequately explains theory-change in the history of science. MHRP would have M_1 being confirmed. If there exists an M_2 which is explanatorily equivalent, except that M_2 was actually employed whereas M_1 was not, MHRP's confirmation procedures must be equally misleading.

The Lakatos-Worrall methodology is inadequate because it makes no reference to the *actual* methodological beliefs of scientists, and requires only that a methodology concur with the 'intuitive' judgements of scientists:

> What is misleading about this picture is that there is no room in it for the scientists' own reasons for preferring one theory over another. They do not simply make intuitive judgements, they standardly give reasons for their preferences. We certainly have not explained their decisions unless we make reference to what they believe, which may not be what we believe. (Newton-Smith [1981], p.97)

An alternative metamethodology is advanced:

> In fact we need first to establish that there has been progress in science without the use of methodological principles. Having done that, we then need carefully to examine the history of science to see what principles have actually been operative in bringing about that progress. That is how one vindicates a methodology; that is, by showing that it encapsulates the principles that have in fact been followed in bringing about progress. ([1981], p.97)

The 'Popperian Dilemma' concludes the critique: Popper, argues Newton-Smith, cannot make MHRP-style appeals to progress because he is anti-induction, and therefore cannot claim that there *has* been progress in science. Although careful to distance himself from Popper on inductivism, Lakatos

> also fails to establish that following his methodology is a means to the aim of science which, with Popper, he takes to be that of increasing verisimilitude. ([1981], p.98)

Newton-Smith discerns an *admission* of this failure in Lakatos' decision to 'accept'—tentatively, of course—the metaphysical inductive principle that increase in *apparent* verisimilitude reflects real progress along the Road to Truth. Citing the famous Russell passage that Lakatos himself invoked in exactly this context, Newton-Smith complains that Lakatos has stolen this result, rather than earned it.

1.4. MSRP AS SOCIAL THEORY

There are parallels between the difficulties for Lakatos' views raised by Newton-Smith and Musgrave: both mix some very astute detections of difficulty with other objections that depend on very *literal* readings of Lakatos' often high-flown rhetoric. More concretely, certain of Newton-Smith's criticisms—particularly regarding heuristic—are very similar to those of Musgrave, and were hopefully answered earlier (in 1.2). Where Newton-Smith's formulation is slightly different to Musgrave's, I will risk repetition. Towards the end of this section, the most interesting of these problems—concerning metamethodology—will be answered by reading MSRP as a certain species of social theory.

First let us turn to the hard core and negative heuristic: Newton-Smith objects to the labels (i) 'dogmatic' and (ii) 'conventional' being applied to scientists' attitudes to the basic metaphysical assumptions underlying their research, because (i) they work on other theories, and (ii) collect evidential support for the core assumptions. Now the 'dogmatic' label was chiefly meant to distance MSRP from Popperianism: the theory-defending tenacity is *relative* to the completely *un*tenacious behaviour that Lakatos felt that Popper unreasonably expected of scientists. For (ii), there is a long tradition in which 'conventional' is applied to principles or beliefs whose provenance we do not wish to discuss for the moment. Popper, for instance, used this device in his [1959] to avoid a discussion of the status of his methodological views, so that their consequences might be explored. It was never likely to be a *stable* position: how could a phenomenon like science—that enjoys a *public* existence—be a matter for *definition*? Lakatos employs the device of conventions to side-step the rather uninteresting question of how new research programmes get to be worked on before there is any evidence in their favour. Newton-Smith's argument then conflates two senses of 'convention'. A statement can *be* a convention, lacking any real content (it doesn't make any *real* difference whether I measure distance in metres or yards). On the other hand, our *reasons* for adopting some contentual sentence might be conventional in the sense that we don't want to discuss its provenance *now*, we want to see where it gets us: *this* is the sense in which commitment to the hard core is conventional in MSRP. Moves of this type are at the heart of the 'quasi-empirical'—as opposed to 'quasi-Euclidean'—structure that Lakatos thought to be applicable to any intellectual endeavour.[4] The standard examples of hard-core assumptions make it clear that they are supposed to express very general *states of affairs*. For instance, the assumption that light consists of a wave process could not *be* a convention: it is plainly of a factual nature, so there *is* some point in gathering evidence for it, and realists would look to the success of its associated research programme for the premises of their evidential arguments. When Newton-Smith observes that

> the scientist's faith is a faith that there is something important in the basic theoretical assumptions and not that those assumptions are exactly right as they stand ([1981], p.84)

he purports to *correct* Lakatos' mistake. Lakatos, however, would have endorsed the sentiment: it sounds much like the faith that at some point the action of the positive heuristic will produce a theory that is correct.

[4] Lakatos [1967b] makes the case for this structure for mathematics, and defines these terms.

Now consider Newton-Smith's accusation that MSRP is false to the historical facts. Newton's earlier models of planetary motion, we are told, were not proposed seriously, and MSRP therefore errs in asserting that anomalies forced him to revise his theories. Unfortunately for Newton-Smith, history disagrees: Newton actually *did* on occasion amend his models in the face of (empirical) objections. So let us emphasise the 'seriously': Newton only amended approximations of his model of which he was *already aware.* Now this could mean either of two things: (i) Newton was already in possession of the more sophisticated models, and comparison with these revealed the crudity of the early models; or (ii) Newton had a plan governing the construction of models such that the improvements would have been made *without* the stimulus of anomaly. (i) raises a historical explanandum: why did Newton fail to publish the more sophisticated models in the first place? (ii) indicates the presence of a *positive heuristic* as described in 1.2. Newton-Smith, however, felt that it was 'implausible in the extreme' to suppose that there are constraints on theory-construction of this order of strength.

The positive heuristic is designed to encapsulate the difference between a *planned* response to anomaly—a response with a theoretical rationale—from an *ad hoc* response. For Newton-Smith this is either *implausible* in its strong form (*à la* Lakatos), or nothing special under its minimal construal (*à la* Worrall and Zahar). When a response to anomaly is appraised in MSRP, one central question is: can the anomaly be explained by correcting some *theoretical* defect of the model *of which we were already aware*? If so, the response is *well-motivated.* This motivating role is clearly what Lakatos had in mind for the positive heuristic when he claimed autonomy for Bohr's development of atomic models (represented by M_1, M_2, ...) in the research programme of 1913:

> The apparent refutation of M_2 turned into a victory for M_3, and it was clear that M_2 and M_3 would have been developed within the research programme—perhaps even M_{17} or M_{20}—without *any* stimulus from observation or experiment. (Lakatos [1970], p.149)

So Newton-Smith may be correct to argue that it would be 'inefficient' to formulate responses to as-yet hypothetical anomalies, but this is *not* how the positive heuristic acts in MSRP. The positive heuristic encodes *theoretical* developments, some of which—if the research programme turns out to be progressive—will imply non-*ad-hoc* responses to as-yet undiscovered anomalies. Whether those anomalies turn up *before* or *after* the addition of the sophistications that would 'explain' them is a matter of little interest. Now in order to know that Bohr's first (static nucleus) model is an idealisation, we first have to know that the nucleus actually *does* move, however insignificantly. Then we need some idea of how that motion could be given mathematical expression. In conclusion, we need a *plan* as to how the inaccuracy can be removed, the plan in question coming from the interpretive analogies that turn a *set of equations* into a *model.* In Bohr's case, the analogue was clearly the solar system, so that the substitution of (for instance) reduced mass for electronic mass was a very natural move. But to possess an implicit plan is not to possess the implicitly-planned: the content of the positive heuristic is formulated *as a set of guidelines* because it would be impracticable to write down descriptions of physical systems—like the solar system—that reflect a realistic interpretation of the core assumptions. Take the inverse square law: suitably interpreted, it 'says' that every free-floating dust particle in the solar system perturbs Mars' orbit. In place of a list of initial conditions of unfeasible length and complexity, MSRP invokes a list of sophistications: allowances for dust-motes would be unlikely ever to reach the top.

To Newton-Smith's complaint that it is the positive heuristic that is doing all the work, and that any success achieved by the programme should accrue to this *instrument for producing good theories* rather than to the factual statements in the hard core, there is the following reply: If the positive heuristic *does* encode factual content (albeit *potential* factual content), as was argued above and in 1.2, its use in MSRP need not be instrumentalist. Admittedly, Lakatos was ambivalent towards the realist construal of scientific theories. As a fallibilist, he thought that 'all hard cores of scientific programmes are likely to be false' ([1971b], p.175), and that science could only *reliably* produce increasing verisimilitude. Elsewhere, however, the *aim* of science was Truth (see his [1974], pp.253-9), so that the link to the *analytic* Popperian virtues was *synthetic*, supported by a "whiff of inductivism".

Newton-Smith's most substantial objection to appraisal in MSRP is that 'the explanation of a known fact can be as important in providing evidence for a theory as the generation of true novel predictions' ([1981], p.87), so that the temporal ordering of theory and observation cannot be significant. He reads Lakatos' acceptance of Zahar's [1973] amendment as removing the *historical* nature of the theory-appraisal calculation: replacing a definition in terms of *temporal* novelty with one in terms of *heuristic* novelty. Newton-Smith then replaces Zahar's version with a system of appraisal that compares the number and variety of successful explanations and predictions, and correctly points out that quantifying such virtues will be of no help when choosing among theories, because of the infinite size—for *all* theories—of the classes to be compared. An initial textual rejoinder to Newton-Smith is that he corrects a position that he *mistakenly* attributes to Zahar. Zahar's amendment did *not* dehistoricise theory-appraisal: it changed the *nature* of the historical context, replacing a temporal ordering with a *heuristic* ordering. Thus (for instance) the precession of Mercury's perihelion *would* count as evidence in favour of general relativity: although it was not *temporally* novel, it was *heuristically* novel with respect to Einstein's construction of relativity, because it played no role in the construction of that theory (see Zahar [1973]).

The next move in Newton-Smith's argument is to raise the problem of underdetermination against one of the few philosophers of science to place a cogent response to it at the heart of his philosophy of science. Lakatos was aware of the 'weight of evidence' problem, which is why he historicised the appraisal of theories in the first place:

> the dogma of independence of evidential support from prehistory is false. It is false because the problem of the *weight of evidence* cannot be solved without historico-methodological criteria for 'collecting' theories and evidence. Both the truth-content and the falsity-content of any theory contains infinitely many propositions. (Lakatos [1968b], p.394)

Lakatos' recognition of the general problems associated with comparisons of verisimilitude is no bar to its being assumed for progress *within* a research programme. Each new theory version T_2 implies the true consequences of its predecessor T_1 within the same research programme. In this case, we don't have to determine the cardinality of their true consequence classes to know that T_2 has greater verisimilitude than T_1; we merely have to note that T_1's truth content is a proper subset of that of T_2.[5] This inclusion relation will not generally hold between theories embedded in different research programmes, but Lakatosian *diachronic* judgements—comparing the relative progression and degeneration achieved by rival *research programmes*—provide a surrogate in such cases. There might be a problem were Newton-Smith to provide a reason to deny *heuristic* novelty a role in appraisal, but there are only the usual logical arguments that *temporal* order cannot alter relations of implication between sentences, and after Zahar's amendment, it is the *heuristic* order that is involved.

Now there *has* been some debate as to the relative value of prediction and accommodation (where 'prediction' includes prediction of *known* facts). Glymour [1980], Zahar [1983], Giere [1984] and Redhead [1986] provide probabilistic arguments to the effect that mere accommodation provides less support to an explanatory hypothesis than prediction (or even none at all). Howson and Urbach have argued on Bayesian grounds for the *independence* of inductive support from heuristic background, and that the correct methodological distinction is between theories with *no* internal rationale and *no* independent support, and theoretical models with a causal structure that *explains* the predicted phenomena, or enjoy support from other sources (Howson and Urbach [1989], pp.275-84). Two things should be noted here. Firstly, Howson and Urbach's distinction is captured in MSRP by the pejorative identification of theories as *ad hoc*$_3$. Secondly, on the Bayesian account, the problem is either pushed back to the 'independent support', or just lumped into the subjective prior probability (this is no problem in itself). Now suppose, for the sake of

[5] This is not to suggest that *Popperian* verisimilitude increases in transition from T_1 to T_2.

argument, that the canons of inductive reasoning *do* provide no reason to differentiate prediction and accommodation (as we have seen this is controversial): this does not imply that it is *incorrect* to take a theory's heuristic background into account, only that *Bayesian inductive logic* provides no reason for doing so. Past episodes of theory construction have *not*, in general, been beset by problems of underdetermination: what strategies for choosing among theories have been used successfully? If there are cases in which heuristic novelty was an issue, and the Howson-Urbach criteria (independent support and internal causal structure) are not applicable, it might be inferred that inductive reasoning *cannot* rationalise these episodes (Bohr's atom might be a case in point: see chapter 3). If such strategies are systematically *successful*, the logico-probabilistic independence arguments would need to be qualified by *ceteris paribus* clauses. In any case, if it is *progress* that is the aim of science, the prediction of *temporally* novel facts will be of pragmatic importance.

There were three things that a methodology should do, and according to Newton-Smith, the second and third tasks—evaluating research programmes and explaining scientific change—are not performed efficiently by MSRP. First consider evaluation: the inability to rationalise the *permanent* rejection of research programmes, I think, reflects a failure *elsewhere* in epistemology: the problem of induction's failure to be soluble. We can never show a research programme to have degenerated irreversibly because we cannot see into the future, and it would be foolish to write off research programmes in this way given their history of reinventing themselves. However, inductive appraisal based on the relative length of the period of degeneration can be given: it is pretty certain that (for instance) a revived phlogistonist research programme would not enjoy success. The same response serves for Newton-Smith's observation that Lakatos fails to establish increasing verisimilitude as a result of the application of MSRP's principles. What can be discerned in MSRP, however, is a rationale for continuing research on a programme that *is* producing increasing *apparent* verisimilitude, plus an explanation of how communities of researchers come to work on such programmes. There can be no *guarantees* that such success will be achieved, or that it must always signal an approach to the ontological order.

This last claim brings us to MSRP's performance in the third role: explaining scientific change. Newton-Smith's most interesting objection is that a certain criterion of explanation is not satisfied by Lakatos' rational reconstructions: that if we wish a methodology to explain specific theory-changes in the history of science, we need to know that the relevant protagonists *actually did* adhere to it. This sounds reasonable: how can a methodology M explain a scientist's transition from (say) T_1 to T_2 unless we have reason to believe that she or he passed through intermediate belief-states that correspond to appraisal in M? There are actually *two* requirements being made here: (i) *we* must see how belief in M ensured the transition, for instance by showing that T_2 is preferable to T_1—according to M—in the evidential circumstances known to obtain at the time of the transition; and (ii) we must know that the relevant scientist *believed in M*, so that the explicit adoption of *M* was the actual (or quasi-causal) driving force in the transition from a cognitive state corresponding to acceptance of T_1 to a cognitive state corresponding to acceptance of T_2. Note that both of Newton-Smith's counter-examples to MHRP hinge on the *latter* requirement.

Ryle ([1949], pp.28-32) made a helpful distinction: people know *how* (for instance) to ride bicycles without being aware *that* certain physical laws underpin balance. This does not stop those theories explaining why people are able to balance in the way they do. The thesis that *real* knowledge is knowledge *that*—which Newton-Smith appears to assume—is criticised by Ryle as 'intellectualism'. Most of the human population make inferences from *P* and 'if *P*, then *Q*' to *Q* without knowing that 'if (if *P*, then *Q*) and *P*, then *Q*' is a logical truth of the propositional calculus, or having any idea what *modus ponens* might be. Elsewhere in his [1981], Newton-Smith characterises scientific judgement as a *skill*: surely one can have a skill while entertaining *false* beliefs about how that skill works: one might believe *false* theories of how bicycles are balanced without this affecting one's ability to ride them. Returning to science, Newton claimed to have deduced the law of universal gravitation from the phenomena, but Duhem [1914] has argued on logical grounds that this was impossible. On a similar note, Koyré's [1965] study of Newton's disputes with the Cartesians reveals that his interpretational attitude to the law of gravitation was far removed from his espoused instrumentalist stance to 'final causes'. Now recent efforts have been made to rehabilitate Newton's explicit methodology (for instance Glymour [1980] and Laymon [1983]), but the aim of these reconstructions is to bring a subset of Newton's methodological utterances into line with *modern* canons, given his actions; explanations of type (i) therefore seem to be primary. Newton may just have been *misguided* about the processes by which he arrived at his theories. This is not to say that his inferences were 'unfounded' or 'unscientific', merely that his methodological utterances might have been a poor guide to the methodology he was *implicitly* following. In a similar context, Lakatos summarises the point eloquently:

> This little story, I think, bears out my pet thesis that most scientists tend to understand little more *about* science than fish about hydrodynamics. ([1970], p.148)

A distinction between the two requirements is therefore crucial. Newton-Smith is quite correct in pointing out that Lakatos' metamethodology requires only (i). Assimilate, for the moment, rationality to *instrumental* rationality: to vindicate *M* for some aim *A*, we might look for an instance of theory-change that promoted *A*, and show that following *M* would have resulted in the *same* transition, in that the successor theory *was* preferable to its predecessor according to *M*. The separate problem of explaining how the transition *actually* occurred would be solved by adding to these considerations a *social* explanation of the social fact that an *implicit* methodology that promoted A came to be adopted among scientists during the relevant period.[6] Would we also need to know that individual members of the scientific community consciously sought to promote A by explicitly adopting M? A useful analogy might be drawn with functionalist explanations of the role of religion in traditional societies: according to their *adherents*, the occurrence of religious rituals would be explained by the truth of the religions in question. For the *anthropologist*, however, the proper explanation— from the point of view of social theory—would invoke the social *function* of rituals, perhaps in promoting social cohesion. The explanation is completed by a description of some process of selection that acts to *spread* cohesion-promoting rituals among such societies. Newton-Smith just *assumes* that to explain the actions of individuals, we have to invoke their intentions. This requirement *rules out* explanations of action that appeal to false consciousness, and Newton-Smith presents no conclusive argument for such general preconditions on social explanation.

Turning finally to the two counter-examples to MHRP, Newton-Smith objects to a metamethodology that disconfirms a methodology M_1 (*qua* meta-historical thesis) where it fails to rationalise an episode in the history of science, because we might disagree *a priori* with the relevant scientists. Now the relevant question is: was that episode successful? Did it produce growth in knowledge? If not, then we can write the episode off (along with M_2) as either (i) *bad science* if the aim was progress, but progress failed to be achieved; or (ii) *some other activity* if the aim was different. If the episode *was* successful, and following M_1 would *not* have produced such success, we really *should* examine M_1, *and* its *a priori* rationale. After all, what use is a methodology that does not deliver progress? The intuition behind Newton-Smith's second counter-example is that M_1 should not be confirmed if another methodology M_2 was *actually* employed. One might wonder how we are supposed to have *access* to the relevant actors' heads. Given that we do not, we have to infer from the actions and utterances of the actors which out of M_1 and M_2 produced the progress. If M_1 would have produced the same advantageous result, there is no reason why it should *not* enjoy some confirmation: M_2 will in any case. The situation is analogous to familiar cases in which two rival theories enjoy support from an experimental outcome that both predicted. The solution, as ever, is to wait and see, hoping that an explanatory divergence will appear. In any case, Newton-Smith's intuition that one can adhere to a methodology that has been refuted by the history of science on the basis of an *a priori* rationale—the disagreement with MHRP that motivates his first counter-example—contradicts the naive meta-inductivism of his alternative metamethodology:

[6] See Hesse [1988] for a fuller discussion.

> we need first to establish that there has been progress in science without the use of methodological principles. Having done that, we then need carefully to examine the history of science to see what principles have actually been operative in bringing about that progress. ([1981], p.97)

1.5. PAPINEAU ON LAKATOS

In his [1988], Papineau argues that traditional 'Cartesian' epistemologies give an account of the correctness of beliefs such that 'rationality' does not amount to *getting the world right*, and that they are consequently unable to counter relativist critiques issuing from Kuhn, Feyerabend and recent social studies of science. The quote marks indicate that 'Cartesianism' is used here in a technical sense, signifying a nest of theses that usually imply the existence of universal standards of reason—over and above how people *actually* reason—that are accessible *a priori* to the unaided intellect. In his [1989], Papineau directs these arguments more specifically at Popper and Lakatos. Now it may seem curious for Lakatos to be included in the set $\{x: x \text{ is a Cartesian}\}$, even for the sake of an argument. After all, in his papers on the philosophy of mathematics, Lakatos set out to *oppose* what he called 'quasi-Euclideanism' (see his [1962] and [1967b]). Likewise, 'justificationism' was the target of extensive criticism in the papers on the methodology of the natural sciences that introduced MSRP, and in his [1971a] he proposed a 'quasi-empirical' methodology for methodology. However, Papineau has a specific argument that Lakatos' views on metamethodology do not make sense unless certain 'Cartesian' theses are assumed. If correct, the claim would show that MSRP is subject to the same relativist challenge as traditional theories of knowledge.

In his [1987], Papineau sets out a taxonomy for epistemological positions under which it is possible to be a scientific realist while adopting an anti-realist position at the level of *method.* Tensions that inevitably afflict such a position, however, are at the centre of Papineau's relativist challenge. For naturalists, the mind is a 'normal part of the natural world' (p.ix). The first disagreement with Cartesianism, then, concerns dualism: cut off from the natural world, the Cartesian mind has to spin a web of knowledge using only the 'givens'—reason and experience—to which it is granted infallible access. Given this isolation, 'rationality' is granted only to those belief-forming processes—arguments, in common parlance—that can be constructed from 'given' inferential steps. This engenders an 'anti-realism of method', according to which

> the principles by which people *arrive* at beliefs are in no need of justification. According to anti-realism of method it is acceptable for people to adopt false beliefs, and therefore perfectly sensible to ask, at that level, whether one person's beliefs are more correct than another's. But anti-realists of method

> insist that there is no sense to questions as to whether the standards by which people decide such questions are themselves legitimate as good methods for arriving at correct beliefs. ([1987], p.13)

If the 'correctness' of a belief is assimilated to its provenance in a preferred method, the key difference between the Cartesian and the naturalised epistemologist will be the grounds on which they prefer methods:

> The *Cartesian* theory of epistemology recommends that we should get our beliefs from *good arguments*. We should assent only to those beliefs that have been generated by logically valid steps from secure premises. An actual belief is justified just in case it issues from such an argument. ([1988], p.39)

On a naturalised account, however,

> the right technique for acquiring beliefs is simply to be a reliable belief-former, that is, to have belief-forming processes that generally produce true beliefs. Concerned believers should try to ensure that all their beliefs come from belief-forming processes that are reliable in this sense. ([1988], p.40)

There is a connection between this anti-realism of method and *traditional* instrumentalism: Papineau detects a 'strong tendency for Cartesians to reject *realism*' (p.46), because:

> They have to show from first principles that the standards of reason which (a) recommend themselves to conscious human minds are (b) guaranteed to produce beliefs that correspond to an independent reality. It is not at all clear how to show this. So the natural move is to reject (b) and embrace *anti-realism*. ([1988], p.46)

Now this anti-realist move makes '*reason* prior to *truth*' (p.46) because we now define success in judgement to be having good *Cartesian* reasons for a belief, rather than in the content of that belief corresponding to an external reality. In the absence of such an anti-realist move, the Cartesian must appeal to some mechanism which links up reason with truth-as-correspondence. Thus, for instance, Descartes himself invoked a non-deceiving God (for whose existence there was a *separate* argument, of course) to warrant reliance on 'clear and distinct ideas'. If we *do* follow the anti-realist move, there *are* no other criteria for beliefs being the right ones, other than that they are held for the right reasons, or have been generated by methods that answer to some universal rationality. This, however, is the problem: if other communities have *other* standards of rationality, it looks like there *are* no universal standards of human rationality, so in what way is this rationality *objective*? The Cartesian anti-realist, Papineau argues, must simply insist that there cannot be alternative rationalities. However,

> this blunt denial of any possibility of alternative rationalities will seem too quick to anybody working in the philosophy of science, to anybody concerned specifically with the rationality of scientific theory

> choice. For it is manifestly *possible* for people to suppose that modern astrology, or creationism, are supported by reason. And it would be tendentious at best to insist that such people must be crazy. ([1988], p.48)

Then, Papineau continues, a natural move is to restrict the sample. If scientists are typically 'mature, serious thinkers' (p.48), a theory of rationality could be distilled from the history of science:

> If there isn't anything more to the 'right' standards of rationality than simply the standards that come naturally to mature human thinkers, then how else should we identify those standards except by familiarizing ourselves with the habits of thought of the central figures in our scientific tradition? ([1988], p.49)

This rather 'puzzling' approach—a view especially associated with Lakatos and Laudan—only makes 'perfect sense when seen against the background of Cartesian anti-realism.' (p.49) From this position, Papineau concludes, we can see that recent sociology of science (e.g. Pickering [1984]), that sees complex *social* interaction and negotiation where rationalism sees scientific reasoning, is going to be a threat to this programme: here are other, 'ulterior' motives that might cloud the judgements of people we had hoped to use in a *definition* of disinterested rationality. If social pressures are effective causal influences on scientific judgement, the purity of scientific reason is at risk, and so, therefore, is the objectivity of rationality. It would be useful at this stage to rehearse the steps in the argument:

(1) Cartesians must explain the link between self-conscious rationality and truth as correspondence, which is a difficulty unless realism is dropped or a link-mechanism found.

(2) For anti-realists, reason is prior to truth, leading a Cartesian to equate the 'right' beliefs with those recommended by the 'universal standards of rationality'.

(3) If there are no *universal* standards of rationality, then in the absence of an external (i.e. truth-linked) characterisation of when a belief is 'right', the Cartesian must resort to the reading of any such standards from an elite: Great Reasoners in the history of science.

(4) A severe problem arises for this programme if sociologists can show that other (social) processes 'interfere' with the cogitations of the Great Reasoners, making the purity of scientific reason a myth.

In his [1989], Papineau pins the 'Cartesian' label specifically on Lakatos. He does this by arguing that MHRP *implicitly* assumes the Cartesian thesis that is central to his critique: that success-in-judgement is conformity to standards of rationality that are accessible to *a priori* reasoning. Popper is presented as the source of Lakatos' implicit anti-realism, where, to repeat a point made earlier, the anti-realism in question is *not* one side of the usual realism-instrumentalism dichotomy. Instead, it is a thesis about what success-in-judgement *is*. To be a realist in this sense is to think there is nothing more to good methodology than the description of a process in which minds—as natural entities—reliably make transitions from states in which their beliefs have *lower* truth content to belief-states with *higher* truth content, where 'truth' is not to be assimilated to warranted assertibility. Anti-realists in this sense think that something crucial is missing from the previous characterisation: *rationality*. Thus for Popper, the rational choice among unfalsified theories is—other things being equal—the most falsifiable, and therefore the *least* probable. Now Popper's attitude to methodological prescriptions underwent a transition from an early conventionalism to a later—albeit unsuccessful—historical meta-empiricism. Even success for the Popperian historiographical programme could only show that past science conforms to Popperian standards; we are still in the dark as to whether that past science came up with any *truths*. Thus the process of taking one's rationality from a revisable identification of good science might be internally coherent and non-circular, but 'the price of this strategy is abandoning all vestiges of scientific realism.' (Papineau [1989], p.436) Turning to Lakatos' vision of metamethodology, Papineau argues that it too is prey to the central objection: anti-realism of method.

First note that each methodology picks out a package of allowed methods or inferential moves. According to MHRP, each methodology will make its own distinction between: (i) episodes in the history of science in which theory-change outcomes were the product of *external* forces (inferential moves that are *not* contained in its package); (ii) those that were governed only by the 'rational' processes that *can* be broken down into 'allowed' moves. The methodology that 'rationalises' the greatest proportion of previous 'successful' science in this way is therefore the one to choose. If we *do* assimilate rationality to rationality-according-to-the-progressive-methodology, we must be equating rationality with the methods used by scientists during 'internal' episodes in the history of science. This is to assume that scientists are 'rational', except when social, political or economic circumstance perturbs the process. Lakatos must therefore be assuming that rationality is what minds exhibit when left to themselves:

> It's not just any human beings whose intellectual inclinations are self-justifying. Some humans will have their thinking distorted by ideology, or politics, or perhaps by simple shortage of time and resources. So to identify the ideal rational standards for human beings we need to turn away from such

> cases of external distortion, and look at how humans think when they are free of such handicaps. It is the pattern of thought displayed by the human mind when it is free from external influences that is constitutive of rationality. (Papineau [1989], pp.438-9)

Now remember that Papineau understands 'rationality' to be the choice of belief-forming processes that reliably conduce to truth. We can therefore ask whether the methodology chosen under MHRP will produce beliefs that are 'rational' in this sense. In other words, would following it be a reliable source of beliefs that correspond to reality? Of course there is no such guarantee, and so

> Lakatos's proposed strategy for evaluating methodologies makes little sense from a realist viewpoint. But if we reject any concern with truth, and look at things from an explicitly anti-realist perspective, then it seems perfectly cogent. ([1989], p.439)

Divorced from truth this way, methodologies that are approved by MHRP will still be subject to the relativist challenge that follows from (1) to (4).

1.6. MSRP NATURALISED

From an early stage in his development of a distinctive methodology, Lakatos deployed historiographical arguments in his debates with Feyerabend, Kuhn and Popper. In his [1971a], he proposed a metamethodological framework in which the argumentative strategies that he had hitherto used implicitly were provided with an *explicit* motivation. Now Papineau criticised Lakatos for turning to the intuitive judgements of a rational scientific elite for the foundation of his view of scientific objectivity. In what follows, I will give an account of Lakatos' theory that will attempt to show how it can be made consistent with Papineau's key requirement that methodology be amenable to realist construal.

Is naturalised epistemology in a better position than Cartesianism? For Papineau, a belief was *justified* just in case it was the product of a reliable belief-forming process. Such a process will be reliable if it typically arranges that the occurrence of a belief is brought about by its being *true*. It is not important whether *some* of the processes that place beliefs in scientific heads are social; we merely have to be sure that the relevant state of affairs in fact *did* have a hand in it:

> So there is a minimal demand to be made by the naturalized friend of science: the explanation of creditable beliefs needs to differ from those of beliefs in general at least to the extent of allowing that amongst the causes of those beliefs are the truth conditions of those beliefs. ([1988], p.52)

In one sense, this is a form of empiricism, except that we claim "I know that Uranium-235 decays with such-and-such a half-life because I arranged for my beliefs on this matter to become an effect of some actual episodes of Uranium-235 decay", instead of "I know that Uranium-235 decays with such-and-such a half-life because I *looked*". Papineau then argues that social constructivists fail to discredit science if they cannot show that social influences were in general *sufficient* for the acceptance of scientific theories. The causal challenge arises at two levels. Where large-scale political pressures have been sufficient for the acceptance of a theory—for instance when Lysenko dominated Soviet biology—we should indeed be suspicious of the theory in question. At the micro-sociological level, we need not worry that personal ambition, rather than a dispassionate and critical quest for truth, motivates scientific enquiry:

> No doubt it is true that scientists are often primarily concerned to attach their names to facts, and to build up the scientific credibility that will enable them to do this. But consider how they go about doing this. A prime concern, if they are to persuade others, is to ensure that their opponents won't be able to pick holes in their published claims. So scientists take care that their experiments are repeatable. They try to design experiments so that the move from the observed results to the desired theoretical conclusions depends on as few disputable auxiliary hypotheses as possible. ([1988], p.53)

At the third stage of his argument, Papineau accuses Cartesians of having to give a 'blunt denial of any possibility of alternative rationalities' (p.48), which is 'too quick', but in what sense is it possible for astrology or creationism to be supported by reason? Papineau objects to the use of the term 'irrational' to signal that anyone who disagrees with us about 'rationality' must *necessarily* be mad. Papineau's intuition is that where there is disagreement about which beliefs are rationally acceptable, there is probably also disagreement as to *facts*, but which *kind* of facts?

Suppose that there are reasons for a belief in astrology or creationism. These might be *rational* (that is, not obviously inconsistent): for instance, the belief that everything in the bible is literally true would be a good enough reason, if one believed it, for a belief in creationism. However, on Papineau's [1987] account, these criteria will fail to cohere with other beliefs that creationists (who are also ordinary people who make everyday inferences) are likely to hold concerning which belief-forming processes are *reliable*. What kinds of belief are relevant here? Given *true* beliefs about the history of the creationist and astrological research programmes, their proponents could not argue that they have enjoyed a history of pragmatic success. It may not be *irrational* to side with the creationists, but a creationist must surely recognise the long-term historical degeneration of creationism as a *scientific research programme*, rather than as an attempt to interpret the natural world through biblical doctrines. Obviously there might be a disagreement over ends: a creationist might not see pragmatic failure—or degeneration—as a *problem*, and there might be other—value-based or theological—reasons for believing everything that is in the bible (see Hesse [1980]). However, if astrology and creationism are supported by 'reason', it is surely not the *same* kind of 'reasoning' as that which supports a science that has manifested pragmatic success, as the other two miscreants have not.

In his [1982], Newton-Smith provides a closely-argued critique of relativist theses about reason and logic as explanatory with respect to the behaviour of others. His point is simple: that relativity of reason implies relativity of facts, which cannot be coherently conjoined with truth-as-correspondence without an inflation of different worlds to which different conceptual systems correspond. Perhaps I should fill in the arguments. Relativism about reasoning can either be relativism about logic, or relativism about 'reasons for thinking that *p*' in the wider sense. The implication is immediate for relativism about logic, because for every valid inference there is a logical truth. Relativism about reason in general is formulated schematically by Newton-Smith as an assertion of the form '*R* is a reason for holding *p* to be true for ψ while *R* is not a reason for holding *p* to be true for ϕ' (Newton-Smith [1982], p.110). To make such an assertion interesting (i.e. to avoid triviality), the situations must be relevantly similar. Low-level 'reasons' then cannot vary, because if I hold *R* to be a reason to believe *p*, then I think there to be a truth linking *R* and *p*. Taking Newton-Smith's example, litmus paper turning red is taken to be a good reason to think that acid is present just because it is thought to be *true* that acid causes litmus paper to turn red. So it appears that relativism about reason(s) implies relativism about truth. The contrapositive of this conclusion—that non-relativism with respect to the facts that support 'good reasons' implies non-relativism with respect to those reasons—suggests an argument against the cultural relativity of reason and method.

In conclusion, it seems that low-level facts may impinge on such high-level debates in two ways. Firstly, there are the causal relations invoked by Newton-Smith that link up the facts whose (low-level) basis we are examining, with those that we hope to adduce in their support. If it is not plausible that incompatible states of affairs simultaneously obtain in this one world, then there must be something to choose between contradictory opinions about such facts. Non-relativism about facts of a different kind provides a second argument against relativism of reasons. These are the *historical* facts: the occurrence of processes (in some scientific institutions) that have delivered growth in low-level truths of a certain kind. If historical records are correct, more is known today about the workings of everyday objects like bridges and telephones than in earlier ages. After all, there were times—quite recently in the case of telephones—when such objects did not exist. The existence of these recently acquired low-level truths then supports the thesis that science has progressed; the attraction of such a strategy is that low-level facts about bridges and telephones seem to be less subject to plausible relativist theses, being comfortably trans-cultural. Given that such progress has occurred, we can investigate the processes that brought it about.

Now MSRP purports to provide a retrospective explanation of the acceptance of (say) evolutionary theory, given its history of pragmatic success. Does this explanation link up to the discovery of the sort of truths outlined above? Summarised briefly, Lakatos' proposal for a metamethodology is as follows: (1) Methodologies of science provide a demarcation between 'rational' and 'non-rational' procedures. (2) When applied to the history of science, this distinction, particular to each individual methodology, sorts episodes in the history of science into two kinds, corresponding to the above demarcation: those to be given a rational (i.e. internal or evidential) explanation; and those to be given a non-rational (i.e. external or socio-psychological) one. (3) We can assess a methodology as a theory of the history of science, by seeing how many pragmatically successful episodes in the history of science are saved as 'rational'. (4) On what basis do we assess the different accounts given by each methodology? In other words, what *meta*methodology do we apply to judge the relationship between the methodological theory and the methodological—historical—data? Falsificationism as a meta-criterion would require the rejection of *all* methodologies:

> *All* methodologies, *all* rational reconstructions can be historiographically 'falsified': science *is* rational, but its general laws cannot be subsumed under the general laws of any methodology.
> (Lakatos [1971a], p.115)

In place of falsificationism, Lakatos—perhaps unsurprisingly—proposes a methodology of historiographical research programmes (MHRP). (5) Leaving aside the yawning infinite regress which seems to open up, if we apply MHRP to each of a number of rival methodologies: (naive) falsificationism, inductivism, conventionalism, and of course MSRP, we can assess their progressive or degenerating character as historiographical research programmes.

So we have the Lakatos recipe for constructing a historical—and retrospective—account of scientific rationality within MHRP, via the process of rational reconstruction. This amounts to setting methodologies the task of giving an internal (scientifically rational) account of as much as possible of the growth of knowledge that the history of science exhibits. MHRP applied to MSRP yields a historiographical conjecture: The unit of growth in knowledge is the research programme. Some of these progress, some degenerate. Knowledge grows by the triumph—perhaps after a 'war of attrition' ([1971a], p.118)—of the progressive over the degenerating. However, MSRP also

> predicts the existence of hordes of known anomalies in research programmes progressing on possibly inconsistent foundations. ([1971a], p.118)

Now to the process of deciding whether this conjecture will yield a progressive historiographical research programme. *First*, pick an episode in the history of science which is generally admired by scientists, in that it is thought to exemplify growth-in-knowledge. These judgements need not *necessarily* be taken on trust: the adoption of some particular theory might have prompted the discovery of a rash of the 'little technological facts' invoked earlier. The discovery of such facts can be recognised even by those outside the scientific community. *Second*, identify the research programme(s) involved, and find the texts germane to it (or them). Within these texts, we should be able to pick out those parts which illustrate the scientific content of the programme, leaving out any irrelevant feelings expressed by the protagonists, descriptions of their breakfasts, and other external chaff. We now have the internal history, for which an MSRP-rational story can be constructed. If we find we are offending too many historians of science, in that we are unable to make a good fit between our story and the historical facts, we can either: re-analyse the history; revise our conjecture; or conclude that this episode does not instantiate 'growth in knowledge'. The third option, as Hacking [1979] rightly observes, looks much like 'monster-barring', *unless we require that an appropriate—and progressive—external explanation be given.* A well-known example of the last option is the invocation of the action on the minds of Soviet biologists in the 1950s of Lysenko's imprisonment of his scientific opponents, to explain the dominance of the (degenerating) Lamarckian programme over the (progressive) Mendelian one. Independent corroboration of such explanations can then be sought, in the form of 'novel' historical facts:

> *Thus progress in the theory of scientific rationality is marked by discoveries of novel historical facts, by the reconstruction of a growing bulk of value-impregnated history as rational*
> (Lakatos [1971a], p.118)

We can see straight away that Feyerabend's famous charge of elitism doesn't stick, because although the basic judgements of a scientific elite are taken as a starting point, they need not be not carried over unquestioningly into the historical account of rationality, because occasionally the 'scientists' judgement fails'. Nor do such judgements *constitute* the growth of knowledge. Firstly, we might have some antecedent notion of progress or growth in knowledge (see above) with which to measure the scientific community's judgements. Secondly, in Lakatos' own words, MHRP implies a 'pluralistic system of authority' ([1971a], p.121). He continues:

> I disagree ... both with those philosophers of science who have taken it for granted that general scientific standards are immutable and reason can recognise them *a priori*, and with those who have thought that the light of reason illuminates only particular cases. The methodology of historiographical research programmes specifies ways both for the philosopher of science to learn from the historian of science and *vice versa*. ([1971a] p.121)

How did MSRP fare under the criticism of methodologies encoded in MHRP? It would surely be non-trivial for MSRP to be successful by its own lights, that is, to provide the hard core for a historiographical research programme that is *progressive* according to MHRP. In an otherwise relentless political and philosophical critique of the very notion of scientific rationality, Feyerabend [1976] endorses the progressiveness of the historical essays that appear in the same volume *as historical studies*. As sociological studies they are 'excellent' (p.202), and as historical studies 'We see at once that they are superior to earlier studies of the same kind' (p.220). In his review of the same volume, Kuhn [1980] observes that 'In the event these historical chapters are a considerable success, and I have not been altogether able to set aside the question how this could be so.' ([1980], p.181) In contrast, Kuhn's *objections* to MSRP's performance as historical description have the flavour of the *a priori* misgivings of a historical purist about the very notion of *rational* reconstruction: 'As a motive for doing history that one seems to me a likely invitation to disaster' (p.182); 'History done for the sake of philosophy is often scarcely history at all' (p.183). When the programme gets to work, however, Kuhn admits that historians 'should not ignore' (p.185) the findings. The historical studies inspired by MSRP are an *advance* on earlier such accounts. While they might exist in a sea of philosophical and historical anomalies, this—under MHRP—is the best that could be hoped for them.

I claim that Lakatos' metamethodological vision is a naturalised one. Firstly, methodologies are interpreted partly *descriptively*. Prescriptions are parasitic on the historiographical reconstructions. Secondly, the standards that are taken to govern the appraisal of methodologies are those that—it is argued—govern the growth of scientific knowledge itself. In MSRP, the objectivity of scientific judgements is argued for by pointing to the growth in knowledge that flows from such judgements. In MHRP, methodologies were appraised by their ability to provide rational reconstructions of historical episodes in which growth in knowledge occurred. The starting point and foundation is the growth of knowledge. Suppose that we have historical arguments showing that where MSRP *has* been followed, growth in knowledge has been achieved: MSRP must encode reliable *means* for achieving pragmatic progress as an *end*. So the crucial question is whether growth in knowledge does answer to truth. Papineau argues that it does not, and attributes his reading to Hacking [1979], a paper to which we will turn in the next section. Papineau's second objection centres on his construal of the distinction between external and internal history. If 'rationality' is assimilated to what people do when left unperturbed by 'corrupting' influences, where is the guarantee that this sort of rationality is *conducive to truth*? First identify *internal* episodes with episodes in which the scientific belief-outcomes were primarily caused by reliable (although perhaps indirect) links to the natural world, or MSRP-rational inferences from the products of such mechanisms. The rest are *external* episodes. On this construal, a reasonable answer to Papineau's query is that there *is* no guarantee that 'internal' episodes will deliver true beliefs. However, there are *a posteriori* reasons to think that this '*ceteris paribus*' rationality is conducive to truth, if previous episodes of unmolested science have, in fact, been productive of *truths*.[7]

1.7. HACKING ON LAKATOS

In a review of the Lakatos' posthumously collected papers (Lakatos [1978a] and [1978b]), Hacking [1979] has provided an intriguing interpretive hypothesis based on Lakatos' philosophical background. On the strength of this reading, Hacking attributes to Lakatos an implicit philosophical agenda that suggests a particular interpretation of his methodology and metamethodology. Now Hacking's criticisms of MSRP and MHRP (revised and repeated in his [1983], chapter 8) depend *crucially* on this attribution, which has been influential: the review has been cited with approval by Papineau, Fine [1986], Leplin [1984b], and Giere [1988]. I will argue that Hacking's is not the only possible reading of Lakatos, and that Lakatos' methodology and metamethodology might find a home in an altogether *different* project. I will therefore begin by examining Hacking's exegesis.

