Philosophy of Science

June, 1974

THE RATIONALITY OF SCIENTIFIC DISCOVERY PART I: THE TRADITIONAL RATIONALITY PROBLEM*

NICHOLAS MAXWELL†

University of London

The basic task of the essay is to exhibit science as a rational enterprise. I argue that in order to do this we need to change quite fundamentally our whole conception of science.

Today it is rather generally taken for granted that a precondition for science to be rational is that in science we do not make substantial assumptions about the world, or about the phenomena we are investigating, which are held permanently immune from empirical appraisal. According to this standard view, science is rational precisely because science does not make *a priori* metaphysical presuppositions about the world forever preserved from possible empirical refutation. It is of course accepted that an individual scientist, developing a new theory, may well be influenced by his own metaphysical presuppositions. In addition, it is acknowledged that a successful scientific theory, within the context of a particular research program, may be protected for a while from refutation, thus acquiring a kind of temporary metaphysical status, as long as the program continues to be empirically progressive. All such views unite, however, in maintaining that science cannot make permanent metaphysical presuppositions, held permanently immune from objective empirical evaluation. According to this standard view, the rationality of science arises, not from the way in which new theories are discovered, but rather from the way in which already formulated theories are appraised in the light of empirical considerations. And the fundamental problem of the rationality of science—the Humean problem of induction—concerns precisely the crucial issue of the rationality of accepting theories in the light of evidence.

In this essay I argue that this widely accepted standard conception of science must be completely rejected if we are to see science as a rational enterprise. In order to assess the rationality of accepting a theory in the light of evidence it is essential to consider the ultimate *aims* of science. This is because adopting different aims for science will lead us, quite rationally, to accept different theories in the light of evidence. I argue that a basic aim of science is to explain. At the outset science simply presupposes, in a completely *a priori* fashion, that explanations can be found, that the world is ultimately intelligible or simple. In other words, science simply presupposes in an *a priori* way the metaphysical thesis that the world is intelligible, and then seeks to convert this presupposed metaphysical theory into a testable scientific theory. Scientific theories are only accepted insofar as they promise to help us realize this fundamental aim.

At once a crucial problem arises. If scientific theories are only accepted insofar as they promise to lead us towards articulating a presupposed metaphysical theory, it is clearly essential that we can choose rationally, in an *a priori* way, between all the very different possible metaphysical theories that can be thought up, all the very different ways in which the universe might ultimately be intelligible. For holding different aims, accepting different metaphysical theories conceived of as blueprints for future scientific theories will, quite rationally, lead us to accept different scientific theories. Thus it is only if we can choose rationally between conflicting metaphysical blueprints for future scientific theories that

* Received April, 1973. **Part II** of this essay will appear in the succeeding issue of *Philosophy of Science.*

† I wish to express my thanks to L. Briskman, D. Crowell and J. Shearmur for their helpful comments on earlier drafts of this essay.

we will be in a position to appraise rationally the acceptability of our present day scientific theories. We thus face the crucial problem: How can we choose rationally between conflicting possible aims for science, conflicting metaphysical blueprints for future scientific theories? It is only if we can solve this fundamental problem concerning the aims of science that we can be in a position to appraise rationally the acceptability of existing scientific theories.

There is a further point here. If we could choose rationally between rival aims, rival metaphysical blueprints for future scientific theories, then we would in effect have a rational method for the *discovery* of new scientific theories! Thus we reach the result: there is only a rational method for the appraisal of existing scientific theories if there is a rational method of discovery.

I shall argue that the aim-oriented theory of scientific inquiry to be advocated here succeeds in exhibiting science as a rational enterprise in that it succeeds in providing a rational procedure for choosing between rival metaphysical blueprints: it thus provides a rational, if fallible, method of discovery, and a rational method for the appraisal of existing scientific theories—thus resolving the Humean problem.

In **Part I** of the essay I argue that the orthodox conception of science fails to exhibit science as a rational enterprise because it fails to solve the Humean problem of induction. The presuppositional view advocated here does however succeed in resolving the Humean problem. In **Part II** of the essay I spell out the new aim-oriented theory of scientific method that becomes inevitable once we accept the basic presuppositional viewpoint. I argue that this new aim oriented conception of scientific method is essentially a rational method of scientific discovery, and that the theory has important implications for scientific practice.

1. Introduction. In this essay my concern is to exhibit science as a rational enterprise. That is, my concern is to develop a theory of scientific inquiry which successfully overcomes the kind of sceptical arguments so vividly articulated by Hume [8], and which is thus able to provide a rationale for the claim that the procedures of science lead to authentic knowledge. The basic problem before us is, in other words, in essence the traditional problem of induction.

I shall argue that in order to resolve this traditional problem of the rationality of science we need to jettison a quite fundamental assumption about the basic *aims* of science that up till now has been rather widely taken for granted. The assumption I have in mind is the idea that, at the most basic level of all, the aim of science is simply to discover more and more about the world, or about the phenomena under investigation, *whatever the world or the phenomena may turn out to be like.* I shall argue that as long as we hold on to this standard conception of the aims of science, we must forego all hope of resolving the Humean problem of induction. Humean sceptical arguments in effect decisively *refute* the idea that the basic aim of science is to discover more and more about the world whatever the world may turn out to be like. In order to overcome Humean sceptical arguments we need to reject entirely this traditional conception of the aims of science, and adopt instead the kind of view most notably advocated by Einstein. We need to uphold the Einsteinian idea that at the most fundamental level of all, science constitutes a search for an underlying simplicity, unity, harmony, order, coherence, beauty or *intelligibility* which we conjecture to be inherent in the universe, new scientific hypotheses only constituting *contributions to our knowledge* to the extent that they promise to help us towards realizing this basic aim. My basic thesis, then, is just that this Einsteinian conception of science succeeds in resolving the Humean rationality problem, whereas more orthodox views about science all fail to solve this problem. In addi-

tion I shall argue that according to this Einsteinian view, scientific method is in essence a method of *discovery*, a *rational procedure for the discovery, invention or development of good new scientific theories.* It will turn out that according to this view there is only a rational method of justification or theory-acceptance to the extent that there is something approaching a rational method of discovery. I shall argue that this theory of rational scientific discovery has important implications not only for our understanding of science but also for *scientific practice* itself. Let me quickly add however that there is nothing *infallible* or *mechanical* about the method of discovery to be advocated here. According to the aim oriented conception of scientific inquiry to be proposed here, discovery, though rational, is both fallible and nonmechanical—in fact much like the rest of science.

2. Standard Empiricism. Above, I have asserted that it is widely accepted today that the basic aim of science is simply to discover more and more about the world whatever the world may turn out to be like. At once let me add that this assumption is usually hedged about with all sorts of qualifications and reservations. Thus it is widely acknowledged that in seeking to develop new scientific hypotheses, *in the context of discovery*, scientists habitually aim at developing theories which are in accordance with more or less specific metaphysical ideas about the world. A whole science community may in fact seek to develop new theories which are in accordance with certain shared metaphysical preconceptions about the nature of the universe, or the nature of the phenomena under investigation. According to the standard viewpoint it is then at most only when the question arises as to which theories should be *accepted*, which *rejected*, that science seeks to choose those theories which do the best justice to the facts *entirely independently of whether or not these theories* fit in with *metaphysical preconceptions.* Thus, according to the standard viewpoint, metaphysical preconceptions may well influence the kind of new hypotheses we seek to *develop*; such preconceptions do not, however, influence the hypotheses we ultimately decide to *accept* in science.

Some proponents of the standard viewpoint are, however, willing to qualify even this last point. Thus, Lakatos [15], has argued that an empirically successful scientific theory may acquire, within the context of a particular research program, a temporary metaphysical status, so that ostensible refutations are ignored, and auxiliary hypotheses are accepted only if they are compatible with the "core" theory. Lakatos acknowledges however that in the end, the acceptability of such a "core" theory, and the research program which it gives rise to, will depend on purely empirical factors, such as the ability of the program to be more empirically progressive than rival programs. Thus for Lakatos our choice of theory in science is not permanently influenced by our metaphysical preconceptions about the world.

Other writers (Mach [17], Duhem [3], Kuhn [12], Goodman [6], Scheffler [29], Rudner [28]) have emphasized the importance of simplicity criteria in addition to empirical criteria when it comes to choosing between rival theories in science. These writers are, however, anxious to *deny* that our preference for simple theories in science in any way commits us to upholding the metaphysical thesis that the *world* is simple, or that the phenomena under investigation are inherently simple. Thus

these writers, too, uphold the idea that at the most fundamental level of all, science seeks to discover more and more about phenomena without in any way prejudging, in a permanent, *a priori* fashion, what the phenomena under investigation are like.

There is a very good reason why it is so widely held that, at the most fundamental level of all, science seeks to discover more and more about phenomena without making permanent presuppositions about the nature of the phenomena under investigation. According to conventional ideas, the very existence of science as a rational procedure leading to authentic knowledge is at stake here. Nonscientific and pseudoscientific disciplines and creeds—such as Christian theology, astrology or scientology—may well make all sorts of metaphysical presuppositions about the nature of reality held permanently immune from empirical appraisal. But science cannot do this if it is to retain its claim to be "scientific," to be a discipline leading to the attainment of genuine knowledge. For we cannot at the outset, as it were, by pure thought alone, know anything about the world. It is only by subjecting all our guesses about phenomena to constant, ruthless empirical appraisal that we can hope to improve these guesses to the extent of evolving genuine scientific knowledge. A science which upholds theses about phenomena in an entirely *a priori* fashion, entirely independent of all empirical considerations, can only be a species of dogma, or of theology, propounding a *faith*, but not genuine *knowledge*. A willingness to submit all our assumptions about the world to empirical appraisal is, according to the standard viewpoint, the very heart of the scientific attitude.