Hacking's chapter on MSRP in his [1983] is entitled "A surrogate for truth". His central thesis is that Lakatos' education within the Hungarian 'Hegelian and Marxist' (p.112) tradition lead him to 'take for granted the post-Kantian, Hegelian, demolition' (p.112-3) of correspondence theories of truth. He was, however, keen on the objectivity of science. Now Putnam has followed a 'simple Piercean' course that attempts to secure objectivity for science by invoking a universal scientific method, so that everyone will eventually agree to accept its products. This consensual vision is therefore *forward looking*. Lakatos, in contrast, embarked on a more radical programme that *denied* a forward-looking rationality. He did not redefine truth to be 'whatever it is rational to believe', because

> Lakatos is no born-again pragmatist. He is down on truth, not just a particular theory of truth. He does not want a replacement for the correspondence theory, but a replacement for truth itself. ([1983], p.119)

For Lakatos, the objectivity of present knowledge can be assessed only in *retrospect*, by inspection of the growth of knowledge through history:

> The one fixed point in Lakatos's endeavour is the simple fact that knowledge does grow. Upon this he tries to build his philosophy without representation, starting from the fact that one can see that knowledge grows whatever we think about 'truth' or 'reality' ([1983], p.119)

[7] Note: conducive to *truths* (plural). We have not yet discussed whether science approaches 'the' truth.

A theory of reason is built up in which past episodes of growth of knowledge—where knowledge is identified internally to the enterprise of science—demarcate the rational from the irrational. But how can *objective* knowledge grow except by the addition of new truths to the known? After all, commentaries on the Talmud have grown. Therefore the account

> must, if it is of worth, effect a distinction not between the rational and reasoning, and the irrational and unreasoning, but between those reasonings which lead to what Popper and Lakatos call objective knowledge and those which pursue different aims and have different intellectual trajectories. ([1983], pp.120-1)

This is where what Newton-Smith ([1981], p.99) calls the "shy Hegelian" thesis comes in. Lakatos made the distinction between *internal* and *external* history central to MHRP: external history includes the history of what people actually believed, and their—good *or* bad—reasons for those beliefs. Internal history, in contrast, only encodes the growth of *objective* knowledge:

> It is to exclude anything in the subjective or personal domain. What people actually believed is irrelevant: it is to be a history of some sort of abstraction. It is, in short, to be a history of Hegelian alienated knowledge, the history of anonymous and autonomous research programmes. ([1983], p.122)

Hacking looks back to *Proofs and Refutations* for an explicit example of a Hegelian gloss to this alienation, but there is another model in Popper's three 'worlds': the physical; the mental (containing consciousness and, crucially, *belief*); and that of ideas and knowledge. The difference between scientific and Talmudic knowledge lies in the loci of their growth, but it should not be forgotten that the criterion of objectivity is *internal* to the enterprise. So Lakatos sets out 'to characterize the growth of knowledge internally by analysing examples of growth' (Hacking [1983], p.124), so that both science and its objectivity would be safe from the cataclysms in knowledge that he thought were implied by Kuhnian irrational change between paradigms. What is required is a *rationale* for the replacement of research programmes, a rationale to be provided by MSRP in terms of progression and degeneration. Growth of knowledge replaces truth as the proper aim for rational enquiry; process stands in for its product. Now MSRP may be a 'good stab' (p.127) at providing the required rationale, but Hacking has reservations. The first of these is that, like Feyerabend, he suspects that Lakatos' analysis is accurately applicable only to the growth of knowledge during the 'last couple of hundred years', because it presupposes a hypothetico-deductive model for reasoning between theories and lower-level facts. Without any external connections (such as truth-as-correspondence), objectivity must be tied inextricably to a particular mode of reasoning. So if that mode of reasoning is shown to be culturally or historically bounded, the same can be inferred for the corresponding notion of 'objectivity'. MSRP offers only a parochial message: a description of the process by which objectivity is achieved *within* the recent 'style of reasoning' called hypothetico-deductivism. That message *cannot* be generalised to 'timeless knowledge and timeless reason' (p.127). Once we observe, with Hacking, that styles of reasoning actually *do* change, *eternal* objectivity is lost, to be replaced by some (disembodied) species of internal consensus. Now changes of style might be *cumulative*, so that the style of reasoning described by MSRP is a clear *advance* on earlier styles, and will be *preserved* through future changes of style. However, 'these are matters which are only recently broached, and are utterly ill understood.' (p.128) We should therefore be

> chary of an account of reality and objectivity which starts from the growth of knowledge, when the kind of growth described turns out to concern chiefly a particular knowledge achieved by a particular style of reasoning. ([1983] p.128)

The second problem follows on from the first: 'a style of reasoning may determine the very nature of the knowledge it produces' (p.128), by dictating (for instance) which *kinds* of sentence are even *candidates* for truth or falsity. Lakatos concentrates on the issues that have dominated recent Western science: high-level laws, their empirical consequences and the theoretical entities they invoke. So for Hacking, MSRP is a competent—but limited—attempt to 'characterize certain objective values of Western science without an appeal to copy theories of truth' (p.128). However, we are left with no 'external way to evaluate our own tradition', but the writer is Ian Hacking, so the conclusion is 'why should we want that?' (p.128)

So is Hacking *correct* to attribute an implicitly coherentist view of rationality to Lakatos? There are two intimately connected issues that need to be separated here: (i) whether Lakatos really *can* be read as a 'shy Hegelian' who nurtures a deep hostility to the correspondence theory of truth; (ii) whether or not this *matters*, i.e. whether or not a definition of 'truth' is *relevant* in MSRP. Addressing the first question, I think that Hacking mistakenly reads agnosticism as atheism. According to Hacking, Lakatos left truth-as-correspondence out of his account of scientific objectivity because he was hostile to it. The Hegelianism is crucial, then, because it *motivates* this hostility. Now Lakatos' philosophy of science was conceived at the London School of Economics in the 1960s, although it was, of course, the product of many influences. Surely, then, another philosophical background that is relevant to Lakatos' development was the *Popperian-fallibilist matrix*; this relevance is reflected in the number of references to Popper in Lakatos' writings on science. In Popperian methodology, correspondence to external reality is the *aspiration*, and it is in correspondence that objectivity *inheres*—hence the preoccupation with the simple application of *modus tollens*:

> It is only the idea of truth which allows us to speak sensibly of mistakes and of rational criticism, and which makes rational discussion possible—that is to say, critical discussion in search of mistakes with the serious purpose of eliminating as many of these as we can, in order to get nearer to the truth. Thus the very idea of error—and of fallibility—involves the idea of an objective truth of which we may fall short. (Popper [1963], p.229)

Now Lakatos criticised Popper for many things, but among them we do not find Popper's adoption of correspondence as the relation that holds between theories and world (in the most favourable cases). However, in a fallibilist setting, truth cannot provide a criterion for *acceptability* because it notoriously fails to be manifest: therefore in appraising theories, science must turn to *pragmatic* virtues, and agnosticism must reign as to the *attainment* of truth. The only reason for fallibilism of this sort is the gulf between the world as truth-maker and its appearances. Now Hacking *discounts* Lakatos' playful references to 'Truth', putting the playfulness down to Lakatos' antipathy to the very notion. But Hacking's reading renders inexplicable Lakatos' famous [1974] plea to Popper for a "whiff of inductivism" that would turn a 'game' of science into a more serious business. If, however, Lakatos is read as a *fallibilist*, the plea—and the playful references to 'Truth'—make perfect sense, reflecting Lakatos' recognition that the approach of scientific knowledge to the Truth About the World can only be a *metaphysical conjecture.*

Truth-as-correspondence can be what scientific objectivity *inheres* in—granted the conjectured metaphysical principle of induction—but we must judge theories by their pragmatic virtues. Appraisal must therefore be *fallible*:

> While upholding the view that the supreme aim of science is the pursuit of truth, one must be aware that the path towards Truth leads through ever-improving false theories. It is therefore naive to believe either that one particular step is already part of the Truth or even that one is on the right path. (Lakatos [1971b], p.175)

A note to that passage reads:

> In these matters I follow Bolzano, Frege and Popper. (Lakatos [1971b], p.182)

Is the definition of truth relevant to appraisal MSRP? Not as Lakatos formulated it. Hacking accurately points out that the grounding for rationality in MSRP is the growth of knowledge, and that 'knowledge' is a *primitive.* MSRP describes the means by which such growth is achieved. Now Hacking assimilates 'growth in knowledge' to the growth of disembodied 'third world' knowledge, and cuts this growth off from its 'first world' by-products. This, however, is unjustified. The Hegelian tag is a red herring because it purports to motivate Lakatos' silence on the issue of correspondence between theories and external reality, and thereby obscures the main advantage of Lakatos' pragmatic account of objectivity over the Putnam-Pierce version: that it does *not* adopt an account of truth-as-coherence or truth-by-consensus. In MSRP, it seems, truth does not appear in the story—quite reasonably, given the fallibilist setting—and is therefore not antecedently defined; *in one sense, any definition of truth can be adopted.* But one theory of truth motivates the whole fallibilist enterprise: truth as correspondence. This points to a reply to the main charge against MSRP: its historical specificity. Hacking's version of how objectivity is obtained in MSRP is perceptive, but the account of objectivity is mistakenly read as a *definition.* It should instead be read as an objective *acceptability criterion*, or as a story about how scientific judgements are judged to be objective, a story that is told within the hypothetico-deductive style of reasoning. When challenged to give a constitutive account of scientific objectivity, realists are as free as ever to appeal to a (conjectured) Truth that is external to their particular style of reasoning. Anyway, a more homely appeal was hinted at in the previous section: The first world by-products of pragmatic progress are technological artefacts like bridges and telephones. If our reconstructions of history show that the knowledge that made telephones possible was uncovered in accordance with MSRP, the objectivity of Lakatosian progress is no more specific to one style of reasoning than are the myriad episodes of the correct functioning of telephones.

Conclusion

In conclusion, the final sections of this chapter have tried to provide a grounding for Lakatosian methodology—the growth in easily accessible technological facts—that is not plausibly subject to relativist challenge. MSRP provides methodological imperatives, which correspond to facts about correlations between means and ends. If (for instance) theory proliferation is a 'good thing', this is because it is an efficient means to an end such as scientific progress: those scientific communities in which many incompatible theoretical frameworks are developed will progress fastest. Does this ensure 'progress'? Taking the methodological 'norms' to mean standards that are—as a matter of social fact—shared by a community, we need some virtue of sets of theoretical assumptions by which their acceptance by scientists is 'hooked up' to scientific progress. In other words, there must be some property shared by the theories that are accepted by a community that *explains* that acceptance, given the criteria that are *applied*—rather than explicitly adhered to—by that community. If MSRP makes factual assertions about science and its history, it implies that that property is *fruitfulness*. But what eventually decides whether a research programmes *is* 'fruitful'? The discovery of *novel facts*. According to a causal theory of scientific knowledge, communities of scientists interact with the world: it is *nature* that determines the acceptance and rejection of research programmes. The causal interaction consists of the many (messy and interpreted) *experiences* that some scientists accumulate during their careers, so the 'bottom line' for MSRP must therefore be *empirical* fruitfulness. Now a theory's claim to (approximate) truth or reference can only begin to be assessed in the long term historical run of things, in the light of the structure of successor theories: how else are we supposed to make such a judgement? We *cannot fail* to be influenced by our current favourite theories. In many cases, this would mean that a fallibilist realist's attitude should be analogous to Mao Tse Tung's when asked about the historical significance of the French Revolution: *It is too early to tell.* But for *current* theories, the inference from empirical adequacy to approximate truth is a separate matter: scientific realism is independent of MSRP, and Lakatos recognised this when he made the convergence of science to Truth the subject of a separate, *metaphysical* conjecture.

2

Realism as a Methodological Thesis

Bold explorations which have contributed greatly to the progress of geography are due to adventurers who were looking for the golden land—that is not a sufficient reason for inscribing 'El Dorado' on our maps of the globe. (Duhem [1914], pp.31-2)

Introduction

For as long as realists and instrumentalists have disagreed, partisans of both sides have pointed in argument to the actions and sayings of scientists. Realists in particular have often drawn comfort from the *literal* understanding given to theories by those who are paid to deploy them. The scientists' realism, according to the realist, is not an idle commitment: a literal understanding of past and present theories and concepts underwrites their employment in the construction of *new* theories. *New* theories point out—and explain—*new* phenomena. So realism, claim the realists, is at the heart of science's achievement of what Bacon identified as science's aim: *new* knowledge offering *new* powers. How does this become an argument for realism? Scientific realism enters the story twice: (tacitly) adopted by *scientists*, it motivates scientific practice, while the success of the practice might support realism as a *philosophical* view of science. To fill the story in, we need to know what the realist view is, and an account of how a scientist who accepts it would behave differently from one who does not. In 2.1, below, I will set out three ways realists and anti-realists differ in their interpretation of science: (i) realists and anti-realists may adopt different semantics for scientific theories; (ii) they may apply different epistemologies (some realists take it that the success of a theory is *ipso facto* a reason to think it true, or referentially successful); and (iii) they may identify different aims for science *as an activity* (discovery and explanation, rather than construction and saving the phenomena).

With an eye on the three components of realism, we can see that some feature of scientific practice might commit practitioners to realism if it reveals realist semantics, presumes that some theory is *true* on account of its predictive success and explanatory power, or is appropriate only to realist aims. Now let *methodological realism* be the claim that some practices that are central to the success of science reveal realist commitments in any of these ways. Arguments for scientific realism from methodological realism have been presented in two forms: the *explanationist* argument due to Richard Boyd and Hilary Putnam, and the vindicationist version presented by R.I.G. Hughes. In 2.2 I will examine the explanationist version which has it that only realism *itself* can explain the success of the realist practices. The vindicationist argument eschews explanation: some realistic commitments are *presupposed* by scientific practice, and where the practice is successful, both practice and presupposition are *justified*: this argument will occupy 2.3. Central to the examination of these arguments are the following issues: (i) whether—and in what way—features of scientific theorising commit scientists to realism; (ii) whether these features contribute to success; (iii) whether a convincing argument for realism can be grounded in positive answers to (i) and (ii). In 2.4 I will argue for a positive answer to (i) and (ii), but 2.2 and 2.3 suggest a qualified *no* to (iii): some of those who reject realism can accept—and understand—the implicit recommendation in methodological realism. In 2.5 I will fill out the argument for (i) and (ii) with some examples. In 2.6, methodological realism is applied to theory construction, and it is argued that the methodology of scientific research programmes provides an account of this activity which *embodies* MR. At the end of the chapter are two appendices that constitute a *digression*. The first appendix explores what follows if the practice-centred notion of model used in 2.6 is explicated in terms of the model of mathematical logic. The second appendix follows this up with an examination of Putnam's model-theoretic argument against metaphysical realism.

2.1. FORMULATING SCIENTIFIC REALISM

Scientific realism is a thesis about science *as an activity*. Put naively, scientific realists see science as an activity best described as *discovery* and *explanation*—scientists *discover* entities and laws that were 'there before', and *explain* the phenomena. Anti-realists tend to see science as a process of *construction* (of theories and concepts) and of *saving* the phenomena. The realist view of science involves a number of theses, of which the following is a representative sample:[1] (Semantics) Theoretical discourse is to be construed *literally*: uses of theoretical terms are putative references to theory-independent entities—whether observable or not. Literally construed, theoretical claims can be true or false independently of our interests and commitments. (Epistemology) Given the success of our best theories, we have reason to believe they are (approximately) *true* and referentially successful, not just empirically adequate. (Aims) Science aims to provide literally true

[1] Here I (roughly) follow van Fraassen [1980], ch.2. The list is not meant to be independent, exhaustive or acceptable to all realists: I have tried to select *methodologically relevant* claims.

stories about the world, to be a process of *discovery* and of *explanation* rather than one of *construction* and of *saving the phenomena*. The different components work together: in accepting a theory, we take it that it advances our scientific aims, so the realist's acceptance of a theory must involve the (tentative and qualified) belief that it is *true*. The realist's deeper commitment reflects *stricter* criteria of adequacy: explanatory power *as well as* empirical adequacy. Thus some realists present their deeper commitment as the result of *inference to the best explanation*. Given that we *do* have predictively successful explanatory theories, we have reasons to think them (approximately) *true*. In what follows I will explore the content of each of the components of realism by looking at the different reasons that have been found to reject them. The purpose in doing this is not to argue for or against any aspect of scientific realism, but rather to investigate *what it is* that a scientist is committed to by being a scientific realist. In later sections I will investigate the expression these commitments may have in scientific method.

The Semantic-Metaphysical Component

The are two parts to the realist's construal of scientific discourse. The first is that theories are to be construed literally, rather in terms of any *translation* or *analysis* that fails to preserve logical form. The second is that, literally construed, statements in scientific discourse can be true or false independently of our ability to know their truth values. This expresses the metaphysical idea behind realism that science investigates a world that is *independent* of our knowledge, commitments and experience, and the semantic thesis that scientific statements are true or false by virtue of some semantic relation to this world.

There are a number of different commitments that might be involved in saying that theories are to be construed literally. (i) Theories genuinely *make statements*, they are not just (non-propositional) *tools* or *instruments* for making predictions or classifying the phenomena. This rules out Duhemian *conventionalism* as well as those forms of *instrumentalism* that derive semantic theses from methodological theses concerning how theories are used in science. (ii) Theoretical statements cannot be translated without loss of content into statements about phenomena. This rules out the *phenomenalistic reductions* conceived by Mach, and early logical positivism. (iii) Terms in scientific theories are to be construed as expressions that putatively refer to *theory-independent* entities, rather than uses of *contextually-defined theoretical concepts* which appear in our theories only to help us predict phenomena in convenient ways. For instance, the theoretical assertion "the force on a body of mass *m* at distance *r* from the centre of the earth is ..." should be taken to imply that *there are such things as forces* that act on bodies, causing their motion (this rules out the *law-cluster theory of meaning* for theoretical terms proposed by the later Rudolf Carnap).

The second part of the realist's account of scientific discourse concerns the theory of truth. According to the realist, theoretical statements—literally construed—can be true or false independently of our ability to know their truth value: realists adopt a *non-epistemic theory of truth* for theoretical statements. This formulation is weaker than that proposed by Dummett [1963], who defines realism about a discourse as acceptance of the principle of bivalence (every statement is either true or false) for statements of that discourse. The bivalence formulation is probably too strong for there are reasons to reject bivalence that are *not* also arguments against realism.[2] Dummett claims that the bivalence formulation captures the intuitive content of metaphysical realism—the existence of a mind-independent world—which at best has a 'metaphorical' meaning. Now one might quibble with this claim, because bivalence is a *semantic* thesis: it seems to express only the *semantic role* of the mind-independent world. Arguments concerning metaphysical realism have not always centred on semantics: sometimes (as in the theory of perception) it is the *causal* role of a mind-independent world that is at issue, which suggests that the existential commitment of metaphysical realism (to a mind-independent reality) could play different roles in different areas and should therefore be separated from its role in semantics. However, it is important for the weaker formulation to capture whatever the bivalence formulation does of a non-epistemic theory of truth. The weak formulation *does* do this, for it is incompatible—as it should be—with two key formulations of *anti*-realism that have been proposed in opposition to the bivalence version of metaphysical realism, and is subject to the same anti-realist objections as the bivalence version.

Firstly, there is the *pragmatism* associated with Pierce and (one stage of) Putnam. The pragmatic theory of truth *defines* truth in terms of the theory that would be accepted in the ideal limit of scientific enquiry. Putnam's argument against metaphysical realism envisages an ideal theory that satisfies all epistemic constraints: according to metaphysical realism, this ideal theory could be false, because truth is independent of rational acceptability, even rational acceptability in the ideal limit of enquiry. Putnam then presents an argument to the effect that it is *inconceivable* that this theory could be false.[3] Putnam's argument, it should

[2] There are, for instance, theories of reference (like Strawson's) according to which a statement can fail to have a truth value because it contains a descriptive term, and the existential presuppositions involved in making an assertion involving that descriptive term are false. Also, according to one interpretation of quantum mechanics, if a statement ascribing a definite momentum is true, then a statement ascribing a particular position is neither true nor false. Yet this is not an *anti-realist* construal of quantum mechanics, for it is consistent with the existence of a mind-independent microphysical realm of which quantum mechanics is *true* (see van Fraassen [1980], p.38).

[3] See Putnam [1980] and [1981]. This argument is also discussed in appendix 1.

be noted, is directed at the *possibility* of the ideal theory's being true or false, and so can be construed as an attack on the *weak* formulation of realism. The second kind of anti-realism relevant here is Dummett's own anti-realism. Dummett's anti-realism, and his arguments against realism, are founded on the intimate connection Dummett sees between *meaning* and *use* (see Dummett [1963], [1975] and [1976]). A theory of meaning must provide an account of the use of sentences, and therefore an account of what a competent language user knows when she knows how to use—i.e. *understand*—a sentence. The competent language user could have learned how to use any particular declarative sentence only by learning how to apply it when certain conditions verifiably obtain. Thus the theory of meaning—as a theory of *understanding*—must explicate meaning (and truth) in terms of *verifiability conditions*. The realist, in contrast, explicates meaning in terms of *truth conditions*, where those truth conditions may outrun our ability to verify whether or not they obtain because they can be true or false independently of our ability to *know* whether they are true or false. Dummett argues that the realist will find it difficult to explain how competent language users come by the knowledge that constitutes their understanding of sentences, if this knowledge involves knowledge of when truth conditions obtain, and those truth conditions outrun our ability to verify whether or not they obtain. Note that Dummett's argument—as given here—turns on the *impossibility* of our coming to understand a sentence in terms of verification-transcendent truth conditions. It therefore works against the *weaker* version of realism.

The Epistemic Component

The epistemic component of scientific realism is the thesis that theories—even where they invoke unobservable entities and processes—can be confirmed as true by the ordinary methods of science. In accepting a theory on account of its empirical success, the realist accepts the theory *as true*, and the entities it invokes *as real*: the realist's acceptance of a theory involves the belief that the theory is true and referentially successful. This rules out (for instance) van Fraassen's constructive empiricism (van Fraassen [1980]), according to which empirical evidence can select among theories only up to *empirical adequacy*. Thus for van Fraassen, acceptance of a theory involves only the belief that it is empirically adequate.[4] Constructive empiricism turns on a general epistemic distinction between observable and unobservable entities, and offers an attitude towards theories—belief in empirical adequacy—to replace that offered by the realist (i.e. belief). There are, however,

[4] Note that the difference between the realist and non-realist accounts of acceptance need not be construed in terms of the difference between belief and acceptance. Realists and anti-realists can each formulate their positions in terms of *either*, but will differ over the proper *reach* of the relevant propositional attitude, be it belief *or* acceptance.

ways to reject realism that *do not* presume an epistemologically important observable-unobservable distinction, *or* offer another general claim as to what is the correct attitude to successful theories. For instance Laudan [1984] and Cartwright [1983] claim (in their different ways) that a theory need not be true to make successful predictions. Their claims depend on particular analyses of the history of science (in Laudan's case) or of the application of theories to particular systems (in Cartwright's case). In addition, there are those, like Fine ([1984] and [1986]), who are suspicious of *any* general philosophical account of science. For Fine, the provision of *either* an external semantics *or* a particular epistemology for science is 'inflationary'. In his deflationary approach to science, Fine refuses to say anything about either truth *or* the inferences involved in acceptance of a successful theory. If scientists say electrons exist, then they exist: end of story.

Defences of (the epistemic component of) scientific realism typically centre on the *inference to the best explanation* (sometimes also called the *abductive* inference). This form of inference is used all the time (argues the realist): providing the best explanation constitutes a reason to *believe* the assumptions that provide it. Inference to the best explanation can be applied at different levels to provide distinct arguments for realism. (i) Firstly, it can be applied directly to unobservable entities: the existence of electrons as described in our best theories is the *best explanation* of the behaviour of cloud chambers, mass spectrometers and so on. Therefore electrons exist. (ii) At a higher semantic level, we could say that the *truth* of a theory is the best explanation of its predictive success (alternatively, the theory *itself* provides this explanation, and so must be true). (iii) Thirdly, some realists would generalise the last argument. Consider scientific realism as a (quasi-scientific) hypothesis about science: the best theories we have *accurately represent* the 'deep structure' of the world and the unobservable entities that populate it (that is, they are true and referentially successful). Scientific realism provides the *best explanation* of the power to predict phenomena with which science furnishes us. This power would be a *miracle* if the theories were not true or if the entities did not exist. Hacking [1983] and Cartwright [1983] have presented arguments for entity realism that, they claim, do *not* involve inference to the best explanation. Hacking's version is that where scientists (appear to) use one kind of entity to investigate the properties of another, the relevant experimental set-ups must be able to be manipulated *reliably*, and so there must *be* some set of entities that have stable causal properties that *underwrite* the scientists' ability to manipulate. Cartwright (for a number of reasons) rejects inference to the best explanation *in general*, but accepts such inferences where a causal explanation is given. Thus she presents *inference to the most probable cause* as the inference that underlies the argument for entity realism: the stable dispositions of electrons *cause* the phenomena that are cited in evidence for their existence. If electrons are the *most probable cause* of the phenomena, electrons must exist.

Objections to these arguments have been various. (i) Some (e.g. van Fraassen [1980], [1985] and [1989]) criticise inference to the best explanation *generally*, where the best explanation in question 'trades in unobservables'. First note that theories are underdetermined by evidence: different *incompatible* theories can make the same predictions. Where one is empirically adequate, so will be its empirical equivalents. The realist will reply that not all theories *explain* the phenomena they predict. The non-realist replies that what counts as an explanation reflects *our* interests: why is it that what *we* think is the best explanation must be *true*? Secondly, we choose the best explanation from a *historically restricted* pool of theories: we have no reason to think that the *true* theory must be among them. (ii) In his [1984], Laudan attacks the second-level argument: the realist does *not* provide a good explanation of predictive success. There have been many past theories, successful in their time, that we now know to be either false or referentially unsuccessful (the phlogiston theory of combustion and the caloric theory of heat are among Laudan's examples). Conversely, there have been (approximately) true and referentially successful theories that have *failed* to be predictively successful. Thus the implication between truth and referential success (on the one hand) and predictive success (on the other) holds in *neither* direction. Neither truth nor referential success explain a theory's predictive success. Cartwright limits her critique of realism to truth, for she *accepts* that referential success can (on occasion) explain predictive success. Analysing the application of theories within physical models, Cartwright ([1983], pp.44-73) notes that the truth of a theory is not a good explanation of its use in the construction of successful models, for models are piecemeal affairs that seldom satisfy the equations of the theories that (ostensibly) are applied through them. (iii) Replying to the 'no miracles' argument (at level (iii), last paragraph), van Fraassen argues that in the predictive success of science generally, there is no miracle to explain: our best theories are successful because we *systematically reject* those that are not. Finally, the realist arguments at the higher levels have been held to be *circular* if intended as defences—against non-realist critiques of inference to the best explanation—of the ground-level argument: the higher level arguments use the same form of inference, and will therefore fail to impress the non-realist. Whether the (causal) arguments for *entity* realism are subject to the same kinds of objection depends on how far inference to the most probable cause can be distinguished from inference to the best explanation.

It should be noted here that according to the formulations given above, the epistemic and semantic components of realism are *independent*. There are those, like Putnam, who have accepted the epistemic component while rejecting metaphysical realism (in its bivalence guise). Indeed Putnam rejects metaphysical realism precisely because he seeks to *defend* the epistemic component of realism (which he calls just 'scientific realism'). On the other hand there are those, like van Fraassen, who accept that theories can be true or false independently of our ability to know their truth values, but deny our ability to know anything more than that a theory is empirically adequate. There is, however, a sense in which the issue of whether theories can be confirmed as true arises in an interesting way *only if* a non-epistemic theory of truth is agreed upon: certainly van Fraassen's critique of the epistemic component seems to presuppose one. Putnam's acceptance of the epistemic component is rendered *Pickwickian*, perhaps, by its identification of truth with rational acceptability in the ideal limit of enquiry.

The Aspirational Component

In discussing science as an activity, we have not yet addressed its *aims*. This is not to seek an account of the motives of individual scientists, but rather to ask what counts as success in science.[5] According to the realist, science aims to discover (entities and processes) and to explain (the phenomena), rather than to *construct* and to *predict*. Realists see constraints on theory acceptance—such as explanatory power—as central to the achievement of the aims of science, where the non-realist—for whom the aim of science is the *construction of empirically adequate theories*—interprets explanatory power as a pragmatic (non-evidential) element of theory choice. This component of realism is required to distinguish between constructive empiricists (like van Fraassen) from *critical* or *fallibilist* realists such as Popper (or, perhaps, Lakatos). Popper and van Fraassen both reject the epistemological component of realism, although Popper's reasons for rejecting it are broader than are van Fraassen's. Popper's rejection of epistemological realism rests on his *general* stance on ampliative inferences,[6] while van Fraassen distinguishes between ampliative inferences that involve unobservables and those that do not ([1980], pp.19-20). Popper, however, accepts the aspirational component, while van Fraassen rejects it. Popper adopts aspirational realism as a kind of categorical epistemological imperative that motivates the methods of science. For van Fraassen, the methods presumably determine the aims: we should trim the latter to suit what can be achieved with the former.

2.2. THE EXPLANATIONIST ARGUMENT

In his middle period, Putnam often adduced scientific practice in his arguments against positivist philosophy of science. In 'Explanation and Reference', for instance, he charged that positivist analyses of meaning fail to do justice to scientific *usage* of either theoretical

[5] In van Fraassen's example, the aim of chess is to checkmate one's opponent, while individual chess players may seek fame or fortune.

[6] See Popper [1956] and [1983], pp.110-11 and 131-46.

terms or the theories in which they are embedded. To understand science, claimed Putnam, we need realist semantics and a realist account of theory acceptance.

To take one of Putnam's examples,[7] it is common to conjoin accepted theories and look to their joint consequences for novel predictions. The realist—for whom acceptance means *acceptance-as-true*—finds this easy to rationalise: if T_1 and T_2 are true, then so must be T_1 & T_2. For the non-realist, however, acceptance means only *acceptance-as-empirically-adequate*, and it is less than obvious that the empirical adequacy of T_1 and T_2 separately must imply the empirical adequacy of T_1 & T_2: indeed, T_1 and T_2 might be *inconsistent*, and therefore *trivially* empirically inadequate. So the pooling of explanatory and predictive power—central to the very cumulativity of science—seems inexplicable on any non-realist account of acceptance. Putnam's other examples continue in similar vein: in a good theory, scientists seek the realist's *theoretical plausibility* rather than the positivist's *simplicity*; auxiliary hypotheses are not the *minor premises* in *deductions of observational consequences*, but rather *further facts* to be filled in, so the aim of research must be *fact-finding* rather than *theory-testing*.

So where are these examples leading? In 'Explanation and Reference', Putnam is content to show that the positivists failed to do justice to scientific practice, but in 'The Meaning of "Meaning"', there is the outline of a *positive* argument for realism:

> It is beyond question that scientists use terms as if the associated criteria were not *necessary and sufficient conditions*, but rather *approximately* correct characterizations of some world of theory-independent entities, and that they talk as if later theories in a mature science were, in general, *better* descriptions of the same entities that earlier theories referred to. In my opinion the hypothesis that this is *right* is the only hypothesis that can account for the communicability of scientific results, the closure of acceptable theories under first-order logic, and many other features of the scientific method. ([1975], p.155)

The force of the argument has to be in the *centrality* of the realist features of practice to the success of science, whether in communication or prediction: otherwise the realist practices—and the semantics—might only be so much decorative embroidery on the tapestry of science. In *Meaning and the Moral Sciences*, Putnam fills out the argument (but famously re-interprets it), avowedly following Boyd in the details.

[7] See Putnam [1973], pp.210-11.

Boyd's defence of realism is set within a naturalistic epistemology that provides a detailed mechanism by which realist methods contribute to the success of science.[8] First define the *instrumental reliability* of a method in terms of the instrumental reliability of theories:

> Call a theory instrumentally reliable if it makes approximately true predictions about observable phenomena. Call a methodology instrumentally reliable if it is a reliable guide to the acceptance of theories which are themselves instrumentally reliable. (Boyd [1985], p.4)

Boyd's explanationist strategy is as follows: (i) Identify a reliable methodological principle or strategy of theory construction, and show that the principle can be rationalised only given realist inferences and a realist construal of theories. (ii) Show that the employment of that principle 'contributes to the likelihood that accepted theories will be good predictors of the behaviour of observables' ([1973], p.9). (iii) Claim that realism provides the only plausible explanation for the *reliability* of that principle.[9] Boyd's various examples of realist methods—united under a theme of the unity of science—arise in all areas of scientific practice: construction of theories, design of experiments, and assessments of the degree to which a given body of evidence supports a theory.

Theory, Evidence and Non-experimental Criteria: Phenomena can evidentially support a theory only if *explained* by it, but explanatory power is appraised by comparing the *structure* of a theory with previously accepted theories. For instance, a theory only *explains* a phenomenon if it accounts for that phenomenon by postulating a process that is relevantly similar to processes that are postulated by previously accepted theories (eschewing universal forces, for example: see Boyd [1973], pp.7-9). So considerations of explanatory power, based on intertheoretic plausibility considerations, count as evidential. On the realist view, it is easy to see why this should be so: only theories that are *plausible in the light of background theoretical knowledge* should be constructed and considered as candidates for confirmation by experimental evidence. This radically reduces the infinite pool that are consistent with any finite body of evidence. Anti-realists see explanatory power as a (non-evidential) *pragmatic* consideration, but how could pragmatic considerations contribute to the instrumental reliability of our standards of theory choice? If they do do so, empiricist claims that theories are underdetermined by evidence must be *false*.[10]

[8] See Boyd [1973], [1981], [1984] and [1985].

[9] The procedure in (i) and (ii) might be the same: realism *motivates* a method by showing that it is likely to produce success, while on at least one view of explanation to show that success is to be expected is to *explain* it.

[10] The evidential underdetermination thesis is formulated by Boyd as the claim that evidence for a theory is evidence *of equal force* for its empirical equivalents ([1973], p.2).

Experimental Design: Typically, theories are tested under experimental conditions in which they are most likely to fail, if they are false. These conditions are identified by comparing the *causal structure* of the theory under test with other theories that postulate (theoretically) relevant processes: we would not consider a theory to have been *tested* were there *another* theory that provides an alternative explanation of a successful outcome for the theory under test. (See Boyd [1973], pp.10-11 for a detailed example.)

Cumulativity and Retention: Commitments which hang over from our previous acceptance of now-refuted theories constrain theory construction: we consider only that 'small handful' of theories which (partially) preserve the ontology and mechanisms of previous theories. This would not make sense unless we regarded the success of the old theories as indicators of their successful reference and their approximate truth. (See Putnam [1978], p.21 and Boyd [1981], p.619).

Reference and Univocality: Realists take it that the theoretical terms of successful theories refer to theory-independent entities. Thus where two theories invoke (say) atoms, it makes perfect sense to apply the claims made by one theory about atoms to the atoms invoked by the other theory. This would mean, for instance, that it is *unacceptable* for the two theories to make incompatible claims about atoms, even where they cover different domains. Again, the realist view—that success indicates successful reference—motivates the strategy and explains its success (See Boyd [1981], sec. 2.4). On an anti-realist view, in contrast, theoretical entities are but players in empirically adequate yet *fictional* stories. Assuming univocality on *this* view would be as much of a mistake as worrying that a character plays the violin in one novel, while in another—unrelated—work of fiction, someone with the same name is tone deaf.

In his [1984], Boyd challenges the chief anti-realist traditions—empiricism and (social) constructivism—to *explain* the instrumental reliability of the above methods. Take empiricism first: non-experimental criteria of theory acceptance like unity, simplicity and explanatory power are *non-evidential* according to the empiricist, and could not contribute to a theory's future fulfilment of *evidential* criteria of theory acceptance such as logical consistency and empirical adequacy. The fact that 'pragmatic' considerations *do* so contribute is inexplicable to the empiricist. Now the theory-dependence of appraisal criteria has long been recognised in the *constructivist tradition*. It is no surprise to the constructivist if our best theories fit 'reality': the theory-dependent methods are construed as procedures for the social *construction* of reality. The problem for the constructivist account of method is the *manifest* reliability of the technological products of scientific advance:

> It is ... evident that theory dependent technological progress (the most striking example of the instrumental reliability of scientific *methods* as well as theories) cannot be explained by an appeal to

> social construction of reality. It cannot be that the explanation for the fact that airplanes, whose design rests upon enormously sophisticated theory, do not often crash is that the paradigm *defines* the concept of an airplane in terms of crash resistance. (Boyd [1984], p.60)

Realism, in contrast, can both motivate the methods *and* explain their instrumental reliability. According to the realist, background theories suffuse the very methods of science. If such 'collateral' theories are (approximately) true, it is easy to see how they contribute to the likelihood that new theories—constructed and appraised in line with the methods—will be approximately true, and therefore instrumentally reliable. The planes we make reliably fly because they have been designed with the help of reliable theories. We have reliable theories because we appraise theories in the light of criteria of appraisal that, although theory dependent, are reliable. Our criteria of appraisal are reliable because we have background theoretical *knowledge*.

Do the premises of the explanationist methodological argument, even if true, allow us to infer the conclusions intended by their authors? Among the premises were two claims: (i) that realism is *embodied* in scientific practices; (ii) that these practices have met with instrumental success. Realism is then inferred as the best explanatory (meta-)thesis. Opponents of realism object to this argument at every possible juncture. Firstly, one can dispute whether realist strategies of theory construction like ontological conservatism (Laudan [1984], pp.235-9) or selection for unity and explanation (Cartwright [1983], pp.44-53) generally *are* followed, and where followed, whether they *have* typically issued in success. In effect, one might doubt whether the 'reliability' of realist methods *requires* explanation (I will return to these issues in 2.5). Secondly, one might provide a motivation for supposedly 'realist' methodological principles that appeals only to *empirical adequacy* as aim and epistemic attitude. These first two objections attack the premise—methodological realism—directly.[11] One might, however, admit the realist motives but attack the *explanation* (as van Fraassen has *also* done), supplying an alternative explanation of the success of the realist methods.[12] Lastly, one can object to the very *structure* of the argument: an inference to scientific realism as the best explanation of methodological realism.

The structural objection runs as follows: realists present theories—literally construed—as the best explanations of the phenomena they explain. Anti-realists object to inferences-to-the-best-explanation where the best explanation in question 'traffics in unobservables' (as

[11] See for instance Fine on the 'small handful' strategy ([1984], pp.87-9), and van Fraassen on the conjunction argument ([1980], pp.83-7). See also 2.5, below.

[12] See van Fraassen [1980], pp.93-4 and 2.3, below.

Lipton [1991] has put it), for the following reasons. Explanations that traffic in unobservables are underdetermined by empirical evidence. In making an inference to the best explanation, we must choose the best explanation from among a necessarily restricted but historically contingent pool of theories: those that *happen* to have been constructed. The point is that there will *always* be further explanatory theories that we have not considered. If providing a good explanation were to provide a reason to *believe*, we would need (i) a reason to think that the *correct* explanation must be among those that have been *proposed*,[13] and (ii) a reason to think that what *we* count as the best explanation is the likeliest to be true. We have neither. If the anti-realist objection to realism rests on the evidential underdetermination thesis, it is easy to see why anti-realists *fail* to be impressed with the classic defence of realism in which the truth (or successful reference) of a theory is presented as the best explanation of its predictive success. The best explanation again 'traffics in unobservables'—this time the (approximate) truth of a theory's claims about unobservable entities and processes—and is as subject to underdetermination as its ground-level counterpart.[14] The explanationist methodological argument has the same objectionable inference at its heart, arguing as it does to realism as the best explanation of the reliability of realist methods, and so *it* need not trouble the anti-realist who rejects inference to the best explanation.[15] Compare the realist explanatory claims at different levels:

Level 0: The best explanation of the behaviour of cloud chambers is provided by the existence of electrons as described in our best theories.

Level 1: The best explanation of the success of our predictions of the behaviour of cloud chambers is provided by the (i) successful reference to real entities and (ii) approximate truth of our best theories about electrons.

Level M: *Aspirational version*: The best explanation of our successful use, in learning how to manipulate the world to our advantage, of methods that are appropriate to discovering truths about real entities is provided by the (i) successful reference and (ii) approximate truth of our best theories.

Semantic-epistemic version: The best explanation of the reliability of methods that presume our best theories to be referentially successful and approximately true is provided by the (i) successful reference and (ii) approximate truth of our best theories.

[13] 'We can watch no contest of the theories we have so painfully struggled to formulate, with those no one has proposed.' van Fraassen [1989], p.143.

[14] Even supposing that the truth of a theory *can* explain anything.

[15] See Fine [1984]: Laudan ([1984], pp.242-3) puts the same argument in trenchant style, although Laudan's reasons for rejecting inference to the best explanation do not turn on underdetermination (see 2.1).

Anti-realists deny that a claim about explanation like that at level 0—even if accepted—could launch an inference to the existence of electrons. The arguments at levels 1 and M are of the same form, and will not convince there, either: they are *parasitic* on the argument at level 0. Evidently the main issue is *not* the methodological claim, but the argument's central inference. No matter *how* good our evidence for methodological realism, we cannot launch an argument for realism that would convince anti-realists from *this* pad.[16] If this sounds like acquittal on a technicality, I think there is a substantive point here: to the anti-realist—convinced of the underdetermination thesis—it is axiomatic that the world would look *just the same* if all our presently-accepted theories and background assumptions were *false*, and some other—empirically adequate, yet perhaps unborn—set *true*. This *includes* the history of advances in science, and the application of theory-dependent methods in the achievement of those advances. Thus the anti-realist could just *re-apply* the underdetermination thesis to the realist's explanation at the methodological level.[17]

2.3. THE VINDICATIONIST ARGUMENT

In 'The Bohr Atom, Models and Realism', R.I.G. Hughes presents a methodological argument for scientific realism that replaces the inference to the best explanation with a direct *vindication*. Central to Hughes' argument is the *use* of physical models in probing the content of theories: he makes a distinction between 'surface' models, which merely 'map...the phenomenal terrain' ([1990], p.74) and those that (in Hughes' analogy) are more like subway maps for those without a ticket, positing unseen *underground* connections between isolated surface phenomena. A historical example of this distinction is then given: Kepler's 'purely kinematic' (p.74) model of planetary motions, contrasted with Newton's dynamical model.[18] Now surface models are comparatively rare and less

[16] Such arguments could appeal only to those who *already* entertain realist intuitions. Now Lipton ([1991], chapter 9) sees *some* value in arguments that have this property: convinced realists might justifiably use such arguments to preach to those who accept the premises and rules of inference. This is a sad end for an argument of high ambition.

[17] If the realist objects that the evidential underdetermination thesis has been *refuted* by Boyd, the anti-realist could point out that Boyd's argument showed only how *pragmatic* considerations dissolve the *practical* underdetermination 'problem'. Boyd's argument that these considerations are truly *evidential* rather than merely pragmatic employed something like the inference to the best explanation at level M. That, of course, begs the question at issue here.

[18] This contrast is open to a historical objection: Kepler *did* embed his model in a wider cosmological framework (see Kuhn [1957], pp.209-19). Nor was his model kinematic in the technical sense: his second

successful, which indicates to Hughes that the enhanced explanatory and predictive power enjoyed by the Newtonian (compared to the Keplerian) model arises from its embedding within a wider dynamical theory that is interpreted realistically.