It is interesting to note that some scientists and historians of science (Einstein [4], Meyerson [21], Koyré [11]) have in effect suggested that science does make more or less permanent metaphysical presuppositions about the nature of reality. Such thinkers, are, however, at a loss to explain how it can be *rational* and *scientific* to do this. The assumption seems to be that to the extent that science does do this, science is less rational, less "scientifically respectable" than some philosophers might hope. (Einstein, for example, calls his viewpoint a "miracle creed," and a matter of "faith.") This attitude reveals that even these writers uphold the standard viewpoint as an *ideal* of rationality, an *ideal* of scientific propriety, even if they deny that science can in practice satisfy this ideal.

The thesis that I wish to defend in this essay is that even as an *ideal* of rationality, of scientific propriety, standard empiricism—as we may call the view we have been discussing—is utterly untenable and unacceptable. My point is not just that, in practice, science does not conform to the ideal of standard empiricism. Rather, my point is that science *must not* conform to standard empiricism if science is to be a rational procedure leading to authentic knowledge. Standard empiricism upholds a *false* ideal of rationality and scientific propriety, and is to be rejected on that account, quite independently of the fact that it fails to fit the realities of scientific practice.

Before I discuss the basic inadequacies of standard empiricism, there is one important implication of this view that I wish to stress. Standard empiricism implies, in a straightforward fashion, that there can be no rational method of discovery, no rational procedure for the invention or development of good, new

scientific hypotheses. For what would a rational method of discovery amount to? It would, I suggest, provide us with a way of choosing rationally between conflicting, more or less vague, metaphysical ideas for future scientific theories, future lines of development. If we were able to choose rationally between rival embryonic ideas for future scientific theories, then we could go on choosing between more and more specific versions of such ideas until we finished up with a theory so precise, so specific, that it had actually become empirically testable! We would have converted the embryo of a theory into a fully formed testable theory by a purely rational procedure of choosing between many rival, untestable ideas.

It is clear, however, that if we accept standard empiricism no such rational procedure can be possible. For according to standard empiricism the only way in which we can in the end make a rational choice between conflicting ideas about the world is to compare these ideas with our experience. According to standard empiricism nonempirical or *a priori* considerations alone cannot provide a basis for choosing rationally between different ideas about the world. Thus standard empiricism rules out all hope of arriving at good new scientific hypotheses in a rational manner by means of the above kind of *a priori* procedure.

At most, according to standard empiricism, we can say that it will be rational to choose those metaphysical theses which are at least *compatible* with our best present scientific theories. But it is precisely these metaphysical theses which we ought to *disregard* if our aim is to develop fundamentally new scientific theories. For almost invariably fundamentally new theories in science have a conceptual or metaphysical framework or background which is entirely different from, and incompatible with, already accepted scientific theories. Thus, if anything, the metaphysical ideas that standard empiricism asserts to be the most rational to accept are precisely the ideas we ought *not* to be considering if our aim is to develop a new and improved scientific theory.

I conclude that once we accept standard empiricism we are obliged also to accept that the invention of new fundamental theories in science is an irrational (or nonrational) process. It is noteworthy that most scientists and philosophers of science do in fact appear to maintain that discovery is not rational (Popper [23], p. 31, Reichenbach [26], p. 231). We should also note that if standard empiricism turns out to be *wrong*, then a major objection to the possibility of there being a rational method of discovery will have vanished.

3. Hume refutes Standard Empiricism. The fundamental inadequacy of standard empiricism is that it seriously misrepresents the aims that we *ought* to be pursuing in attempting to improve our knowledge by undertaking scientific inquiry. The most blatant inadequacy of standard empiricism is, however, perhaps that it fails to provide an answer to Humean scepticism about the possibility of acquiring genuine scientific knowledge. Let us now look at this second point.

Let us suppose that in some scientific discipline there is a theory T which has met with absolutely staggering empirical success, whatever precisely our criteria of empirical success may be. T, we may suppose, has successfully predicted an incredible wealth of new phenomena; it has trounced all its rival theories, and has not

faced one single refutation or anomaly. T is, in short, more "firmly established" on empirical grounds than any actual theory in the entire history of science.