In keeping with his paper's title, Hughes' central example is the development by Bohr of the atomic model that bore his name. Bohr constructed his atomic model within a generally classical framework, his chief explanatory problem being a theoretical one: how could a 'sun and planet' picture of the atom like Rutherford's be stable with respect to mechanical disturbance? Bohr *was* able to explain this, but at the expense of assuming *ad hoc* that only a *countable* infinity of orbits around the nucleus were available to an electron: in classical mechanics there is no reason why a *continuum* of allowed trajectories should not be possible. From the start, Bohr used his model to construct qualitative explanations and predictions covering a wide range of physical and chemical phenomena. Unfortunately, the *quantitative* versions of these theories turned out to be unforthcoming (the calculations were too difficult or problematic) or else empirically inadequate. However, at an advanced stage in the writing of his 1913 papers, Bohr added a new feature to the theory: a mechanism for the emission of radiation, and therefore an account of atomic spectra (as late as January 1913, Bohr explicitly *excluded* such a development). When Bohr performed the relevant calculations for the hydrogen atom, he was able to predict the gross structure of the absorption spectrum for atomic hydrogen—including several previously unobserved series of spectral lines—and the Rydberg constant for hydrogen R_{H} to within 7% of its empirical value.

Hughes makes much of the fact that it was predictions that were *unforeseen* at the time of the model's construction that were—famously—corroborated. Now Bohr would not, perhaps, have so eagerly extended the explanatory domain of his model from the intended problem area (atomic stability) to include unrelated phenomena (spectroscopy), had there not been the possibility that it captured *some* aspect of reality. However, it would have been entirely reasonable to expect such a model—*realistically interpreted*—to furnish explanations of apparently unrelated phenomena that were, at the time, thought to be atomic in origin.

Now Hughes rejects the traditional argument from the empirical success of a model to the existence of the entities appearing therein: such an argument must appeal to an inference of the form 'same phenomenal structure, therefore same internal structure'. This would be unjustifiable given that two entities with different internal structures may exhibit the same

law was derived by considering the action on the planets of a *driving force* originating in the sun. Perhaps these aspects are ignored because only the *surface* of his model lived on as an *advance*.

behaviour: 'To take an obvious example, behind the same software can lurk many different kinds of hardware.' (Hughes [1990], note 26) A conclusion that *can* be drawn, he claims, is the existence of the *subject* of the model—the atom—rather than its purported constituents (the electrons and nuclei). The argument runs as follows: the practice of model-building presupposes the existence of its subject. Model-building is sometimes a successful scientific endeavour. Where successful, model-building is vindicated. The building of Bohr's model was successful, and was therefore vindicated. So, therefore, is the assumption on which this activity is predicated: to wit, (in this case) the existence of atoms. Hughes eschews the *explanationist* argument, preferring what I have called the *vindicationist* variant:

> I am proposing a simple criterion of justification, applicable in all spheres of practical reason. If by adopting a certain practice we are led to success, then in this case the practice is justified. If not, not. ([1990], p.81)

Now we see why the existence of the *constituents* is not inferred: the activity of model-building does not presuppose *that*:

> The practice we are looking for is that of building a constitutive model. It involves two things: (1) assuming that an entity exists; (2) modelling its behavior in a particular way. If the model is successful, then both elements of the practice have been justified. The justification extends no further than the actions described. ([1990], p.81)

He concludes that

> what has been justified is precisely the assumption that a particular kind of entity, exhibiting a certain kind of behavior, exists. ([1990], p.81)

Now the *strength* of the realist thesis that would be justified by this argument is unclear. There are two possible readings of the phrase 'exhibiting a certain kind of behavior': the weak reading makes the conclusion uncontroversial (but not an especially *realist* one), but on the strong reading it is obviously false. Hughes baulked at the inference to a realist thesis about the inner structure of a model from its external—observable—structure. Perhaps what has been justified, then, is the existence of something that exhibits the observable behaviour of the model. The weak version is this: let us infer that something exists and call it 'atoms', where 'atoms' invokes only the regularities that realists cite as evidence for the existence of something *behind* the phenomena.[19] Thus Hughes quotes Suppe:

[19] In terms of the *semantic* view of theories, the weak view of what is justified is that part of reality is

> A theory has as its intended scope a natural kind of class of phenomenal systems. ... In propounding a theory, one commits oneself to the existence of the phenomenal systems within the theory's scope. (Suppe, quoted in Hughes [1990], note 28)

Would this commitment be peculiar to the realist? Surely not: the conclusion would be *banal*, because knowledge that 'atoms' existed would not imply our possession of any knowledge that could be subject to empiricist objections. For to say that the *empirical sub-model* of the model of a theory has a counterpart in reality is just to say that the theory is *empirically adequate*. Hughes stresses the realist credentials of his conclusion, indicating a stronger reading:

> Simply put, ... the assumptions at work in Bohr's theory, and justified by its success, are that atoms exist, that they are stable, that those of a given element are uniform in size, etc.—and that only an anti-realist axe-grinder would describe matters otherwise. ([1990], p.81)

Now there is no evidence that Bohr *himself* saw any distinction between the *non*-essential (internal) structure of his model, and the existential assumptions that were presupposed by its very construction. Nor do we have any reason to think that the facts about atoms—stability, uniformity of size—that have been established to Hughes' satisfaction were presupposed by Bohr's model constructing. Maybe what Bohr presupposed was that atoms *under his theoretical description* existed, but it has been clear for some time that no set of real entities *do* have all the required properties.[20] These, however, are relatively minor matters, for there is a more serious problem: whatever its strength, the conclusion cannot follow from the premises of the argument, for it trades on a conflation.

The conflation concerns two notions of justification. Hughes' 'simple criterion of justification' acts in the field of *practical reason*, and justifies a *practice*—in this case the building of models and using them *realistically*—as efficient means to some such desirable end as instrumental success. This criterion is eminently reasonable: we are surely justified in pursuing our ends via means that have previously satisfied those aims. In the argument, however, the justification is transferred to the realistic presuppositions of the practice, and read as *epistemic* justification. Even supposing the transfer to be possible—that having a belief or making a presupposition is the kind of thing that can be vindicated with respect to an aim as if it were a voluntary *action*—it is the wrong *kind* of justification. In practical

correctly described by the empirical *sub-model* of the successful model.

[20] In contrast to the causal arguments for entity realism advanced by Hacking [1983] and Cartwright [1983], we are given no special feature that picks out some subset of our theoretical claims about atoms for special attention.

reasoning, we judge the rationality of a practice by the degree to which it helps us to achieve some given aim. So the *vindication* has to be with respect to some desired *aim.*[21] That is why Hughes' argument cannot work: when we achieve our aims by actions that only make sense if we have certain beliefs, it does not follow that the presuppositions are true.[22] We could *at best* conclude that we had to have those beliefs in order to achieve those aims by those means, not that we must have those beliefs *simpliciter.*[23]

2.4. METHODOLOGICAL REALISM

So far, my conclusions have been predominantly *negative*: neither the explanationist nor the vindicationist attempts to fill out the methodological argument for realism straightforwardly support their intended conclusion. In this section, however, I will explore the common premise of these arguments (methodological realism, henceforth: MR), remaining *neutral* on scientific realism (henceforth: SR). MR is the claim that the adoption—by scientists—of realist aims, methods and inferences is (or has been) central to their construction and acceptance of (what turned out to be) successful theories. It is a small step from this to a *recommendation* that scientists be realists. This raises two groups of issues. Firstly, what *is* the logical relation between MR and SR? Does MR make *sense* as a recommendation outside the realist view of science?[24] Secondly, there is MR itself: I noted at the end of the first section that some non-realist critics (Fine and Laudan, for instance) chose to attack the

[21] Giere [1989] seems to offer an argument with a similar structure: he raises the issue of realism, giving a methodological answer (in DNA research, a realist programme yielded the greatest 'payoff'). What is unclear is whether we are meant to conclude that therefore DNA exists and has the structure attributed to it by Crick and Watson, or just that, with hindsight, their methods yielded the greatest payoff.

[22] Unless, of course, we know more about the *mechanism* by which the means are appropriate to the ends. If we knew the mechanism, and it required that the presuppositions were *true*, we could say that the truth of the presuppositions was the *best explanation* of our achieving our aims. That, however, is another argument.

[23] As ever, Duhem provides an apposite quote: 'Chimerical hopes may have incited admirable discoveries without those discoveries embodying the chimeras that gave birth to them. Bold explorations which have contributed greatly to the progress of geography are due to adventurers who were looking for the golden land—that is not a sufficient reason for inscribing "El Dorado" on our maps of the globe.' (Duhem [1914], pp31-2)

[24] Leplin [1986] defends a position that is similar to MR (see 2.5 below), but does not consider its *logical* relation to scientific realism.

premises of the methodological arguments (MR, that is). So *is* MR in fact a plausible claim when applied (for instance) to the history of science?

Consider first whether we can defend MR independently of a realistic view of the aims, methods and products of science. The problem is this: even though we have seen the failure of some of the arguments for SR from MR, isn't MR in some sense *incoherent* given an anti-realist view of science? According to SR the aim of science is truth, so that success in science is the possession of theories that are (approximately) true. Not only that, but SR claims those aims to be *achievable*: in all probability we advance toward their fulfilment with every new predictively successful and explanatorily powerful theory. MR is then almost trivial: it recommends the adoption of methods that are *appropriate* to realist aims. In the conjunction of SR and MR, then, we have an agreeable confluence of aims and methods. Anti-realists, in contrast, *deny* that we could have good reasons to think any but the *empirical* claims our best theories make are true. Our aims should be limited to what we can achieve. Now MR commends methods that are appropriate to *realist* aims, methods that also presume that those realist aims can be (and in fact have been) partially achieved by previously successful theories. Surely it is folly to adopt methods that are reasonable only on the assumption that we can know what—in principle—we *cannot* know.

This, however, is too quick: all that the harmony between MR and SR can show is their consistency (or perhaps that MR is a natural consequence of SR), but the methodological arguments seek to establish SR on the basis of MR. To the earlier criticisms of these arguments I would like to add the claim that SR is in fact *independent* of MR. This I will argue for by showing that MR can be motivated within either realist *or* anti-realist views of science.[25] If the consistency of MR and SR has already been established, it remains only for me to remove the air of paradox from the adoption of MR within an *anti*-realist framework. This proves to be surprisingly easy, for in van Fraassen's empiricist account of science we have a ready-made candidate.

First consider again the three realist theses set out in 2.1, concerning: (i) aims; (ii) semantics; and (iii) epistemic attitude. How might these theses, adopted by scientists, affect their practice? Realists often claim that, at the level of practice, realist aims counsel a search for theories that *explain*, rather than merely *save* the phenomena. The semantic and epistemological claims work together: a realistic construal of our best theories directs our interest to the consequences of conjunctions of theories covering disparate domains, and also licenses univocality assumptions and intertheoretic plausibility considerations (see

[25] In effect I will take a cue from elementary logic: if I exhibit models of $(A \ \& \ B)$ and $(A \ \& \ \neg B)$, I have shown that B is *independent* of A.

2.2). The epistemological claim rationalises the retention of (portions of) previously successful—but now refuted—theories in future theories.

Compare this with van Fraassen's constructive empiricism: 'Science aims to give us theories that are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate' ([1980], p.12). Now van Fraassen *accepts* the realist's semantic claim: 'After deciding that the language of science must be literally understood, we can still say that there is no need to believe good theories to be true, nor to believe *ipso facto* that the entities they postulate are real' ([1980], pp.11-12). So van Fraassen rejects realist theses (i) and (iii), but accepts (ii). Now for van Fraassen, epistemology is unlike totalitarian codes of law: it allows what it doesn't specifically prohibit, and it does not prohibit what it does not specifically allow ([1989], pp.171-6). The content of the realist's *belief* in a theory is logically stronger than the constructive empiricist's acceptance of it,[26] but belief in the extra content is not *irrational*, for the contents of the two beliefs are *empirically equivalent*, the extra realist belief could not make us more vulnerable to empirical surprises, and that is what counts ([1985], p.255).

This covers only *static* (epistemic) features of acceptance, but the happy effects of realist commitment outlined by Boyd and Putnam concerned *dynamic* features of science: how would our previous acceptance (in van Fraassen's sense) of a theory affect the shape of theories that we build and accept *in future*? Here van Fraassen turns to the pragmatics of theory acceptance: acceptance of a theory may involve *not only* the belief that it is empirically adequate, but also a commitment to a research programme, and to framing future explanations in its terms. Acceptance of a theory may also involve immersing ourselves in its world picture, letting it constrain the vocabulary *and* grammar of our theoretical discourse. Thus: 'to some extent, adherents of a theory must talk just as if they believed it to be true' ([1980], p.202). A survey of van Fraassen's answers to particular realist arguments concerning scientific practice underlines this *methodological indistinguishability* of acceptance and belief.[27] However, one does get the curious feeling that the actions of the realist are here just *re-described* in empiricist terms rather than explained or motivated—here we are talking, theorising and predicting as if we are realists, but with our anti-realist hearts pure. I will return to this point in 2.5.

[26] In terms of the semantic view of theories: belief in the empirical adequacy of a theory is just the belief that part of one of its models—the empirical sub-model—corresponds to part of the world.

[27] See his discussion of acceptance ([1980], pp.12-13), and of Putnam on conjunctions and Boyd on experimental design ([1980], chapter 4).

In any case, a more interesting consequence of van Fraassen's separation of evidential and pragmatic aspects of theory choice (and therefore of epistemology and methodology) allows him to endorse the realist search for *explanatory* theories. The problem is this: the history of science is littered with highly successful theories for which their creators sternly held out, seeking explanation, when there were extant theories that *saved* the phenomena. Answering Feyerabend's charge that a search for empirical adequacy *alone* might therefore hinder scientific progress[28]—the search for explanation has 'paid off handsomely'—van Fraassen argues:

> Paid off handsomely, how? Paid off in new theories we have more reason to believe empirically adequate. But in that case even the anti-realist, when asked questions about *methodology* will *ex cathedra* counsel the search for explanation! We might even suggest a loyalty oath for scientists, if realism is so efficacious ([1980], p.93).

In effect, van Fraassen admits that *if* realist aims (in particular, the aim to explain) contribute to the possession of empirically adequate theories, then one who rejects the adoption of explanation as an aim in favour of empirical adequacy might nevertheless endorse its adoption *by scientists.*

The anti-realist then has a choice: faced with some successful feature of scientific practice whose plausibility appears to depend on acceptance of aspirational realism, she can *either* construct an alternative empiricist rationale for the practice *or* accept its apparent realistic commitment, but endorse the method as conducive to *empiricist* aims. But does not the second strategy collapse into the first? No: the force of Feyerabend's argument was that the search for explanation might not be so fruitful were it not pursued *as an end in itself.* There is a parallel in ethical theory: utilitarians endorse the acceptance *by others* of a non-utilitarian rule if its compliance utility is higher than that of a corresponding utilitarian rule.[29] In both ethical and methodological cases, the possibility of such an endorsement is a consequence of the consequentialist's separation of motivation and appraisal. Now van Fraassen accommodates only the realist *motives* by this mechanism, but perhaps it is

[28] Feyerabend [1964] cites the impetus to developments in dynamics provided by difficulties encountered by the Copernican system—*realistically construed*—against the background of the prevailing Aristotelian dynamics.

[29] 'Thus a Utilitarian may reasonably desire, on Utilitarian principles, that some of his conclusions should be rejected by mankind generally; or even that the vulgar should keep aloof from his system as a whole, in so far as the inevitable indefiniteness and complexity of its calculations render it likely to lead to bad results in their hands' (Sidgwick [1877], pp.448-9).

possible that anti-realists who reject *other* elements of the realist position—concerning semantics and patterns of commitment—might endorse *their* adoption in the same way.[30]

Take the epistemic component of realism first: the anti-realist must weigh up the positive and negative pragmatic consequences of scientists' having beliefs that (according to the epistemic anti-realist) are unsupported by evidence. Now van Fraassen *himself* argues convincingly that anyone who accepts the thesis of the underdetermination of theories by evidence *must* accept that there could be no negative consequences (in terms of nasty empirical surprises) to which scientists will be subject because they believe a theory to be *true* on account of its predictive success, rather than empirically adequate. On the other hand, there may well be positive *heuristic* consequences of believing a theory to be true or, perhaps more accurately, there might be negative heuristic consequences associated with the scientist's limiting her commitment to the empiricist's weaker belief that the theory is empirically adequate: that is the force of the next section.

Turn now to the semantic component of realism: MR does seem to be consistent with *some* forms of semantic anti-realism. The consistency (or otherwise) depends on the closeness of the relationship that the particular anti-realist sees between meaning and use. On the one hand there are the logical positivists who were willing to admit that, besides *cognitive* significance, a statement might have *emotive* significance.[31] For the positivist, the cognitive meaning of a theory *T* is given by its translation into statements concerning actual or possible observations. The extra-empirical 'content' that the metaphysically minded interpreter of *T* claims it to have could not be *propositional*, given the verifiability criterion of meaningfulness, and so must be *emotive*: associated feelings and images. It is perfectly consistent within this view (although perhaps unlikely) for the associated images to be suggestive in heuristically fruitful ways, and therefore for the super-empirical 'content' to play a *positive* role in science. In contrast, for the anti-realist who, like Dummett, gives a *genetic* account of the connection between meaning and use (that is, one in terms of language learning), it is difficult to see how a *stable* set of connotations could become associated with a sentence, or set of terms appearing therein, such that these connotations contribute to an understanding of the sentence in terms of extra-empirical truth conditions. Since it would be from such stable connotations that any positive heuristic consequences would flow, we might conclude that a Dummett-style semantic anti-realism does seem to

[30] This need be no self-defeating attempt at self-deception: the appraiser can stand outside the community—scientists—whose beliefs are appraised on their effects. There is a parallel here with the functionalist explanation of religion in primitive societies (also drawn by Elkana [1981], pp.42-4).

[31] See for instance Carnap [1932], section 7.

rule out MR. Thus there are forms of semantic anti-realism with which MR is consistent, and others with which it seems to be *in*consistent.

In conclusion, both the realist and the anti-realist might accept MR's recommendation: in both settings, the plausibility of MR depends crucially on claims about past and present scientific practice, although the appraisals will differ with respect to aims and the efficacy will be explained by different mechanisms. *Anyone* who takes discovery and explanation to be the aims of science will find it probable that the best way to achieve those aims is for them to be pursued directly. In Boyd's account of science in particular, however, the plausibility of the realist view of science as a whole depends on its explanatory power with respect to the reliability of realist methods. The key evidence for the picture as a whole would be a historical record showing the successful application of Boyd's realist methods in science. For the anti-realist (of whatever stripe), the plausibility of MR must depend more directly on a meta-induction. So *have* realist methods been widely adopted? Where adopted, *have* they been associated with success? In 2.5 I will explore some realist methods, and provide a few examples of their successful application along the way. In chapters 3, 4 and 5 I will pursue some lengthier case studies. However, if these examples provide a relatively poor base for induction, one could at least claim that MR is supported by historical counterfactuals: had realist methods *not* been adopted by the creators of theories in some key episodes in the history of science, then some of the theories[32] which have been the basis of great progress over the last century would never have appeared in the form that they did.

2.5. THE CONTENT OF METHODOLOGICAL REALISM

In this section I will explore the content of MR: the methods that, it is argued, are open only to realists. In an article whose spirit is very close to the present project, Leplin [1986] has argued that many research efforts would have no obvious rationale if the aims of the researchers were those endorsed by empiricists: to test the empirical adequacy of theories. His examples include Millikan's determination of electronic charge (the electron was not tied to any particular theory) and the investigation by contemporary astronomy of open versus closed models of the universe (general relativity is consistent with all current models). Modern particle physics is another case:

> The motivating idea of the search for new particles had nothing to do with theory assessment. It was simply that if there really are these quarks then there ought to be combinations of them with properties

[32] Such as Bohr's atomic model (see chapter 3) and quantum mechanics (see chapter 4).

> different from those of known particles. So there was an opportunity for discovery. The purpose of the search was not to test or extend theory, but to learn how quarks combine. (Leplin [1986], p.38)

More helpful than examples, however, is Leplin's separation of two explananda for MR: the 'realist' *practices* and their *success*. He claims explanatory power for methodological realism only with respect to the former, arguing that empiricist criticisms of the methodological arguments for realism too often 'confuse methods with results' (p.39). To see the distinction more clearly, it would be helpful to examine Leplin's two intimately related types of 'realist' activity: (i) attempts to replace sets of theories that are empirically adequate over disparate phenomenal domains with a unified successor; and (ii) the ontological conservatism of amendments to theories in the face of empirical refutation.

First note that 'Attempts at unification pervade science.' (p.39) Aspirations to empirical adequacy cannot be the motive, because the *previous* theories were empirically adequate. If, however, the previous theories were *true*, it is obvious that a *theoretical unification* will be successful over their joint domains. So SR explains the strategy and its success. Problem: moderate realists admit that successful theories might only be *approximately* true:

> Truth, but not necessarily approximate truth, is preserved under the operation of unification. A true unifying theory should be as successful or more so than the collection of theories unified. But the logic of approximate truth suggests that the unifying theory will, if anything, be less successful. Why should not the degrees to which the theories unified fall short of truth be additive or worse? ([1986], p.40)

This blocks the argument for SR, but the *interest* in unification remains to be explained:

> The attribution of realist assumptions to the theoretician purports to explain not the success of the theory he produces but the existence of a methodological prescription to seek a unifying theory where our current theoretical account of a domain of phenomena is far too eclectic to be believed. Methodological realism appeals to truth as a goal. ([1986], p.40)

The unifying impulse is not amenable to an *empiricist* rationale however, for it is far from clear that candidate pragmatic theoretical virtues (like convenience, simplicity or economy) in fact *do* increase under unification; mathematical tractability almost certainly will not. Another typical realist move is ontological conservatism across theory change. The methodological arguments for SR assume that only the approximate truth of the old theory can explain the empirical success of this type of response to anomaly. Unfortunately, anti-realists point out that *approximate* truth doesn't explain anything of the sort, and that the argument is *circular* if the putative conclusion is SR, because the realist first has to show that we can infer the old theory's approximate truth from *its* predictive success. Leplin therefore shrinks the explanandum:

> The fact that theoreticians attack anomalies by means of natural extrapolations of successful ontologies, distrusting solutions involving radical changes, is evidence of their realist aims. If the incumbent ontology is but convenient fiction, there is no reason to expect better results from its extension and refinement than from revolution. ([1986], p.42)

The instrumentalist's rationale for ontological conservatism is mathematical convenience: old, familiar theories yield old, familiar equations with old, familiar solutions. Leplin's reply is more adamant than cogent:

> The point, however, is that short of realism there is no reason to suppose outstanding problems tractable in terms of the incumbent ontology, however successful it may be in other areas. If a theoretical mechanism works as well as it does just by chance rather than in virtue of its reflecting the actual mechanisms responsible for predicted effects, it is simply irrational to believe in its extendability. Yet it is the practice of science to assume the extendability of successful mechanisms, to assume that outstanding problems are more likely to be solved by means of existing mechanisms than by means of alternatives that do not yet claim empirical success. ([1986], p.42)

Now two issues have been run together here: the discussion started with (i) theoretical responses to anomaly ('outstanding problems'), but then turned suddenly to (ii) the extension of successful theories into new areas ('extendability'). These are better separated.

Ontological Conservatism: Leplin has given *no* reason why an old (refuted) theory should not—with empiricist blessing—be fixed with a 'fudge-factor' to cover refutation. A further problem is that, sometimes, radical changes in ontology *do* occur, with great success. Leplin's rationale for ontological conservatism seems to leave no room for this second kind of change. The first problem is at the heart of Fine's [1984] criticism of the realist motivation for ontological conservatism, which he calls the 'small handful' strategy. Fine claims that while realism *fails* to explain (and therefore motivate) the 'small handful' strategy, instrumentalism *succeeds* in providing such a motivation. At any given time, so the (realist) argument goes, instead of the infinite host invoked by underdetermination arguments, only a *small handful* of possible theories are considered by scientists: those that are related to the most successful refuted theory in certain special ways. This is reasonable because the old theory was approximately true, and our attention might fruitfully be confined to those theories which stand in some correspondence relation or have a 'family resemblance' to it at the ontological level. If one is a realist, one can expect that the new theory will inherit the happy characteristic of approximate truth. The explanationist claim is that SR both rationalises the strategy and provides the only explanation of its success. Fine counters that SR doesn't, in fact fulfil either of these claims. The explananda are three: the small size of the handful, the family resemblance, and the success of the strategy. On the first question, the realist would still be faced with an infinity of choices. On the second question, SR again fails, because we can't infer the approximate truth of the *old* theory from its previous empirical success: *that* is as suspect as any other realist inference. Therefore the *new* theory's success cannot be explained on the basis of its transferred approximate truth: for all we know there is none to transfer. Worse, even if the old theory *was* approximately true, the element of truth that it captured might not be passed on to the new theory, such is the logic of approximate truth. The third explanandum requires no explanation, because it is *false*: the small handful strategy too often *fails*.

On the first two questions, Fine claims, the methodological instrumentalist—one who claims that successful theories are mostly constructed with purely pragmatic virtues in mind—has an explanation that enjoys the virtues of the realist's, while dispensing with the dubious allusion to approximate truth. Consider the 'small handful' part: constructing theories that conform to known empirical constraints is difficult, but we might as well focus on the few theories that satisfy the following pragmatic considerations. Firstly, it is quite reasonable to keep the highly confirmed bits of an old theory, while tacking on some new bits to cover any evidence that refuted it. Retention of familiar and tractable mathematical structures has the same rationale, which explains the family resemblance. Thus the conservatism of the small handful strategy is easily explained given that empirical adequacy is the aim. If it often fails, this reflects the 'trial and error' nature of the exercise. SR, meanwhile, has 'struck out' ([1984], p.89).

Now Fine does not distinguish between SR and MR, and some of his criticisms of the SR explanation of ontological conservatism need not threaten MR, as Leplin has pointed out: it doesn't matter whether *we* think that the old theory was approximately true, for what counts is whether the *scientist* thinks that it is (in some specified respect). But what of *radical* theory changes? The instrumentalist version of the 'small handful' strategy is shown to be *too permissive*: it indicates that the instrumentalist, like the bad builder, will opt for the easy, patched-together option. Thus it falls foul of Feyerabend's objection: so much the worse for subsequent science if the *Copernicans* had been happy with one of Fine's botched-together theories instead of holding out for a unified theory that was capable of *explaining* the phenomena it saved. Now one typical reply here (advanced by Duhem, for instance), invokes convenience: when so many patch-jobs have been carried out that the theory begins to be mathematically unwieldy, it might be time for radical change. That answer will not work for the creators of quantum mechanics: if Pauli, Heisenberg and the rest had acted like the instrumentalists they are often portrayed as, they should have been *perfectly satisfied* with the old quantum theory, which could be made empirically adequate with suitable *ad hoc* adjustments. Instead they waited for the long-heralded quantum mechanics. Nor is there any reason to think that mathematical convenience could have been a very important factor in their reception of the new theory: Heisenberg's matrix mechanics was famously *in*tractable, but was nevertheless hailed by Pauli—the arch-critic of the old quantum theory—as the long-awaited advance.

Thus there *is* a realist response to Fine's instrumentalist rationale for ontological conservatism, but more complex methodology is required than Leplin deployed: a two-track methodology. Sometimes, an ontologically conservative theory change *is* the right move to make: such moves can be given both realist and instrumentalist rationales. However, such responses to anomaly are not always appropriate: where a series of conservative moves seems to have produced a regressive series of *ad hoc* adjustments, it is time for a radical change. But the choice of *when* to make the radical change seems to be governed by realist criteria of theory choice—coherence and explanatory power—for the instrumentalist analogues—simplicity and predictive power—will not necessarily favour the radical new theory, at least in the short run. So although an empirically adequate but *ad hoc* theory is better than none if there is no alternative, the instrumentalist rationale for the 'small handful' strategy is *too* permissive.

The above reasoning is subject to a two-pronged criticism. Laudan [1984] directly confronts the realist's contention that ontological conservatism is a common feature of theory construction: he *denies* that new and old theories are generally related in ways that are consistent with the realist rationale for retention. Cartwright, taking as her examples quantum-mechanical models that are actually *used*, urges that models—the vehicles of scientific achievement—rarely bear the right relations to the fundamental theories in which they are purportedly embedded. Theories don't *explain* within models, and the success of the models won't support the theories as *factual* claims.[33] Thus the realist tells the wrong story about the majority of cases of model-building. If Laudan is right, ontological conservatism is a myth. If Cartwright is right, explanatory power is a myth (for theories) and coherence is a myth (for models). Two responses are possible here: either (i) get into an argument about the relative numbers of cases of theory-construction that fit the realist and non-realist templates, or (ii) limit the realist claim to the 'commanding heights' of theorising. The rationale is as follows: when applying quantum mechanics to lasers, numerical accuracy is important: model-building will be a messy and pragmatic business. However, think of the contrast with the initial introduction of the *same* theory in 1925: would we really have counselled Pauli, Heisenberg and the rest to be content with the old quantum theory? Instead of *numbers*, the methodological realist turns to historical counterfactuals: would we have had the Newtonian synthesis, Bohr's atom or matrix mechanics without stern demands for explanation?

[33] Cartwright [1983], pp.100-27.

Extending Theories: Leplin failed to give a convincing reason why the instrumentalist *cannot* see how mathematically-tractable theories should be extended into new domains. Leplin correctly argues that there is no reason for the instrumentalist to *expect* theories to be successful (see the last-quoted passage), but *no reason why* is not the same as *a reason why not*, and the latter is what Leplin's inappropriate use of the word 'irrational' implies. Leplin also provides no insight into what it *is* about a theory that makes it a promising candidate—to one who construes it realistically—for extension into a given area. He does cite metaphysical uniformity of nature assumptions, but uniformity of what? Would it be rational to substitute the dynamics of banana prices into the kinematics of electrons? *Some* theoretical connection is surely required between otherwise disparate phenomenal areas to motivate the extension of a theory associated with one into the other, at least something more substantial than vague 'principles of the long-term reliability of evidence, and the underlying unity of nature and concordance of natural law' (Leplin [1986], p.43). In fact the required connection *can* be identified given a realistic interpretation of the theory whose domain of applicability is to be extended: part of the theoretical structure of the theory to be extended might (for instance) invoke some process or entity that is relevant to another domain, given other theories in that domain; alternatively, common terms might appear in two theories associated with different domains, leading to a transfer of theoretical structure between the two domains. I will explore these possibilities below.

In 2.2, we saw Putnam and Boyd claim that certain uses of scientific theories implied their realistic interpretation by scientists. Putnam pointed to the communicability of scientific results, and the closure of scientific theories under deduction. Boyd raised the assumed *univocality* of occurrences of the same term in different theories. Are these features of scientific practice explicable only for the realist? How do they contribute to scientific progress? First consider closure under deduction: scientists, a realist might argue, would only be interested in the logical consequences of theories if they thought the premises—theories—to be *true*. Why would empiricists be interested in deduction rather than some other inference engine? After all, deduction is sound for *truth*, rather than empirical adequacy. An instrumentalist might reply that science is primarily concerned only with the *observational* consequences of theories. We might *design* theories that tersely summarise a large number of *observation conditionals*, but a largely fictional non-observational edifice might facilitate the deductions and make the theory simpler to use and easier to remember.

But there is a problem here: the methodological instrumentalist's picture could surely only rationalise an interest in the *intended* empirical consequences of theories, but it has often happened that the greatest successes for theories *as instruments for prediction* have been in *unintended* domains, Bohr's extension of his atomic theory of 1913 to cover spectroscopy being a particularly clear example. Now the prediction of hitherto *unknown* kinds of facts is extension into unintended areas *par excellence*: Popper has emphasised this kind of prediction. Instrumentalism accounts for prediction of events of a kind which is known, but theory-led discovery of new types of fact is a mystery to the instrumentalist, because

> if theories are instruments for prediction, then we must assume that their purpose must be determined in advance, as with other instruments. Predictions of the second kind can be fully understood only as discoveries. (Popper [1963], p.118)

In similar vein, Zahar ([1989], pp.38-9) has catalogued some of the ways that a realistic interpretation of even quite exotic parts of a theory's mathematical structure can lead directly to the discovery of new kinds of facts, one example being Dirac's physical interpretation of the negative energy solutions to his relativistic equation, and the subsequent discovery of the positron. If realistic interpretation fosters an interest in the unintended consequences of one's theories, it thereby increases their empirical content, suggesting the formulation of new conjectures, thereby promoting new discoveries. It might be objected that new uses may be found for old instruments, but if 'instruments' are things that are re-employed as often as theories are fruitfully extended into new domains, methodological instrumentalism must lose its distinctiveness as a methodology. Interpreted thus loosely, there may be nothing in methodological instrumentalism to *prohibit* extension into new areas, but that is the problem: no *positive* reason is provided by a theory's success in one area to think that it might be successful in another. On an instrumentalist interpretation of a theory, the correctness of its unintended empirical consequences would be supererogatory, while on a realistic interpretation, the correctness of unintended consequences would constitute a criterion of adequacy. Thus where scientists are found worrying about the correctness of the empirical consequences of their theories outside the domain for which the theory was constructed, we have evidence that the theories in question are interpreted realistically and realist aims pursued.

So is *univocality* a realist phenomenon? In 2.2, we saw Boyd claim univocality for realism: where two distinct theories are taken to invoke the *same* entity, we can no longer say that their non-observational content is mere theoretical superstructure, whose only role is the *internal* one of supporting deductions of observational consequences. Instead, theoretical terms must refer to *theory-independent* entities. Consider again Bohr's use of his atomic model to derive a theory of line spectra: (i) The initial theory of the Rutherford atom's stability, T_A, and the background observational theories concerning atomic spectra, T_B, both mention 'atoms'. (ii) Now unless the 'atoms' in T_A can be identified with those in T_B, there is no reason to expect that T_A might explain T_B (or exhibit any other *cognitive* connection). (iii) Taking both the 'T_A-atoms' and the 'T_B-atoms' to be references to *real* atoms—rather than fictions particular to those two theories—would suggest their identification. Thus Bohr's extension of his theory suggests that he assumed the occurrences of the term 'atom' in his own theory and in spectroscopy to refer univocally to theoretical entities of *independent* status. Since it was the extension that earned the model its plaudits, even the instrumentalist must admit that Bohr was *instrumentally vindicated.*

A possible counter is that the identification was between two episodes of the observed behaviour of hydrogen gas. This will not do: there was no obvious observational connection between the evidence for Rutherford's structure—the statistics of α-particle scattering by thin gold foil—and the spectroscopic experiments on hydrogen in discharge tubes to which Bohr later turned his explanatory attentions. In any case, Bohr's initial calculations on the Rutherford structure sought to answer a *theoretical* puzzle concerning stability, rather than an empirical one. The connection assumed by Bohr's extension of his theory must have been at a deeper theoretical level: the enduring atoms that individuate the elements and whose properties explain their chemistry. In contrast, for the methodological instrumentalist the natural vision for the unity of science is surely Neurath's: unification at the point of application.

A subtler example of the heuristic utility of realistic construal—connected with univocality assumptions—is provided by Zahar [1989]. Suppose some equation H is re-expressed in a suggestive—but equivalent—mathematical form $H^*(t)$. Under a realistic interpretation, term t falls under a philosophical category which is subject to some philosophical opinion (such as a symmetry or conservation principle) held in high regard by the theoretician. If $H^*(t)$ violates the principle in question, we can expect her to *modify* the theory so that it is satisfied. Zahar contends that it is moves such as this—that assume univocality at an altogether more abstract level—that typify the heuristic behind Einstein's relativistic research programme, and that it was this heuristic that differentiated this programme from its (empirically equivalent) contemporary rivals. Another striking example of this kind of theoretical constraint on the acceptance of theories—inspired by a realistic construal—is provided by Heisenberg's supposedly instrumentalist construction of matrix mechanics. On the *same night* that he wrote down the equations that were to become matrix mechanics, Heisenberg checked that they satisfied the principle of the conservation of energy.

Now it would appear that most of these criticisms of methodological instrumentalism would be relevant only to those forms of instrumentalism that reject *semantic realism* for theories, on the basis that theories are just *tools for prediction* rather than factual statements about the world. Given this appearance, van Fraassen might reply that since he *accepts* semantic realism, the above arguments could not show that the adoption by scientists of the above methods of theory construction would constitute evidence against the thesis that scientists are constructive empiricists (this latter thesis being *methodological constructive empiricism*). Under van Fraassen's account of acceptance, the only *belief* involved in acceptance is in the empirical adequacy of a theory, but *pragmatic* features of the acceptance of a theory include scientists *immersing themselves* in its world picture, committing themselves to using it to answer requests for explanation, and allowing its terms to suffuse scientific discourse ([1980], pp.12-13). For van Fraassen, acceptance *includes* a commitment to the (non-observational) structure of present theories being a constraint on the construction of future theories (no matter what the domain, it seems). It is this last commitment that would underwrite the methodological constructive empiricist's ability to explain features of theory construction such as ontological conservatism and the extension of theories into new domains. As I noted in 2.4, the problem is that this commitment—crucial as it is to van Fraassen's ability to give a plausible account of theory construction—seems to be *entirely unmotivated.* Given that the empiricist's acceptance of a theory involves only the belief that it is empirically adequate, we can see why the empiricist would wish to preserve the structure of the *empirical sub-models* of the models of a theory, but why should the *non*-observational structure of those models provide a constraint on future theorising? Just because we know that a theory is empirically adequate over *one* domain, why should we think that it (rather than its empirical equivalents in that domain) should be empirically adequate in *another* domain, when the alternatives may be incompatible with respect to the new domain? Although methodological constructive empiricism is *consistent* with the application of methods that are founded on these assumptions, it hardly *explains* them. The same goes for the search for explanation: van Fraassen countenances the search for explanatory theories as reflecting a pragmatic requirement. This would appear to make explanation *supererogatory*, but explanation is often a *duty* of theory. Methodological realism, in contrast, seems perfectly able to explain these features of scientific practice.

2.6. INTENDED INTERPRETATIONS: METHODOLOGICAL REALISM AND MSRP

In this section I will draw on further historiographical support for MR by showing that Lakatos' MSRP *embodies* MR. Now Lakatos was committed to a normative version of MR in an uninteresting way in so far as he accepted what in 2.1 I called *aspirational realism.* Normative aspirational realism is associated with MR as any aim is associated with methods that are directed towards that aim. However, the commitment to MR I would like to explore here is a deeper one: MSRP *embodies* the content of MR (explored in the last section), because one of the central notions of its apparatus of appraisal—the positive heuristic—can be explicated in terms of explanatory power and theoretical unity, the theoretical virtues that—according to MR—are directly pursued in successful research. In so far as MSRP is supported as a descriptive account of scientific research by its relation to the history of science, MR will also be supported if MSRP does in fact embody MR. In 1.2, it was argued that the construction of refutable variants in line with the *positive heuristic* of a research programme could be identified with model construction using analogical reasoning. On this reading, a feature of a theory is *ad hoc*$_3$ if it fails to 'fit' with the analogies that underwrite the positive heuristic. It now remains to explore the ways in which realist commitments can be implicit in the use of analogies to extend theories.

A model is a theoretical representation of a particular kind of system. For Hesse, models are based on analogies with more familiar theories or structures, and are indispensable devices for the pursuit and understanding of less familiar theories. They also provide the 'open texture' by which the equations which express physical theories are tested, extended and modified (Hesse [1953], [1961] and [1966]). A model is an *indispensable* heuristic tool because it

> can be generalised, extended and tested, and if necessary modified, as a purely deductive system cannot. The model can be tested, because it is a system of entities and processes whose behaviour is already known apart from the new experimental facts which it is being used to explain. ([1961], p.21)

Where a theory is interpreted through an analogy that is known to hold in *some* respects, we can ask *how much further* the analogy obtains. Answering theoretical questions of this kind—suggested by the analogy—enables the model to be 'fleshed out'. A good example of this heuristic role is the 'billiard ball' model for gases:

> Further questions can be asked, such as 'Are gas molecules like rigid balls or like elastic ones?', 'What is their diameter?', and so on, and the theory is tested and developed by devising experiments to answer questions like these suggested by the model. ([1961], p.21)

Now models need not be quite so concrete, and indeed may operate at quite abstract levels, for example the analogical interpretation of Riemannian geometry in general relativity through an analogy with the 'geodesic' of two-dimensional geometries projected onto spherical surfaces. An analogy need not be perfect to be useful, and Hesse distinguishes *positive* and *negative* analogy: structural similarities and dissimilarities between the analogue and the system to be described. There is also *neutral* analogy: those features of the analogue that are *not presently known* to have counterparts in the real system. Two aspects of Hesse's account are crucial to the present discussion. Firstly, an analogy can encode *factual* information about a system that is described with its help: what was 'precarious theory' suggested by analogy can be elevated by the passage of successful research to the status of accepted fact. The crucial logical point is that models can be known to be *false*:

> And if they were, in fact, false then they could, logically, have been true, and this is sufficient to place all such theory-models in the category of factual statements, and to enable us to make finer distinctions between those which were better or worse approximations to the truth. ([1961], p.26)

The second point of importance follows from this: the distinction Hesse makes between models that are used *realistically* to extend or modify a mathematical theory, and those that are designed for, or consigned *to*, non-realistic use, aiding the understanding or manipulation of mathematically intractable theories.[34] To make distinctions between more and less realistic analogical models suggests that models—and the analogical relations through which they are filled out—are appraised for their representational content, not only their heuristic value. This is further underlined if certain kinds of model are identified explicitly as *non*-representational, for it would not make sense to mark off non-representational models if *all* models were non-representational.

This distinction—and its relevance to methodological realism—is supported by a closer examination of model construction in science. Redhead [1980] supplies a realist analysis and descriptive account of the use and status of physical models. An important distinction Redhead makes is between 'impoverishment' and 'enrichment' models. An enrichment model is the vehicle for the application of a general theory to a particular case: extra information about the case at hand is added to what the general theory says by virtue of its applying to *any* system. Take for example the 'filling in' of dynamical details—the form of the potential function—within quantum mechanics' kinematical framework. The 'filled-in' detail comes from background knowledge of the entities and interactions that are present in the subject system. Thus when a quantum-mechanical model of a molecule is constructed, the Hamiltonian is written down in terms of the charges and masses of constituent species, and the forces taken to govern their interactions. Another example is the Newtonian model of the solar system, in which (in the first, idealised instance) a system of point masses is specified, dynamical details (gravitational interaction) filled in, and the kinematics (the three laws of motion) applied. Such a model does not necessarily *misrepresent* the subject system, or contradict the general theory, but it will often be idealised in the interests of mathematical tractability: thus non-electrostatic interaction terms in molecular Hamiltonians (that quantum mechanics *qua* general theory 'says' will occur) are often neglected, as are gravitational interactions between planets in the solar-system case. Redhead also notes that those parts of a model that are fixed by the general theory (and are therefore *model-independent*) are more 'highly regarded' than model-dependent parts, an observation that is in line with the Lakatosian distinction between hard core and auxiliary hypotheses. If the process of enrichment produces insoluble equations, an *impoverished* model can be proposed, playing a role similar to Hesse's *analogue machines*. Impoverished models are *assumed* not to correspond to reality: such models will (for instance) not satisfy the equations for an exact model. Redhead also notes that such approximate theories specify

[34] Hesse [1961], pp.26-7 describes four types of non-realistic model. See also Leplin [1986], pp.44-9.

different sets of possible points in the relevant state space to the general theories they ostensibly approximate, and are therefore *different theories* in a technical sense outlined elsewhere (see Redhead [1975]).

Redhead also distinguishes between 'floating models'[35]—and the *ad hoc* adjustments that are needed to bring them into line with experiment—and models that *do* have a theoretical rationale. An example of this distinction is the use of models of molecular structure in calculations of the infrared spectra of organic molecules. The 'received' view of how to apply a theory mirrors Hempel's 'covering law' account of explanation: we apply a theory to a particular system by seeing how the system 'falls under' the laws of the theory, as if the *representation* of the system were already given. Thus, in the case of quantum mechanics and molecules, we write down the Hamiltonian for the molecule, solve the resultant eigenvalue equation, and separate out wavefunctions for different motions. One major difficulty makes this approach impracticable in real life: the eigenvalue equations for interesting molecules are insoluble. Instead, the Hamiltonian for particular motions of the molecule are filled in piecemeal via idealised molecular models drawn from classical chemistry. Such models are analogical in so far as we interpret them through our knowledge of macroscopic systems of balls and springs, and use our knowledge of these analogues to extend the models (for instance to account for stearic hindrance in reaction mechanisms). They are, moreover, well motivated (see 5.4): even where we are forced by intractable mathematics to use impoverished models, we can see how some such models are motivated, and can therefore make distinctions between *representational* and *non-representational* impoverished models. Note, however, that the rationale that grounds the use of a well-motivated model need not be *internal* to the general theory in which the model is embedded: its justification may arise from background theories, although background theories will impinge on the construction of the model via constraints that are expressed *in terms of* the general theory.

Redhead observes that a pejorative distinction is often made by scientists between 'theories' and 'models': what begins as a 'model' can, if successful, come to be regarded as a proper 'theory' (the history of stereochemistry being a case in point). Conversely, an over-simplified, unsuccessful theory can be downgraded: thus, for example, the early kinetic theory later became a mere "billiard-ball model of a gas". Now this usage of the word 'model' differs from that previously adopted in the present discussion, in which models are *structures* representing real systems *within* general theories. However, the pejorative usage aligns with the above distinctions between different types of 'model' in the standard sense: (i) those whose positive analogy to the system to be described is taken to be merely

[35] The term, due to Post [1974], applies to models that are free of either empirical *or* theoretical grounding.

suggestive—that is, an aid to the understanding or manipulation of a theory's mathematical structure (Hesse calls this 'dead metaphor'); and (ii) those whose structural analogy could be *deeper* than is presently known to be the case, and might therefore be used to extend the theories that are interpreted through them. Now if this kind of distinction *is* made between the two types of analogy in the context of theory construction—the fact that there have been *transitions* between them implies this—and if analogies that are at the heart of progressive research *do* thereby earn the status of type (ii) analogies, it is hard to see how this would be rationalised by the instrumentalist. For the instrumentalist, all models should be of type (i): why would one *extend* a theory based on 'mere' analogy? Model-use for the instrumentalist (such as Duhem) is an *external* matter of suggesting equations to weak minds that cannot manipulate a bare mathematical formalism without heuristic aid.[36] The realist, however, can have analogies at the *heart* of theory construction.

To conclude the discussion of model-building: the use of analogies provides a mechanism for the extension of theories that explains the *open texture* of physical theorising. Only if an anologue is thought to provide clues as to the deep structure of the target system could it make sense to use that analogy to extend or enrich a model of the target system, in the hope that neutral analogy becomes positive analogy.[37] Also, if scientists make methodological distinctions between *representational* and *non-representational* models, and if these distinctions are reflected in their *use* of such models, the challenge is for the instrumentalist to rationalise this distinction if—on *general* arguments—*no* analogical models are taken to be representational. Of course it is consistent for scientists to extend models using analogical reasoning, and perhaps even to make *distinctions* between different analogical models while keeping their instrumentalist hearts pure, but this leaves their behaviour unmotivated and unexplained. The methodological realist can challenge the methodological instrumentalist to give a *positive* rationale for these methods that would explain their *systematic* deployment.

How do these observation relate to MSRP? In response to the Duhem-Quine thesis, Lakatos suggested that theory change be modelled on the diachronic appraisal of series of theories rather than the synchronic appraisal of single theories. The new scheme of appraisal was ***methodological***, in that attention was focused on the motivation of amendments to refutable variants, which is a function of the context of theory-construction. Methodological appraisal of individual theories is the process of applying the Lakatos-

[36] See Mellor [1968] for a discussion of Duhem's attitude to models.

[37] The instrumentalist might reply here that analogies are just one tool for theory construction among many, but the point here is that analogies are used *systematically* to extend theories. An instrumentalist rationale would give analogies only a *trial and error* application in theory construction.

Zahar criteria of *ad-hoc*ness (Lakatos [1971a]; Zahar [1973] and [1978]). The definitions of what it is for a theory to be *ad hoc*$_1$ and *ad hoc*$_2$ invoke a direct relation between a theory and the phenomena it was designed to save: a theory is *ad hoc*$_1$ if it can explain no more than that for which it was constructed; it is *ad hoc*$_2$ when its excess content over the problem situation for which it was constructed has not yet been corroborated, or has been falsified. The definition of what it is for a theory to be *ad hoc*$_3$ is qualitatively different, in that it turns on its relationship to the *positive heuristic* of the research programme of which it is a product: a theory is *ad hoc*$_3$ if it is not constructed in the spirit of the positive heuristic. This last notion has enjoyed more examples than explication. Now the first two pejorative categories—*ad hoc*$_1$ and *ad hoc*$_2$—concern the circumstances of the *proposal* of individual refutable variants. For *ad hoc*$_3$ theories, however, the problem concerns a *standing relationship* between the individual theory and the positive heuristic. If we explicate the notion of the positive heuristic in terms of analogy, then one way that a theory can be *ad hoc*$_3$ is that it fails to fit the analogies that drive the positive heuristic. Now for a refutable variant to 'fail to fit' the analogies that drive the positive heuristic is for some feature of that theory not to have a natural interpretation in terms of those analogies. To require that a research programme is theoretically progressive is therefore to require that, associated with the research programme, there is a set of analogies through which (features of) individual theories that arise in that research programme are interpreted:

> the—well planned—building of pigeon holes must proceed much faster than the recording of facts which are to be housed in them. (Lakatos [1970], p.188)

Now in 1.2 it was noted that the positive heuristic can work at different levels. It would be worthwhile exploring the operation of interpretive analogies at these different levels and inter-relations between them. On the one hand—at the level of theories concerning specific systems—we have Lakatos' example of the Newtonian model of the solar system, and quantum-mechanical models of molecules. Given an analogical model at this level, we have a recipe for applying the equations of the general theory to a particular system that *fixes* both the initial (idealised) treatments and (possibilities for) further refinements. Individual refinements at any particular stage will be executed through the application of a stock of mathematical methods (or exemplars) that are found to be associated with presentations of theories in textbooks of the mathematised sciences.[38] What governs the application of these mathematical methods will be the view of what the target system is *like*. This is encoded in the positive heuristic's analogies. If we were to isolate any particular stage in this refinement, it might not be clear from the mathematics *alone* what kind of system the equations represent, and therefore how more refined versions are to be obtained, for that is

[38] Compare the notion of exemplar used by Kuhn ([1977], p.229), and Giere ([1989], pp.68ff).

determined by the positive heuristic. Thus the positive heuristic encodes an *intended interpretation* for the refutable variant. As an example, the motion of a pendulum-type system might be treated—to a first approximation—as the undamped harmonic motion of the bob (that is, as simple harmonic motion). However, when further refinements are to be considered—such as the effects of friction—we need to know about the arm of the pendulum, which was absent from the initial equation. We can also explain distinctions between explanatory and *non*-explanatory theories on this account: a theory *explains* a phenomenon if calculations concerning that phenomenon draw on features of the theory that are interpreted via the positive heuristic and as such invoke *causal processes* that are relevant to the phenomenon. Otherwise we have only a *calculation*, not an explanation. Hence the realist intuition that intra-theoretic coherence and explanatory power are linked.

There are also analogies that operate through the general equations of the theory. I have already mentioned non-planar models of subsets of Euclid's axioms and non-Euclidean geometry. Another example—to be followed up in chapter 4—is Schrödinger's wave interpretation of quantum mechanics. How the equations are interpreted at the general level will determine relations with other general (physical or metaphysical) principles. In the case of quantum mechanics, although wave and matrix mechanics had (purportedly) been shown to be *mathematically equivalent*, Schrödinger constructed a mechanism for the interaction of matter and radiation that was distinct from that envisaged by Heisenberg *et al.*[39] Different mechanisms for radiation implied *different* relations between quantum mechanics—the kinematics of the microphysical world—and (for instance) electromagnetic theory and spectroscopy. The different mechanisms were the *direct expression* of the different interpretations of what (given the equivalence proofs) might be the 'same' general equations (see 4.4 for further details).

So we have seen how intended interpretations drive the positive heuristic—constraining theory-construction and intertheoretic relations—at the specific and the general levels. What of the interplay *between* the levels? The interpretation of the general equations will of course constrain how they are to be *applied*, that is, the specific form of equations that are written down for specific systems. But constraints run both ways: successful applications are the *cash value* of a particular interpretation of the general equations. In a sense a general interpretation lives or dies by its ability to motivate particular models: Schrödinger's wave interpretation, for instance, was discarded at least partly *because of* its inability to provide a mechanism for radiation that worked in *detailed* examples. So the intended interpretation of

[39] Schrödinger's mechanism involved resonant vibration of matter and field, and avoided the 'damned quantum jumps' of the Copenhagen camp.

the general equations may change under the influence of insights gained through particular applications of the general equations.

In terms of Lakatos' apparatus of appraisal, a specific application is *ad hoc*$_3$ if it fails to cohere with the intended interpretation of the general equations, while the general interpretation can be criticised if too many (successful) applications are *ad hoc*$_3$ with respect to it: the search would be on for a more helpful interpretation. Of course a particular refutable variant can be *ad hoc*$_3$ in another way: by failing to fit with the intended interpretation of features of a model that are peculiar to the model at hand.[40] The requirement that models not be *ad hoc*$_3$ therefore comes down to two requirements of model building: (i) that the model 'fits' the (general) intended interpretation of the equations (although *this* may change as research progresses); (ii) that the model 'fits' the (specific) view of what the target system is like (as given by the interpretive analogies) and therefore of what an *entirely accurate* model of the target system would be like (although *this* may change as research progresses). These requirements presuppose that there are such intended interpretations, and that they are not *themselves* incoherent.

Making a connected point, the early Feyerabend [1964] argued that we should expect theories and models to be amenable to realistic interpretation against background knowledge—effectively that entire theoretical schemes should be non-*ad hoc*$_3$, in Lakatos' terminology. This requirement governs both the construction of physical models and their relationships to their predecessors and background theories. Feyerabend also cites examples that illustrate the *value* of this requirement: Copernicans, faced with Aristotelian objections that objects would fly off the surface of a moving earth, might make one of two possible responses. (i) The *Bellarmine-Osiander option* is to accept that the inconsistency with observation only arises if the Copernican system is interpreted *realistically* in conjunction with (realistically-interpreted) Aristotelian physics, and retreat to an instrumentalist interpretation of the Copernican system. (ii) The *realist option* is to work towards the *overthrow* of the Aristotelian mechanics, according to which birds would be left behind by a moving earth, and search for a terrestrial physics that *explains* their failure to do so. Given the subsequent history of science, surely the realists were *vindicated.* Contrast Feyerabend's requirement with the much weaker *Duhemian* requirement of logical consistency, and with Laudan's [1984] directive to accept any empirically adequate theory ([1984], p.235) and the instrumentalist rationale constructed by Fine for the 'small handful' strategy (see 2.5). For realists, explanation and intra-theoretic coherence are duties, not supererogatory virtues that reflect our pragmatic interests. Where there are scientists

[40] This distinction between different ways of being *ad hoc*$_3$ follows immediately from the distinction between heuristics at different levels.

holding out for either (in the face of empirically-adequate theories that do neither), there are scientists acting like realists.

CONCLUSION

This chapter began, in 2.2 and 2.3, with an examination of two specific arguments for scientific realism. These were *methodological* arguments, because their premises were that the methods of science *presuppose* scientific realism, while they were intended to support realism conceived of as incorporating the semantic and epistemological components of 2.1. Reasons were found to reject those arguments *as arguments for realism*, but in 2.4 their premises—versions of *methodological* realism—were separated out and appraised in their own right. My claims concerning methodological realism were as follows. (i) Scientific realism is *independent* of methodological realism: perhaps then the unsatisfactory nature of the methodological arguments was to be expected. (ii) The non-realist who rejects a particular realist thesis (be it semantic, epistemological or aspirational) can nevertheless endorse the corresponding *methodological realism*, *if* a suitable pragmatic rationale can be found. (iii) What realists tend to think of as *methodological evidence* for realism, anti-realists who accept the endorsement in (ii), above, will see as providing an inductive-pragmatic rationale for the adoption by scientists of realism. In 2.5. and 2.6, the content of methodological realism—the specific methods that are available only to the realist—was explored. Wherever these methods are used *successfully*, methodological realism is supported *inductively*. In 2.6, it was argued that Lakatos' methodology of scientific research programmes (MSRP) *embodies* methodological realism, in that the methods endorsed by MSRP can be rationalised as the pursuit of realist aims. Thus methodological realism inherits any historiographical support enjoyed by MSRP. If this is a poor base for induction, methodological realism is supported by *historical counterfactuals*: there have been episodes in the history of science in which the application of realist methods produced theories that have been the basis of progress during long subsequent periods of research. Were realist methods not to have been adopted, there is no reason to think that these important theories would have emerged in the way they did. Specific historical counterfactuals of this kind will be the subject of the following chapters.

APPENDIX 1: MODELS IN PRACTICE AND IN LOGIC

In 2.6, the Lakatosian positive heuristic was explicated in terms of the intended interpretation (or model) of the equations that express theories, either at the particular level of the refutable variants or the general level of the hard core. The phrase 'intended interpretation', however, suggests an identification that need not have been made: the notion of model explored in 2.6 was the *practice-centred* notion from Hesse and Redhead, while 'intended interpretation' invokes the models of *metamathematics*. These are different notions. Under the practice-centred conception, a model is individuated by its (intended) representation *of a real system*, much as a model of the Eiffel Tower *is* a model of the Eiffel Tower because of the intention of its maker *to represent the Eiffel Tower*. The model of metamathematics is individuated by the axioms or equations it *satisfies*. The relations are different (representation *versus* satisfaction), as are the relata (things, processes or events *versus* axioms or equations). The logician's notion is applied to the understanding of scientific theories in the semantic and structuralist conceptions of theories. Applying this conception of theories to the description of actual scientific *practice* (rather than to the 'reconstruction' of science) has often involved identifying the models of practice with the models of logic.[41] This identification—although non-trivial—does seem plausible given (for instance) Giere's insights into the models that are found in textbooks of the mathematical sciences: the harmonic oscillator *does* satisfy well-defined sets of equations, and it is possible to see how it might also play a role in a *representative* model. But *playing a role in* a representational model and *being* one are different, as are textbooks and empirical research. Cartwright's [1983] account of the construction of models presents an argument against the identification: Cartwright urges that the models that embody our causal knowledge of concrete systems (the models that would do any *representing*) do *not* in general satisfy the equations of the general theories of which, ostensibly, they are applications. This suggests that models *of theories* do not represent real systems directly.

However, in what follows it will be assumed for the sake of an argument that the practice-centred notion *can* be explicated in terms of the logician's, so that a theorem of mathematical logic might be appended to the discussion of models in 2.6. In 2.6, physical theorising was characterised as the production of series of equations which, with increasing closeness, capture something which is not *itself* an equation, but which guides the writing down of equations. Could this process in principle come to an end with the capture of the model by the equations? According to the Löwenheim-Skolem theorem, every satisfiable first order theory (of finite *or* infinite length) that admits infinite models has models of *different* infinite cardinalities. Since models with different cardinalities are non-isomorphic, every satisfiable theory with infinite models will have *non-standard* models: those that are not isomorphic to the standard model, and yet satisfy the axioms. Call the set of models that satisfy a theory its *satisfaction set*. The isomorphicity relation is reflexive, symmetric and transitive; it can therefore be used to partition the satisfaction set into *equivalence*

[41] As do, for instance, van Fraassen [1980] and Giere [1988].

classes. The Löwenheim-Skolem theorem shows that there is more than one such non-empty equivalence class in the satisfaction set of any interesting theory.[42] This can be put in terms of *categoricity*: consistent first order theories fail to be categorical, for no matter how many their axioms, there is always something about the intended model that they will fail to 'capture'. Thinking back to the earlier process—of the interpretation successively augmenting the theory—at any *finite* stage of the development of a theory, constraints that issue from the intended interpretation are formalised (if formalisable), and pick out a new satisfaction set. The *new* satisfaction set will be a proper subset of its predecessor: the intended interpretation encodes *excess* or *extra-theoretical content* over the theory version it augments. The Löwenheim-Skolem theorem shows, however, that the process of augmentation will never end in a *categorical* theory.

Now theories are not developed in isolation, for their authors are influenced by opinions of all kinds: metaphysical, epistemological or aesthetic. Theorists also entertain other scientific theories, and will subject their models to empirical constraints. For example, among the interpretations of quantum mechanics are some that involve instantaneous action-at-a-distance, which would be unacceptable in conjunction with the special theory of relativity. Thus many possible interpretations of one's favourite theories, although *logically* permissible, could be rejected on other grounds. When all the unacceptable interpretations are discarded, there will *still* be non-isomorphic models, even if it is possible to have an *infinite* number of empirical, methodological, scientific and metaphysical constraints. This would follow from the consistency of their formalisation in the first order predicate calculus and addition to the axioms of the theory: the Löwenheim-Skolem theorem will still apply, invoking a new underdetermination.

If there is a privileged interpretation, only the interpreter could have access to it: theories are public property, while an intended interpretation could never be captured by anyone's description of it, supposing that natural languages suffer the same limitations as first order logic. What is it, then, to understand and accept a physical theory *qua* set of equations? To interpret it via a member of its satisfaction set. It follows that there is more than one non-equivalent way to understand any theory. To accept a theory as *true of* a particular system is to take it that the satisfaction set has a member that correctly portrays the system in question. Therefore two scientists may accept the *same* equations but different theories, by interpreting the equations via non-isomorphic models. If the formal theory were to be conservatively extended in order to capture some aspect of the models on which it was hitherto silent, the two theorists might propose mutually incompatible extensions. To speak of *the* intended interpretation, however, could be misleading, suggesting something static.

[42] In order to simplify what follows, let a model stand for others to which it is isomorphic.

In 2.6, it was argued that the intended interpretation at the general level could change under the influence of particular applications. This could be the consequence of an empirically-inspired amendment to a set of applied equations, such that the new equations are no longer compatible with their intended interpretation. The new equations would be *ad hoc*$_3$ if the situation were left there, but a change in the intended interpretation at the specific level—and a consequent change at the general level if necessary—would remove the problem. Thus an intended interpretation itself may evolve under empirical or theoretical selection as research continues.

APPENDIX 2: PUTNAM ON SKOLEMISATION

In Appendix 1, the Löwenheim-Skolem theorem was used to illustrate the failure of categoricity. In his [1980], Putnam extends Skolem's [1922] reading of the same theorem in a way that seems to rule out my use of the notion of an intended interpretation to explicate theoretical development, unless access to the intended interpretation is thought to involve access to Platonic forms. Putnam's question is: how is reference fixed? A theory of reference provides a relation that holds between referring terms and referents, and Putnam presents three alternatives: extreme Platonism (reference is a special kind of relation that allows the mind to 'grasp' the referent), moderate metaphysical realism (referential relations are grounded in ordinary physical facts) and his own internal realism (referential relations are relative to *theory*). His conclusion is that only *moderate* metaphysical realism is in trouble from the Löwenheim-Skolem theorem.

The argument runs as follows. Consider the totality of constraints that could be placed on theory-acceptance by all *possible* experiments and theoretical opinions. Then consider the first-order axiomatisation of an *ideal* (i.e. complete) total theory T_I that would satisfy such constraints.[43] If T_I is consistent, its axiomatisation would be subject to the Completeness and Löwenheim-Skolem theorems. There will therefore be a denumerable infinity of non-isomorphic interpretations which satisfy T_I. This is because T_I is an interesting theory, and 'no interesting theory (in the sense of first order theory) can, in and of itself, determine its own objects up to isomorphism' (Putnam, [1980], p.442). Metaphysical realists posit a mind-independent world of which theories can be true or false independently of our knowledge, their truth and falsity being determined by referential relations between terms that appear in them and entities that populate the world. This is why they divorce the

[43] If there is doubt about the possibility of such a formalisation, Putnam invokes schemata that purportedly 'first-orderize' any other logic in which T_I did turn out to be formalisable.

notions of truth and rational acceptability: a rationally acceptable theory can be false in virtue of referential relations, and the way the world is. In particular, the metaphysical realist has it that T_I could be false.

This, claims Putnam, is where the problems arise. Firstly, T_I encodes all physical facts (the realist will want to add: *to which we could possibly have epistemic access*). If T_I doesn't fix reference, nothing else that is epistemically accessible will do so. Thus we are forced either to admit that reference is fixed by *non*-physical facts (presumably ones to which we could have no ordinary *empirical* access), or to relativise referential relations to theory. But if reference is relativised to theory, metaphysical realism is in trouble, for we must then internalise *truth* also. On Putnam's view, T_I maps out—within its own models—referential relations *for itself.* According to the Löwenheim-Skolem theorem, T_I will have non-isomorphic models. If T_I comes out *true* in some model, it doesn't make sense to say that it could be false. Thus metaphysical realism is false. If the realist replies that the reference relations under which T_I is true might not be the *intended* ones, since the model in question is not the intended one (so that T_I might be false), Putnam will ask what *else* fixes referential relations: it would have to be some special class of facts to which the language user must have access independently of the usual methods embodied in the epistemic constraints that helped to select T_I.

As we saw, Putnam's argument against metaphysical realism is blocked if 'object-grasping' powers are attributed to minds—allowing theories to be understood by the direct apprehension of their intended interpretations—to fix reference in addition to operational and theoretical constraints. All that the Löwenheim-Skolem result would then show was that Platonic heaven was not amenable to precise formal capture. For naturalistic philosophers, however, 'the postulation of unexplained mental faculties' is 'unhelpful epistemology and almost certainly bad science' (Putnam, [1980], p.433). Overall, the 'Skolemization of absolutely everything' (p.434)—including an ideal, complete science—forces the theory of meaning onto a fork: accept Platonism or drop metaphysical realism. Moderate metaphysical realism is untenable. If Platonism is unsavoury, the utopian idea that science investigates entities that populate a reality independent of our conceptual system should be dropped. Instead truth itself must be relativised to T_I, and truth equated with rational acceptability. It is not conceivable that T_I could be false, since it satisfies (by hypothesis) all operational *and semantic* constraints.

Some objections can be raised to Putnam's argument.[44] Firstly, it fails to address the externalism implicit in his own causal theory of reference. It doesn't matter how *users* of

[44] Lepore and Loewer [1987] provide a survey of objections.

terms understand referential relations if these are causal relations that hold independently of T_I. Secondly, Putnam seems to beg the question against (moderate) metaphysical realism in his claim that there could be no physical facts that could render T_I false. In any case, one line of thought suggests that the account of model building given in Appendix 1 is untouched by Putnam's argument, indicating that the methodological realism embedded in that account is compatible with Putnam's internal realism. Opinions—whether general and central or particular and peripheral—impinge on theory construction in ones and twos. In practice, a scientist's views will never be exhausted in the way that Putnam's easy quantification over 'empirical and theoretical constraints' indicates. We need not attribute non-natural Platonic 'object-grasping' powers to the minds of the authors of theories in order for intended interpretations to play a role in theory construction, for *dispositions to augment the equations* (i.e. enrich the formal theory) would suffice. Putnam's argument *might* indicate (if one accepted its conclusion) that there was no fact of the matter which of the models of the final (completed) theory that would end all theory-development represents 'the world' in that theory. There might even be no fact of the matter *which* model we are *presently* using to develop our equations. This still presents no problem: suppose there are different non-isomorphic models that satisfy our equations. Either we have opinions about which is the model we are presently using, or we do not. If we *do*, we can use them to select one or another, building this selection into our equations. If we do not, the underdetermination cannot affect our practice.

However, there is a possible reply to Putnam's sceptical use of the Löwenheim-Skolem theorem that is analogous to a Lakatosian response to *empirical* underdetermination. In the case of the empirical equivalence of incompatible theories, MSRP looks to the heuristic power of the research programmes in which the two theories are embedded. This is reflected in the retrospective judgements as to whether the programmes' refutable variants are *ad hoc* or *well motivated.* Instrumentalists sometimes argue that the existence of empirically equivalent theories shows that we can have no (*evidential*, rather than *pragmatic*) reason to prefer one over another, but the Lakatosian can reply that one of the theories will be *embedded in our scientific history*: the one that lead us to *discover* some of the empirical regularities that constitute its supporting evidence. We can prefer it because *it*—rather than any of the hordes with which it is empirically equivalent—directed us to new knowledge. This preference can be read as *either* a pragmatic choice—we prefer *this* theory because it has a proven track-record in leading us to new discoveries—or as providing further evidence that the theory is *true*. The instrumentalist might argue that any of the empirical equivalents would have had a similarly illustrious role in scientific progress, for they made the same predictions. However, this isn't a fair comparison. When we started out on the research during which all the valuable discoveries were made, the well-developed theory we have now—which is empirically adequate—was not available.

There would only have been a sketch of an empirically *in*adequate theory, plus some ideas on how to improve it. Applying these ideas to the sketch (that is, following the programme of research) led us systematically to new discoveries. The instrumentalist must therefore argue that the empirically equivalent theories cited in the underdetermination argument would have occurred within *equally fruitful* research programmes, if they are to be real alternatives.

Given underdetermination at a higher (semantic) level, one can similarly invoke previous extensions of the equations that embody our theories. Given that the Löwenheim-Skolem theorem applies, there will of course be an infinite array of unintended models which satisfy the equations we currently accept, analogously to the empirical underdetermination case, so theoretical underdetermination will *always*, *in principle* be with us. That doesn't mean, of course, that we have to throw up our hands in horror, and forsake all reference and all ontological commitment. 'Snapshot' views of 'total constraints'—like Putnam's—should be ruled out altogether. Other interpretations can be ignored because they did not, as a matter of historical fact, give rise to the formal theory as it presently stands. This kind of historical specificity perhaps explains why no-one *really* adopts denumerable models as their interpretation of axioms that were 'really' designed for the real numbers, even though the Löwenheim-Skolem theorem shows this to be logically permissible. The real numbers are whatever was associated with the historical process that was the investigation of the mathematics of real numbers.

3

THE BOHR ATOM, REALISM AND INCONSISTENCY

The quantum theory gives me a feeling very much like yours. One really ought to be ashamed of its success, because it has been obtained with the Jesuit maxim "Let not thy left hand know what thy right hand doeth". (Einstein: Letter to Born)

INTRODUCTION

In the hope that history of science may enlighten their discourse, philosophers of science have often chosen Bohr's 1913 model of the atom to illustrate their diverse methodological lessons. A common theme in the growth of this folklore—the roots of which can be traced back to the methodological judgements of Bohr's contemporaries—has been that the theory suffered from some sort of inconsistency. Conclusions of two opposing kinds were drawn: firstly that the inconsistency *explained* the adverse methodological judgements passed down on the theory by contemporaries; secondly, that a theory could contribute to scientific progress *despite* being logically inconsistent. Taking the latter view, Lakatos argued that because the theory bore so much heuristic fruit, a 'logician's proof' of inconsistency could not provide a knock-down reason for the rejection of a research programme, although scientific honesty would demand its public recording. Feyerabend agreed that logical dogmas may hinder growth in knowledge:

> Scientists proposing theories with logical faults and obtaining interesting results with their help (for example: ... the predictions of the older quantum theory and of early forms of the quantum theory of radiation—and so on) evidently proceed according to different rules. (Feyerabend, [1988], p.15).

Lakatos' methodology was subsequently declared to be empty, because it placed no significant constraints on scientifically honest behaviour: only the ornament of rationalist rhetoric distinguished it from his own epistemological anarchism. Philosophers have not been alone in spotting an inconsistency: in a standard historical text, Jammer has claimed that the quantum theory of poly-electronic systems that grew out of Bohr's atomic model 'lacked two essential characteristics of a full-fledged scientific theory, conceptual autonomy and logical consistency' (Jammer [1966], p.196).

This scientific episode has been the subject of an extensive historical literature, and I introduce no new historical facts. Rather, the historical content of the chapter is heavily indebted to two key secondary texts: Rosenfeld [1963] and Heilbron and Kuhn [1969], as well, of course to Bohr's papers. What this chapter *does* hope to offer is a new interpretation of Bohr's progress, and therefore fresh narrative. In 3.1 to 3.4, I will cover the background to Bohr's construction of the most celebrated part of his theory: the model of radiative emission in one-electron atoms. This is central to the appraisal because it was the theory's success in this phenomenal region alone that ensured its fame. In 3.5, the accusation of logical inconsistency will be assessed: I will argue that there is no historiographical motivation for Lakatos' indulgent attitude to contradictions. Since the *logical* arguments against inconsistency are as strong as ever (even if 'childish', according to Feyerabend), there seems no reason to free the growth of knowledge from the 'tyranny of logic'. Instead, an inconsistency of *interpretation* will be identified as the perceived problem. *Ad-hoc*ness will occupy 3.6, where a careful reading of Bohr's progress will reveal his theory to fall under one of Lakatos' subdivisions of the notion of *ad-hoc*ness, which can, however, be thought of as a (weaker) form of inconsistency after all. Although many sections of the chapter *are* historical in character, they are present as substrate for a traditional philosophical project: hunt the inconsistency.

3.1. BACKGROUND

Extensive empirical information on the emission and absorption spectra of the elements was gathered during the nineteenth century, following the discovery of dark lines in the solar spectrum in 1802. By 1860 it was recognised that each element emits a characteristic discrete spectrum. Attempts were made to discern some order among the phenomenal chaos, notably in 1884, with Balmer's representation of the wavelengths λ of the first four lines in the visible spectrum of hydrogen as the difference of two integer terms:

$$\lambda = k\left[\frac{n^2}{n^2 - 4}\right] \qquad (k \text{ is a constant}, n = 3,4,5,6)$$

This formula can be expressed in a more familiar form in terms of radiation frequencies ν ($\nu = c/\lambda$, c is the speed of light *in vacuo* and R_H is the Rydberg constant for hydrogen) as:

$$\nu = R_H\left[\frac{1}{4} - \frac{1}{n^2}\right]$$

Runge (in 1888) and later Rydberg (in 1890) produced more general formulae designed to accommodate the more complicated spectra of the alkali metals, among other elements:

Balmer's formula was a special case of these. In fact Rydberg claimed to have been in possession of Balmer's formula prior to its publication. In later years, Rydberg (1900) speculated that every spectral line of every element could be captured as the difference between two integer 'spectral' terms, as did Ritz (1908: the 'combination principle').

At the beginning of this century, there were many competing attempts to elucidate atomic structure following the negatively charged electron's discovery in 1897. According to classical electromagnetic theory, the emission of radiation originated in the acceleration of charged particles, and so it was natural to expect any viable atomic model to account for spectral phenomena. However, the complexity and variety of spectral equations and the rough qualitative status of most atomic models inspired a certain pessimism about the prospects of such explanation. Classically, it was thought that line spectra arose from the natural periodic oscillations or vibrations of electrons in atoms occupying stable states: it was a consequence of electromagnetic theory that oscillating charged particles would emit radiation at an optical frequency equal to the mechanical frequency. In 1907, Conway contended that single atoms could produce only a single spectral line at one time requiring a large number of atoms in a range of states to produce the whole spectrum. Also, Conway argued that spectral lines arose from atoms remaining in excited states long enough for a single electron to be stimulated to produce the requisite wave trains. Conway's assertion was supported by Bevan's (1910) work with the anomalous dispersion of certain red lines in the potassium spectrum by potassium vapour, the explanation of which would require the presence of far too many electrons per atom if each were to produce all the lines in the relevant spectrum at the same time.

A connection had also been made between the discrete frequencies of spectral lines and the old quantum theory's ascription of integral multiples of energy packets to oscillators. At the first Solvay conference Lorentz conjectured a relationship between atomic structure and the quantum of action *h*, amid inconclusive debate about what form this relationship might take. In 1910, Haas attempted to introduce quantisation into a theory based on Thomson's model of the atom, in which positive and negative charge were distributed throughout the atom. Haas showed the Rydberg constant to be expressible in terms of the constants *e*, *m* and *h*, obtaining a value $R = 16\pi^2e^4m/h^3$ which differed by a simple numerical factor from the empirical one. However, Whittaker [1953] points out that since e^4m/h^3 is the only product of powers of these quantities with the correct dimensions, the result was perhaps merely fortuitous. Nicholson published a pair of papers in 1911 in which he outlined a rather arcane theory invoking a set of otherwise elusive new elements, although it appeared to meet with some rather impressive quantitative empirical success. Notable in the theory was the quantisation of electron ring angular momenta in units of $h/2\pi$ and its basis in Rutherford's proposed atomic structure of the same year. Rutherford's atom consisted of negatively-charged electrons orbiting a small and dense positively-charged nucleus, in which most of the atomic mass was concentrated. However, although influenced by Conway in his notion of atoms existing in distinct states corresponding to Planck's oscillatory energy levels, Nicholson failed to employ the idea of *single* spectral lines arising from *single* electrons, instead studying vibrations of *sets* of electrons grouped in rings. Nicholson was not the only researcher to quantise angular momentum prior to 1913: Bjerrum explained the absorption spectra of gaseous hydrogen bromide and chloride by introducing that assumption into his 1912 theory based on a conjunction of vibrational and rotational molecular motions.

3.2. BOHR'S THINKING 1911-1913

The development of Bohr's thinking in the period up to the publication of his famous trilogy has been well documented elsewhere, notably in Heilbron and Kuhn [1969] and Rosenfeld [1963], the latter benefiting from the author's long friendship with Bohr. In what follows, I will summarise the developmental stages of Bohr's atomic models in order that the inspiration of his novel assumptions is clear.

According to Heilbron and Kuhn, it was clear from Bohr's (1911) doctoral thesis—on the electron theory of metals—that he was aware at the time that a coherent theory of atomic structure would require a radical departure from classical mechanics. He had also identified Planck's quantum theory as the basis for this departure and was ready to require of the new theory some limiting relation of correspondence with classical description. However, Bohr was not particularly interested in atomic models at this stage, preoccupied as he was with subjects closer to his thesis area. After acquiring his doctorate, Bohr travelled to Cambridge to work with Thomson in September 1911. Following a period there during which he failed to achieve the publication of his doctoral thesis in English, Bohr moved to Manchester in March 1912. Working at first on some laboratory exercises regarding the absorption of α- and β-rays, Bohr began some experiments involving radium that had been suggested by Rutherford. The long-term value of this work, claim Heilbron and Kuhn, was social rather than scientific: it threw Bohr into regular contact with other researchers, notably Hevesy and Darwin. Bohr's mental efforts, however, still centred on the electron theory of metals.

It was on 12 June 1912 that the first surviving record of an interest in atomic models on Bohr's part was written: a letter to his brother Harald. On reading a paper of Darwin's on the content of Rutherford's model with respect to the energy lost by α-particles passing through thin metal sheets, Bohr was critical of some of the simplifying assumptions that had been made. From then on, work on Rutherford-type atomic models successively replaced his earlier preoccupations. Just one week later (19 June), Bohr again wrote to his brother, giving an indication that he had achieved some useful results in his atom-model work: 'It could be that I've perhaps found a little bit about the structure of atoms' (Bohr, quoted in Heilbron and Kuhn, pp.238-9). This 'little piece of reality' was a product of his criticisms of Darwin's paper: Bohr's use of the phrase indicated his new-found commitment to the Rutherford model.

So what initially fired Bohr's interest in the Rutherford atom? Rosenfeld attributes the interest to Bohr's 'dialectical turn of mind' ([1963], p.xv), since 'the stability of the Rutherford atom was beyond the scope of classical theories', but Heilbron and Kuhn point out that *radiative* instability would have afflicted *any* classical model postulating electrons in motion. Since *mechanical* stability is a problem for the Rutherford atom only, Heilbron and Kuhn attribute to it the relevant heuristic role in Bohr's reasoning ([1969], p.241, fn.81). We shall see that both the radiative *and* the mechanical stability questions turned out to be heuristically important.

In June and July 1912, Bohr hand-wrote a memorandum to Rutherford in which many of the qualitative aspects of the published model are anticipated (for a partial reprint see Rosenfeld [1963], pp.xxi-xxviii). Present in this document are prototype versions of arguments presented in parts II and III of the trilogy. On the first page of the memorandum, Bohr compares the mechanical stability of the Rutherford and Thomson atoms. The solution to the Rutherford atom's mechanical stability problem, notes Bohr, suggests an explanation of chemical periodicity, but in a footnote he claims:

> The difference in this respect between the atom-model considered, and J.J. Thomson's atom-model is very striking, and seems to make it impossible, to give a satisfactory explanation of the periodic law from the last mentioned atom-model. (Bohr [1912], p.A1)

The heuristic importance of the Rutherford atom's mechanical stability problem is then clear: its solution suggests a promising explanation of chemical periodicity. On the second page of the memorandum, another explanandum is raised. Considering the Rutherford atom, Bohr notes

> that a ring, if only the strength of the central charge and the number of electrons in the ring are given, can rotate with an infinitely great number of different times of rotation, according to the assumed radius of the ring. (Bohr [1912], p.A2)

In other words, on the basis of the Rutherford atom *classically* treated there is no hope that characteristic orbital quantities can be calculated, so that some constraint on their allowed values is necessary. The required constraint, Bohr promises, will take the form of the "hypothesis" that there be a definite ratio between the kinetic energy and frequency of rotation for any stable ring. Furthermore there will be no 'attempt of a mechanical foundation (as it seems hopeless)' (Bohr [1912], p.A2). The constant of this proportionality, k, had not yet been fixed; to yield the final, published form of the theory, Bohr would need to put $k = h/2$. Rosenfeld ([1963], p.xxxi) calculates that Bohr was working with an implicit value of $k \approx 0.6h$, and Heilbron and Kuhn ([1969], p.262) note that an empirical source cited elsewhere by Bohr—the resonance frequency of helium atoms—could have yielded $k \approx 0.4h$. The passage quoted indicates the heuristic importance of the *radiative* stability problem: *its* solution (which also, of course, solves the problem of the continuity of ring radii) is the inspiration for the introduction of the quantum condition, from which the heuristic power of Bohr's model flows. Bohr then sets out an explanatory agenda, the main items being the periodicity of atomic volume, the relative stability and the heats of combination exhibited by the 'single compounds' (i.e. homonuclear diatoms) H_2, $[He_2]$[1] and O_2. Much of the rest of the memorandum is taken up with some promising calculations connected with these topics. Entirely absent, however, is any mention of emission spectra, which figure so large in the first section of the published trilogy. Up to this point, Bohr had been pessimistic about the prospect of constructing a theory of emission spectra on the basis of atomic structure.

So if Bohr was working with a particular atomic model, why was he now committed to it? There were *three* positive arguments for the Rutherford structure (and against the Thomson one) alluded to by Bohr: a natural explanation of the large-angle scattering of α-particles (Rutherford's work), heuristic promise with respect to periodicity (mentioned above) and isotopy. The Rutherford atom, in which the notions of atomic number (equal to the number of electrons, determining chemical behaviour) and atomic weight were naturally separate, would be amenable to an account of chemically-identical atoms with different atomic weights. According to Thomson's model, however, the number of electrons determined both an atom's weight *and* its chemical properties, so no such account would be available.

After writing the memorandum, Bohr returned to Copenhagen, got married and taught a course in thermodynamics. Towards the end of 1912, or in early 1913, Bohr read the papers of Nicholson mentioned in the last section: the spectral aspect of his atomic model was born. At first dismissive of Nicholson's theory (as he had been on earlier contact with Nicholson), Bohr was later to be impressed by its strikingly accurate predictions of spectral lines. Initially, Bohr thought this model to be incompatible with his own, but later, in a letter to Rutherford of 31 January 1913, he argued that the two *were* consistent, but described different states of the same system:

[1] The brackets indicate the non-existence of this species, qualitatively explained by Bohr.

> ... the state of the systems considered in my calculations is to be identified with that of the atoms in their permanent (natural) state.
>
> ... the states of the systems considered by Nicholson are, [on the] contrary, of a less stable character; they are states passed during the formation of the atoms, *and are the states in which the energy corresponding to the lines of the spectrum characteristic for the element in question is radiated out.* From this point of view systems of a state as that considered by Nicholson are only present in sensible amount in places in which atoms are constantly broken up and formed again; i.e. in places such as excited vacuum tubes or stellar nebulæ. (See Rosenfeld [1963], pp.xxxvi-vii, my italics)

This letter illustrates Bohr's original mechanism—inspired by Nicholson's model—for the production of spectral line emissions: they arise from the combination of positive ions and electrons to reform neutral atoms. As the electron 'falls' from infinity to a permanent stable orbit, it passes through many (Nicholson-type) excited states, corresponding to different modes of vibration of a Planck oscillator, it either causes electrons in the higher energy orbits to release discrete quanta of radiation at their resonance frequencies, or itself resonates at these frequencies. This mechanism, likened to 'a finger drawn across the strings of a harp' in Heilbron and Kuhn's colourful phrase ([1969], p.263), is still essentially classical in that radiation of a given frequency must be emitted by a mechanical oscillator of the *same* frequency. While not uncontested, the connection between radiative emission and reformation of atoms from ions was widely accepted. Although it would preclude the derivation of the Balmer formula which made his name, Bohr still entertains this mechanism in Part I of the trilogy. Also of interest in the letter is Bohr still ruling out a theory of atomic spectra:

> I must however remark that the considerations sketched here play no essential part of the investigations in my paper. I do not at all deal with the question of calculation of the frequencies corresponding to the lines in the visible spectrum. (see Rosenfeld [1963], p.xxxvii)

In the next five weeks, Bohr acquainted himself with the spectral equations described above (3.1). In particular, his reaction to Balmer's formula is recorded in his oft-quoted recollection 'As soon as I saw Balmer's formula, the whole thing was immediately clear to me' (Rosenfeld [1963], p.xxxix). Bohr now recognised that radiative emission must arise from transitions *between* stationary states, which can themselves be identified with the spectral terms appearing in the Balmer formula. The *correct* constant of proportionality ($k = h/2$) between kinetic energy and frequency of radiation was also now in place. These points were *not settled* until Bohr had seen the Balmer formula, which must therefore have played a more significant role in Bohr's working than that attributed to it by Lakatos (see for instance his [1970], p.147), who was more interested in rubbishing the Baconian 'inductive ascent' account of Bohr's reasoning than in strict historical accuracy. In early

March, Bohr sent Rutherford a draft of Part I of the trilogy which differed only in minor respects from the published version. The speed with which the paper was completed is reflected in a curious—but revealing—ambiguity in the paper, as we shall see in the next section.