Now the brutal fact is this. Whatever precisely T may be, and whatever precisely our criteria of empirical success may be, we can easily invent as many theories, T_1, T_2, . . . as we please, all of which differ drastically from T and from each other as far as empirical consequences are concerned, but each of which meets our criteria of empirical success just as well as T does. A general procedure for the construction of rival theories to T which are at least as successful empirically as T is, can be specified like this. First we pick out some specific type of experiment which has, we may suppose, in essence been performed endless times already in laboratories throughout the world. We may suppose that T (perhaps in conjunction with certain auxiliary hypotheses) successfully predicts that if the experiment is performed the outcome is P. We now add to the specification of the experiment certain entirely bizarre, ludicrous details which would ordinarily be judged to be entirely irrelevant to the outcome of the experiment. Thus we might stipulate that the apparatus be painted red, or that sulphur powder be sprinkled in a circle around the apparatus, or that the sounds "Abracadabra" be made in the vicinity of the apparatus during the experiment. In this way we specify in *purely universal terms* a type of experimental set up (essentially, a set of initial conditions) which we can be reasonably sure has never obtained anywhere on earth or probably anywhere in the universe. We can now construct empirically successful rivals to T by means of the following rubric: "As long as E does not occur, everything occurs in accordance with T; if E occurs then the outcome is Q." We are here free to choose Q as we please. Q may differ only slightly from P (the prediction of T), or Q might be some quite drastic assertion such as that all phenomena in the universe occur in accordance with such and such a set of laws which are very different from our present physical theories (in which case our rival theory to T would in effect assert that if E occurs the laws of nature change!). By dreaming up different experimental set ups E and different outcomes Q we can very easily invent an unlimited number of *aberrant* versions of T—as we may call these theories—each of which will differ in predictive content from T, and yet will be just as empirically successful as T.[1]

[1] The problem of the rationality of science, as formulated here, rather different from anything to be found in Hume [8], is, I wish to claim, a considerable improvement over Hume's original formulation of the problem, for the following reasons. (1) Hume's formulation of the problem lays considerable emphasis on the *semantic* or *ontological* thesis that it is not possible for there to be logically necessary connections between successive events. But, as I have shown elsewhere (see [18], reprinted in [31]), this semantic or ontological thesis of Hume's is in fact false. This does not, however, affect the main thrust of Hume's epistemological argument (see [18], pp. 3–4 for this point). The formulation of the problem given here is then an improvement over Hume's formulation in that here no appeal is made to the false semantic or ontological thesis. (2) As Hume formulates the problem, it looks as if only inductive or justificationist conceptions of scientific knowledge are threatened by the basic argument, the problem being evaded by the kind of conjectural, falsificationist conception of knowledge advocated by Popper. But as formulated here, it is clear that Popper's view does not succeed in avoiding, or resolving the problem. The aberrant theories considered in the text are just as highly falsifiable and just as well corroborated as T itself is; hence Popper's theory does not succeed in providing a rationale for preferring T to any of the aberrant theories. The problem, as formulated here, is then more *general* than Hume's formulation of the problem, in that it poses a problem for a wider range of concepts of knowledge; it is thus an improved formulation. (3) The problem as formulated

The decisive objection to all versions of standard empiricism can now be put like this. In actual scientific practice we entirely ignore the infinite ocean of aberrant rivals to our best scientific theories in an entirely *a priori*[2] fashion. Not for one moment do we set out to refute experimentally such theories. If we tried to do this, scientific progress would come to an immediate halt, just because there will always be an endless supply of aberrant versions of our best scientific theories. As quickly as we refuted these aberrant theories, new unrefuted, highly confirmed or corroborated aberrant theories could always be dreamed up to take their place. Thus if science is to avoid becoming completely stultified, it is essential that in any actual scientific situation infinitely many highly confirmed aberrant theories are dismissed from all consideration in an entirely *a priori* fashion, independently of all empirical considerations. But in dismissing all empirically successful *aberrant* theories in this wholly *a priori* fashion we are in effect presupposing that the world itself is not aberrant. We are prejudging the outcome of infinitely many possible experiments on grounds wholly independent of all empirical considerations. Given the potentially infinite class of aberrant experiments E_1, E_2, . . ., we prejudge in a wholly *a priori*