3.3. The Hydrogen Atom

The first section of Bohr's [1913] attacks the stability problem associated with Rutherford's atomic structure. Bohr considers a completely classical treatment and deduces the inevitable radiative collapse of such a system. In Bohr's version, electrons were assumed to move in elliptical orbits around positively-charged nuclei. A one-electron system was considered in the first instance, and the treatment made a number of simplifying assumptions: firstly, that the nucleus is stationary (nuclear mass being large compared to electronic mass, m); secondly, that the electron velocity v is small compared to the speed of light; and thirdly that electrons describe circular orbits (this makes no difference for one-electron systems). If no radiation was emitted, the electron would describe a stable orbit (radius r) and there would be no net forces on the electron, allowing the equation of inward (electrostatic) and outward (centrifugal) forces, the former described by Coulomb's Law. So we could write, for circular orbits:

(1) $$\frac{mv^2}{r} = \frac{Ze}{r^2}$$ (where Z = nuclear charge)

but Total Energy E_{tot} = Kinetic Energy + Potential Energy

i.e. $$E_{tot} = \frac{mv^2}{2} + \int_{\infty}^{r} \frac{Ze}{r^2}\, dr$$

Integrating and substituting from (1):

(2) $$E_{tot} = -\frac{Ze}{2r} = -\frac{mv^2}{2} = -W,$$

where W is the energy that needs to be added to the system to remove the electron to an infinite distance from the nucleus. Rearrangement of (2) gives:

(3) $$r = \frac{Ze}{2W} \quad \text{and} \quad v = \left[\frac{2W}{m}\right]^{1/2}.$$

Now if the frequency of orbit is ω, then for circular orbits:

$$\omega = \frac{v}{2\pi r}\,, \tag{4}$$

and substitution of (3) into (4) gives:

$$\omega = \frac{2^{1/2}W^{3/2}}{\pi m^{1/2}Ze}\,. \tag{5}$$

So far so classical, but if we follow classical electromagnetic theory, the continually-accelerating electron will radiate energy, so E_{tot} will decrease continuously and W will increase, as will orbital frequency ω. From (2) we can see that r will gradually decrease and the electron will spiral into the nucleus: this is the famous radiative collapse. Furthermore, the above equations hold out no hope for the calculation of characteristic atomic values for the dynamical variables. So if atoms are stable and exhibit constant characteristic dimensions far greater than those of the nucleus (as we have seen, Rutherford provided the evidence for this), the selection of a *denumerable* set of allowed orbits would be required of a viable atomic model. Also, an entirely classical treatment would predict a continuous emission spectrum, since the frequency of emitted radiation depends on the continuously varying frequency of orbit. Over the large number of atoms in a sample of the gas, electron-nucleus distances (and therefore frequencies of orbit) would be distributed over all values. The discreteness of spectral lines requires again that these quantities assume characteristic discrete values for the system.

So far, the structure of Bohr's argument has been clear: (1) Full electromagnetic theory predicts the radiative collapse of the atom. (2) Therefore consider a model in which only Coulomb's law is applied, rather than full electromagnetic theory, according to which Coulomb's law holds only approximately. (3) In such a model, if the frequency of emitted radiation depends on frequency of electronic orbit (a hitherto mathematically continuous quantity), a continuous spectrum will be predicted. Also, characteristic atomic dimensions will be difficult to explain. (4) Therefore, a characteristic discreteness should be introduced into the mechanical model.

In Part I of the trilogy, Bohr presents two different *and incompatible* ways to introduce the required discreteness. In the first, characteristic energy values are specified as the answer to the stability problem. He does this by considering the process of an electron moving from infinite distance (for which, trivially, $W = \omega = 0$) to a stable orbit of radius r, for which equations (1) to (5) hold. According to Planck's theory, an oscillator of frequency ω will radiate an amount of energy $nh\omega$ (where n is integral) in a 'distinctly separated emission'. The energy lost in this process is assumed to be emitted as radiation *of homogeneous frequency* v. He secondly assumes that this frequency is equal to half the frequency of the final, stable orbit, so $v = \omega/2$. Bohr then argues 'If we assume that the radiation emitted is homogeneous, the second assumption concerning the frequency of the radiation suggests itself, since the frequency of revolution of the electron at the beginning of the emission is 0.' (Bohr, [1913], p.5) I can only take this to mean that Bohr saw the emitted radiation frequency an 'average' of the initial and final orbital frequencies: it could be a natural consequence of the rather obscure harp analogy. Now Planck's formula is:

$$W = nh\nu , \qquad \text{(where } n \text{ is integral)}$$

so we can put:

$$(6) \qquad W = \frac{nh\omega}{2} .$$

Bohr's use of Planck's formula as an analogical precedent, and his argument for introducing (6), suggest that (6) is to be interpreted (for the purposes of *this* derivation) in the same way that Planck interpreted *his* formula. It therefore represents a process in which an oscillator emits n quanta of energy $h\omega/2$, the total emission having energy W. Rearranging (6) for ω, equating the result with (5) and solving for W gives:

$$(7) \qquad W = \frac{2\pi^2 mZ^2 e^2}{n^2h^2}$$

Different integer values for n yield a hierarchy of energy states with corresponding values for ω and r. Furthermore, we can now calculate the atom's characteristic dimensions. For instance, by equations (3) and (7) we can write:

$$r = \frac{n^2h^2}{4\pi^2 mZe} .$$

Putting $n = 1$, Bohr calculates the radius of the atom (among other characteristic constants) in the lowest energy state, a value which was 'of the same order of magnitude' ([1913], p.5) as that obtained by other means. He has therefore solved the problems associated with the atom's stability and characteristic dimensions.

Before Bohr can go on to derive the Balmer formula, however, he must amend his mechanism of radiation and therefore his interpretation of (6), so that it represents the (more familiar) emission of one quantum of frequency $nh\omega/2$. In so doing the connection between optical and mechanical frequencies is broken at last. This he does in a passage in which he criticises Nicholson's theory and, implicitly, his own earlier assumptions. He notes that Nicholson's model 'does not seem to be able to account for the well-known laws of Balmer and Rydberg' ([1913], p.7). A more serious objection to Nicholson's atom is that:

> systems like those considered, in which the frequency is a function of the energy, cannot emit a finite amount of a homogeneous radiation; for, as soon as the emission of radiation is started, the energy and also the frequency of the system are altered. (Bohr, [1913], p.7)

If this is the case, the system could not emit *n* quanta of the same frequency, because as soon as *one* was emitted, the mechanical frequency of the system would change and the *next* quantum would have to have a different frequency. The answer is to *drop* the connection between mechanical and radiative frequencies *and* re-interpret equation (6), which until now has meant that 'the different stationary states correspond to the emission of a different number of Planck's energy-quanta' (Bohr, [1913], p.8). Bohr then promises that (7) can be arrived at using 'special assumptions' of 'somewhat different form', and that this second derivation will be given later (section 3 of his paper). Bohr's presentation of all this illustrates his awareness that the radiation mechanism behind his first derivation of the energy levels was not compatible with the account of the Balmer lines he wanted to give.

A picture emerges of Bohr's atom being a quasi-classical system which for some reason has a restricted set of possible stationary states: instead of a continuum of them, there would only be countably many. To these stationary states would correspond a set of stable orbits in which no radiation would be emitted, contrary to electromagnetic theory. So when and how does radiative emission happen? Before answering this question, Bohr expresses a version of the agnosticism of the later principle of complementarity, arguing:

> (1) That the dynamical equilibrium of the systems in the stationary states can be discussed by help of the ordinary mechanics, while the passing of the systems between different stationary states cannot be treated on that basis.
> (2) That the latter process is followed by the emission of a *homogeneous* radiation, for which the relation between the frequency and the amount of energy emitted is the one given by Planck's theory. (Bohr [1913], p.7)

These considerations allow derivation of the Balmer formula. We ignore the question of the *detailed* mechanics of the emission process and consider only the energy change, ΔW, which accompanies an atom's transition between states. For a transition between states with $n = n_1$ and $n = n_2$, we can write:

$$\Delta W = W_{n2} - W_{n1}$$

$$\Delta W = \frac{2\pi^2 mZ^2e^2}{h^2}\left[\frac{1}{n_2^2} - \frac{1}{n_1^2}\right] .$$

Dividing through by Planck's constant to give the frequency of emitted radiation yields:

(8) $$\nu = \frac{2\pi^2 mZ^2e^2}{h^3}\left[\frac{1}{n_2^2} - \frac{1}{n_1^2}\right],$$

which can be compared with the formula for the Balmer series:

$$\nu = R_H\left[\frac{1}{4} - \frac{1}{n^2}\right]$$

(where R_H is the Rydberg constant for hydrogen), suggesting that, since $Z = -e$ for hydrogen:

(9) $$R_H = \frac{2\pi^2 mZ^2e^2}{h^3} = \frac{2\pi^2 me^4}{h^3}$$

The second derivation of the energy levels is supplied as promised in section 3 of the paper, where Bohr presents what he takes to be a more rigorous discussion of the assumptions required to derive the above equations. First, he explicitly rejects the previous interpretation of (6), and therefore the analogy with Planck's theory, on the grounds stated above. He therefore has the problem of fixing the ratio of the energy of the system to the electron's frequency of revolution by some other means. The argument starts with the requirement of discrete energy states of the atom (see the argument from stability and constant atomic dimensions, above). Bohr now wants a formula expressing the discreteness *generally*, rather than in any particular form, because its exact form will be filled in later. To this end, instead of (6), he writes:

(10) $$W = f(n)h\omega$$

By substitution into (5) as before, he gets:

(11) $$W_n = \frac{\pi^2 mZ^2e^2}{2h^2f^2(n)} \quad \text{and} \quad \omega_n = \frac{\pi^2 mZ^2e^2}{2h^3f^3(n)}$$

If we again consider the change in the system's energy due to a transition between two states and divide the resultant equation through by Planck's constant to give the frequency of the radiation emitted in this process, we have:

(12) $$\nu = \frac{\pi^2 mZ^2e^2}{2h^3}\left[\frac{1}{f^2(n_2)} - \frac{1}{f^2(n_1)}\right].$$

The energy states can then be identified with the integer terms in the Balmer formula, which allows the derivation of the*form* of function $f(n)$: by inspection it must be $f(n) = cn$. Constant c can be calculated using an argument which applies an early but recognisable form of the famous correspondence principle. Consider the system passing between two neighbouring stationary states, $n = N$ and $n = N - 1$. Calculated from (11), the frequency of emitted radiation will be:

$$(13) \qquad \nu = \frac{\pi^2 m Z^2 e^2}{2c^2 h^3} \cdot \frac{2N - 1}{N^2(N - 1)^2},$$

whereas the mechanical frequencies for the two states are:

$$(14) \qquad \omega_N = \frac{\pi^2 m Z^2 e^2}{2c^2 h^3 N^3} \quad \text{and} \quad \omega_{N-1} = \frac{\pi^2 m Z^2 e^2}{2c^2 h^3 (N - 1)^3}$$

Bohr then argues:

> If N is great the ratio between the frequency before and after the emission will be very near equal to 1; and according to the ordinary electrodynamics we should therefore expect that the ratio between the frequency of radiation and the frequency of revolution is very nearly equal to 1. This condition will only be satisfied if $c = {}^1/2$. (Bohr [1913], p.13)

Substituting $f(n) = n/2$ into (10), it is easy to see that Bohr has re-derived equation (6), by (i) working backwards from the Balmer formula to yield the correct *form* of quantisation, (ii) invoking the correspondence principle in order to fix the constant. Note also that in order to reach (12), Bohr has already re-interpreted the emission process so that (6) now represents the emission of *one* Planck energy quantum of frequency $nh\omega/2$; in subsequent papers the quantum condition will always be interpreted similarly. This, together with the use of the correspondence principle, illustrates that this second derivation is in much closer conformity to the later development of the old quantum theory.

Now Bohr's quantum condition, equation (6), is famously equivalent to the quantisation of angular momentum (for circular orbits) conjectured by Nicholson and Bjerrum. Bohr, however, doesn't make this observation until later in the paper ([1913], p.15), where he presents it as a 'simple interpretation' of his quantum condition. It is easy to see why Bohr would want a simple *mechanical* expression of his quantum condition, rather than a condition restricting radiative emission (although he rules out a "mechanical *foundation*"). Firstly, only a *mechanical* quantum condition will answer the stability and atomic size problems. Secondly, if the quantum condition could be simply expressed mechanically (although its exact form was derived with the help of the Balmer formula), the theory might appear to be a coherent atomic model from which an adequate spectral theory is *derivable*. Otherwise, his explanation of line spectrum formulae would look irredeemably *ad hoc* (or at least semi-empirical), because he would be using an *ad hoc* restriction on radiative emission from an otherwise classical atom to *explain* radiative emission. So how does Bohr attempt to derive this mechanical quantum condition? He writes, for *circular* orbits:

$$(15) \qquad \pi M = \frac{T}{\omega},$$

where T is kinetic energy, M angular momentum and ω as before. Equation (2) gives:

$$T = W,$$

by equation (6) it is easily seen that:

$$(16) \qquad M = \frac{nh}{2\pi}$$

Thus Bohr's derivation of the quantisation of angular momentum from equation (6) relies on the previous quantisation of another constant of circular motion: kinetic energy. Only *average* kinetic energy is a constant of elliptical orbits, however, and Bohr thought electronic orbits in general to be elliptical. Since quantising the *average value of a variable quantity* makes much less sense than quantising a constant *proper*, it seems clear why, in later expositions of his theory, Bohr continued to express the quantisation via equation (6), even though he might have preferred (for the above reasons) to quantise a mechanical quantity. This approach, Heilbron and Kuhn point out, lasts until Sommerfeld's formulation, in 1915, of phase-integral quantum conditions applicable also to elliptical orbits. For these reasons, Heilbron and Kuhn label the importance of Bohr's quantisation of angular momentum in the derivation of the Balmer formula as a 'myth' (Heilbron and Kuhn, [1969], p.280). Whether or not Heilbron and Kuhn are right about this, Bohr's *aspiration* to quantise angular momentum is surely significant. I will return to this later.

In later accounts of his theory, Bohr entirely drops the close analogy with Planck's radiation process and constructs the system more in line with the second derivation. Significantly, it will be noticed that in the first derivation, Bohr gives no compelling rationale for introducing the particular form of equation (6) that he did, apart from the rather lame aside that it was 'suggested' by the homogeneity of emitted radiation and the analogy with Planck's theory. Indeed, Rosenfeld's account of Bohr's working seems to indicate the first derivation to be something of a *post-hoc* rationalisation on Bohr's part. Heilbron and Kuhn similarly downgrade the first derivation. However, a brief summary of Bohr's progress towards the published version of his atomic model will invite a different conclusion.

As we saw in the last section, Bohr had good reasons to begin work on the Rutherford atom. Once he had done so, the arguments from the stability and characteristic constant dimensions of the atom suggested the introduction of discrete stationary states, but their exact form was unclear. On reading Nicholson's papers, Bohr was led to the correct quantisation via an argument akin to the first derivation of the energy levels, therefore invoking the earlier radiation process based on the reformation of atoms from ions. At this stage, Bohr did not envisage a theoretical account of spectral lines. Sight of the Balmer formula would immediately have suggested the identification of his energy levels with the famous integer terms. *Explanation* of the Balmer formula, however, required the revision of this process of radiative emission and the final cutting of the (classical) link between optical and mechanical frequencies. The energy levels would then have to be obtained on the basis of the correct radiation theory via the second derivation, the (now *derived*) quantum condition being re-interpreted in the process. The price of this theoretical correctness was that the Balmer formula became a vital stage in the derivation, and his reasoning had a consequent *ad hoc* air. On the plus side, Bohr was forced to drop the heuristic analogy with Planck's interpretation, instead using the much more powerful correspondence principle to take him from the classical equations to their quantum successors. The first derivation was no '*post hoc* rationalisation', but a genuine stage in Bohr's reasoning, albeit one which was incorrect from the point of view of the later theory.

3.4. SUCCESS

Bohr's derived optical frequency equation predicted a series of spectral series, corresponding to integer values $n_2 = 1, 2, 3,$ *etc.* and $n_1 = (n_2 + 1), (n_2 + 2),$ *etc.* Calculations in which empirical values were put in for constants appearing in his equations would have yielded two previously observed spectral series: the Balmer series in the visible region (with $n_2 = 2$); and the other (with $n_2 = 3$) observed by Paschen in the near infra-red in 1908. More interestingly, *unobserved* series were predicted: putting $n_2 = 1$ gave a series in the ultra-violet region (observed by Lyman in 1914); putting $n_2 = 4$ and 5 gave two infra-red series (observed by Brackett in 1922 and Pfund in 1924 respectively). Furthermore, the theoretical value for the empirical Rydberg constant suggested by equation (9) was found to be fairly accurate; Bohr gave figures of 3.1×10^{15} and 3.290×10^{15} respectively, the difference being within the limits of observational error in the values of the relevant constants (Bohr [1913], p.9).

More striking (and, for Lakatos, methodologically revealing) empirical success came with Bohr's application of his model in the same paper to singly-ionised Helium, He^+. A series of spectral lines, observed by Pickering in the spectrum of the star ζ Puppis, had previously been identified by Rydberg (1896) as the sharp subordinate series of atomic hydrogen. Reasoning by analogy with the spectra of the alkali metals, the Balmer series would have been its diffuse counterpart. From Rydberg's spectral equation, the wavelengths of the corresponding *principal* series could therefore be calculated. A line close to the first line of the predicted series was observed in the spectrum of ζ Puppis and the solar chromosphere (Fowler, 1898). In 1912, Fowler used a hydrogen-helium mixture in a discharge tube to produce lines close to the required wavelengths for the rest of the series, as well as a series in the ultraviolet region identified as the second principal series of hydrogen. There remained the unexplained problem of small but systematic differences between the predicted and empirical wavelengths.

The failure of Bohr's model to predict these lines would have been problematic, were they really to have arisen from hydrogen transitions. However, Bohr calculated that the lines could be accounted for using a version of equation (9) for helium (i.e. by putting $Z = 2e$), removing the hydrogen model's apparent omission. This contention was supported by Evans, who, at Bjerrum's suggestion (related by Bohr) produced the lines in question in a tube filled only with helium and chlorine (present to catalyse the ionisation of the helium). Fowler, however, objected that the accuracy of Bohr's calculations was no better than Rydberg's (hydrogen based) account, there still being discrepancies. These objections were withdrawn when Bohr improved the accuracy of his electronic energy values by replacing electronic mass m in the formula for the Rydberg constant with reduced mass μ, to give:

$$R_{He} = \frac{8\pi^2\mu Z^2 e^4}{h^3},$$

where $$\mu = \frac{m m_N}{m_N + m}. \qquad (m_N = \text{nuclear mass})$$

This replacement amounted to taking into account the motion of the helium nucleus, the original formula corresponding to the (counterfactual!) infinite nuclear mass approximation, and, when supported by the experiments of Evans, Fowler and Paschen, corroborated Bohr's theory and destroyed the generality of Rydberg's spectral equations. Lakatos' famous [1970] interpretation of this sequence of events was that Bohr's new improved model *corrected* a 'lower level' empirical law, converting a possible 'refutation' into predictive success. As significant, however, was the *way* that Bohr achieved this success. An apparent anomaly was explained by making the model more accurate. The increased accuracy was achieved by correcting an idealisation that was *already* known to make the original model counterfactual. Thus the correction would have been made at some point *anyway*, for reasons *internal* to the theory: this is the *autonomy of theoretical development* that Lakatos claimed was so typical of research driven by powerful heuristic. The heuristic power of a research programme arises in part from the illustrative analogies that are drawn between structures known only by mathematical description and those of more familiar acquaintance. That Bohr was readily able to calculate the difference that would be made by a moving nucleus underlines the close analogy between his atomic model and the Newtonian picture of the solar system. It was this analogy and the mathematical techniques developed for the latter that provided the former with its heuristic power. (For further discussion on this point see 3.6).

Parts II and III of the trilogy were much in line with the programme outlined in Bohr's memorandum to Rutherford of 1912, and had therefore been conceived much earlier. However, compared with the startling success of Part I, the later instalments were mostly qualitative and programmatic, with figures and calculations appearing mainly in the form of order-of-magnitude plausibility arguments. The chief problem was, of course, the extreme difficulty of the calculations. Most of the support the model *publicly* enjoyed arose from the explanations it allowed of the line spectra of simple atoms.

3.5. Was the Bohr Atom Inconsistent?

In what follows, I will consider some of the methodological judgements pronounced on the Bohr atom. Firstly, there is the inconsistency accusation, stated by Einstein (among many others) at the time, Jammer in a historical account, and Lakatos and Feyerabend in methodological contexts. The natural *strong* sharpening of the charge of 'inconsistency' is to one of *logical* inconsistency. Lakatos makes this accusation a number of times (for instance [1970], [1971a] and [1974]), and even draws a methodological lesson from it, in his attempt to falsify Popperian falsificationism:

> [T]his does not mean that the discovery of an inconsistency—or of an anomaly—must *immediately* stop the development of a programme: it may be rational to put the inconsistency into some temporary, *ad hoc* quarantine, and carry on with the positive heuristic of the programme. (Lakatos, [1970], p.143)

> There can be no 'instant rationality'. *Neither the logician's proof of inconsistency nor the experimental scientist's verdict of anomaly can defeat a research programme at one blow.* (Lakatos [1974], p.249)

Taking Bohr's atom as historiographical evidence, Lakatos judged it scientifically honest to 'quarantine' an inconsistency—its existence publicly recorded, of course—while further developing the research programme. This relaxation of what is usually taken to be a fairly weak requirement of rationality—that of logical consistency—seems *itself* inconsistent with Lakatos' own comment that in a *progressive* research programme, theories 'should be largely built according to a preconceived unifying idea, laid down in advance in the positive heuristic' (Lakatos [1974], p.249). Unless, of course, an inconsistent set of assumptions can serve as the 'unifying idea', in which case his requirement is not the strict one he claimed it to be. Feyerabend [1988] likewise saw the Bohr atom (along with the early infinitesimal calculus) as evidence of the growth of knowledge flouting the "tyranny of logic". In fact, Lakatos' judgement on inconsistency formed part of Feyerabend's [1976] critique of the standards of scientific honesty embodied in the methodology of scientific research programmes. Feyerabend observed that to countenance as honest the 'quarantining' of an inconsistency or empirical anomaly, given that its presence *is* recorded, is to do the same for a thief who openly declares his thievery.

To be quite fair to Lakatos, he did go to some trouble in his [1970] to distinguish *strong* and *weak* inconsistencies. According to this distinction, two propositions are inconsistent in the *strong* sense if their conjunction has no model. Two propositions are *weakly* inconsistent if they are inconsistent when the terms appearing therein are given a fixed interpretation, but do share some (unintended) models. The importance of the inconsistency depends on whether the relevant terms are *formative* or merely descriptive. For Lakatos, Bohr acknowledged the inconsistency by formulating the (in)famous complementarity principle, according to which there are parts of the world whose complete understanding requires different *incompatible* descriptions. Lakatos then accused him of elevating inconsistency to the status of a heuristic principle. He also declared that

> *consistency*—in the strong sense of the term—*must remain an important regulative principle* ... and inconsistencies (including anomalies) *must* be seen as problems. The reason is simple. If science aims at truth, it must aim at consistency; if it resigns consistency, it resigns truth. To claim that 'we must be modest in our demands', that we must resign ourselves to—weak or strong—inconsistencies, remains a methodological vice. (Lakatos, [1970], p.143)

On the other hand, Lakatos drew no *methodological* distinction between the temporary toleration of strong and weak inconsistencies, claiming that proof of inconsistency—weak or strong—could not provide grounds for 'instant rationality'. Furthermore, he never explicitly claimed that Bohr's theory suffered from weak (rather than strong) inconsistency, and often cited it as a historiographical example in the same paragraphs as the inconsistency of the early infinitesimal calculus and Fregean logicism after Russell's paradox (see for instance Lakatos [1971a], p.112-3 and [1974], p.248). I think that the distinction between the strong and weak forms of inconsistency is crucial, and that there is a fundamental *dis*analogy between the methodological standing of the Bohr atom and the early infinitesimal calculus on the one hand, and Fregean logicism at the time of Russell's paradox on the other.

It is not clear on what grounds Lakatos would have identified a *logical* inconsistency in Bohr's theory, apart from its being the natural reading a philosopher would give to Einstein's oft-quoted suspicions of incoherence behind Bohr's assumptions:

> That this insecure and contradictory foundation was sufficient to enable a man of Bohr's unique instinct and tact to discover the major laws of the spectral lines and of the electron shells of the atoms, together with their significance for chemistry appeared to me like a miracle—and appears to me as a miracle even today. (Einstein, [1951], p.83)

Now Lakatos set out to explain the intuitive judgements of scientists; to rationalise an *Einsteinian* judgement must have been of primary importance. There is, however, no trace of a contemporary "logician's proof" of the inconsistency of Bohr's atomic model. Such a negative existential statement cannot of course be proved, but I will make, as a pedantic point of elementary logic, the observation that from any logically inconsistent set of sentences can be derived p & $\neg p$, for any proposition p. In particular, this means that for any observation statement that was seen at the time to be hypothetico-deductive evidence for Bohr's theory, both it *and its negation* would have been derivable from Bohr's assumptions. This would have been well known in 1913, a number of years after Russell's deep 'intellectual sorrow' (Russell [1959], p.73) at the detection of antinomies at the heart of Frege's logicist edifice. It therefore seems inconceivable that the corroborating evidence cited in 3.4 would have been recognised as such, were there to have been a proof of the logical inconsistency of the theory. As it stands, this is a weak historical argument based on (what I take to be) intellectual standards in force in 1913. The weakness is that it merely *assumes* that the communities of logicians and physicists would share such a desire for consistency. However, the argument can be turned round: one starts with the 'principle of charity' when interpreting other people's statements. Thus contemporary methodological utterances are interpreted in such a way that they express at least *partial* truth, where possible. To this is added the observation that there is *no evidence* that the community of physicists of 1913 would have tolerated logical inconsistency *in general*. The clinching argument would be that there is *no evidence* for *either* of the following particular historical theses: (i) that Bohr's theory actually was logically inconsistent; and (ii) that it was thought to be so at the time.

So where did the logical inconsistency claim come from? Lakatos [1970] characterised Bohr's theory as quantum stipulation 'grafted' on to Maxwell's theory of electromagnetism. He was mistaken, however, in thinking that the whole Maxwell edifice was required for Bohr's successful predictions. As their derivation in 3.3 shows, one can get by with just the electrostatic Coulomb's Law to give the initial equation of forces. This point is crucial, because it might explain the otherwise incomprehensible inconsistency accusation. If, onto a theory which predicts p (i.e. radiative collapse of the Rutherford atom) one grafts some assumptions, and the extended theory allows deduction of $\neg p$ (*modulo* auxiliaries, of course), then *prima facie* one has a case of logical inconsistency. It depends, however, what you mean by 'grafting on'. If Lakatos meant simple conjunction he was right about the inconsistency, but wrong if he thought he was still referring to Bohr's theory. In the crucial derivation of the optical frequencies, Bohr needed to assume only an *approximation* to full electromagnetic theory—Coulomb's Law—that was furthermore strictly inconsistent with it (see equation 1). Furthermore, as we saw in 3.3, Bohr *explicitly rejected* the application of classical electromagnetic theory to his model. Thus historical thesis (i) is false, but one can see why Lakatos might have believed it. The *strong* reading of the inconsistency charge is historically groundless: wherever the inconsistency in Bohr's theory lies, it is surely not in its logical structure.

Physical theories are not, however, uninterpreted syntactical structures. When a physical theory is constructed, the author has in mind an intended interpretation. If she did not, it would not be possible to make the theory predictive, since there would be no explanatory domain; it would not be *about* anything. The intended interpretation encodes the general theoretical framework through which the theory 'says something' about the world. Its content arises from background theories, the author's metaphysical prejudices and any mathematical analogies that can be drawn between the theory at hand and better-understood theories. Therefore this intended interpretation will have excess content (I do not mean *empirical* content) with respect to the mathematical theory which is an attempt to formalise it. This is provable if the theory is formalisable in the first order predicate calculus. From the Löwenheim-Skolem theorem it follows that every first order system which allows infinite models has non-standard models—models which satisfy sentences which are false in the standard interpretation. There are, of course, more interesting ways to prove the failure of categoricity than via cardinality, as the Löwenheim-Skolem theorem does, but cardinality suffices. A natural interpretation of this theorem is that the formal theory fails to characterise fully the structure of the standard model, which *therefore has some excess content not captured by the theory*. In other words, the theory fails to be categorical. What the formal theory does capture, of course, is just that structure which the standard and non-standard models share. Quine, of course, is another precedent for this view of ontology and sentences, although he would no doubt be impatient with the mentalistic connotations of the phrase 'intended interpretation'.

If, as I argued above, the Bohr atom's perceived inconsistency does *not* lie in its syntactical structure—the equations—then perhaps it is located in the intended interpretation. The *excess content* account of the Bohr atom's failings would be that the mathematical part of the theory was inspired by two different and *incompatible* intended models, one being the mathematically continuous world of classical electrodynamics and the Newtonian solar system, the other the (rather ill-formed as yet) Planck-Einstein ontology of quantised oscillators. On this story, the perceived inconsistency occurred when equations arising naturally as formalisations of the two informal theories were conjoined to yield a theory which, although logically consistent (there being a logical interpretation which satisfied it), was not satisfied by any *single* physical model available at the time, or at least had no *single* natural associated interpretation. Bohr's theory must have appeared unfounded and fantastic—when realistically interpreted—against the towering theoretical edifice that was classical electromagnetic theory. There was, as yet, no suitable theoretical background in which to embed it, although efforts to construct one were an important part of the historical process that eventually *overthrew* Maxwell's theory.

The perceived inconsistency then turns out to be semantical and historically bounded. It is a historical fact about the context of the theory's proposal, rather than a timeless property of the theory's logical structure. It might be pointed out that any consistent first order theory has a model, and so interpretations would have been there for the constructing (*if* Bohr's theory *was* consistent: see above). However, not all of them would have been permissible as *physical* models against the background of available (meta-)physical theories. Logical constructibility implies neither possession *nor* scientific viability. This account, I think, is corroborated by Bohr's search for a coherent interpretation of his quantum condition. He re-interpreted it when he changed the theoretical radiation process, *meanwhile keeping the same equations*. This reinterpretation allowed him to explain the Balmer formula, an explanation that would have been unavailable under the previous interpretation. In a sense, Bohr was changing his theory at this point, but the change was between two theories that shared the same equations.

Feyerabend makes a useful connected point in his [1964], at a time when he was collecting the historical exceptions to methodological rules that later convinced him that there *are* no exceptionless methodologies. He begins by observing that as far as he is concerned, the general methodological arguments against instrumentalism are telling and conclusive. It is therefore preferable (for heuristic and predictive reasons) to interpret a theory realistically. However, there *are* times when, given background theories, the theory one accepts becomes false when this is done. The bulk of his paper is spent in consideration of two historical examples held to bear out this thesis. The first of these is quantum mechanics, which is false if considered in conjunction (interpreted realistically, of course) with General Relativity—hence the (apparent) instrumentalism of Copenhagen. His second example, the heliocentric Copernican cosmology against an Aristotelian background, is even more telling for a realist, because of the final outcome of the historical process. At the time of that episode, it will be remembered, the motion of the earth was widely considered to be physically laughable, because the prevailing Aristotelian physics—in which an object left to itself will remain at rest—predicted that objects thrown into the air would get left behind by the earth's motion and would therefore appear to fly off into space. Given that they don't, the natural response to this theoretical-empirical argument would be to give an instrumental interpretation to a theory which assumed the earth's motion. Thus in a certain restricted sense Osiander and Cardinal Bellarmine were correct. However, Feyerabend argues that

> a realist cannot rest content with the general remark that theories just *are* descriptions and not merely instruments. He must then also revise the accepted *physics* in such a manner that the inconsistency is removed; i.e. he must actively contribute to the *development* of factual knowledge rather than make comments ... about the *results* of this development. In addition he must offer methodological considerations as to why one should change successful theories in order to accommodate new and strange points of view. An excellent example of this situation is provided by the arguments against the realistic interpretation of the Copernican hypothesis and by the attempts that were made in order to overcome these arguments. (Feyerabend, [1964], p.177)

Thus if one *does* wish to interpret one's favourite theories realistically, one will not be content with sets of theories which are not consistent when so interpreted. Realist Copernicans should therefore have endeavoured to *replace* the Aristotelian kinematics, which is just what they did over the next few centuries. The fact that *Bohr* strove for a theoretical background which would allow his equations to *explain* atomic line spectra is evidence that he put the same constraint on his own theorising.

Now we can draw a connection between Feyerabend's comments, Lakatos' promulgation of weak consistency as a regulative principle, and my central argument. Each of Feyerabend's historical examples cited a theory which, when interpreted realistically, contradicted relevant accepted background theories. I have been arguing that the Bohr atom's spectral explanations and predictions were seen as methodologically suspect at the time *just because* there was no readily-available interpretation of the mathematical theory that was either consistent with—or capable of entirely replacing—relevant background theories and *their* associated interpretations. Instead, different parts of the theory were drawn from theoretical edifices with contradictory interpretive traditions. Whatever the reason, the Bohr theory's lack of a natural interpretation was methodologically problematic at the time, and made its eventual replacement (or that of the background theories) inevitable: in Lakatos' terminology, the Bohr atom was *weakly inconsistent* with the background theories of the time. The similarity between Feyerabend's account and mine is obvious when we note that they share a fundamental methodological intuition.[2] The common intuition is the requirement that *every theoretical complex has a coherent intended interpretation in line with which it was constructed.* Furthermore, this favoured model should be consistent with any (suitably interpreted) background theories which impinge on its domain of applicability. This last requirement is not, of course, a nascent criterion of

[2] Given that Feyerabend entertained such things in 1964.

demarcation, but rather a cross between pious hope, historiographical thesis and human necessity. Neither is it original: Lakatos formulated the same thing in the vocabulary of his methodology (see below). The intuition—a realist adage of long standing—is that instrumental interpretation can only ever be a temporary response to interpretive incoherence, never an end in itself. Instrumentalism as an *end* would be the enemy of progress; it is the push to overcome interpretive and theoretical (as much as empirical) difficulties that drives theoretical development.

3.6. WAS THE BOHR ATOM *AD HOC*?

Having located the inconsistency in Bohr's theory, another methodological issue arises: was it *ad hoc*? Heilbron and Kuhn ([1969], p.266) say so, but don't expand on the claim or comment on its methodological import. In this section I will answer the question with reference to the precise formulation of the phrase '*ad hoc*' provided by Lakatos, in the process unearthing another precedent for the interpretive requirement outlined at the end of the last section. In his [1971a], Lakatos attempted to formalise the intuitive notion of *ad-hoc*ness in line with his methodology of scientific research programmes (MSRP), in the process distinguishing three distinct varieties. This classification was amended by Zahar [1973]. Originally defined by Lakatos to be a relationship between a theory and its predecessors in the research programme of which it is a product, *ad-hoc*ness was later redefined (Zahar, [1978]) so that it corresponded to a three-place relation between a problem situation, a theory and a heuristic. Although philosophically novel, this redefinition makes no difference at the *historical* level, because the heuristic is mainly identified *as a historical item* through its influence on the construction of a series of theories. My main interest in this aspect of Bohr's reasoning will be in this historical sense and the redefinition anyway makes my task easier. A theory is said to be *ad hoc*$_1$ if it has no empirical content over the problem situation which it was constructed to explain. A theory is *ad hoc*$_2$—with respect to a particular explanandum or problem situation—if it does have such excess empirical content, but this content has not, however, been empirically tested, or has been found to conflict with some observations. Lastly (and most crucially for this chapter) a theory is *ad hoc*$_3$ if the assumptions brought in to deal with the problem situation 'sit uneasily' with the hard core, invoking a different (perhaps inconsistent) metaphysics to it. Thus they are not constructed in the spirit of the explanatory technique of the relevant research programme and therefore 'violate the heuristic'. One example of this last might include ascription of wave-like properties in a theory arising from a corpuscularist research programme, as in Newton's famous 'fits' of easy transmission, used by him to explain the increased (rather than decreased) speed of light inside solid objects. Another example of an *ad hoc*3 move relevant to this chapter is the quantisation of hitherto continuous dynamical quantities in a mathematically continuous research programme.

The above definitions of what it is for a theory to be *ad hoc* allude to a 'problem situation', a 'heuristic' and the 'relevant research programme', therefore to answer the question in hand, one must identify these items. Bohr knew that his theory was a decisive break with the classical tradition and constantly referred to the 'limited validity of ordinary mechanics'. However, his starting point in the 1913 paper was the classical instability of the Rutherford model. Furthermore, this was no mere rhetorical device, because we saw in 3.3 that stability initially *was* Bohr's chief perceived explanandum. So it is surely reasonable to infer that the *perceived problem situation*, which Bohr's theory was to explain, was the instability (both radiative and mechanical) of the classical Rutherford atom. Heilbron and Kuhn argue that the Balmer formula (along the rest of radiation theory) was also heuristically important, or, in the present terminology, that it was among the perceived problem situations the theory was constructed to accommodate. However, although *historically* correct, in that Bohr arrived at the final form of his theory only after seeing the Balmer formula, the role the empirical formula played isn't that great, since the only formal difference it made was fixing the form of a quantum condition whose necessity had already been argued for *on independent grounds*. On the other hand, Lakatos is clearly mistaken when he claims that 'Bohr hád not even heard of these formulae before he wrote the first version of his paper' (Lakatos, [1970], p.147), because according to Bohr's own testimony (and the timetable of the trilogy's writing bears this out), the Balmer formula *was* instrumental in his coming to the final interpretation of the quantum condition. Hitherto, as I argued in 3.4, Bohr had been working with a radiation mechanism which was in far closer analogy to Planck's conception, but the sight of the Balmer formula led him to the 'correct' interpretation, in which transitions *between* stable stationary states were the origin of spectral emission. The assumption brought in to deal with the original problem situation is then easily identified as the quantum condition, which after all was what was necessary to avoid the problem of radiative collapse.

So what was the relevant research programme? This is a more difficult question to answer, because Lakatos' methodological superstructure becomes rather cumbersome in just those situations where there is no clear candidate for this role, defined as it is in terms of 'heuristic' and 'hard core'. These items regulate the construction of theories in a research programme, being the MSRP answer to tacking-type paradoxes, among other problems. As I argued earlier, the problem with the Bohr atom is just that it lacked such unifying principles which would (ideally) supply the extra-theoretical content in line with which later, more sophisticated, versions were to be constructed. However, some information on this point can be gleaned from Bohr's reply to Fowler's reference to the Pickering lines (see

3.5, above). Bohr, it will be remembered, improved the accuracy of his model with the typically classical replacement of electron mass by reduced mass, indicating that he had Keplerian orbits in mind for the stationary states. Furthermore, there is nothing in Bohr's equations *themselves* that would indicate how they might be made more accurate, but central to the Rutherford model of the atom was the analogy with the solar system. Bohr applied mathematical techniques (for instance perturbation theory) that had been developed over the previous centuries as a by-product of the construction of extremely accurate planetary orbits. Of course, this classical and continuous structure is overwritten by the quantum condition. Overall, the theory has a classical flavour mitigated by the agnostic stance towards the mechanism of transition between the stationary states which hid behind the correspondence principle. Although he was aware that classical mechanics was ripe for replacement, Bohr's research programme had this classical aspect just because in 1913 there was no appropriate theoretical background with which to motivate his theory-construction, apart from the (higher order) correspondence principle. The metaphysical background, then, is classical *by default* and the theory's truly novel content—the quantum condition—was physically uninterpreted. Although Bohr made a number of comments about the breakdown of classical concepts, this breakdown had, as yet, only a limited application in a formal mathematical theory. So Bohr simply *had no choice* but to advance his programme according to a classical plan.

We can now answer the question of whether Bohr's theory was *ad hoc* under any of the above categories. It is obvious that Bohr's theory was neither *ad hoc*$_1$ nor *ad hoc*$_2$. Firstly, it could not have been *ad hoc*$_1$, since it did have excess empirical content beyond assumptions (the quantisation of electron orbits) brought in to deal with the problem situation for which the theory was constructed (the radiative collapse of the classical Rutherford atom). This excess empirical content was, of course, embodied in the spectroscopic predictions for non-Balmer hydrogen series and the spectral series of other one-electron atoms. So was it *ad hoc*$_2$? The equation for the energy differences between states, which invites comparison with the Balmer formula, follow—in the first derivation—from the assumptions introduced by Bohr to solve the stability problem (i.e. for independent reasons). We saw in 3.3 that the other derivation used the Balmer formula to fix the exact form of the quantisation, although Bohr had independent reasons for introducing a quantum condition of *some* form: the problem of stability. However, this is to ignore the heuristic importance of the sight of the Balmer formula in bringing Bohr to the correct radiation process. But this is all beside the point, because there was excess empirical content (the Paschen series and the value for the Rydberg constant) even beyond the Balmer formula. This information played no part in the construction of the theory and was furthermore confirmed *before Bohr even constructed his theory*, so it was born corroborated *even if* it was constructed via *ad hoc*2 reasoning from the Balmer formula. Bohr's theory was not *ad hoc*2 from the moment it hit the page.

So was the theory *ad hoc*3? Its structure—Coulomb's Law, classical kinematics and quantisation—is a curious and certainly not *unified*, mix of the continuous and the discrete. Via the correspondence principle, Bohr quantised only those quantities for which continuity was thought to fail: no overarching rationale for quantisation was produced. The theory consequently looks *ad hoc*3. At this point, the connection can be drawn between my account of the inconsistency in Bohr's theory lying in its excess content and its being *ad hoc*3. The two are intimately connected. To be *ad hoc*3 is to lack a coherent guiding metaphysics. Where is this guiding metaphysics located? In the excess content of the intended interpretation over the formalised theory.

CONCLUSION

So what was the promised methodological principle which the Bohr atom transgresses, which was expressed in the adverse methodological comments directed at Bohr's theory and which ensured its eventual replacement? Bohr's theory lacked a unified guiding metaphysics in line with which the formal theory was interpreted and further theory versions constructed. Therefore the sensible thing for Bohr to do, as the Feyerabend of 1964 argued, was to interpret part of his theory—the quantisation and consequent jumps between states—instrumentally, *although Bohr later added a philosophical superstructure to this stratagem.* This instrumental interpretation amounts to the 'quarantining' alluded to by Lakatos. So in the ideal limit, a powerful—progressive in Lakatos terminology—research programme will be one that enjoys the heuristic resource of and understandable and unified metaphysics superimposed on the bare formal theory. However, mindful of the heuristic role that this metaphysics is supposed to play, we can see that not just any ontology will do: there are constraints on useful interpretations. My guess is that the research programmes that progress fastest are those which impose a realistic interpretation on their successive theories. It is significant that Bohr expended much time and effort ensuring that the derivation he supplied for the quantum condition was *consistent with the physical process he thought it represented.* He already *had* the equation and knew that it could account for the Balmer formula if interpreted correctly, but this was not good enough: he wanted the equation to arise naturally from its interpretation. This suggests that he was implicitly applying a rule much like that presented here: *interpret your equations in coherent fashion.*

A historical prediction follows: when no such *coherent* interpretation is available and (partial) instrumentalism is necessary as a temporary response to theoretical difficulty, it will be recorded *as a failing* by contemporaries (cf. Einstein's comments). In the long run, the interpretational problem will inspire either the overthrow of the background theory or the construction of a suitable interpretation, which might facilitate a new overarching theory (*viz.* quantum mechanics). This realist requirement is entirely methodological: it is a recognition that the aspiration to interpret one's favourite theories realistically motivates the construction of unified pictures of the world, and that these have constituted some of the best examples of the growth of scientific knowledge. It does not imply the truth of such theories.

4

SCHRÖDINGER *VS.* HEISENBERG *OR* THE IMPORTANCE OF INTENDED INTERPRETATIONS

> After a further ten minutes of hard climbing we were standing in the sun—at saddle height above the sea of fog. To the south we could see the peaks of the Sonnwend Mountains and beyond them the snowy tops of the Central Alps, and we all breathed a sigh of relief. In atomic physics, likewise, the winter of 1924-1925 had obviously brought us to a realm where the fog was thick but where some light had begun to filter through and held out the promise of exciting new vistas. (Heisenberg [1971], pp.59-60)

INTRODUCTION

The circumstances surrounding the proposal of the two formalisms of quantum mechanics in 1925 and 1926 have inspired many a critique of scientific realism. Fine [1984] makes his attack a methodological one. The chief thrust of the argument is that the progress that has been achieved in the quantum domain since the 1930s *might not have happened* had physicists in general held up the development of the theory until Einstein and other dogmatic realists were satisfied that the theory was consistent with a metaphysics that realists would find attractive. Not only was realism illegitimate, it was also *unhealthy*.

In this chapter, I will take issue with Fine's *historical* claim that quantum mechanics was the product of a research effort with purely instrumentalist aims. In 4.1, Fine's argument will be examined and the historical theses identified. These will be compared with the actual history in 4.2 and 4.3, which deal with the progress of Heisenberg and Schrödinger towards their respective formalisms. Neither these developments nor the authors' subsequent attitudes to their creations, it will be argued, make sense given a purely instrumentalist rationale. Section 4.4 can be read as an existence claim for *intended physical interpretations*, especially on Schrödinger's part, in that the calculi of the two theories were equivalent, but the theories themselves could be distinguished by their authors' intended interpretations. It is argued that Schrödinger in particular envisaged developments of his theory that were peculiarly dependent on his interpretation.

4.1. Fine, Abductive Inferences and Twentieth-Century Physics

What has killed realism, argues Fine in his [1984], is not only the neopositivists' ability to

> accept all the results of science, including all the members of the scientific zoo, and still declare that the questions raised by the existence claims of realism were mere pseudoquestions. (Fine [1984], p.83)

There was an accomplice in quantum mechanics:

> Its [i.e. realism's] death was hastened by the debates over the interpretation of quantum theory, where Bohr's nonrealist philosophy was seen to win out over Einstein's passionate realism. Its death was certified, finally, as the last two generations of physical scientists turned their backs on realism and have managed, nevertheless, to do science without it. ([1984], p.83)

In other words, the history of quantum mechanics shows that scientists need not construct and interpret their theories in accordance with realist claims. It is one of the chief aims of this chapter to take issue with Fine's brief portrayal of the history of twentieth-century physics as an enterprise whose aims and methods are primarily those of the instrumentalist. First, however, it would be useful to reiterate Fine's attack on realism, so that the differences between 'realist' and 'non-realist' methodological claims are manifest.

In 2.2, we saw Fine validate the death certificate for realism by pointing to the poverty of the *explanationist* argument that realism is the best explanation of the success of science, which he argued to require an inference that itself is at the heart of anti-realist worries concerning realism. The methodological twist—that realism provides the best explanation of the success of *realist strategies* of theory construction—was argued by Fine to be subject to the same objection. Thus *even if* the premises of the methodological arguments *were* correct, they would still be of no use: realism would be no more attractive to sceptical eyes. Call this the *circularity objection.* Fine added to this the claim that the strategies invoked by the methodological twist—the 'small handful' being the example—could not in any case be rationalised on the basis of realism, in other words that *methodological realism* (MR) could not explain the prevalence of such strategies. The thesis that theories might be constructed with purely instrumentalist virtues in mind—*methodological instrumentalism*—was held to explain the small handful strategy more efficiently (see 2.4 to 2.6 for a rebuttal).

Concluding that MR cannot support scientific realism (SR), and that MR cannot explain scientific practice, Fine goes on to argue that 'realism has not always been a progressive factor in the development of science' (Fine [1984], p.91). In other words, MR is actually false. Two historical examples are cited that purportedly refute MR. Firstly, there is Einstein's much-cited Machian construction of relativity. The young Einstein, according to Fine, applied Occam's razor to absolute space and time with a zeal only available to paid-up positivists. Einstein's 'Machist line', which was 'always used to deny that some concept has a real referent' ([1984], p.92), was indispensable to Einstein's discovery of special relativity. As this young Einstein got older, he underwent a 'philosophical conversion', so that by 1920

> Einstein wanted to claim genuine reality for the central theoretical entities of the general theory, the four-dimensional space-time manifold and associated tensor fields. ([1984], p.92)

This leap of faith cannot be a precondition of the development of relativistic theories, however, because

> the majority opinion among working, knowledgeable scientists is that general relativity provides a magnificent organizing tool for treating certain gravitational problems in astrophysics and cosmology. ([1984], p.92)

Fine concludes that

> For relativistic physics, then, it appears that a nonrealist attitude was important in its development, that the founder nevertheless espoused a realist attitude to the finished product, but that most who actually use it think of the theory as a powerful instrument, rather than as expressing a 'big truth'. ([1984], p.92)

Fine's second example is Heisenberg's construction of matrix mechanics, and Schrödinger's of wave mechanics. Heisenberg, of course, began his [1925] paper with a famously positivistic abstract:

> In this paper an attempt will be made to obtain bases for a quantum theoretical mechanics based exclusively on relations between quantities observable in principle. (Heisenberg [1925], p.261)

Heisenberg consciously rejected 'the very idea that one should try to form any idea of a reality underlying his mechanics' (Fine [1984], p.92). Schrödinger, meanwhile, briefly entertained a 'vague picture of an underlying wavelike reality for his own equation.' ([1984], p.92-3), but difficulties and objections forced him to abandon 'the attempt to interpolate any reference to reality.' ([1984], p.93) Heisenberg's and Schrödinger's 'instrumentalist moves, away from a realist construal of the emerging quantum theory' ([1984], p.93) were capped by the triumph of Bohr's interpretation of the new theory at the Solvay conference of 1927. This 'quantum nonrealism' has provided the 'conceptual backdrop' for fifty years' progress in microphysics. Furthermore, it is now standardly imbibed by students with the milk of their lecture notes, and must therefore have inspired the growth of physical knowledge from 1927 onwards.

For Fine, the subsequent debate between Bohr and Einstein over the correct interpretation of quantum theory was more than just an 'idle intellectual exercise' ([1984], p.93); it was 'an important endeavor undertaken by Bohr on behalf of the enterprise of physics as a progressive science.' ([1984], p.93) Now Fine refers to the dispute in a sweeping historical story whose chief aim is to provide an argument against realism and its methodological place in science. Elsewhere in the article, Fine took the problem with realism to be the invalidity of abductive inferences. For the reference to Bohr and Einstein to be *relevant* in that context Fine must be making a distinction between two possible readings of their debate. (i) On the one hand, consider a history in which Bohr and Einstein argue over which interpretational superstructure should be superimposed on a *true* quantum mechanics: in Fine's terms, this would have been 'just a sideshow' ([1984], p.93), because *both* would have been making abductive inferences, and the dispute would have concerned extra-scientific—interpretational—matters. (ii) On the other hand, if the protracted exchanges signalled a disagreement over the legitimacy of *any inference at all* about reality, they would be a crucial part of the *methodological* fight between *progressive instrumentalism* and *reactionary realism*: a fight for the hearts and minds of the practitioners of the new physics. Note that, according to history (ii), Heisenberg, Bohr *et al.* should not have made any abductive inference to the (approximate) truth of a theory from its empirical adequacy. Now Fine takes Bohr's victory in the 'war between Einstein, the realist and Bohr, the nonrealist' ([1984], p.93)—and the subsequent history of progress in physics—to be both *relevant*—i.e. to be a good stick with which to beat realism—and to have been 'an important endeavour' on Bohr's part: (ii) must therefore be Fine's favoured reading.

This second reading is supported by the nature of the conclusions drawn by Fine. The nonrealists have been vindicated historically, he claims, because realist programmes—Einstein's fields, de Broglie's pilot waves and Bohm's hidden variables—have been relatively *sterile* when compared with the smooth progress inspired by Copenhagen agnosticism. Instead of seeking *foundations* for quantisation, the nonrealists of twentieth-century physics have taken the quantum as *basic*, and teased out its consequences:

> The task of 'explaining' the quantum, of course, is the realist program for identifying a reality underlying the formulas of the theory and thereby explaining the predictive success of the formulas as approximately true descriptions of this reality. ([1984], p.93)

In contrast, the agnostic or nonrealist attitude has been *methodologically vindicated*:

> One can hardly doubt the importance of a nonrealist attitude for the development and practically infinite success of the quantum theory. Historical counterfactuals are always tricky, but the sterility of actual realist programs in this area at least suggests that Bohr and company were right in believing that the

> road to scientific progress here would have been blocked by realism. The founders of quantum theory never turned on the nonrealist attitude that served them so well. ([1984], p.93)

At an earlier point in the article, Fine remarks that he and Putnam have 'fought a battle to show that the quantum theory is at least consistent with some kind of underlying reality' (Fine [1984], p.94), despite the 'charybdis of realism' that is the problem of Bell correlations. As an aside, I think this comment indicates that two realist issues have been run together. The first issue concerns SR and abductive inferences: whether or not the predictive success of the quantum mechanical formalism—or any other theory—licenses *any* view about the world beyond its empirical adequacy. The second issue concerns the consistency with the formalism of a particular *type* of interpretation. It is a *realist* issue because the interpretations in question are those in which quantities that correspond to certain physical concepts (such as position and momentum, or trajectory) possess determinate values, which values may therefore be considered physically *real*, if possession of determinate values be the criterion of reality. The second question surely becomes interesting only if the first is settled in favour of SR.

Let us return to the central issue, that of SR and MR: anti-realists were people who worried about the (in)validity of abductive inferences. Realists, on the other hand, would typically see inferences to the approximate truth from previous success of even *refuted* theories as *necessary* in specific cases, even though they may not be amenable to a general justification (*that* would be to solve the problem of induction). Now of course Fine did not separate MR from SR: MR was merely a dubious lemma in an invalid argument. When they *are* separated, however, the argument is clearly directed at both. Fine's argument against MR turned on the failure of the *abductive inference* to rationalise the small handful strategy. MR being the thesis that scientists often act like realists, an argument that mentions abductive inferences in this way can only make sense *as an argument against MR* if abductive inferences are, in fact, *typically realist moves*. Thus Fine—and leading anti-realists who make an issue of the validity of the abductive inference (see for instance Laudan [1984])—cede this ground to the methodological realist. This suggests a historiographical hint: if one wishes to look to scientific practice or history for support for MR, one might look out for abductive inferences. Now Zahar [1989] has doubted the depth of the young Einstein's positivism. There *is* a realist rationale for Einstein's progress. Removing absolute space and time from the scientific ontology might be *metaphysically* motivated: one cannot believe that nature contains entities whose existence is concealed by physical conspiracy. In other words, Einstein might have been applying metaphysical constraints on the acceptability of scientific theories.

To sum up, Fine attacks methodological arguments for SR on two separate counts: their validity and the soundness of the premises. (i) He has shown that the inference from MR to

SR must be invalid: this much follows from their logical independence. (ii) He has argued that methodological instrumentalism can better explain typically 'realist' strategies of theory construction and choice. (iii) Finally, he has argued that the discoverers of relativity and quantum mechanics did not act in accordance with MR anyway, so MR is false. I will remain silent on the first group of claims, but to demonstrate the falsity of the last group of claims is the main aim of this chapter. In the next two sections, I will therefore attempt to sketch the background to the construction of the formalisms of matrix and wave mechanics. The ease of the explanation in (ii) will turn out to be the explanatory downfall of the instrumentalist account, because neither Schrödinger nor Heisenberg were satisfied with mere empirical adequacy. They restricted their attention to those theories which could give a non-*ad hoc* explanation for relevant theoretical facts.

4.2. Heisenberg's Probabilistic Agnosticism

Heisenberg had been interested in the quantum theory of the atom since his earliest postgraduate work under Sommerfeld in Munich. One of the failures of the old quantum theory had always been its inability to yield the *intensities* of atomic spectral emission lines: Bohr had more or less rendered the *frequencies* of the principal spectral lines of simple atoms in 1913. Heisenberg, among many others, attempted—unsuccessfully—to derive expressions for line intensities by constructing quasi-mechanical orbital models of the atom, and modelling their interaction with radiation via the classical perturbation theory that had been developed for the gravitational interactions between planets (see Cassidy [1979] and Hendry [1984], chapter 4).

In his [1913], Bohr required of the quantum-theoretic equations he was endeavouring to derive that in the limit of large quantum numbers they approach the relevant classical laws. This was the first proto-formulation of the correspondence principle. This principle, further elucidated in Bohr [1918], turned out to be a powerful heuristic device in that it put constraints on the equations of the old quantum theory, allowing "systematic guessing" of their correct form. Another important conceptual strand—the roots of which can be traced back to Planck's early model for black body radiation—begins in 1900 with Drude's classical theory of dispersion. In order to capture the atom's interaction with incident radiation, Drude modelled the atom as an array of oscillators with frequencies v_i, the v_i being the atom's characteristic absorption frequencies. The classical dispersion equations were obtained by modelling these oscillators as oscillating charged particles, deriving and solving a second-order differential equation for the electric moment vector P for each frequency, then summing over frequencies to get the expression for the whole atom:

$$P = \alpha E,$$

where $$\alpha = \sum_i \frac{e^2}{4\pi^2 m} \cdot \frac{f_i}{\nu_i^2 - \nu^2}$$

The constant f_i appearing in the derived expression represents the *strength* of the oscillator, and is therefore a measure of the atom's absorption at this frequency.

Ladenburg [1921] produced a quantum-theoretic re-interpretation of Drude's equations, again modelling the atom as a set of 'virtual' oscillators of frequencies equal to its absorption frequencies. Since Bohr's 1913 theory, such absorption frequencies had been understood to represent transitions between *two stationary states* of the atom. Ladenburg therefore represented the oscillator strength f_i as a measure of the 'transition amplitude' between the two states, which would be a measure of the probability of a transition between them. His formula, however, described only absorption processes, and was therefore fully applicable only to atoms in their ground state.

In the January 1924 edition of *Nature*, Slater published a letter outlining an extension of the virtual oscillator idea: the virtual radiation field. The most interesting part of the letter reads:

> Any atom may, in fact, be supposed to communicate with other atoms all the time it is in a stationary state, by means of a virtual field of radiation originating from oscillators having the frequencies of possible quantum transitions and the function of which is to provide for statistical conservation of energy and momentum by determining the probabilities for quantum transitions. The part of the field originating from the given atom itself is supposed to induce a probability that that atom lose energy spontaneously, while radiation from external sources is regarded as inducing additional probabilities that it may gain or lose energy, much as Einstein has suggested. (Slater [1924], pp.307-8)

Now Slater's idea was built in to the ill-fated Bohr-Kramers-Slater (BKS) theory (see Hendry [1981]). The refutation of BKS, via the Bothe-Geiger experiments of 1924, was blamed on an assumption that Bohr and Kramers persuaded Slater to conjoin with his innovation, namely statistical (rather than strict) conservation of energy and momentum. Slater's idea was originally that the virtual field would guide quanta of energy emitted and absorbed by atoms—conservation of quanta would guarantee conservation of energy. However, Bohr and Kramers persuaded Slater, against his "better judgement" that light quanta shouldn't enter the theory as real entities. However, the BKS theory should be credited with two important developments: the advancement of Kramers' quantum theoretic dispersion theory, and the endowment of the concept of probability with physical reality.

In later papers, Kramers developed a theory of dispersion based on Slater's idea *only*, that amounted to another quantum-theoretic version of Drude's theory. Kramers [1924] noted that the earlier quantum version of Drude's theory, although empirically adequate, failed to satisfy the correspondence principle, in that it did not converge to classical theory in the limit of large quantum numbers. In an attempt to generalise Ladenburg's formula to atoms in states other than the ground state, he introduced a set of terms in the dispersion expression which represented "negative oscillators" with negative strengths, corresponding to emission frequencies. A more rigorous derivation of the formula followed in later papers, including a joint paper with Heisenberg. Kuhn and Thomas, meanwhile, proved in separate papers that the transition amplitudes f_i satisfied the "f-Summensatz" or "f-summation theorem":

$$\Sigma_k f_{ki} - \Sigma_{k'} f_{ik'} = 1.$$

The next significant development was Born's [1924] attempt to construct a quantum mechanics—the first time this phrase had been used—to replace classical perturbation theory in the calculation of interactions between electrons in the same atom. Electrons and radiation, he noted, interacted non-classically; he therefore applied the method of quantisation used by Kramers in the construction of his dispersion theory to interactions between electrons. Born—assisted in writing the paper by Heisenberg—first shows classical dispersion theory to be an application of classical perturbation theory: what is needed is the *quantum*-theoretic perturbation theory of which the quantum-theoretic dispersion theory is a special case. Kramers had reached his quantum theory of dispersion by replacing certain differential terms in the classical formulae by difference terms. Born does the same with a formula derived from classical perturbation theory to get a "quantum mechanical" perturbation formula. Jammer ([1966], p.193) calls this process 'Born's correspondence rule', it being a formal special case of the correspondence rule of Bohr's [1913] and [1918].

In the spring of 1925, Heisenberg made a number of attempts to derive formulae for the intensity of lines in the hydrogen spectrum by considering the Fourier expansion of the electron's Keplerian orbit. After a (now famous) attack of hay-fever, Heisenberg retreated to an island without vegetation, Heligoland, and overnight had a burst of inspiration. For once, it is not simplistic to say that the fruit of this labour, published as his [1925], was the decisive breakthrough to the long-foreseen quantum mechanics. In the paper, Heisenberg attributes the frustrating lack of success in dispersion theory to the *incorrectness of classical kinematics*. Heisenberg accepts the classical equation of motion for an electron:

$$d^2x/dt^2 + f(x) = 0,$$

except that the function $x(t)$ no longer represents the spatial position of the electron as a function of time. In classical mechanics, a periodic motion $x(t)$ can be expanded as a

Fourier expansion. Quantum theory makes the frequency of the periodic motion, and the coefficients a_α, depend on the quantum number *n*:

$$x(t) = \sum a_\alpha(n)e^{i\alpha\omega t} .$$

Heisenberg replaces the Fourier terms by quantum mechanical terms of the form:

$$a(n, n-\alpha)e^{i\omega(n, n-\alpha)t},$$

the coefficients of which, the *a(n, n–α)*, again represented "transition quantities", which in turn were measures of transition probability, which, in turn, Einstein had linked to line intensities. These would be the new quantum-mechanical quantities, manipulations of which would replace the manipulation of position and momentum variables.

There were a number of reasons to proceed with the construction of a kinematics in terms of line intensities and frequencies. Firstly, they were *observable,* unlike the classical kinematics of position and momentum variables, and thus had two possible advantages: all attempts at a theory of dispersion founded on a *mechanical* model of the atom had failed: one way to interpret this failure was as a refutation of the very *notion* of orbit. Jammer reads Heisenberg's move thus:

> Basically, Heisenberg's attitude, in this respect, resembled that of Einstein, for whom the concept of Newtonian time had lost its physical significance not only, as he showed in his analysis of the simultaneity of spatially separated events, because of its insusceptibility to operational determination but also because classical physics which assumed this concept as observable conflicted with experience. ([1966], p.199)

Heisenberg had been impressed by Einstein's denial of any sort of physical reality to absolute time: a calculus of observables might therefore have appeared attractive. A last (crucial) advantage was that for these quantities, a reliable version of the correspondence principle had been developed by Kramers and Born. Heisenberg built the correspondence principle (with its obvious heuristic advantages) into the new mechanics, to replace the ad hoc "systematic guessing" that had been the hallmark of the old quantum theory's application of the same principle.

To construct a *kinematics*—his avowed aim—Heisenberg would have to derive a multiplication law for the all-important coefficients, which turned out to be:

$$a(n, n-\alpha) = a(n, n-\alpha')a(n-\alpha', n-\alpha),$$

which Born recognised to be the rule for the multiplication of matrices:

$$(xy)_{mn} = \Sigma_k x_{mk} y_{kn}.$$

Born then proceeded to develop a coherent matrix-mechanical formalism in joint papers with Jordan [1925] and Heisenberg and Jordan [1926].

Cassidy [1991] identifies at least three stages in Heisenberg's interpretational views: (i) an initial instrumentalism with respect to the *equations* of quantum mechanics, and the probabilities it appeared to invoke (what was real was only the non-classical discreteness); (ii) acceptance of the complementarity of the wave and particle pictures under pressure from Bohr; (iii) a shift of emphasis with the Kantian version of the Copenhagen interpretation under the influence of his student von Weizsäcker. It should be stressed, however, that at each stage the interpretation appears to have been seen by Heisenberg as a *natural reading* of the lessons that quantum mechanics can provide as to the nature of the physical world, rather than a position with a more general *epistemological* rationale. The indeterminacy surrounding the classical states arises from the *structure* of quantum mechanics: to call this position 'instrumentalist' is to get the inference the wrong way round (see Redhead [1987], pp.50-1 and Krips [1987], chapter 1). Heisenberg argued that the duality of the description afforded by his interpretation was a property of nature itself, as were the probabilities appearing therein: 'the laws of nature determine not the occurrence of an event, but the probability that an event will take place' (Heisenberg [1958]). As is also well known, Heisenberg linked these probabilities to the *possibilities* or *potentia* of Aristotelian metaphysics. The subject-dependent element enters only when we note that 'the determinateness of phenomena exists only insofar as they are described with the concepts of classical physics' (Heisenberg [1958]). This raises the question of why the concepts of classical physics are not just abandoned, to be replaced by the new ones:

> Here it first of all necessary to stress, as von Weizsäcker has done, that the concepts of classical physics play a role in the interpretation of the quantum theory similar to that of the *a priori* forms of perception in the philosophy of Kant. Just as Kant explains the concepts of space and time or causality aprioristically, because they already formed the premises of all experiences and could therefore not be considered as the result of experience, so also the concepts of classical physics form an *a priori* basis for experiments in quantum theory, because we can conduct experiments in the atomic field only by using these concepts of classical physics. (Heisenberg [1958])

For Heisenberg, then, the formalism of the matrix mechanics was interpreted partly realistically, and partly conventionally, the conventionalism having a Kantian gloss. The indeterminacies were real, in whatever sense the Aristotelian *potentia* were, but our inability to describe the outcomes of experiments other than in classical terms—and the apriority of causality—yields a version of Kant's transcendental conventionalism with respect to the concepts of classical physics. The interpretation had a realistic heart, but with a pessimistic overlay that denied the knowing subject's ability to depart from the classical physics of everyday experience, a departure that would be necessary for the comprehension of events at the level of applicability of his matrix mechanics.

To conclude, Heisenberg's progress fails to fit Fine's description in the following ways. Firstly, Heisenberg applied non-empirical (and only tenuously empirical) constraints on the construction of his new theory: he carried over structural features of earlier quantum theories (the correspondence principle, for instance) and interpreted the failure of the BKS theory *abductively* as a demonstration of the conservation of energy for individual events. Secondly, he took the success of quantum mechanics to be *evidence* for the Copenhagen interpretation, an interpretation whose content surely transcends the evidence in its favour. This bears all the signs of an abductive inference to the correctness of quantum mechanics under the Copenhagen interpretation.

4.3. Schrödinger's Realistic Analogy

Three distinct approaches to the construction of the wave mechanics can be given. The first is a historical account of the influences on Schrödinger's thought, and his strikingly relevant work in connected fields before 1926. The others were presented explicitly as derivations in published papers: I will rehearse the historical development first. Schrödinger's published derivations, and considerations on their relevance to the circumstances of the construction of wave mechanics, will follow.

Schrödinger studied at the University of Vienna from 1906 until receipt of his Ph.D. in 1910. He had entered the university just after the suicide of Ludwig Boltzmann, whose tradition in statistical physics was therefore still dominant. Later, he was to call Boltzmann's line of thought his 'first love in science' (Schrödinger [1935], p.13). Much of Schrödinger's work before 1926 was concerned with the kinetic theory of gases and statistical mechanics; this aspect of his work has been covered in detail by Hanle [1977a], Wessels [1977], and Moore [1989]. After serving in what is now Slovenia during the first world war, Schrödinger published a number of papers on the statistics of ideal gases. There had been much debate among physicists in the period between 1918 and 1924 regarding calculations of the absolute entropy of an ideal gas. In classical theory, the natural way to proceed was count the number of ways that energy states could be assigned to the N molecules of the gas, to give W, the number of *complexions* the gas could take. The entropy S could then be found from $S = k \ln W$. Sackur and Tetrode had independently realised that entropy so calculated would fail to be an *extensive* state function: it would not satisfy $S(X_{1+2}) = S(X_1) + S(X_2)$, where X_{1+2} is the supersystem composed of two subsystems X_1 and X_2. The resultant statistics therefore had to be corrected, effectively by dividing the number of arrangements W by a factor $N!$—a correction widely thought *ad hoc*. In 1921 Planck proposed to interpret this correction as an acknowledgement of the indistinguishability of assignments of energy states to gases that differed only by the exchange of identical molecules.

In 1924, S.N. Bose sent Einstein a copy of a manuscript—following its rejection by the *Philosophical Magazine*—in which he constructed a novel statistics for light quanta in a new derivation of Planck's radiation law. Einstein did two things: arranged for the paper to be published, and himself applied the new statistics to ideal gases. The result was the Bose-Einstein statistics, in which the $N!$ correction was found to be unnecessary, for the new statistics effectively enumerated the different the ways that cells in the energy phase-space could be assigned occupancies by molecules. Thus indistinguishable complexions were not counted more than once. The physical foundation of the new statistics was unclear, however, and there were many objections, notably from Planck and Ehrenfest, who defended the motivation of that the $N!$ correction within the classical treatment. In some early papers on the subject, Schrödinger attempted to give a motivation or *physical foundation* to the application of the Bose-Einstein statistics. Schrödinger argued that the physical background to Planck's definition of entropy would, if applied consistently, imply Einstein's statistics rather than Planck's, for Planck's motivation for the Sackur-Tetrode correction amounted to the attribution of individuality to molecules followed by a denial of the distinguishability of states of the system that differed by the exchange of identical molecules. It would be better, argued Schrödinger, to deny identical molecules individuality in the first place, and apply classical statistics only to the gas as a whole. This would result in a procedure that was mathematically equivalent to that of Bose and Einstein, but would—physically—make the same sense as the classical Boltzmann-Gibbs statistics.

In the second of two papers on the subject (Einstein [1925]), Einstein cited the ideas of Louis de Broglie, whose Ph.D. thesis of November 1924 proposed a complete parallel between matter and light. Material particles were assigned a "phase wave" of frequency $\nu_{d.b.}$, where $\nu_{d.b.} = E/h = mc^2/h$, and $\lambda_{d.b.} = h/p$.

Shortly after reading Einstein's paper (on 3 November 1925), Schrödinger wrote to him:

> I have read with greatest interest a few days ago the ingenious thesis of Louis de Broglie, which I finally got hold of; with it also ¶8 of your second degeneracy work has become completely clear to me for the first time. The de Broglie interpretation of the quantum rules seems to be related in some ways to my note in the [Zeitschrift für Physik] 12, (p.) 13, 1922, where a remarkable property of the Weyl 'gauge factor' $e^{-\int \phi_i dx_i}$ along each quasi-period is shown. The mathematical situation is, as far as I can

> see, the same, only from me much more formal, less elegant, and not really shown generally.
> (Schrödinger, letter to Einstein, quoted in Hanle [1979])

In 1922, Schrödinger had published a paper called "On a Remarkable Property of the Quantized Orbits of a Single Electron" (Schrödinger [1922]), in which he noted that if the quantum conditions are applied to a model of the hydrogen atom in which only electrostatic forces are taken into account, and a vector was associated with the electron (which is in a closed Bohr orbit) such that the vector was always displaced parallel with itself, the tract of this vector would, on completion of one orbit, be multiplied by an integral power of $e^{-h/\gamma}$. Schrödinger's final paper on gas statistics—his last publication before the construction of wave mechanics—contains some striking echoes of wave mechanics.

Bloch [1976] relates an anecdote concerning the social events that put Schrödinger on the last steps of the route to wave mechanics. In early 1926, Bloch, a student at the E.T.H. (Zurich's Federal Institute of Technology), was in the habit of attending a physics colloquium, which although run by Debye of the E.T.H., was held jointly with the University of Zurich, where Schrödinger taught. According to Bloch, Debye suggested that Schrödinger give a presentation on de Broglie's thesis. Schrödinger gave a 'beautifully clear account' (Bloch [1976], p.23) of de Broglie's association of a wave with each particle, and the consequent derivation of the Bohr-Sommerfeld quantisation rules. Apparently, Debye commented that the de Broglie picture was 'childish', and suggested that Schrödinger set up a wave equation. Schrödinger began his next colloquium presentation— which was only 'a few weeks' later—with the comment 'My colleague Debye suggested that one should have a wave equation; well, I have found one!' The presentation that followed was essentially what was about to be published as the first of the papers on wave mechanics.

There are a number of reasons for being sceptical of the strict historical accuracy of Bloch's account. Firstly, although Bloch is not specific about the dates of his story (earlier in the article he refers to 'early 1926', at which point Schrödinger must have been well on the way to wave mechanics), and so one cannot be sure that it *contradicts* other historical evidence (Schrödinger's letter to Einstein of late 1925, for instance). However, the account suggests that Schrödinger's only, or main, reason for reading de Broglie's thesis was Debye's suggestion, which is implausible given that Schrödinger had *many* reasons for such an interest, including his earlier work on the Weyl 'Tract factor', Einstein's reference to de Broglie and his own work on gas statistics. Secondly, Debye later claimed to fail to remember the incident. Now Bloch takes this to mean that Debye had expunged the incident from his memory in regret at not carrying through the suggestion himself. On the other hand, Debye's 'lapse' might indicate that the episode should be interpreted as a contribution to scientific folklore, rather than scientific history. Erwin Fues, an assistant of

Schrödinger's, remembers only that Debye made suggestions as to the *presentation* of the wave mechanical formalism in the first paper (Schrödinger [1926b]). Although of questionable accuracy, Bloch's story does bring into sharp relief the factors that were available to Schrödinger: firstly de Broglie's relation; secondly the crucial switch from model of a moving wave associated with a single electron to the construction of a suitable wave equation for the system as a whole.

A short and highly appealing construction of Schrödinger's wave equation appears in Moore's biography of Schrödinger (Moore [1989], p.197-9). Although it never appeared in print, it is historically apposite in two ways. Firstly, it reflects considerations that were uppermost in Schrödinger's mind. Secondly, something similar—a relativistic version—appears in one of Schrödinger's notebooks from the time. Interestingly, this derivation yields an equation equivalent to the Klein-Gordon equation, which Schrödinger mistakenly discarded. The *non*-relativistic procedure amounts to the simple-minded substitution of the de Broglie relation into an ordinary steady-state equation describing wave motion, in line with Debye's reported suggestion. It begins with the classical equation for the amplitude ψ of a wave motion:

$$\Delta\psi + k^2\psi = 0,$$

where $$k = 2\pi/\lambda,$$

and $$\Delta = \sum_i \delta^2/\delta q_i^2.$$

Into the second formula substitute the de Broglie relation:

$$\lambda = h/mv;$$

rearranging and substituting $E - V = T = mv^2/2$ yields the wave equation:

$$\Delta\psi + 8\pi^2m^2/h^2(E - V)\psi = 0.$$

The first *published* derivation is presented by Schrödinger in the first of the "Quantization as an Eigenvalue Problem" papers (Schrödinger [1926b]), and proceeds via the famous Hamilton-Jacobi extremum argument. *Ostensibly*, the paper is neutral with respect to interpretation: the presentation turns on the application of purely mathematical constraints to a function whose intended physical interpretation is not explicitly stated, although it is clear from some of Schrödinger's comments that he does have an interpretation in mind.

A second derivation (Schrödinger [1926c] and [1928b]) proceeds explicitly via the wave interpretation, surprisingly enough given its short and turbulent history. Contrary to Fine's claim that Schrödinger quickly dropped any attempt at realistic interpretation, it is in line with the foundations for wave mechanics that Schrödinger constructed later, and referred to in his debates with the Copenhagen opposition. Much later, (in his [1952]) Schrödinger notes that, given the quantisation of the storage of energy by matter, and of electromagnetic radiation itself, it was a 'really quite obvious deduction' that 'with a particle mass m which, according to Einstein has an energy mc^2 (where c = the velocity of light), there must be associated a wave process of frequency mc^2/h', a deduction first drawn by Louis de Broglie in 1925. To this de Broglie frequency corresponds the de Broglie wavelength, $\lambda_{d.b.}$. Furthermore, the theoretical "electron waves" were experimentally demonstrated only a few years later, in diffraction and interference experiments. He continues: 'This was the point of departure for the early recognition ... that everything—*really everything*—is both particle and wave field' (Schrödinger [1952]).

In his [1928b], Schrödinger presents a number of arguments for this realistic and non-probabilistic interpretation of the wavefunction ψ. Although by no means always convincing, they do underline the importance of the wave picture in the construction of Schrödinger's version of quantum mechanics, and his continuing adherence to a realistic interpretation. MacKinnon [1980] argues that this interpretation-lead derivation was a mere *reconstruction* on Schrödinger's part, historically more important being the first derivation via the extremum principle as presented in the first of the wave mechanics paper (Schrödinger [1926b]). There are, however, a number of reasons to think that the Hamilton analogy *was* an integral part of the theory's genesis: Wessels [1980] sets out the case in a reply to MacKinnon's paper, the above historical account underlining this. In any case, the standing wave model is mentioned even in the supposedly interpretationally neutral paper (Schrödinger [1926b]), where Schrödinger argues that his theory *explains* the Bohr-Sommerfeld quantisation rules—unexplained in the old quantum theory—much as the presence of a standing wave explains the discrete modes of vibration in a string. The [1928b] presentation is the most elegant, beginning with an analogy:

As:	Geometrical Optics	is to	Undulatory Optics
so:	Ordinary Mechanics	is to	Wave Mechanics

The analogical argument runs as follows: When Huyghens' undulatory optics replaced Newton's corpuscular theory, it was realised that geometrical optics, with its definite rays of light, approximates to the undulatory theory. The correspondence between the two theories can be illustrated by comparing motion in the two theories, as Schrödinger does in his [1952], following Debye's deduction of ray optics from wave optics and Hamilton's optical-mechanical analogy. Perpendicular to the wave surface of a moving wave (i.e. in the direction of motion) are the wave normals, which correspond to the "rays" of geometrical optics. The motion along the normal of a small part of the wave front—the rest of which is damped—approximates the motion of a corpuscle along the path defined by the ray, with the same velocity as the spreading of the corresponding wave front. The geometrical description holds only when the optical phenomena under consideration are coarse in comparison with the wavelength of the relevant light. It breaks down, however, for 'minute' optical phenomena. It was the failure of the corpuscularist picture in these cases that resulted in light being regarded as *constituted by waves.*

Schrödinger then considers the correspondence between classical and quantum mechanics. The classical equations of motion break down for 'minute' phenomena, although adequate for 'macro-mechanical' phenomena. Empirical problems for the classical theory occur in just those situations where the associated (de Broglie) wavelength becomes a non-negligible quantity. The analogy between the failure in the two cases of the corpuscular picture suggested a second order differential equation, used in classical physics to describe wave motion, for the motion of (for instance) electrons. The solution to this equation—the wavefunction ψ—would represent the motion of the system, capturing its undulatory nature. The rest (as they say) was history: 'we are led to describe what really happens in such a system by a wave motion in the generalised space of its coordinates (q-space).' (Schrödinger [1928b], p.160)

In a footnote in his [1926d], Schrödinger paid tribute to the influence of Einstein and de Broglie:

> My theory was inspired by L. de Broglie, ..., and by brief, yet infinitely far-seeing remarks of A. Einstein (Schrödinger [1926d], p.46)

Raman and Forman [1969] have traced the reasons why it was Schrödinger—rather than anyone else—who developed de Broglie's ideas into a mechanics of matter waves. They note three relevant factors. Firstly, Schrödinger appreciated the relativistic framing of de Broglie's theory, which would have been congenial to the author of Schrödinger's [1922]. Secondly, he was outside the Copenhagen-Göttingen circle of physicists influenced by Bohr and Sommerfeld, who had two reasons to be suspicious of de Broglie's work: its naïve faith in definite electronic orbits, and de Broglie's poor personal reputation. This infamy was partly the result of an acrimonious priority dispute between de Broglie's collaborator Dauvillier and the Copenhagen group over the discovery of element 72,[1] and partly the result of de Broglie's temerity in interpreting the correspondence principle in a fashion unacceptable to its creator (see Raman and Forman [1969], pp.294-5). A third factor would have been the reference to de Broglie's work by Einstein—a major influence on Schrödinger. There is direct evidence for the importance of both the first and last factors

[1] The Copenhagen claim won out: the element was named Hafnium, for Copenhagen.

in Schrödinger's letter to Einstein of 3 November 1925 (see above), written at the time that Schrödinger was first working on matter waves.

However, another letter to Einstein, this time of 23 April 1926 seems to deny the "gauge factor" theory any importance:

> the whole thing [i.e. wave mechanics] would certainly not yet, and perhaps never, have been developed (I mean not by me) if the importance of de Broglie's ideas had not been put right ùnder my nose by your second paper on gas degeneracy. (Schrödinger, letter to Einstein, quoted in Raman and Forman [1969], p.311)

The three different factors appear to invoke different conceptual routes into wave mechanics. The first two suggest that Schrödinger, someone who was interested in atomic structure and spectroscopy, was uniquely suited—by his lack of intellectual antipathy—to build on de Broglie's work. Thus it was '*qua* theoretical spectroscopist' that Schrödinger developed wave mechanics. The third factor, however, implies that wave mechanics was an attempt to provide a foundation for the theory of matter waves that Schrödinger thought was a necessary motivation for the work on gas theory. In stressing the importance of the 1922 paper—i.e. the *first* factor—Raman and Forman were correcting Klein's [1964] emphasis on the gas theory work as the precursor to wave mechanics. Hanle [1977b] supports Raman and Forman's reading over Klein's, on the evidence of the letter of 3 November 1925 (see above). Since there is evidence for *both* historical accounts, perhaps they both contain an element of truth. Schrödinger was caused to read de Broglie's thesis because of his work on gas statistics, and Einstein's consequent influence. However, what he saw *of value* in the thesis—the possibility for a wave theory of the atom—was determined by both the "Remarkable Property" paper, *and* the need for a physical foundation for the Bose-Einstein statistics. As often with major scientific advances, wave mechanics was the fruit of many different conceptual strands.

Common to all the possible routes to wave mechanics, however, is the indispensable role that physical intuition played in guiding Schrödinger's mathematical work. Wessels [1977] has charted the importance of Schrödinger's thoroughgoing realism on his attitude to contemporary scientific developments. For instance, in a letter to Bohr of May 1924, he commented on the recently published Bohr-Kramers-Slater theory:

> I cannot completely go along with you when you keep calling [this radiation] 'virtual', ... For what is 'real' radiation if it is not that which causes transitions, i.e. that which creates the transition probabilities? ... one might even venture to wonder which of the two electronic systems has a greater reality—the 'real one' which describes the stationary states or the 'virtual one' that supplies the virtual

> radiation and scatters impinging virtual radiation. (Schrödinger, letter to Bohr, quoted in Wessels [1977], p.313)

In a published paper on the subject, Schrödinger interpreted the theory realistically, rather than via the notion of "virtual" radiation.

The old question of what medium the wave motion was a fluctuation *in* remained, however, and Schrödinger had no clear answer. He did have a number of rather murky arguments for his realistic interpretation of the ψ-function. One of these was that quantum mechanics gave the wrong results if interchange between particles of the same type in a system was neglected, leading one to the conclusion that such particles could not be unambiguously assigned individual identities (a position Schrödinger had earlier taken in his work on gas statistics). The identity of a particular wave train, in contrast, was conceivable and sensible within quantum mechanics. A second argument was directed at the probabilistic interpretation, which he pointed out raised the question of what types of events $|\psi|^2$ represented probabilities *of*. The standard answer was, of course, that from the ψ-function one could calculate the probability of finding the system in one of a set of allowed *classical* states as the outcome of a measurement. This Schrödinger later found curious: 'Is it not rather bold to interpret measurements according to a picture which we know to be wrong?' (Schrödinger [1952]) A further perceived advantage of the wave interpretation was its avoidance of such 'irrational' features of the theory as 'quantum jumps' which could be given no mechanical description. In wave mechanics, the discreteness embodied in the quantum jumps were readily explained.

In summary, it seems that Schrödinger both constructed and interpreted his wave mechanics *realistically*, in terms of his (admittedly sketchy) wave/field picture. Not only was this interpretation important heuristically, but it continued to be Schrödinger's intended interpretation, notwithstanding the difficulties in its reconciliation with the apparently discrete experience of everyday life. Schrödinger also recognised that *if* the Copenhagen interpretation *were* correct, and quantum mechanics better understood formalistically, then the wave picture would be just misleading, despite the formal equivalence of its associated formal theory. However, characteristic of Schrödinger's route to wave mechanics was his concern with more than mere empirical adequacy. He was fully aware of the importance of the intended interpretation which turned a set of equations into a physical theory:

> Physics does not only consist of atomic research, science does not only consist of physics, and life does not only consist of science. The aim of atomic research is to fit our empirical knowledge concerning it into our other thinking. All of this other thinking, so far as it concerns the outer world, is active in space and time. If it cannot be fitted into space and time, then it fails in its whole aim and

> one does not know what purpose it really serves. (Schrödinger, letter to Wien of 25 August 1926, quoted in Moore [1989], p.226)

4.4. The Two Were Not the Same

In the early months of 1926, there appeared a number of proofs which were taken at the time to show that matrix and wave mechanics could be identified. This might have been unexpected at the time, given the backgrounds of the two theories. Jammer echoes this sentiment:

> Heisenberg's was a mathematical calculus, involving non-commutative quantities and computation rules, rarely encountered before, which defied any pictorial interpretation; it was an *algebraic* approach which, proceeding from the observed discreteness of spectral lines, emphasized the element of *discontinuity*; in spite of its renunciation of classical description in space and time it was ultimately a theory whose basic conception was the *corpuscle*. Schrödinger's, in contrast, was based on the familiar apparatus of differential equations, akin to the classical mechanics of fluids and suggestive of an easily visualizable representation; it was an *analytical* approach which, proceeding from a generalization of the classical laws of motion, stressed the element of *continuity*; and as its name indicates, it was a theory whose basic conception was the *wave*. (Jammer [1966], pp.271-2)

Jammer's comments invite a question: how is it that two theories with such disparate conceptual origins could have turned out to be so intimately connected? We have seen that they were distinct, at least from the point of view of their construction and the intended interpretations of their authors. Proofs of their equivalence were published by Schrödinger [1926d], Eckart [1926], and presented in a letter from Pauli to Jordan (Pauli [1926]), written *before* the publication of Schrödinger's proof. So did the proofs work, and what *did* they prove?