---

here is more general than Hume's formulation in another way as well. Hume in effect considers only one kind of aberrant theory, namely a theory which specifies a sudden change in the laws of nature at some specific time or place. Such nonuniversal aberrant theories can, however, always be reformulated as purely *universal* aberrant theories, of the type considered in the text, by specifying in universal terms some unique event E, and then asserting the universal aberrant law: "t years after the occurrence of E, the laws of nature change suddenly." Thus the kind of aberrant theory considered by Hume is but a very special case of the kind of aberrant theory considered here. I should perhaps add that both Goodman ([5], [6]) and Scheffler [29], have stressed that the Humean problem has nothing to do specifically with theories that make an assertion about some specific time t. The particular *example* chosen by Goodman is, however, I believe, invalid for this purpose. As I have argued elsewhere ([18], pp. 10–17), a genuine property term is linked analytically to other property terms, and contains, inherent in its meaning, some more or less vague stipulation as to how the property is "conserved" in relation to the conservation of other properties. Once such conservation stipulations of necessity inherent in the meaning of 'grue' are made explicit it will become obvious that 'grue' contains an implicit reference to a specific time t, whereas 'blue' and 'green' do not. One might try to overcome this objection by introducing a whole system of grue-type predicates analytically interrelated in such a way that within this system no implicit temporality emerges. But once we allow such systems of predicates to be legitimate we run into the difficulty of knowing on what grounds other people are not employing ordinary language in just such a way. For all I can know 'grue' might be highly entrenched in that for most people 'green' means what I mean by 'grue'. Goodman's assumption that 'grue' is *not* highly entrenched in effect rules out such bizarre possibilities. (4) The rationality problem as formulated here makes clear that the problem is to provide a reason, a rationale, for excluding complex, ugly, grotesque, *ad hoc* theories on purely *a priori* grounds. The problem, in other words, is to provide a rationale for *a priori* simplicity criteria. At once it becomes clear that the Humean problem concerns an issue that is of central importance both for our understanding of science, and for scientific practice itself, since simplicity criteria are widely believed to play an important role in science even though no one has succeeded in specifying what these criteria are. None of this is clear from Hume's original formulation of the problem, nor from formulations usually provided by philosophers today, such as Goodman's formulation in terms of scientifically absurd predicates such as 'grue'.

[2] Throughout the essay *a priori* simply means 'nonempirical'. *A priori*, as used here, does not carry connotations of absolute certainty or indubitability, as, for example, Kant's use of the term does. Thus a theory quite legitimately accepted on "*a priori*" grounds, is not thereby necessarily "known to be true with certainty."

way that the outcome of all these experiments will be that predicted by T rather than that predicted by any of the aberrant rivals to T. In dismissing on nonempirical grounds all the aberrant rival theories to T we are, in other words, taking for granted a rather general assumption about the nature of the world—or the nature of the phenomena we are investigating—in a wholly *a priori* way, entirely independently of empirical considerations. The decisive difference between T on the one hand, and all the aberrant versions of T (just as well supported empirically as T) on the other hand, is that whereas T will have a certain inherent coherence, unity, simplicity, or intelligibility, all the aberrant theories will be quite grotesquely *ad hoc*, incoherent, ugly, complex, arbitrary, or unintelligible, postulating as they do abrupt, inexplicable, *ad hoc* disruptions in the harmony and coherence of natural law. In choosing T in preference to all the aberrant rivals to T we are thus in effect assuming that *Nature herself* is nonaberrant. We are assuming that the world (or the domain of phenomena we are investigating) is more likely to be coherent, simple, harmonious, nonarbitrary, unified, intelligible, or in a word *nonaberrant*, rather than incoherent, arbitrary, inexplicable, disunited, unintelligible, or *aberrant*. And we are upholding this general assumption that the phenomena are intelligible in an entirely *a priori* fashion, entirely independently of all empirical considerations. Theories which *conflict* with this general assumption that the phenomena are intelligible will be rejected out of hand even though such theories are highly successful empirically.

The actual situation is in fact much worse than the above argument suggests. So far we have made the wholly unrealistic assumption that the chosen scientific theory, T, faces no empirical difficulties. In practice, even our best scientific theories invariably face empirical problems or anomalies—that is, ostensible refutations. Let us then make the realistic assumption that T faces n ostensible experimental refutations. We can now easily construct all the aberrant rivals to T in such a way that these rival theories predict—in an entirely *ad hoc* fashion—the "correct" results of the n experiments that refute T. Thus the infinity of aberrant theories will be more successful empirically than T is; and, nevertheless, in scientific practice we dismiss all such aberrant, unintelligible theories in an entirely *a priori* fashion.

The situation is even worse still. For suppose T fails to predict the obtained result of just *one* kind of experiment E^*. We can then easily transform this one kind of experiment into as many different kinds of experiments E_1^*, E_2^*, ... as we please, merely by adding on different bizarre details to E^*, which would ordinarily be considered to be wholly irrelevant to the outcome of the experiment. In this way, we can ensure that if T is ostensibly refuted by just *one* kind of experiment, then it is refuted by N different kinds of experiments, where N is as large a number as we please. All our aberrant rivals to T can, however, be constructed (in an entirely *ad hoc* manner) to predict the obtained results of these N experiments. In this way we will always be able to construct as many aberrant theories as we please which will be more or less infinitely more successful empirically than our best scientific theory T (as $N \to \infty$), and nevertheless T will be preferred to any of these aberrant theories. Theories which conflict with our *a priori* assumption that the phenomena we are investigating are intelligible or nonaberrant are, in other words, rejected out of

hand even if these theories are infinitely more successful empirically than our chosen scientific theory T which *is* compatible with our basic intelligibility assumption.

But the situation is even worse than this still. For it turns out that in science the *a priori* assumption that the phenomena are intelligible not only leads us to reject *theories* that are enormously successful empirically which conflict with this assumption: far worse, experimental *data* are not accepted unless they harmonize with the basic intelligibility assumption. The essentially metaphysical thesis that the phenomena under investigation are intelligible (in some way or another) is upheld in such a dogmatic, *a priori* fashion that the very *evidence* is twisted and contorted until it can be made to harmonize with this thesis.

Let us suppose that some aberrant theory T_1 predicts some such effect as that muttering "Abracadabra" has a radical effect on the outcome of an experiment E. And let us suppose that this prediction is actually corroborated! If this were to happen, the science community would not only continue to ignore T_1, but even worse would ignore the experimental result successfully predicted by T_1! Aberrant experimental results, of a degree of absurdity almost as high as the above, are obtained every day in laboratories throughout the world. And even if such aberrant results are *repeatedly* obtained, experimental scientists persistently subject their experiments to devastating criticism and analysis until the ostensibly aberrant result has been eliminated, or transformed into a more or less intelligible result. For an experimental result to be publishable, and accepted as a *bona fide* result, it is, in short, not enough that a repeatable effect be obtained: it is essential that the result harmonizes with the *a priori* metaphysical assumption that the phenomena under investigation are more or less intelligible.

The basic thesis of standard empiricism—namely that science does not make permanent metaphysical assumptions about the world (or about the phenomena) upheld in an entirely *a priori* fashion—is thus wholly untenable. Not only do scientists invariably prefer to hold onto the idea that the phenomena are intelligible even to the extent of rejecting out of hand all *theories* which conflict with this assumption, even though these theories may be infinitely more successful empirically than any theory which is compatible with the assumption: even worse, scientists reject *experimental data* which conflict with the assumption that the phenomena are more or less intelligible.

One niggling doubt may perhaps be raised in passing about this conclusion that science presupposes that Nature herself is intelligible, nonaberrant or simple. It has been argued by Goodman ([6], p. 337) that the notion of simplicity can hardly be applied to the world itself, but only to the world relative to some type of systematization or organization. In itself, the world is neither simple nor complex. In view of this argument, what becomes of the idea that science is committed to the idea that the world itself is simple? We can, I suggest, rescue this idea by interpreting it to assert: the world is such that it is in principle possible to formulate a theory that is both true *and* simple. It cannot be assumed that it will always be in principle possible to do this for all possible worlds.

4. Failure of Attempts to Salvage Standard Empiricism. The above considerations

appear to show conclusively that science must take for granted some such metaphysical conjecture as that the world is more or less intelligible or nonaberrant, to the point of rejecting all empirically successful theories and all well established empirical data that conflict with this basic conjecture, *even though we can have not the slightest reason for supposing that the basic conjecture is true.* As long as we hold onto the standard empiricist conception of knowledge outlined above, this state of affairs must be anathema. It can only mean that science is a species of theology, of pseudoscience, or of disreputable *a prioristic* metaphysics. Thus, ever since Hume [8], first highlighted the problem, or the *scandal,* as one might call it, numerous attempts have been made to try to show that the real situation is not quite as described above, either because we do after all have good reasons for supposing that the basic metaphysical conjecture underlying science is true (or will inevitably be "borne out" by our experience (cf. Kant [10]), or because science after all is not *really* committed to upholding in a permanent fashion metaphysical conjectures about the world, despite appearances to the contrary.

It is my contention that all these attempts to rescue the standard empiricist conception of scientific knowledge in the end fail. There is, however, insufficient space here to examine this point in detail, and we must confine ourselves to a brief consideration of one or two of the main general viewpoints upheld today.

One traditional approach is to argue that despite the Humean considerations outlined above, it nevertheless remains the case that in science we ultimately choose that theory which is, quite simply, the best supported by the evidence. According to this verificationist viewpoint, if we are presented with a number of rival theories all of which superficially fit the available evidence equally well, then it is, in reality, the simplest, most intelligible theory, the least aberrant theory, which receives the greatest degree of support from the evidence, and can be regarded as being the best verified (Jeffreys [9], Barker [2]). Thus, in choosing the least aberrant theory available in science, sometimes even to the point of ignoring aberrant theories which *prima facie* fit the available data rather better, we are in reality basing our choice solely on a consideration of which theory is the best verified by the facts.