I will follow Pauli's version, which is by far the most elegant. Let $\psi_1, \psi_2, ..., \psi_n, ...$ be a complete orthonormal set of functions satisfying a one-dimensional Schrödinger equation, so that:

$$\int_{-\infty}^{\infty} \psi_n \psi_m \, dx = \delta_{nm},$$

where

$$\delta_{nm} = \begin{cases} 1 \text{ for } n = m \\ 0 \text{ for } n \neq m \end{cases}$$

Pauli leaves out the complex conjugates ψ^*_n because in the one dimensional case, the eigenfunctions are single and real. Given that the set is *complete*, any arbitrary function of x can be expressed as a linear combination of the ψ_n. Consider in particular

$$x\psi_n(x) = \Sigma_m \, x_{nm}\psi_m(x),$$

so that
$$x_{nm} = \int_{-\infty}^{\infty} x\psi_n\psi_m \,\mathrm{dx}.$$

With the comment 'One also puts', Pauli *stipulates* that

$$(p_x)_{nm} = iK\int_{-\infty}^{\infty} \frac{\delta\psi_n}{\delta x}\,\psi_m \,\mathrm{dx}.$$

so that
$$iK\frac{\delta\psi_n}{\delta x} = \Sigma_m \,(p_x)_{nm}\psi_m(x). \qquad (\mathrm{K} = \mathrm{h}/2\pi)$$

Observing that '$x_{nm} = x_{mn}$ is real, $(p_x)_{nm} = (p_x)_{mn}$ is purely imaginary', Pauli states:

> It can be shown without difficulty that the matrices for x and p_x thus defined satisfy the equations of the Göttingen Mechanics. (Pauli [1926], p.281)

In other words if **x** and $\mathbf{p}_x$ are matrices formed from elements x_{nm} and $(p_x)_{nm}$ respectively, they will satisfy the matrix-mechanical equations of motion:

$$\mathbf{p}_x\mathbf{x} - \mathbf{x}\mathbf{p}_x = -iK$$

and
$$\frac{1}{2m}\,\mathbf{p}_x{}^2 + \mathbf{V}(x) = \mathbf{E},$$

where **E** is a diagonal matrix representing the total (non-relativistic) energy, and $\mathbf{V}(x)$ represents the potential energy for the system. Pauli concludes by noting that the rule of multiplication of matrix mechanics implies that for *any* function $F(x)$, the corresponding matrix is given by

$$F_{nm} = \int_{-\infty}^{\infty} F(x)\psi_n\psi_m \,\mathrm{dx}.$$

Thus the proof is general: from the solution to the wave equation can be constructed quantities that describe *any* physical system—regardless of the form of the potential—

which correspond to a matrix that will obey the equations of motion of the 'Göttingen Mechanics'.

Now Hanson [1961] has argued that Schrödinger's and Eckart's published proofs—which are very similar to Pauli's—do not in fact show that matrix and wave mechanics can be *identified*, even though that is how they were read at the time. Hanson's argument turns on what is meant by the phrase 'physical theory':

> A physical theory is, at least, a contingently interpreted formalism—a delicate trinity of algorithm, physical interpretation, and correspondence rules. ([1961], p.405)

Consider first the *formalisms* of matrix and wave mechanics—the algorithms in Hanson's terminology. Hanson argues that the most that could be claimed for the proofs is that they demonstrated an *analogy* between the two formalisms. Consider first Pauli's stipulation that

$$(*) \qquad (p_x)_{nm} = iK \int_{-\infty}^{\infty} \frac{\delta\psi_n}{\delta x} \psi_m \, \mathrm{d}x \, .$$

We have seen that Pauli gave no rationale for this assumption, and *could not* have done. On the left-hand side is a term which characterises the *momentum matrix*, a mathematical object which is basic to the matrix mechanics of Heisenberg, Born and Jordan. On the right-hand side, there appear wavefunctions ψ_m and ψ_n, the *wavefunction* being the central theoretical tool of Schrödinger's wave mechanics. The equation cannot be a *definition*, because the ψ_n and the $(p_x)_{nm}$ have already been defined in their respective calculi. Nor can it be a consequence of either *published* theory: wavefunctions are as absent from the papers on matrix mechanics as matrices are from Schrödinger's papers. Hanson ([1961], p.418) makes the same point with respect to the analogous assumption for position:

$$(**) \qquad x_{nm} = \int_{-\infty}^{\infty} x\psi_n\psi_m \, \mathrm{d}x$$

Perhaps (*) and (**) are *translation rules*, from which it follows that the proof establishes the *intertranslatability* of matrix mechanics and wave mechanics, rather than their *equivalence*. However, (*) and (**) must have been *natural* assumptions to make: both Schrödinger and Eckart, as well as Pauli, made these assumptions for the purposes of their equivalence proofs. So if matrix-mechanical system representatives are translatable into wave-mechanical analogues, what can be inferred about the relationship between the two *physical* theories of which the algorithms were the formal expression? Even if the two *formal* theories were intertranslatable, the two *physical* theories might be distinguished by their *physical interpretation*, if theories are *interpreted* mathematical structures.

That Heisenberg and Schrödinger entertained different physical interpretations is evident from 4.2 and 4.3: the differences continued after the proof of the equivalence of the 'rival' theories. Heisenberg commented that he found wave mechanics 'disgusting' (Heisenberg: letter to Pauli, quoted in Jammer [1966], p.272). Schrödinger was 'discouraged, if not repelled, by what appeared to me as very difficult methods of transcendental algebra, and by the want of perspicuity' of the matrix formalism (Schrödinger [1926d], p.46). If translated into the first-order predicate calculus, Heisenberg's theory would quantify over the discrete states of particles, providing calculation rules for transitions between them. Schrödinger's version, in contrast, speaks of the contingently-quantised motions of wave-groups. So did these interpretations do any scientific work, or did Schrödinger and Heisenberg disagree over an extra-scientific matter? For the moment let the *scientific* be assimilated to the (in principle) *decidable-by-observation.* The question then becomes: could the two theories have been distinguished *observationally*? An answer requires some stage-setting. Schrödinger claimed his proof to be general:

> I will first show how to each function of the position- and momentum-co-ordinates there may be related a matrix in such a manner, that these matrices, *in every case, satisfy* the formal calculating rules of Born and Heisenberg ... This relation of matrices to functions is *general*; it takes no account of the *special* mechanical system considered, but is the same for all mechanical systems. (In other words: the particular Hamiltonian function does not enter into the connecting law.) (Schrödinger [1926d], p.46)

Hanson, however, makes a curious claim: that in fact the proof was *inductive* in character, in the sense that Schrödinger proceeded by showing—problem by problem—that matrix and wave mechanics give the same results:

> A proof showing [two different calculi] C_1 and C_2 mathematically identical would consist not only in the provision of logical transformations converting any selected statement of C_1 into a corresponding statement of C_2. It would also establish that for all *possible* statements within C_1 there is a transformation into a corresponding statement of C_2. Do Eckart and Schrödinger provide such a proof? I think not. What they do is select 'typical' or 'paradigmatic' types of microphysical problems (for example, the harmonic oscillator, Compton scattering, Doublet atoms, etc.), and show that, *via* the operator calculus, every wave-mechanical formulation and solution of these problems has a matrix-mechanical analogue. (Hanson [1961], pp.422-3)

Although correct for Eckart's proof Hanson's claim is strictly false: Schrödinger's and Pauli's proofs gave a *rule* that shows how to construct a matrix corresponding to *any* physical quantity that can be expressed as a function of position and momentum, given an *arbitrary* wavefunction—and therefore the eigenfunction of *any* Hamiltonian. In another sense Hanson has a point, because what the proof shows is how to construct *a* matrix. It does not prove that the matrix so constructed is *the* matrix that would result from the application of matrix mechanics *qua* physical theory. What (*) and (**) provide is an *identification* of wave- and matrix-mechanical terms that are antecedently defined. They do not establish that if wave- and matrix-mechanical descriptions are written down to describe a physical system s, the resultant equations will coincide under the translation scheme defined by (*) and (**). For particular systems considered by Schrödinger and Eckart, this was found to be the case. Suppose we start with an equation $p^{wm}(s)$ that encodes a wave-mechanical description of s. Suppose the *translation* of $p^{wm}(s)$ is $tr(p^{wm}(s))$, an equation of matrix mechanics. Now suppose the 'natural' application of matrix mechanics to s yields $p^{mm}(s)$. If the equivalence of the two theories is to follow from the mathematical intertranslatability of their associated calculi, we would need, *as well as* the translation scheme:

(†) $\forall s(tr(p^{wm}(s)) \Leftrightarrow p^{mm}(s))$ and $\forall s(tr(p^{mm}(s)) \Leftrightarrow p^{wm}(s))$

So how might (†) be argued for? The 'natural' description of a system for a given calculus is fixed by the other parts of Hanson's 'delicate trinity'—the correspondence rules and physical interpretation. Now interpretations are difficult things to *prove* anything about, but *if* the details of both wave- and matrix-mechanical descriptions of physical systems are provided by equivalent translations from classical mechanics, *then* (†) can be assumed to hold. There are reasons to think that the antecedent of this last conditional is false.

Both Hanson and van der Waerden take Schrödinger's comments in the equivalence paper to suggest that Schrödinger thought that his proof would allow the *identification* of the two theories. On the face of it, this is borne out by a remark at the beginning of the final section:

> If the two theories—*I might reasonably have used the singular*—should be tenable in the form just given, *i.e.* for more complicated systems as well, then every discussion of the superiority of the one over the other has only an illusory object, in a certain sense. For they are completely equivalent from the mathematical point of view, and it can only be a question of the subordinate point of convenience of calculation. (Schrödinger [1926d], p.57)

However, in a highly significant footnote, he retreats:

> There is a special reason for leaving this question open. The two theories initially take the energy function over from ordinary mechanics. Now in the cases treated the *potential* energy arises from the interaction of particles, of which perhaps *one*, at least, may be regarded in wave mechanics also as forming a point, on account of its great mass. ... We must take into account the possibility that it is

> no longer permissible to take over from ordinary mechanics the statement for the potential energy, if both 'point charges' are really extended states of vibration, which penetrate each other. ([1926d], p.57)

In the body of the text, Schrödinger later argues that

> the validity of the thesis that mathematical and physical equivalence mean the same thing, must itself be qualified. Let us think, for example, of the two expressions for the electrostatic energy of a system of charged conductors, the space integral $\frac{1}{2}\int E^2 d\tau$ and the sum $\frac{1}{2}\sum e_i V_i$ taken over the conductors. The two expressions are completely equivalent in electrostatics; the one may be derived from the other by integration by parts. Nevertheless we intentionally prefer the first and say that *it* correctly localises the energy in space. In the domain of electrostatics, this preference has admittedly no justification. On the contrary, it is due simply to the fact that the first expression remains useful in electrodynamics also, while the second does not. ([1926d], pp.58-9)

Schrödinger's vision for the ψ-function is just such a role:

> the mechanical field scalar (which I denote by ψ) is perfectly capable of entering into the unchanged Maxwell-Lorentz equations between the electro-magnetic field vectors, as the 'source' of the latter. ([1926d], p.60)

Here is a reading of these comments: Schrödinger is retracting either the equivalence of the two theories, or their joint tenability as they have so far been formulated. The reason for the retraction is that he takes his initial treatment of (for instance) the hydrogen atom to be *approximate*, in that the potential term in the Hamiltonian is based on a Coulomb attraction between point particles. Schrödinger's *intended physical interpretation* of wave mechanics—the picture of particles as wave-crests—implies to him that this description is approximate, and dictates that its form be changed at some later stage in the development of his research programme. Given its roots in the old quantum theory that followed Bohr's 1913 theory, the interpretation of matrix mechanics would, in contrast, seem to dictate that the form of the potential be taken over from classical *point*-mechanics. So in 1926, a heuristic difference has opened up between the two theories: it is not yet expressed in the form of the fundamental equations, but waits in their interpretation. As well as being central to the construction of the two calculi, the intended physical interpretations place constraints on how the presently-approximate descriptions of physical systems are to be made more accurate in future: if the intended interpretations encode the futures of the research programmes in this way, they should not be dismissed as idle metaphysics.

CONCLUSION

A serious objection to the methodological instrumentalist account of Heisenberg's route to matrix mechanics is the millennial reception that even the arch-positivist Pauli gave to the published theory. Heisenberg himself was also aware that the new theory was an *explanatory* breakthrough compared to his earlier work: the last five years of the old quantum theory had been marked by the proliferation of mutually-contradictory *ad hoc* theories each of which, however, was empirically adequate over some well-defined domain. Why was the new mechanics so preferable? Because its *explanations* were better, despite its lack of *Anschaulichkeit*. The lack of easy spatio-temporal models for the new theory was a—possibly temporary—problem only for those realists who insisted on one particular *kind* of realistic interpretation: one that involved the spatio-temporal descriptions appropriate to ordinary experience.

Both Schrödinger and Heisenberg followed typically realist patterns of reasoning when constructing their respective formalisms, by (i) applying constraints based on theories in other domains in their constructions of their nascent theories; (ii) carrying over interpreted theoretical structure from previous theories of atomic structure and of spectroscopy, using correspondence arguments couched at the theoretical level; and (iii) declining to be satisfied with the piecemeal—but empirically adequate—theories embodied in the old quantum theory. They later compounded this with what—to the instrumentalist—is an inexplicable attachment to the interpretations that either inspired their work or arose naturally from the context of the metaphysical and epistemological views they held at the time. Whether or not such conduct was reasonable, my case is that this is what Schrödinger and Heisenberg did, and that had they done anything else, there is no reason to think that quantum mechanics would have appeared in the form that it did. If the Copenhagen interpretation *did* finally triumph, this was not the triumph of instrumentalism over realism, but a victory for one transcendent philosophical superstructure over another.

5

APPROXIMATIONS IN QUANTUM CHEMISTRY

INTRODUCTION

Where smaller entities are to be found among the remains of larger ones, it has often been inferred that good theories of the former should explain good theories of the latter. So when successful theories of the sub-atomic realm began to appear early this century, it seemed natural to expect that out of them would emerge deeper insights into the structure and interactions of molecules. After all, molecules are made of atoms, and atoms of electrons and nuclei. Many of the equations produced by this application of quantum theory to molecules turned out to be insoluble, but methods for their approximate solution were developed by quantum chemists. Philosophical expectations that reductive explanation would be a straightforward deductive affair were protected with the claim that the soluble approximate equations served merely to elucidate the content of the as-yet inscrutable exact ones: "in principle", the approximate proxies introduce nothing new. This assumption has been the subject of critical attention from a number of quarters. Critics have argued that the "approximations" do *not* invoke approximate versions of the exact equations, but introduce new assumptions about molecules that may be incompatible with the reducing theory, quantum mechanics. The chief task for this chapter will be to examine the methodological arguments that underpin these critiques, in order to see whether there is anything in these problems that is peculiar to quantum chemistry.

Therefore in 5.1, I look at the motivation of the traditional view of explanation by high-level physical theories, and the approximate theories that take their place when the *deductive* explanations invoked by philosophers fail to appear. In 5.2, some critiques of these approximate theories are surveyed. 5.3 tries to lay bare the core of methodological requirements—the notions of 'good' and 'bad' explanations—that run through these critiques. Then in 5.4, it is argued that some of these critiques get the methodology the wrong way round: motivation for the approximate theories arises from the local context of the application—specific knowledge of specific molecules—rather than the general principles of quantum mechanics. A methodological realist lesson is drawn.

5.1. Quantum Chemistry in an Ideal World

In this section, I will very briefly outline the standard model of how abstract theories are applied to the explanation of less general facts. Along the way, I will note the conditions that realists might place on explanation, if explanation is to be more than a linguistic exercise. What this view entails for the explanatory relations between physical theories and chemical facts will then be drawn out. The 'standard model' of scientific explanation is the deductive-nomological or covering-law model due to Hempel [1965], comprehensively criticised by Cartwright [1983]. According to Hempel's model, the explanation of facts about a particular system by a more general theory draws on two types of statement: *internal* principles and *bridge* principles. Internal principles (like Newton's laws or the Schrödinger equation) are general statements which are tied to the specific case at hand by bridge principles, which supply an *enrichment* model in the sense of Redhead [1980]. In any explanation, we want the explanans to show why the explanandum is true. One way to ensure this connection is for the derivation to be *logically sound*, so that truth flows from the true explanans along truth-preserving channels to the explanandum. Now we cannot be sure of the truth of the general laws, in fact realists often want to argue for the truth of a theory from the number and variety of facts it can explain, making explanation and theory-testing two aspects of the same procedure. That the connection between explanans and explanandum is truth-preserving is therefore crucial, and a good explanation must *at least* be a valid deduction in which the internal principles are the major premises, the bridge principles the minor premises, and the explanandum the conclusion.[1]

For realists, the aim of explanation is to understand salient features of a system's behaviour on the basis of a physical theory *coherently applied to it*. By 'coherently applied', I mean that properties and relations are attributed to the constituent parts of the system—via the bridge principles—in a consistent, rather than *ad hoc*, manner in line with the putative explanatory theory. The motions of electrons, for instance, should be subject to the same general constraints wherever they occur. The complex of theory and bridge principle will determine the state of the system and its evolution. If the theory is *not* coherently applied in this sense, the enriched explanatory model might not be consistent with the standard interpretation of the reducing theory. This would amount to a theoretical difficulty, in need of either 'quarantining' (Lakatos [1970]) or instrumentalist interpretation (Feyerabend [1964]).[2] This account of explanation is, of course, inimical to instrumentalism, for which

[1] Leaving aside the complications associated with statistical explanation.

[2] For this Feyerabend, "one should not rest content with a theory which at most admits of an instrumental interpretation but becomes false when interpreted realistically" ([1964], p.190).

explanation and intertheoretic reduction are matters of administrative convenience in the business of saving the phenomena. It is an old realist accusation that the Ptolemaic system 'explained' planetary motion under the instrumentalist criterion of explanation, but became creakingly improbable when its epicycles were interpreted realistically as crystalline spheres. Instrumentalism, realists argue, over-emphasises the quantitative aspects of explanation at the expense of qualitative factors. A *methodological* commitment to realism requires more of explanation than calculation: explanatory models in different areas should make mutual sense against the background of theory, and so must be amenable to the same realistic interpretation. Primas [1975] calls this a requirement of *interpretive* connection, which makes clear why it is not usually articulated separately from the requirement of valid deducibility: the soundness theorem for first order logic assures us of the interpretive compatibility of a statement with its logical consequences.

If we wish to explain some chemical facts, we enter a two-stage process of theoretical description (Cartwright [1983] provides a detailed discussion of what she calls *theory-entry*). First we decide on a set of bridge principles which describe the system under study within the general theory. We then write down the equations yielded by the theory for this description. In quantum chemistry, this is the familiar process of enumerating the particles in the molecule, and writing down the Hamiltonian in terms of their charges, masses and any incident potentials. Textbooks of quantum chemistry typically consider only Hamiltonians containing Coulombic potential terms (see for instance Atkins [1983], or Szabo and Ostlund [1982]). It should be noted that a certain amount of idealisation has already taken place, because we have written down a Hamiltonian that we *know* is incorrect for any real system: relativistic effects have been ignored, and the usual Hamiltonians are relevant only to isolated molecules, of which there are none in the real world.

The second stage should be the unfolding of the logical consequences of the application of the theory to this description, beginning with the solution of the Schrödinger equation. However, further idealisation becomes necessary if it is insoluble. Useful idealisations might falsify the interactions between subsystems of a composite system. In molecular calculations, nuclear and electronic motions are almost universally separated in this way through the *adiabatic approximation*. The Born-Oppenheimer approximation compounds this by setting the nuclei instantaneously at rest; electrons move in the resultant nuclear potential. We could then explain structural features of the molecule by calculating the effect of change in nuclear configuration on electronic energy. How is the electronic energy calculated? The Hartree-Fock procedure replaces a multi-electron wavefunction $\Psi(x_1, x_2, ..., x_n)$ with a product $\psi_1(x_1)\psi_2(x_2)\dots\psi_n(x_n)$ of single-electron 'orbitals' representing a system of non-interacting electrons. This distortion can be calculated away using increasingly-sophisticated mathematical devices to model the electronic interactions.

A different approach is the direct quantisation of the motion of subsystems of the molecule. For instance, the action of carbon dioxide as an atmospheric greenhouse gas is explained by associating its infrared absorption with its rotational and vibrational energy levels. Without solving the Schrödinger equation for CO_2, we know that it is a linear triatomic molecule. By analogy with macroscopic systems of three balls attached by springs, we can imagine the types of motion that such a system would exhibit. The molecule is then treated as if it were a coupled system of the quantised oscillators and rigid rotators that appear in any undergraduate quantum mechanics course. Similar explanations appear in the spectroscopy of more complicated organic molecules (see for instance Silverstein, Bassler and Morrill [1981], pp.95-6). The problem is then to derive the molecular structure from the fundamental equations that quantum mechanics gives as the most general descriptions of such systems.

These idealisations would, at first sight, present a serious obstacle to explanation and theory-testing under the standard model. If one of the premises of a deduction is known to be false, there is no reason for the explanans or prediction itself to be true: the theory has neither explained nor been tested. Laymon [1987] puts the problem succinctly: 'If we conceive of our theory T as justifying the material conditional x & $T \supset y$, where y is the theory output, then the fact that x is false relieves T from the responsibility of yielding a true prediction' ([1987], p.210). Arguments such as this have often provided the starting-point for instrumentalist contentions that the truth of abstract theories is irrelevant to their predictive power and scientific worth, and is therefore otiose as a presumed theoretical virtue (Duhem, for instance, takes this line). The task for realists is then to construct appraisal criteria applicable to approximate theories which will nevertheless support the abductive inferences they wish to make.

Leaving aside the difficulties associated with the first stage of idealisation, we will merely note the reasons why idealisations at the second stage are not usually thought to present a problem. When an approximate model is introduced, mathematical or physical arguments are presented that purport to show that it can be mapped on to the 'exact' treatment it replaces by some continuous transformation. The deductive nature of the explanation can be defended if the explanatorily-relevant features of the approximate solution can be shown to be possessed by the exact one.

5.2. Critiques of Approximate Models

Earlier, I outlined the standard view that the idealised models of quantum chemistry are mathematical devices designed to approximate the exact equations embodied in isolated molecular or atomic Hamiltonians. Over the last two or three decades, a relatively small number of articles have appeared in which this interpretation has been questioned. The various critics detect problems in different parts of the body of approximate methods, and for varying reasons. Their suggested solutions are just as diverse, ranging from the revision of classical notions that molecular structure is a property *possessed* by molecules, to the suspension of the application of quantum mechanics in large areas of chemistry. Claverie and Diner [1980], Weininger [1984], and Woolley [1985] provide something approaching surveys of this literature. In what follows, I will give only a brief selective sample of these critical views.

Hans Primas has been one of the longest-serving critics. His views are set out most fully in his [1983]. A more succinct statement is his [1975], on which the following discussion will concentrate. Primas' starting point is the *holism* that is associated with quantum mechanics. Theoretical propositions in physics are typically formulated with *closed* systems in mind, and then applied in modified—perturbed—form to open systems. Even for classical mechanics, background knowledge implies that all systems interact: there *are* no closed systems apart from the universe as a whole. The effect of EPR correlations in quantum mechanics is, however, more dramatic: composite systems must be represented as *single* quantum systems, rather than as networks of coupled subsystems each having a well-defined state. For a given system s composed of two interacting subsystems s_1 and s_2, the state of the composite cannot in general be represented as a simple product of well-defined subsystem states, no matter how weak the interactions:

$$\Phi(s) \neq \Psi(s_1) \otimes \Xi(s_2).$$

For Primas, this holism has some interesting consequences. Firstly, the process of picking out for study a system, and calculating its well-defined quantum state presupposes a separation of the world into system-plus-background. This process—which for Primas is fundamental to the scientific enterprise—therefore necessitates a degree of idealisation: there *are* no exact quantum-mechanical theories[3] that exactly describe real systems. Secondly, if there is only one real closed system, and therefore one system to which one could attribute a well-defined quantum state, 'we need an interpretation of quantum mechanics in which the notation of a world state is conceptually well-defined' ([1975], p.132). Thus 'traditional' epistemic and operational interpretations such as the Copenhagen interpretation, or the von-Neumann-London-Bauer ensemble formulation—which are 'meaningless' when applied to the universe as a whole—are to be rejected in favour of an 'ontic' alternative.

[3] Or at least any that can be written down.

Background knowledge implies universal entanglement, and entanglement implies non-separability: subsystems of the Great System do not *have* quantum states to call their own. It is therefore mistaken to take the central activity of quantum chemistry to be the calculation of the correct quantum states for real molecular systems. Instead, quantum chemistry associates quantum states for model molecular systems with recognised phenomenal *patterns*. Neither the subsystem states nor the phenomenal patterns are given by nature: both are *created*, and essentially interest-dependent. Primas presents an interesting formal characterisation of this process. When we single out a system for study and attempt to calculate its quantum state, we decompose the state space H_w for the world or 'universe of discourse' into a tensor product of system and environment Hilbert spaces H and H_e respectively:

$$H_w = H \otimes H_e.$$

For arbitrary world state $\Phi \in H_w$, there are non-unique expansions:

$$\Phi = \Sigma_j \ \Sigma_k \ c_{jk} \Psi_j \otimes \Xi_k,$$

such that $\Psi_j \in H$ and $\Xi_k \in H_e$. Primas quotes a theorem due to Schmidt that if the coefficients c_{jk} are chosen to be diagonal, and the Ψ_j and Ξ_k are orthonormal, there is (if the c_j are non-degenerate) a *unique* expansion of the form:

$$\Phi = \Sigma_{j=0}^{\infty} \ c_j \Psi_j \otimes \Xi_j,$$

from which it follows that if Φ is normalised:

$$\|\Phi\|^2 = \Sigma_j \ |c_j|^2 = 1$$

and

$$\langle \Psi_j | \Phi \rangle = c_j \, \Xi_j,$$

$$\langle \Xi_j | \Phi \rangle = c_j \, \Psi_j.$$

Now if it is chosen that

$$1 \geq |c_0|^2 \geq |c_1|^2 \geq \ldots \geq 0,$$

then in order to generate an idealised model of a quantum system such as a molecule, the world state Φ is replaced by the particular direct product Φ_0 of system and background states Ψ_0 and Ξ_0:

$$\Phi_0 = \Psi_0 \otimes \Xi_0.$$

Φ_0 then provides the best approximation to the real world-state that can be expressed in product form for this particular decomposition. Now it is obvious that

$$|\langle\Phi|\Phi_0\rangle|^2 = |c_0|^2 \leq 1.$$

The entanglement of all real systems means, however, that $|c_0|^2$ can never reach unity; Primas uses its proximity as a measure of the closeness of model world state Φ_0 to Φ. $|c_0|^2$ also provides a measure of the accuracy with which the exact world state can be treated as a simple product state, that is, the extent to which the system under study is independent of its environment. If $|c_0|^2 \approx 1$, Φ_0 is the *dominant Schmidt state*, and is robust with respect to environmental perturbation. It is, however, 'background-dependent', in that its applicability depends on the particular decomposition of the world state into system and environment, and therefore on our interests. The non-linear process by which the entangled world state is replaced in our calculations by a separable dominant Schmidt state for a system of interest *creates* any properties we attribute to the system on the basis of the representation, with $|c_0|^2$ representing the degree to which the real system could be said to *have* those properties. Molecular structure is one such property: it is an artefact of our separation of the world into molecule-plus-background. Successful quantum chemistry is the construction of robust model states which 'explain' the phenomenal patterns that we read into the structureless quantum world, and label 'molecular'.

For a shorter period—since the 1970s—R.G. Woolley has been publishing articles critical of the 'fundamental dogma of quantum chemistry' (Woolley and Sutcliffe [1977], p.397). The dogma in question is the supposition that the Born-Oppenheimer procedure is a mathematical approximation to the solution of the exact isolated molecular Hamiltonian eigenvalue equation. Woolley and Sutcliffe [1977] present an argument to the effect that Born-Oppenheimer structures do not have the required symmetry properties, which goes as follows. It can be shown that the standard electrostatic Hamiltonian of quantum chemistry commutes with translation and rotation operators **P** and **J**:

$$[\mathbf{H},\mathbf{P}] = [\mathbf{H},\mathbf{J}] = 0.$$

The Born-Oppenheimer procedure picks out a structure in which the nuclei have determinate positions and are at rest. The electrons move in quantum-mechanical molecular orbitals in the field of the nuclei. For the purposes of explaining chemical facts, the most useful such structure will be the most probable one: the equilibrium structure, for which the energy of the system as a whole is a minimum. Momentum-position uncertainty implies that the nuclear positions must be blurred, and adds a zero-point energy Δ_0 to the classical minimum energy E^{cl}. Thus the Born-Oppenheimer energy E^{BO} is given by:

$$E^{\mathrm{BO}} = E^{\mathrm{cl}} + \Delta_0$$

The zero-point energy, Δ_0, may be visualised as corresponding to an 'uncertainty oscillation' about the equilibrium nuclear position. Woolley and Sutcliffe call this model *semi-classical*: a classical structure perturbed by quantum effects. Furthermore, it is of lower symmetry than eigenstates of the isolated-molecule Hamiltonian. Operations on it by **P** and **J** sweep out a six-dimensional hypersurface of constant—albeit blurred—energy (see below for further discussion).

Does this mean that the Born-Oppenheimer procedure is illegitimate? Woolley argues not, and presents a *fundamental discontinuity* between the behaviour of small isolated molecules and large molecules which interact strongly with their environment. In his [1976], Woolley points out that non-adiabatic calculations based on the isolated molecular Hamiltonian are accurate for describing atoms and small molecules in the form of rarefied gases or molecular beams, where the energy levels are probed with high resolution. When care is taken to minimise line-broadening processes such as intermolecular collisions and Doppler effects (see Woolley [1976] and [1977]) and the molecules are illuminated by lasers, the spectra will be very close to those of the isolated molecule Hamiltonian. In such calculations, the notion of molecular structure is not required, and is indeed 'no longer appropriate' ([1976], p.30).

In contrast, when environmental interactions *cannot* be ignored, and for large molecules which are not suitable for preparation in states of high rarefaction, the classical notion of molecular structure—embodied in the Born-Oppenheimer procedure—becomes 'indispensible'. The typically 'chemical' facts whose explanation was considered in 5.1 require determinate *asymmetrical* structures to be attributed to molecules. The history of the investigation and explanation of isomerism and optical activity illustrates that this is *explicitly* the case for some chemical facts.

The above considerations have been an argument to the effect that the properties of molecules have not so far been derived from properties of the full molecular Hamiltonian. Worries were raised that some of the properties of molecules that play central roles in chemical explanations might be arising as artefacts of the approximations. A problem of principle arises when one considers the symmetry properties of the usual electrostatic molecular Hamiltonians. It would appear that some of the explanatorily useful properties are *bound* to vanish as quantum-mechanical exactitude is approached. Woolley [1976] argues that the Hamiltonian for an isolated molecule is fully symmetric, and so, therefore, are its eigenfunctions. One argument for this is as follows: Consider a Hamiltonian **H** for a molecular system with eigenfunctions ψ_n and eigenvalues E_n. Thus:

$$\mathbf{H}\psi_n = E_n\psi_n$$

Now suppose **H** is symmetric with respect to inversion of coordinates. If we consider an operator **R**, which effects such a transformation, we have

$$\mathbf{R}(\mathbf{H}\psi_n) = \mathbf{H}(\mathbf{R}\psi_n)$$

putting these two together:

$$\mathbf{H}(\mathbf{R}\psi_n) = E_n\mathbf{R}\psi_n$$

So if **H** is symmetric, and ψ_n is one of its eigenfunctions, then so is $\mathbf{R}(\psi_n)$ with the same energy eigenvalue. If ψ_n and $\mathbf{R}(\psi_n)$ have some directional property such as a dipole moment, the dipoles will act in opposite directions, but they have the same energy. One could therefore imagine an energy degeneracy among distinct spatial states: any energy eigenstate could be represented as a superposition of different sharp-dipole states. In such a representation, structures with (for instance) equal dipole moments in opposite directions would be equiprobable: ψ_n and $\mathbf{R}(\psi_n)$ would appear *with equal weight* in any expansion of a *general* state in which they appeared. It follows that no such directional properties would be possessed by an isolated molecule in a *general* eigenstate of the Hamiltonian.

Woolley has interpreted these arguments in two quite different ways. The first concentrates on the different *experimental* contexts in which energy eigenvalues and structural properties are measured. Physical concepts such as these are well-defined only for those experimental situations in which quantum systems with sharp values for the corresponding quantities have been prepared. If we cannot prepare molecules with sharp values for both energy and structural quantities, these concepts must be 'complementary in Copenhagen sense' ([1976], p.30). In standard discussions of complementarity between physical quantities, its formal expression is the relation of non-commutivity between the corresponding operators. In his [1976], however, Woolley does not present a proof that the operator for some quantity necessary for the definition of molecular structure fails to commute with the Hamiltonian. In the *second* interpretation with which Woolley glosses his formal arguments, he is much closer to Primas. He takes seriously the lack of classical structure attributed by quantum mechanics to isolated molecules in stationary eigenstates of the Hamiltonian. Molecular structure may be the effect of a *physical* interaction of a molecule with its environment. Molecular properties may therefore be expected to 'disappear abruptly' ([1978], p.1077) as a molecule's interactions with its environment decrease. For Woolley, the future of quantum chemistry lies in the development of rigorous quantum-mechanical (for instance non-adiabatic) treatments of small molecular systems in isolated states, in which molecular structures do not figure.

The aim of Ogilvie's rather more trenchant criticisms (Ogilvie [1990]) is to distinguish between 'what is fundamental and what is artefact' in quantum chemistry ([1990], p.280).

His chief target is the separation of multi-electron wavefunctions into single-electron terms which is basic to the Hartree-Fock procedure. It is not the approximation itself to which he is hostile, but the use of approximate one-electron orbitals in the explanation of chemical facts. The worst such case is the typical account of chemical bonding in terms of hybridisation (the interference of atomic orbitals to form bonding molecular orbitals): the subtitle of Ogilvie [1990] is "There are no such *things* as orbitals". Ogilvie takes the quantum-chemical account of the structure of methane (CH_4) as his main example. It is often claimed in textbooks of chemistry that methane's tetrahedral structure arises from its electronic structure. Eight electrons occupy SP^3-hybridised molecular orbitals, resulting in increased charge density in the space between the carbon nucleus and each of the four hydrogen atoms. This decreases electrostatic repulsion between these nuclei. The remaining two electrons are localised close to the carbon centre, and do not significantly contribute to bonding.

The attack on electronic orbitals has three main prongs. Firstly, the choice of basis orbitals in a molecular orbital expansion is irrelevant to the quality of the final energy values, and is therefore arbitrary and can have no physical meaning. From the success of calculations using atomic orbital bases one therefore *cannot infer* that molecular orbitals are somehow formed from the interactions of atomic orbitals. Secondly, the ten electrons in CH_4 are 'fundamentally identical and indistinguishable' ([1990], p.283). It therefore makes no sense to distinguish qualitatively between them, as their occupancy of single-electron orbitals must.[4] Thirdly, on formal grounds, 'a molecule consists of only electrons and nuclei, certainly not orbitals or even atoms.' ([1990, p.287). Atoms do not exist in molecules *as* atoms: Methane consists of a carbon nucleus, four protons and ten electrons, rather than a carbon atom and four hydrogen atoms. In fact it is a 'category fallacy' ([1990], p.287) to think otherwise. It makes no sense to explain the properties of a system by reference to the properties of entities which are *not* in fact present, so explanations of molecular structure by the interactions of either atoms *or* one-electron orbitals are ruled out.

Historically, Ogilvie argues, molecular structure—and the standard electrostatic accounts of chemical bonding—arose against the background of *classical* physics. There is no reason why these explanatory tools should be transferable into a quantum-mechanical context. According to the arguments set out above, the idealised quantum-mechanical explanations are essentially parasitic upon the classical accounts they are supposed to supersede, because they certainly make no sense against the background of rigorous quantum mechanics. Ogilvie quotes photoelectron studies of methane to argue that the approximate models of

[4] To be fair to the Hartree-Fock method, this artefact *can* be calculated away using exchange integrals.

quantum chemistry are not even empirically adequate; the purported electronic structure is not found. Classical models of molecular structure and bonding may do no worse, and at least have the advantage of conceptual consistency. The valence shell electron pair repulsion theory is cited by Ogilvie as an—admittedly unsatisfactory—starting point.

Scerri [1991] detects conceptual problems in standard quantum-chemical models of the electronic configuration of *atoms*. These difficulties would also infect the accounts of molecular structure in which they are embedded. Scerri points out that these models depend on the *aufbau* principle by which the electronic configurations of successive elements in the periodic table were built up in the quantum theory of atomic structure that developed out of Bohr's atomic model of 1913. Central to the building-up process was the possession by individual electrons of stable stationary states embodied in the attribution of quantum numbers to electrons. Such individual stationary states would of course be perturbed by the addition of further electrons, but the continuity of their existence was a cornerstone of the spectroscopic theory.

The retention of individual electronic stationary states during perturbations of the atom was problematic even within the old quantum theory, claims Scerri, because it relied on the adiabatic principle. This principle, which had been proven to hold for *particular* classes of systems by Ehrenfest, and extended by Burger, had not been shown to hold *generally*. In particular, the principle was not known to hold in aperiodic systems, a class that sadly includes multi-electron atoms.[5] In the context of quantum mechanics, Scerri notes that one-electron angular momentum operators for individual electrons do not commute with the full Hamiltonian, from which it follows that electrons in atoms described by eigenstates of the full Hamiltonian cannot also be in stationary states characterised by the usual quantum numbers. For among the usual quantum numbers are some which correspond to eigenstates of one-electron angular-momentum operators, and non-commuting operators cannot share eigenfunctions. Scerri concludes: 'standard quantum mechanics thus shows that giving electrons individual quantum numbers, or putting them into boxes or orbitals is incorrect', and 'only the atomic system as a whole possesses stationary states' ([1991], p.317). The 'orbital approximation', however, is a useful device for classifying spectroscopic terms, and as a zero-order starting point for more accurate calculations. Scerri concludes by arguing that the approximate orbital models, although indispensible in practice, are

[5] Ehrenfest, following Einstein, in fact called the principle a 'hypothesis' in his [1917]. Bohr, who in his [1918] called the hypothesis the *Principle of Mechanical Transformability*, was aware of its status as an idealisation, and stressed the "limits of its applicability" ([1918], p.102).

'floating' models in the sense of Post [1974]: lacking either theoretical justification or empirical support.

The approximate methods—interpreted realistically—introduce semi-classical rigid structures. The crucial explanatory parts of the model are: the separation the molecular quantum system into the subsystem under study, and the semi-classical balance. When tractable methods are applied to complex molecules, the semi-classical nature of the molecule is *presupposed.* Under the Born-Oppenheimer approximation, the molecular descriptions are more reminiscent of the old quantum theory than of quantum mechanics proper, in their haphazard quantisation of the motion of fast quantum-mechanical electrons perturbed by a framework of classical nuclei. In even the adiabatic approximation, these nuclear configurations are averaged over in an iterated process which hopefully approaches the quantum-mechanical limit. According to Woolley and Primas, there are good reasons to think that some of the structural properties of the approximate solutions that play a crucial role in chemical explanations based on molecular structure are artefacts of the approximations. If this is accepted, molecular structure cannot be said to have arisen as a concept derived from—or as a limiting case of—quantum mechanics applied to molecules.

5.3. Approximation and Explanation

The arguments presented in the previous section share a common structure, and there is a sense in which they express legitimate worries. In this section, I will elucidate this structure. In the final two sections, I will attempt to set out a different conclusion. Consider the application of theory *T* to a physical system *S*. Bridge principles suggest a conjunction of boundary or initial conditions *A*, which invokes a (possibly idealised) model of *S* (call it *M*(*S*)) within *T*. Now suppose that application of *T*'s mathematical apparatus to *M(S)* invokes an equation $E(\alpha)$. The solution, α, describes aspects of the state of *S* relevant to the intended explanatory domain of *T*.

If, as is often the case, $E(\alpha)$ is insoluble directly under present mathematical methods, theorists may overcome the difficulty by:

1. Amending *A*, for instance by constructing a model of a subsystem of *S*, from which salient aspects of *S*'s behaviour may be inferred (an example in quantum chemistry is the calculation of vibronic energy levels for complex molecules by treating certain functional groups *as if* they are isolated quantum-mechanical harmonic oscillators). A *T*-theory of the subsystem follows.

2. Finding some α' such that:

$$\alpha \approx \alpha'$$

This assumption would typically be supported by the construction of a *T*-model of a simpler—but relevantly similar—system *S* ' which under *T* invokes a more readily solved state equation $E'(\alpha')$. There might follow an argument (physical or mathematical) that the difference is not significant. Perhaps a convergent mathematical series might be found for which:

$$\alpha = \alpha_0 + \alpha_1 + ... + \alpha_n + ...$$

i.e. $$\alpha = \Sigma_i \alpha_i$$

or $$\alpha = \lim_{n \to \infty} \sum_{i=0}^{n} \alpha_i .$$

Approximate solution α' would therefore correspond to the sum of a finite initial portion of the series. It is of course possible to mix these approaches, as perturbation theory does. Another possibility is the construction of a *T*-model of the system which is valid only under certain conditions (this is a variant on the first strategy), or over a certain time interval.

Now consider some state of affairs *e* within *T*'s explanatory domain, thought to be attributable to an *S*. Suppose that *e* can be inferred from the mathematical properties of α'. Approximate description α' might correspond to 'simplified' or idealised boundary conditions A_S. The properties of α are unknown (due to the insolubility of equation $E(\alpha)$); *e* has not yet been shown to be derivable from *T* and *A* alone. The logical possibilities are:

a. If, in whichever platonic heaven α can be said to exist, *e* is implied by *T* & *A* (and therefore is implicit in the properties of α) *T* would have explained *e* on the basis of model *M*(*S*). In this case *M*(*S*)—realistically interpreted—represents *S* and evolves in accordance with *T* and its bridge principles. The simplifying assumptions are logically otiose, their only function being to facilitate the calculation of *T*'s consequences with respect to *S*, given *M*(*S*) as the representation of *S* within *T* .

b. If *e* is independent of *T* & *A*, A_S would not be a 'simplifying assumption' at all, but rather an auxiliary assumption which fixes some mathematical property undefined for α (to that of the 'approximate solution' α') to yield *e*. In this case, the simplifying assumptions A_S are an indispensable part of the theoretical complex from which *e* is derived. It is doing a significant amount of the explanatory work. For a methodological realist, whether the complex *T* & A_S yields a coherent picture under a realistic interpretation would then become an issue.

c. If *e*'s falsity can be inferred from T & A, then the theory-model complex is, in some sense, refuted, although the 'simplifying assumption' has somehow overwritten T & A's prediction. I will neglect the possibility that complex T & A_S is inconsistent.