It is difficult to see how this conventional viewpoint can be made even remotely plausible. In a world which we somehow know in advance to be intelligible or nonaberrant, no doubt we would be justified in regarding the simplest or most intelligible theory to be the best verified—even perhaps to the point of disregarding ostensibly better verified unintelligible theories. But if we do not know in advance that the world is intelligible or nonaberrant, what possible rationale can there be for regarding an intelligible theory to be better verified than infinitely many rival aberrant theories, each of which will actually fit the available data far better than our chosen theory? And what rationale can be given for only accepting empirical data which fit in with the metaphysical presupposition that the data are intelligible, if we have no reason to suppose that this presupposition is *true*? This conventional viewpoint is, I suggest, quite unable to provide satisfactory answers to these questions.

Can we argue that our past experience has taught us that nonaberrant theories

are more likely to be empirically successful than aberrant theories? No, because if our experience can be interpreted as confirming the success of "nonaberrant" methodological rules (which favor all nonaberrant theories, other things being equal), then, equally well, our experience can be interpreted as confirming the success of *aberrant* methodological rules, which favor all nonaberrant theories *except for just this aberrant theory* T_1 (or this specific class of aberrant theories C).

A rather different approach to the problem of salvaging standard empiricism has been made by Popper ([23], [24], [25]). According to Popper, granted that the fundamental aim of science is to capture as much Truth about the world as possible, the rational procedure will be to put forward theories which assert as much as possible about the world, and which are as vulnerable as possible to experimental refutation. The best theory in science is then the theory which has the greatest empirical content, and has best survived our most strenuous search for refutations. According to Popper, our best scientific theories, however "well established," remain pure conjectures, pure speculations; nevertheless, such theories can be held to be rationally chosen in that they are theories upheld in such a way that there is the permanent possibility of discovering that they are false. For Popper, the rationality of science arises from the fact that in science we subject our theories to the most devastating form of criticism possible, namely, empirical testing; science is rational just because in science we do our best to maximize the role of empirical considerations over our choice of theory.

Unfortunately, this Popperian defense of standard empiricism does not work. (For a detailed criticism see [19].) Quite apart from the difficulty that Popper's viewpoint hardly seems capable of explaining why the theory chosen for its maximum vulnerability to refutation should be the theory we will be prepared to use for technological applications, there is the major difficulty that aberrant versions of the "accepted" scientific theory T may well have far more empirical content than T does, thus being, according to Popper's position, far more acceptable than T.

In [24] Popper does, it is true, specify a methodological rule which would have the effect of ruling out aberrant theories. Popper asserts that an acceptable "new theory should proceed from some *simple new, and powerful, unifying idea* about some connection or relation (such as gravitational attraction) between hitherto unconnected things (such as planets and apples) or facts (such as inertial and gravitational mass) or new 'theoretical entities' (such as field and particles)" ([24], p. 241). If this simplicity criterion could be sufficiently precisely formulated it might enable us to exclude empirically successful aberrant theories. The trouble is, however, that Popper provides no rationale whatsoever for this simplicity criterion, no explanation as to why scientists ought to prefer simple theories in this sense granted that they are engaged in "the task of getting nearer to the truth" ([24], p. 241). Invariably, to accept only simple, unified theories of the above type is in effect to presuppose, in a manner placed beyond all criticism, that the world itself is simple or unified; Popper, however, fails to acknowledge this, and fails to explain how the *a priori* acceptance of such a metaphysical thesis can be rational, given the general Popperian explication of the rationality of science.

It might be thought that Popper's position, modified somewhat in the direction

advocated by Lakatos, [15], could suffice to solve the problem, in the following manner. One could argue that in science we accept the theory (or research program) which actually itself leads to the greatest empirical growth. Inevitably it will be the nonaberrant theory T which will have successfully predicted new phenomena, not the aberrant rivals to T constructed from T in the manner described above. Thus, according to this Lakatosian view, T is the rational theory to accept rather than any of the aberrant rivals to T. The trouble with this argument however is that it is always possible that a mad genius might invent an aberrant theory T^*, which becomes the "hard core" of an enormously progressive research program (T^* might for example predict, in aberrant fashion, the day of judgement). In this situation the Lakatosian viewpoint would oblige us to accept T^*, whereas we should in fact in science accept the nonaberrant variant of T^*.

A quite different way in which one may seek to salvage standard empiricism is to adopt the kind of conventionalist viewpoint advocated by Duhem [3], and others (for example, Rudner, in [28]). According to this kind of view, science seeks both to discover more and more phenomena *and* to develop theories which organize phenomena into as systematic a deductive structure as possible. Thus in science, other things being equal, *simple* theories are invariably preferred to complex, aberrant theories; but this unswerving preference for simplicity, for systematicity, in no way commits science to the idea that the world is itself simple. The simplicity of our theories is a purely human construction which in no way reflects how things are.