In fact, by assumption, we don't know which of the above three possibilities holds: the insolubility of $E(\alpha)$ by present methods implies our imperfect knowledge of the topology of this area of platonic heaven. The simplifying assumptions are indispensable in practice when constructing a theoretical model whose properties are amenable to mathematical analysis, and it is only *through* such analysis that T's content can be explored. It is theoretical complex T & A_S which is the effective explanans of *e*.

It might be argued that mathematical approximations merely express numerical identities or asymptotic convergence, and the situation described by (a) can be taken to hold. This is unsatisfactory for a number of reasons. Firstly, in one sense it merely imports the instrumentalist 'computation is explanation' thesis, as the following comparison hopefully illustrates. Consider the argument that quantum mechanics 'reduces' to classical mechanics in the limit $h \rightarrow 0$. The dynamical behaviour of a classical theoretical model can be imitated to any desired level of accuracy by setting h equal to a low enough numerical value. However, there is a discontinuity between $h = 0$ and h having an infinitesimal value: the non-zero spacing between adjacent energy levels. The two systems are distinct, though they are not observationally distinguishable: a quantum system can never *be* a classical one, no matter how small the spacing between its energy levels. For those who concentrate on the observational level, however, an infinitesimal difference is effectively no difference at all. So far so unconvincing, but there might be attendant *practical* difficulties. Limit theorems invoked by quantum chemists—even where available—typically concern the relative energies of the approximate and exact solutions. They do not imply that *all* dynamical quantities will smoothly converge to their exact values. In fact it is proverbial among quantum chemists that a wavefunction that gives a good value for energy will usually give *bad* values for other quantities such as charge distribution. Thus one cannot, without further argument, simply identify the exact solutions with limiting cases of increasingly accurate approximate models.

The arguments presented in the previous section, although differing in their conclusions, identified properties of approximate molecular models which could not be attributed to the exact solutions of the Schrödinger equation for an isolated molecule. If they are correct, these models cannot be rationalised as its approximate solutions. Explanations that invoke these properties cannot therefore have the desirable status (outlined in the first section) of deductive inferences from quantum mechanics plus suitable auxiliaries. If a model *does* successfully render some molecule's behaviour accurately, under what conditions can that model be said to have *explained* that system's behaviour? If it is approximate, the coherence of the model invoked by the theory and 'simplifying' assumptions *together* becomes an issue for a realist. This, claim the critics, is just the problem that afflicts quantum chemistry. The exact equations of quantum chemistry are insoluble, and attempts to solve them approximately introduce *ad hoc* assumptions—like the separability of the electronic wavefunction into single-electron molecular orbitals—that are *non-quantum-mechanical.* Worst of all, the rationale for making these assumptions is background knowledge about molecular structure, which is what is supposed to be derived. Consequently, claim the critics, there is an air of methodological scandal.

5.4. APPROXIMATION AND IDEALISED MODELS

It is a commonplace of recent philosophy of science that abstract physical theories do not provide detailed descriptions of the behaviour of interesting systems on their own. Instead, they must be supplemented with some characterisation of the particular facts of the system to be studied of interest to the theory in question. In the covering-law model of explanation, these characterisations are the bridge principles: they originate in that famous catch-all category 'background knowledge'. Together, theory and background determine a model. In her [1983], Nancy Cartwright attempts to inject some realism into our understanding of how abstract physical theories are applied to concrete situations.[6] The covering-law model is, she argues, a 'model for a physics we do not have' ([1983], p.145). When a theoretical account is given of some system's behaviour, we don't just write down a literal description and apply the formalism. Instead, there are two distinct *stages* of theory entry. First imagine a list of everything we know about the system. We cannot go from this to an equation. The description has to be *prepared* for entry into theory: we must pick a description for which the theory has an equation. Now these equations can be tailored to some extent, and the list of them is always growing, but we must realise that they are 'off the peg' rather than bespoke: we fit the facts to the equations, rather than *vice-versa.* How we go from an unprepared description to a prepared one is not dictated by the theory, but there are *good* and *bad* prepared descriptions. The criteria for this assessment are 'rules of thumb, good sense, and, ultimately, the requirement that the equation we end up with must do the job.' ([1983], p.133) The job in question here is a *pragmatic* one, of meeting practical requirements.

[6] By realism, I mean a hard-headed recognition of the actualities of explanation: the account itself is instrumentalist with respect to laws, in that the laws of physics could not be literally true.

The critiques surveyed in the previous two sections proceeded by comparing the models used in quantum chemistry with the solutions to the exact equations they supposedly approximate. The exact solutions were unavailable, of course, but *some* of their features could be discerned with the help of such general considerations as symmetry arguments. We could see that the molecular models were not approximations at all: they differed *qualitatively* from what quantum mechanics says about molecules concerning those features we *can* discern. The critics drew different sorts of conclusion, but the conclusions always bore on the relation between molecular structure as an *explanandum* and quantum mechanics as a putative *explanans*. It is the relevance of the critical arguments to *this* relation that should be qualified in the light of Cartwright's account of how abstract theories are applied to concrete cases. The exact equations did *not* drop fully-formed from the quantum-mechanical formalism: a prepared—fictional—description first had to be given. So the critiques were not comparing the approximate molecular models with what quantum mechanics says about molecules *simpliciter*, but with what quantum mechanics says about molecules *if we pretend molecules are isolated systems of point-mass nuclei and electrons subject only to Coulomb interactions*. This description is *itself* an idealisation: there *are* no isolated molecules; electrons and nuclei experience interactions that are non-Coulombic; nuclei are not without internal structure; and the energies referred to by the Hamiltonian should be relativistic.

So the fact that the approximate molecular models falsify some of the details of real molecules cannot *itself* be a good reason not to use them. The argument now concerns which model Hamiltonians provide the most useful results. An immediate response could be given here: perhaps the molecular models constitute *worse* falsifications than the more exact treatments, maybe by introducing qualitatively different untruths (like the Born-Oppenheimer models) which will further distort our understanding of what molecules are really like. This, however, begs a number of important questions. Firstly, can we order idealisations in this sort of way, so that phrases like 'more exact' have a precise meaning? In his [1987], Laymon presents a formal characterisation of a criterion for the appraisal of theories which are applied to counterfactual initial conditions: *monotonic piecemeal improvability*. This criterion has been argued by Laymon to be operating in a number of historical case-studies (see for instance his [1983]), and underpins his defence of the realist use of the covering-law model of explanation against Cartwright's critique (Laymon, [1989]). In simple form, the reasoning goes as follows. We know that theories are only ever applied to idealised models of real systems, and so we cannot expect them to provide predictions that are precisely true. However, if our idealised descriptions are close to the real systems, the predictions of a good theory will also be close. Thus if we *improve* our models (i.e. make them more realistic) we can expect that the predictions of a *good* theory will also improve. Hence the criterion: a good theory is one whose predictions improve monotonically as the input theoretical descriptions are improved. This account assumes that increased accuracy is necessarily a good thing.

Do more exact treatments necessarily contribute more to our understanding? It depends on what you wish to understand. What we need is a way of deciding when an approximate or idealised theory is good or useful *distinct* from how many lies it tells. Cartwright gives some pragmatic answers: a good approximate theory should have as many of the usual instrumental virtues as is consistent with mathematical tractability. This is why it is a mistake to criticise the standard approximate methods of quantum chemistry by making an unfavourable comparison with the 'exact' equations: being insoluble, they were of *no* help in finding out about molecules. Ramsey [1992] argues that formal accounts like Laymon's miss something important about the construction of idealised theories. They concentrate on the comparison of idealised models to 'exact' descriptions. Ramsey feels that we should also appraise the motivation of such models:

> When praising or blaming an approximate result, it is not sufficient to consider only the magnitude of the discrepancy between the theoretical and experimental results. How we got to the result is just as important as the fact that we got pretty close to where we wanted to go. ([1992], p.162)

I think Ramsey is correct. Woefully inaccurate molecular models may in certain circumstances provide interesting insights, but which circumstances?

Redhead [1980] distinguishes approximation from idealisation. The distinction is not one of principle: Redhead points out that for every approximate solution to an exact equation (an approximation in Redhead's terminology), there is an approximate equation which can be solved exactly (i.e. an idealisation). However, the distinction can be made in practice, and will turn out to be a very useful one. Some approximate methods come with a ready-made physical justification. This specifies a new—and readily interpretable—*model*, which is related to the 'exact' model through the process of idealisation: it is an *impoverishment model* in Redhead's terminology. Now the purely mathematical approximations, being more abstract, will not come with such an intended physical interpretation. Although they will pick out a different set of allowed values in the space of the variables in which they are expressed (see Redhead [1975]), how this 'approximate' set of values differs *physically* from those picked out by the 'exact' theory may not be obvious. In contrast, the *physically* motivated approximate theory will differ *systematically* from its exact partner in ways that can be understood physically, and therefore might allow causal differences between real systems to be inferred. The systematic difference between exact and approximate model might, for instance, be interpreted as a *physical perturbation*.

Del Re [1974] observes that some approximate models are more amenable than others to this type of interpretation. The much-maligned LCAO-MO-CI model is—despite its formal inadequacies—a good basis for chemical explanation, notwithstanding what he calls the 'basis problem' (mentioned also by Ogilvie [1990]). This is because it is readily interpretable in terms of the 'bonds, atoms and simple orbitals' (Del Re [1974], p.95) with which practical chemists work. He has in mind something very similar to the heuristically useful interpretation described above. He also provides ([1974], p.97) an example of the kind of reasoning considered at the end of the last paragraph. When we counterfactually replace a many-electron wavefunction for an organic molecule with a product of individual bond-pair-states, we can envisage separating the pairs into those occupying σ-bonds, and those in π-bonds. If the electrons did not interact, the electronic energy would be the sum of the pair-bond energies. Obviously, molecules in which π-bonds are close together will deviate further from this *sum-energy* than molecules in which they are further apart, because the π-bond electrons will interact more strongly. This is the origin of the (in)famous resonance energy of benzene. Ogilvie singled out this type of explanation for special disapproval. What is the alternative? Insoluble equations have little heuristic value.

Liegener and Del Re [1987] introduce the notion of the 'reverse of reduction': the process whereby the content of the theory which is to be 'reduced' plays a useful role in the elucidation of the theory which is supposed to be doing the reducing. If this is what is going on with models of molecules, it is clear why they such models will offend the expectations embodied in the standard view of scientific explanation: they appear to assume what is supposed to be derived. But if we have explanatorily powerful approximate models that conflict—when interpreted realistically—with 'rigorous' isolated molecule treatments, so much the worse for the latter.

Conclusion

Critics of the approximate methods of quantum chemistry argue that chemical theories such as molecular structure have not been reduced to quantum mechanics. This is true if reduction is deductive, but it is not a state of affairs that is peculiar to quantum chemistry. It was only the standard view of explanation that provided a reason for thinking it would be so. By the same token, Cartwright's numerous examinations of quantum-mechanical models of lasers might suggest that lasers are not quantum-mechanical. Instead, Cartwright took the failure of models of lasers to fit the covering-law template for explanation to be a failure of that template, rather than a failure of the models. By analogy, the arguments presented in 5.2 reflect as much on the structure of explanation in quantum chemistry as on a peculiar irreducibility of chemical facts to physical theory. In any case, there is something funny about the approximate models being criticised from the point of view of the general theory. How can any theory earn support except through its applications? If you chip away at approximate models, at some point you must begin to chip away at the support enjoyed by quantum mechanics *itself* as a theory for molecules.

Conclusion

Methodological realism (MR) is the thesis that some methods of theory construction that are applied successfully in science *presuppose* realism, because they make internal sense *only if* motivated by realist aims, or *if* theories are construed realistically and can be confirmed as true. This thesis, as we saw in 2.3 and 2.4, has been used as the premise of two different kinds of argument for scientific realism (SR). In 2.5, it was argued that SR is independent of MR. Leaving aside the arguments from MR to SR, *and* the status of SR itself, both the realist and the anti-realist can ask whether or not MR is (meta-)inductively supported by the history of science as a plausible description of successful scientific practice. If appraised according to Lakatos' metamethodology (discussed in 1.6 and 1.7), MR would be expected to provide *new insights* into successful episodes in science. By providing those insights, MR would be supported as a description of methods for producing growth in knowledge. But MR's methods are accessible only to realists: practical reason then tells us that realism is an *instrumentally rational* position where growth in knowledge is sought. Growth in knowledge is an uncontroversial aim for both realist and anti-realist, although each will interpret 'knowledge' differently. Thus either can endorse realism as a position for scientists on *pragmatic* grounds, although the realist may propose a particular *explanation* of this pragmatic rationality of realism.

In 2.6, I set out an account of theory construction and development in which the *intended interpretation* of a set of equations systematically enriches them over time. This, it was argued, is a *realist method* in the sense outlined above. The historical case-studies of chapters 3, 4 and 5 were intended to illustrate and support that account of theory construction, and therefore to also to show how realism has been a motivating presupposition in important examples of theorising. Suppose, however that these histories provide too poor an inductive base for this kind of reasoning. Or perhaps that—despite these few cases—scientists too often settle for messy and non-explanatory theories in their creation of predictive knowledge for the methods of 2.6 to provide an accurate description of the use of models in science. Instead of getting into an argument about numbers, the methodological realist could, raise some historical counterfactuals for the methodological instrumentalist. *Some* theories have, after their construction, gone on to provide *long term* frameworks for successful research. They were accepted at the times of their construction for their unity, coherence and explanatory power, and interpreted realistically then and thereafter. Furthermore, there is no reason to think that this progress would have been achieved if scientists had soldiered on with their attempts to make the messy and *ad hoc* predecessors to these theories empirically adequate. Nor is there any reason to think that a *different* set of theories—had *they* been accepted instead—would have inspired this progress. Thus we should be glad that the scientists in question made these choices, never mind what *we* think of the aims, assumptions or inferences that motivated their methods. Anti-realists may see in this a pragmatic rationale for the scientists' adoption of realism in some cases. Realists may see instead an argument for realism: such long-term success would have been a *miracle* unless the presuppositions of the methods in question were correct. Therefore the success would have been a *miracle* unless the theories in question really *did* make some approach to the ontological order. The realist's explanation is, however, outside the argument presented here.

REFERENCES

Asquith, P. and Giere, R. (eds.) [1981]: *PSA 1980* **2**, (East Lansing, Michigan: Philosophy of Science Association)

Atkins, P.W. [1983]: *Molecular Quantum Mechanics*, 2nd Edition (Oxford: Oxford University Press)

Ayer, A.J. (ed.) [1959]: *Logical Positivism* (Glencoe: Free Press)

Benacerraff, P. and Putnam, H. (eds.) [1983]: *Philosophy of Mathematics* (Cambridge: Cambridge University Press)

Bloch, F. [1976]: "Reminiscences of Heisenberg and the Early Days of Quantum Mechanics", *Physics Today*, December 1976, pp. 23-27.

Bohr, N. [1912]: "Memorandum to Rutherford", partially reproduced in Rosenfeld [1963], pp.xxi-xxviii.

Bohr, N. [1913]: "On the Constitution of Atoms and Molecules" *Philosophical Magazine* **26**, I: 1-25, II: 476-502, III: 857-875, (1913). Reprinted in Bohr [1963] (complete), French and Kennedy [1985] and ter Haar [1967] (Part I only).

Bohr, N. [1918]: "On the Quantum Theory of Line Spectra", reprinted in van der Waerden (ed.) [1967], pp.95-138.

Bohr. N. [1963]: *On the Constitution of Atoms and Molecules* Reprint of Bohr [1913], with an introduction by Leon Rosenfeld (New York: Benjamin).

Born, M. [1924]: *Z. Phys.* **26**, p.379, translated as "Quantum Mechanics" in van der Waerden (ed.) [1967], pp.181-198.

Born, M. and Jordan, P. [1925]: *Z. Phys.* **34**, p.858, translated as "On Quantum Mechanics" in van der Waerden (ed.) [1967], pp.277-306.

Born, M., Heisenberg, W. and Jordan, P. [1926]: *Z. Phys.* **35**, p.557, translated as "On Quantum Mechanics" in van der Waerden (ed.) [1967], pp.321-386.

Boyd, R. [1973]: "Realism, Underdetermination and a Causal Theory of Evidence" *Nous* **7**, pp.1-12.

Boyd, R. [1981]: "Scientific Realism and Naturalistic Epistemology" in Asquith and Giere (eds.) [1981], pp.613-62.

Boyd, R. [1984]: "The Current Status of Scientific Realism" in Leplin (ed.) [1984a], pp.41-82.

Boyd, R. [1985]: "Lex Orandi est Lex Credendi" in Churchland and Hooker (eds.) [1985], pp.3-34.

Buck, R.C. and Cohen, R.S. (eds.) [1971]: P.S.A. 1970, *Boston Studies in the Philosophy of Science* **8**, (Dordrecht: Reidel)

Carnap, R. [1932]: "The Elimination of Metaphysics through the Logical Analysis of Language", reprinted in Ayer (ed.) [1959], pp.60-81.

Cartwright, N. [1983]: *How the Laws of Physics Lie* (Oxford: Clarendon)

Cartwright, N. [1989]: *Nature's Capacities and their Measurement* (Oxford: Clarendon)

Cassidy, D. [1979]: "Heisenberg's first core model of the atom: the formation of a profssional style" *Hist. Stud. Phys. Sci.* **10**, pp.123-86.

Cassidy, D. [1991]: *Uncertainty: The Life and Science of Werner Heisenberg* (New York: Freeman)

Churchland, P.M. and Hooker, C.A. (eds.) [1985]: *Images of Science* (Chicago: University of Chicago Press)

Claverie, P. and Diner, S. [1980]: "The Concept of Molecular Structure in Quantum Theory: Interpretation Problems" *Israel J. Chem.* **19**, pp. 54-81.

Cohen, R.S., Feyerabend, P.K., and Wartofsky, M. (eds.) [1976]: *Essays in Memory of Imre Lakatos* (Dordrecht: Reidel)

Del Re, G. [1974]: "Current Problems and Perspectives in the MO-LCAO Theory of Molecules", in P-O Löwdin (ed.) *Advances in Quantum Chemistry* **8**, (New York: Academic Press), pp.95-136.

Duhem, P. [1914]: *La Theorie Physique: Son Objet, Sa Structure* (Second Edition). Page references are to the 1954 translation by P. Wiener as *The Aim and Structure of Physical Theory* (Princeton: Princeton University Press)

Dummett, M. [1963]: "Realism", reprinted in Dummett [1977].

Dummett, M. [1977]: *Truth and Other Enigmas* (London: Duckworth)

Dummett, M. [1975]: "What is a Theory of Meaning (I)?", in Dummett [1993].

Dummett, M. [1976]: "What is a Theory of Meaning (II)?", in Dummett [1993].

Dummett, M. [1983]: *The Seas of Language* (Oxford: Oxford University Press)

Earman, J. (ed.) [1983]: *Minnesota Studies in the Philosophy of Science* **X**, (Minneapolis: University of Minnesota Press)

Eckart, C. [1926]: "Operator Calculus and the Solutions of Quantum Dynamics" *Physical Review* **28**, pp.711-26.

Ehrenfest, P. [1917]: "Adiabatic Invariants and the Theory of Quanta", *Phil. Mag.* **33**, pp.500-13. Reprinted in van der Waerden [1967], pp.79-93.

Einstein, A. [1925]: *Berliner Berichte* (1925), pp.3-14.

Einstein, A. [1951]: "The Advent of the Quantum Theory" *Science* **113**, 82 (1951).

Elkana, Y. [1981]: "A Programmatic Attempt at an Anthropology of Knowledge" E. Mendelsohn and Y. Elkana (eds.) *Sciences and Cultures* (Dordrecht: Reidel)

Feigl, H. and Maxwell, G. (eds.) [1961]: *Current Issues in the Philosophy of Science* (New York: Holt, Rinehart and Wilson).

Feyerabend, P. [1964]: "Realism and Instrumentalism: Comments on the Logic of Factual Support", reprinted in Feyerabend [1981b], pp.176-202.

Feyerabend, P.K. [1970]: "Consolations for the Specialist" in Lakatos and Musgrave (eds.) [1970], pp.197-230, reprinted in Feyerabend [1981c].

Feyerabend, P.K. [1976]: "On the Critique of Scientific Reason" in Howson [1976], pp.309-39. Page numbers refer to the partial reprint as "The Methodology of Scientific Research Programmes".in his [1981c], pp.202-30.

Feyerabend, P.K. [1981a]: "More Clothes from the Emperor's Bargain Basement: A Review of Laudan's *Progress and its Problems*" *Brit. J. Phil. Sci.* **32**, pp.57-71. Page references are to the reprint in Feyerabend [1981b], pp.231-46.

Feyerabend, P.K. [1981b]: *Philosophical Papers Volume I: Realism, Rationalism and Scientific Method* (Cambridge: Cambridge University Press)

Feyerabend, P.K. [1981c]: *Philosophical Papers Volume II: Problems of Empiricism* (Cambridge: Cambridge University Press)

Feyerabend, P.K. [1988]: *Against Method* (Revised Edition) (London: Verso)

Fine, A. [1984]: "The Natural Ontological Attitude" in Leplin (ed.) [1984a], pp.83-107. Reprinted with corrections in Fine [1986], pp.112-35.

Fine, A. [1986]: *The Shaky Game: Einstein, Realism and the Quantum Theory* (Chicago: University of Chicago Press)

French, A.P. & Kennedy, P.J. (eds.) [1985]: *Niels Bohr: A Centenary Volume* (Cambridge: Harvard University Press)

French, P.A., Uehling, T.E. and Wettstein, H.K. (eds.) [1987]: *Realism and Antirealism: Midwest Studies in Philosophy* **XII** (Minneapolis: University of Minnesota Press)

Gavroglu, K., Goudaroulis, Y. and Nicolacopoulos, P. (eds.) [1989]: *Imre Lakatos and Theories of Scientific Change* (Dordrecht: Kluwer)

Giere, R.N. [1984]: *Understanding Scientific Reasoning* Second Edition (New York: Holt, Rinehart and Wilson)

Giere, R.N. [1985a]: "Constructive Realism" in Churchland and Hooker (eds.) [1985].

Giere, R.N. [1985b]: "Philosophy of Science Naturalized" *Philosophy of Science* **52**, pp.331-56.

Giere, R.N. [1988]: *Explaining Science: A Cognitive Approach* (Chicago: University of Chicago Press)

Giere, R.N. [1989]: "Scientific Rationality as Instrumental Rationality" *Stud. Hist. Phil. Sci.* **20**, pp.377-84.

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

Gödel, K. [1947]: "What is Cantor's Continuum Problem?" *American Mathematical Monthly* **54**, pp.515-25, reprinted in expanded form in Benacerraf and Putnam [1983], pp.470-85.

Gunderson, K. (ed.) [1975]: *Language, Mind, and Knowledge: Minnesota Studies in the Philosophy of Science* **7**, (Minneapolis: University of Minnesota Press)

Hacking, I. [1979]: "Imre Lakatos's Philosophy of Science" *Brit. J. Phil. Sci.* **30**, pp.381-402, reprinted in part in Hacking (ed.) [1981].

Hacking, I. (ed.) [1981]: *Scientific Revolutions* (Oxford: Oxford University Press)

Hacking, I. [1983]: *Representing and Intervening* (Cambridge: Cambridge University Press)

Hacking, I. [1984]: "Experimentation and Scientific Realism" in Leplin (ed.) [1984a], pp.154-72.

Hanle, P. [1977a]: "The Coming of Age of Erwin Schrödinger: His Quantum Statistics of Ideal Gases", *Arch. Hist. Exact Sci.* **17** (1977) pp.165-92.

Hanle, P. [1977b]: "Erwin Schrödinger's Reaction to Louis de Broglie's Thesis on the Quantum Theory", *Isis* **68**, pp.606-9.

Hanle, P. [1979]: "The Schrödinger-Einstein Correspondence and the Sources of Wave Mechanics", *Am. J. Phys.* **47**, pp.644-8.

Hanson, N.R. [1961]: "Are Wave Mechanics and Matrix Mechanics Equivalent Theories?" in Feigl and Maxwell (eds.) [1961], pp.401-25.

Heilbron, J. and Kuhn, T. [1969]: "The Genesis of the Bohr Atom" *Historical Studies in the Physical Sciences* **1**, pp.211-290.

Heisenberg, W. [1925]: *Z. Phys.* **33**, p.879, translated as "Quantum Thoeretical Re-Interpretation of Kinematical and Mechanical Relations" in van der Waerden (ed.) [1967], pp.261-76.

Heisenberg, W. [1958]: "Planck's Discovery and the Philosophical Problems of Atomic Physics", reprinted in Heisenberg, Born, Schrödinger, and Auger: *On Modern Physics* (London: Orion, 1961)

Heisenberg, W. [1971]: *Physics and Beyond* (London: Allen and Unwin)

Hempel, C. [1965]: *Aspects of Scientific Explanation* (Glencoe: Free Press)

Hendry, J. [1981]: "Bohr-Kramers-Slater: A Virtual Theory of Virtual Oscillators and its Role in the History of Quantum Mechanics" *Centaurus* **25**, pp.189-221.

Hendry, J. [1984]: *The Creation of Quantum Mechanics and the Bohr-Pauli Dialogue* (Dordrecht: Reidel)

Hesse, M. [1953]: "Models in Physics" *Brit. J. Phil. Sci.* **4**, pp.198-214.

Hesse, M. [1961]: *Forces and Fields* (London: Nelson)

Hesse, M. [1966]: *Models and Analogies in Science* (Indiana: Notre Dame University Press)

Hesse, M. [1980]: "Theory and Value in the Social Sciences" in Hookway and Pettit (eds.) [1980], pp.1-16.

Hesse, M. [1988]: "Socializing Epistemology" in McMullin (ed.) [1988], pp.97-122.

Hollis, M. and Lukes, S. (eds.) [1982]: *Rationality and Relativism* (Oxford: Blackwell)

Hookway, C. and Pettit, P. (eds.) [1980]: *Action and Interpretation: Studies in the Philosophy of the Social Sciences* (Cambridge: Cambridge University Press)

Howson, C. (ed.) [1976]: *Method and Appraisal in the Physical Sciences* (Cambridge: Cambridge University Press)

Howson, C. and Urbach, P. [1989]: *Scientific Reasoning: The Bayesian Approach* (La Salle, Illinois: Open Court)

Hughes, R.I.G. [1990]: "The Bohr Atom, Models, and Realism", *Philosophical Topics* **18**, No.2, pp.71-84.

Jammer, M. [1966]: *The Conceptual Development of Quantum Mechanics* (New York: McGraw-Hill)

Klein, M. [1964]: "Einstein and Wave-Particle Duality", *Nat. Phil.* **3** (1964) pp.3-48.

Koyré, A. [1965]: *Newtonian Studies* (Cambridge: Harvard University Press)

Kramers [1924]: "The Quantum Theory of Dispersion" *Nature* **114**, p.310, reprinted in van der Waerden (ed.) [1967], pp.199-202.

Krips, H. [1987]: *The Metaphysics of Quantum Theory* (Oxford: Clarendon Press)

Kuhn, T.S. [1957]: *The Copernican Revolution* (Cambridge: Harvard University Press)

Kuhn, T.S. [1977]: *The Essential Tension* (Chicago: University of Chicago Press)

Kuhn, T.S. [1980]: "The Halt and the Blind: Philosophy and History of Science" *Brit. J. Phil. Sci.* **31**, pp.181-92.

Ladenburg, R. [1921]: *Z. f. Phys.* **4**, p.451, translated as "The Quantum-Theoretical Interpretation of the Number of Dispersion Electrons" in van der Waerden (ed.) [1967], pp.139-58.

Lakatos, I. [1962]: "Infinite Regress and the Foundations of Mathematics" *Aristotelian Society Supplementary Volume* **36**, pp.155-84, reprinted in Lakatos [1978b].

Lakatos, I. (ed.) [1967a]: *Problems in the Philosophy of Mathematics* (Amsterdam: North Holland)

Lakatos, I. [1967b]: "A Renaissance of Empiricism in the Recent Philosophy of Mathematics?" in Lakatos (ed.) [1967a], pp.199-202, reprinted in Lakatos [1978b].

Lakatos, I. (ed.) [1968a]: *The Problem of Inductive Logic* (Amsterdam: North Holland)

Lakatos, I. [1968b]: "Changes in the Problem of Inductive Logic" in Lakatos (ed.) [1968a], pp.315-417, reprinted in Lakatos [1978b].

Lakatos, I. [1968c]: "Criticism and the Methodology of Scientific Research Programmes" *Proceedings of the Aristotelian Society* **69**, pp.149-86.

Lakatos, I. [1970]: "Falsification and the Methodology of Scientific Research Programmes" in Lakatos and Musgrave (eds.) [1970], pp.91-196, reprinted in Lakatos [1978a].

Lakatos, I. [1971a]: "History of Science and its Rational Reconstructions" in Buck and Cohen (eds.) [1971], pp.91-135, reprinted in Howson (ed.) [1976], Lakatos [1978a] and Hacking (ed.) [1981].

Lakatos, I. [1971b]: "Replies to Critics" in Buck and Cohen (eds.) [1971], pp.174-82.

Lakatos, I. [1974]: "Popper on Demarcation and Induction" in Schilpp (ed.) [1974], pp.241-73.

Lakatos, I. [1978a]: *Philosophical Papers Volume I: The Methodology of Scientific Research Programmes*, edited by J. Worrall and G. Currie (Cambridge: Cambridge University Press)

Lakatos, I. [1978b]: *Philosophical Papers Volume II: Mathematics, Science and Epistemology*, edited by J. Worrall and G. Currie (Cambridge: Cambridge University Press)

Lakatos, I. and Musgrave, A. (eds.) [1970]: *Criticism and the Growth of Knowledge* (Cambridge: Cambridge University Press)

Lakatos, I. and Zahar, E.G. [1976]: "Why did Copernicus's Programme Supersede Ptolemy's?" in Westman (ed.) [1976], pp.354-83.

Laudan, L. [1984]: "A Confutation of Convergent Realism" in Leplin (ed.) [1984a], pp.218-49.

Laymon, R. [1983]: "Newton's Demonstration of Universal Gravitation and Philosophical Theories of Confirmation" in Earman (ed.) [1983], pp.179-99.

Laymon, R. [1987]: "Using Scott Domains to Explicate the Notions of Approximate and Idealised Data" *Philosophy of Science* **54**, pp.194-221.

Laymon, R. [1989]: "Cartwright and the Lying Laws of Physics" *Journal of Philosophy* **86**, pp.353-72.

Leplin, J. (ed.) [1984a]: *Scientific Realism* (Berkeley: University of Califirnia Press)

Leplin, J. [1984b]: "Truth and Scientific Progress" in Leplin (ed.) [1984a], pp.193-217.

Leplin, J. [1986]: "Methodological Realism and Scientific Rationality" *Philosophy of Science* **53**, pp.31-51.

Lepore, E and Loewer, B. [1987]: "A Putnam's Progress" in French, Uehling and Wettstein (eds.) [1987].

Liegener, C. and Del Re, G. [1987]: "The Relation of Chemistry to other Fields of Science: Atomism, Reductionism, and Inversion of Reduction" *Epistemologia* **10**, pp.269-84.

Lipton, P. [1991]: *Inference to the Best Explanation* (London: Routledge)

MacKinnon, E. [1976]: "de Broglie's Thesis: A Critical Retrospective" *Am. J. Phys.* **44**, pp.1047-55.

MacKinnon, E. [1977]: "Heisenberg, Models and the Rise of Matrix Mechanics", *Hist. Stud. Phys. Sci.* **8**, pp.137-88.

MacKinnon, E.[1980]: "The Rise and Fall of the Schrödinger Interpretation" in Suppes (ed.) [1980], pp.1-57.

McMullin, E. [1976]: "The Fertility of Theory and the Unit for Appraisal in Science" in R.S. Cohen *et al.* (eds.) [1976], pp.395-432.

McMullin, E. [1984]: "A Case for Scientific Realism" in Leplin (ed.) [1984a], pp.8-40.

McMullin, E. (ed.) [1988]: *Construction and Constraint: The Shaping of Scientific Rationality* (Notre Dame: University of Notre Dame Press)

Mehra, J. (ed.) [1973]: *The Physicist's Conception of Nature* (Dordrecht: Reidel)

Mellor, D.H. [1968]: "Models and Analogies in Science: Duhem *versus* Campbell?" *Isis* **59**, pp.282-90.

Moore, W. [1989]: *Schrödinger, Life and Thought* (Cambridge: Cambridge University Press)

Musgrave, A [1976]: "Method or Madness?" in R.S. Cohen *et al.* (eds.) [1976], pp.457-91.

Musgrave, A [1978]: "Evidential Support, Falsification, Heuristics and Anarchism" in Radnitzky and Andersson (eds.) [1978], pp.181-201.

Musgrave, A [1989]: "Deductive Heuristics" in Gavroglu *et al.* (eds.) [1989], pp.15-32.

Newton-Smith, W.H. [1981]: *The Rationality of Science* (London: Routledge)

Newton-Smith, W.H. [1982]: "Relativism and the Possibility of Interpretation" in Hollis and Lukes (eds.) [1982], pp.106-122.

Nola, R. (ed.) [1988]: *Relativism and Realism in Science* (Dordrecht: Kluwer)

Ogilvie, J. [1990]: "The Nature of the Chemical Bond—1990" *J. Chem. Ed.* **67**, pp.280-9.

Papineau, D. [1979]: *Theory and meaning* (Oxford: Clarendon Press)

Papineau, D. [1987]: *Reality and Representation* (Oxford: Blackwell)

Papineau, D. [1988]: "Does the Sociology of Science Discredit Science?" in Nola (ed.) [1988], pp.37-57.

Papineau, D. [1989]: "Has Popper Been a Good Thing?" in Gavroglu *et al.* (eds.) [1989], pp.431-40.

Pauli, W. [1926]: Letter to Jordan of 12 April 1926, reprinted and translated in van der Waerden [1973].

Pearce, G. and Maynard, P. (eds.) [1973] *Conceptual Change* (Dordrecht: Reidel)

Pickering, A. [1984]: *Constructing Quarks* (Edinburgh: Edinburgh University Press)

Popper, K.R. [1953]: "Three Views Concerning Human Knowledge". Page references are to the reprint in Popper [1963].

Popper, K.R. [1959]: *The Logic of Scientific Discovery* (London: Hutchinson)

Popper, K.R. [1963]: *Conjectures and Refutations* (London: Routledge and Kegan Paul)

Popper, K.R. [1983]: *Realism and the Aim of Science* (London: Hutchinson)

Post, H.R. [1974]: "Against Ideologies", Inaugaural Lecture, Chelsea College, University of London.

Primas, H. [1975]: "Pattern Recognition in Molecular Quantum Mechanics" *Theor. Chim. Acta* **39**, pp.127-48.

Primas, H. [1983]: *Chemistry, Quantum Mechanics and Reductionism* (Berlin: Springer-Verlag, 1983)

Putnam, H. [1973]: "Explanation and Reference" in Pearce and Maynard (eds.), pp.199-221. Page references are to the reprint in H. Putnam: *Mind Language and Reality* (Cambridge: Cambridge University Press)

Putnam, H. [1975]: "The Meaning of 'Meaning'" in Gunderson (ed.) [1975], pp.131-93.

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

Putnam, H. [1980]: "Models and Reality" *Journal of Symbolic Logic* **45**, pp.464-82. Page refererences are to the reprint in Benacerraf and Putnam (eds.) [1983], pp.421-44.

Putnam, H. [1981]: *Reason, Truth and History* (Cambridge: Cambridge University Press)

Quine, W.v.O. [1953]: "Two Dogmas of Empiricism" in Quine [1980], pp.20-46.

Quine, W.v.O. [1960]: *Word and Object* (Cambridge: M.I.T. Press)

Quine, W.v.O. [1980]: *From a Logical Point of View*, Second Edition, Revised (Cambridge: Harvard University Press)

Radnitzky, G. and Andersson, G. (eds.) [1978]: *Progress and Rationality in Science* (Dordrecht: Reidel)

Raman, V.V. and Forman, P. [1969]: "Why was it Schrödinger who Developed de Broglie's Ideas" *Hist. Stud. Phil. Sci.* **1**, pp.291-314.

Ramsey, J. [1992]: "Towards an Expanded Epistemology for Approximations", in D. Hull, M. Forbes and K. Okruhlik (eds.) *PSA 1992* Volume 1 (East Lansing: PSA, 1992), pp.154-64.

Redhead, M.L.G. [1975]: "Symmetry in Intertheory Relations" *Synthese* **32**, pp.77-112.

Redhead, M.L.G. [1980]: "Models in Physics", *Brit. J. Phil. Sci.* **31**, pp.145-63

Redhead, M.L.G. [1986]: "Novelty and Confirmation" *Brit. J. Phil. Sci.* **37**, pp.115-8.

Redhead, M.L.G. [1987]: *Incompleteness, Nonlocality, and Realism* (Oxford: Clarendon Press)

Rosenfeld, L. [1963]: "Introduction" to Bohr [1963].

Russell, B. [1959]: *My Philosophical Development* (London: Allen and Unwin)

Ryle, G. [1949]: *The Concept of Mind* (London: Hutchinson)

Scerri, E. [1991]: "The Electronic Configuration Model, Quantum Mechanics and Reduction", *Brit. J. Phil. Sci.* **42**, pp.309-25.

Schilpp, P.A. (ed.) [1974]: *The Philosophy of Karl Popper* (La Salle: OpenCourt)

Schrödinger, E. [1922]: *Z. f. Phys.* **12** pp.13-23.

Schrödinger, E. [1926a]: *Phys. Zeit.* **27**, pp.95-101.

Schrödinger, E. [1926b]: *Ann. der Phys.* **79**, pp.361-76, Page references are to the translation as "Quantisation as a Problem of Proper Values" (Part I), in Schrödinger [1928a] and [1982], pp.1-12.

Schrödinger, E. [1926c]: *Ann. der Phys.* **79**, pp.489-527. Page references are to the translation as "Quantisation as a Problem of Proper Values" (Part II), in Schrödinger [1928a] and [1982], pp.13-40.

Schrödinger, E. [1926d]: *Ann. der Phys.* **79**, pp.734-56. Page references are to the translation as "On the Relation between the Quantum Mechanics of Heisenberg, Born, and Jordan, and that of Schrödinger", in Schrödinger [1928a] and [1982], pp.45-61.

Schrödinger, E. [1928a]: *Collected Papers on Wave Mechanics* (London: Blackie, 1928)

Schrödinger, E. [1928b]: *Four Lectures on Wave Mechanics* (London: Blackie, 1928)
Page references are to the reprint in Schrödinger [1982], pp.147-207.

Schrödinger, E. [1935]: *Science and the Human Temperament* (London: Allen and Unwin, 1935)

Schrödinger, E. [1952]: "Our image of matter", reprinted in Heisenberg, Born, Schrödinger, and Auger: *On Modern Physics* (London: Orion, 1961)

Schrödinger, E. [1982]: *Collected Papers on Wave Mechanics*, third (augmented) edition of [1928a] includes reprint of [1928b]

Sidgwick, H. [1877]: *The Methods of Ethics* Second Edition (London: Macmillan)

Silverstein, R.M., Bassler, G.C., and Morrill, T.C. [1981]: *Spectrometric Identification of Organic Compounds* (New York: Wiley, 1981)

Skolem, T. [1922]: "Some Remarks on Axiomatized Set Theory" Reprinted in van Heijenoort (ed.) [1967], pp.290-301.

Slater, J.C. [1924]: "Radiation and Atoms" *Nature* **113**, pp.307-8.

Suppes. P. (ed.) [1980]: *Studies in the Foundations of Quantum Mechanics* (Lansing, Michigan: P.S.A.)

Swinburne, R. (ed.) [1983]: *Space, Time and Causality* (Dordrecht: Reidel)

Szabo, A. and Ostlund, N. [1982]: *Modern Quantum Chemistry* (New York: Macmillan, 1982)

ter Haar, D. [1967]: *The Old Quantum Theory* (Oxford: Pergamon, 1967)

Urbach, P. [1978]: "The Objective Promise of a Research Programme" in Radnitzky and Andersson (eds.) [1978], pp.99-113.

van der Waerden, B.L. (ed.) [1967]: *Sources of Quantum Mechanics* (Amsterdam: North-Holland)

van der Waerden, B.L. [1973]: "From Matrix Mechanics and Wave Mechanics to Unified Quantum Mechanics" in Mehra (ed.) [1973], pp.276-93.

van Fraassen, B.C. [1980]: *The Scientific Image* (Oxford: Oxford University Press, 1980)

van Fraassen, B.C. [1985]: "Empiricism in the Philosophy of Science" in Churchland and Hooker (eds.) [1985], pp.245-308.

van Fraassen, B.C. [1989]: *Laws and Symmetry* (Oxford: Clarendon Press)

van Heijenoort, J. (ed.) [1967]: *From Frege to Gödel* (Cambridge: Harvard University Press, 1967)

Weininger, S. [1984]: "The Molecular Structure Conundrum: Can Classical Chemistry be Reduced to Quantum Chemistry", *J. Chem. Ed.* **61**, pp.939-44.

Wessels, L. [1977]: "Schrödinger's Route to Wave Mechanics", *Stud. Hist. Phil. Sci.* **10**, pp. 311-40.

Wessels, L. [1980]: "The Intellectual Sources of Schrödinger's Interpretation" in Suppes (ed.) [1980], pp.59-76.

Westman, R. (ed.) [1976]: *The Copernican Achievement* (Los Angeles: University of California Press)

Whitt, L.A. [1992]: "Indices of Theory Promise" *Philosophy of Science* **59**, pp.612-34.

Whittaker, E. [1953]: *History of the Theories of Aether and Electricity vol. 2* (London: Nelson, 1953)

Woolley, R. [1976]: "Quantum Theory and Molecular Structure" *Adv. Phys.* **25**, pp. 27-52.

Woolley, R. [1977]: "Must a Molecule have a Shape?" *Chem. Phys. Letts.* **55**, pp.443-6.

Woolley, R. [1978]: "Must a Molecule have a Shape?" *J. Am. Chem. Soc.* **100**, pp.1073-8.

Woolley, R. [1985]: "The Molecular Structure Conundrum" *J. Chem. Ed.* **62**, pp.1082-4.

Woolley, R. and Sutcliffe, B. [1977]: "Molecular Structure and the Born-Oppenheimer Approximation", *Chem. Phys. Letts.* **45**, pp.393-8.

Worrall, J. [1976]: "Thomas Young and the 'Refutation' of Newtonian Optics: a Case Study in the Interaction of Philosophy of Science and History of Science" in Howson (ed.) [1976], pp.107-79.

Worrall, J. [1984]: "An Unreal Image" *Brit. J. Phil. Sci.* **35**, pp.65-80.

Zahar, E.G. [1973]: "Why Did Einstein's Programme Supersede Lorentz's?" *Brit. J. Phil. Sci.* **24**, pp.95-123 and 223-262, reprinted in Howson (ed.) [1976], pp.211-75.

Zahar, E.G. [1978]: "Einstein's Debt to Lorentz: A Reply to Feyerabend and Miller" *Brit. J. Phil. Sci.* **29**, pp.49-60.

Zahar, E.G. [1983]: "Absoluteness and Conspiracy" in Swinburne (ed.) [1983], pp.37-41.

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