This argument might work if the problem were simply to choose between n different theories, $T_1, T_2, \ldots, T_n$, of varying degrees of simplicity, each of which makes precisely the same experimental predictions. In this case we could justifiably claim that our preference for simple, nonaberrant theories in no way involves presupposing anything about the world. The situation discussed above is, however, very different from this. There we had an indefinite number of rival theories $T, T_1, T_2, \ldots$, where T alone is nonaberrant, where $T_1, T_2, \ldots$. actually fit the evidence far better than T, and where each theory makes predictions *different from* all the other theories. In this situation, to choose T and to reject the rest on the grounds that T alone is nonaberrant *is* in effect to decide, on nonempirical grounds, that the phenomena are themselves nonaberrant rather than aberrant. The conventionalist attempt to allow for an *a priori* preference for simple theories in science without thus committing science to the thesis that the phenomena themselves are simple does not, in other words, succeed.

A fourth general approach to the problem is frankly to acknowledge that science does make metaphysical presuppositions about the world (cf. Russell, [27]). However, as long as the basic tenet of standard empiricism is retained—namely, that at the most basic level of all, science seeks merely to discover more and more about the world (or the phenomena) whatever the world (or the phenomena) may turn out to be like—there remains the insoluble problem of how it can be rational for science to make entirely unjustified, unsupported metaphysical presuppositions about the world.

Recently, a number of writers appear to have abandoned the problem of providing a *rationale* for rejecting empirically successful aberrant theories in science, and

have taken the basic problem to be simply one of precisely specifying the rules employed in science governing our choice of theory in the light of evidence (Hempel [7], Swinburne [30], Kuhn [13], Lakatos [16]). Goodman ([5], [6]) in particular, has sought to solve this problem of defining 'confirmable' by means of the suggestion that a "confirmable" hypothesis is to be defined as a hypothesis which is supported, unviolated and unexhausted, and which is such that all hypotheses which conflict with it are less well "entrenched" ([6], p. 393).

There is, however, a basic inadequacy in this approach, if interpreted as a putative solution to the problem of the rationality of science. To begin with, there appear to be serious inadequacies in Goodman's definition of 'confirmable' or 'projectable', if this is interpreted as constituting a first step towards explicating the conditions under which hypotheses are accepted in science. First, refuted theories often continue to be accepted in science (for the want of better alternatives), whereas for Goodman refuted theories are unprojectable. Second, theories are put forward in science even though they conflict with well established, unrefuted theories (an example is perhaps Einstein's 1905 photon hypothesis); according to Goodman, however, such a theory would be unconfirmable. Third, if we accept Goodman's theory, it is very difficult to see how a theory such as the special theory of relativity could ever be accepted. For the special theory of relativity undermines almost all the predicates of classical physics, in that it renders the basic absolute (nonframe-dependent) notions of mass, length and time inapplicable. These classical predicates, and the maze of highly confirmed low level laws which use them, could hardly be more firmly entrenched. The classical laws could not be held to be *refuted*, since the development of an ether theory which parodied the predictions of special relativity could always have been manufactured. In these circumstances Goodman's theory would seem to characterize the special theory of relativity as irredeemably unprojectable, in that it is irredeemably "overridden." And yet the special theory of relativity was accepted (and was even upheld by Einstein in the teeth of ostensible refutation).

A more fundamental objection to Goodman's approach is that even if Goodman were to succeed in explicating a notion of relative confirmability which more or less mirrored scientific practice, the problem of *why* we accept theories in this way would remain untouched. I suggest that the primary problem that Humean considerations ought to prompt us to face is just this: What are the basic desiderata of scientific inquiry? We have, in effect, seen that if we accept the standard empiricist idea that the basic desideratum of science is merely to discover more and more about the phenomena whatever the phenomena may turn out to be like, then it is very difficult to make rational sense of the choices we do in fact make in science. The conclusion we should draw from this is that we have perhaps misconceived what the basic *aim* of science is. Once we have a clearer idea about what the basic aims of science ought to be, then we may hope to see theory choices in science as rather natural, rational choices, given these aims. In other words, instead of abandoning the rationality problem, and concentrating on the problem of providing a definition of 'confirmable' (as Goodman advocates) we should rather, I suggest, concentrate on trying to specify an acceptable, desirable, fundamental *aim* for science, and a

set of rules which give us the most rational hope of realizing this aim, and *then* see whether these rules reflect what we do in science. Until we know what the *aims* of science ought to be, it hardly makes sense to speculate at length about what the procedural *rules* of science ought to be.

5. Adequacy Criteria for Solution to the Problem of the Rationality of Science. In order to know whether it is rational to accept a given theory T in the light of evidence it is essential that we know what our basic *aim* is in considering whether or not to accept the theory. Given one aim, it may be quite rational to accept T; given a rather different aim it may be entirely irrational to accept T. Thus in order to solve the problem of rational acceptance in science it is clearly essential to specify a definite, unambiguous basic aim for science.

It might be thought that the basic aim of science is simply to procure, or improve, knowledge. But unfortunately this fails completely to specify an unambiguous aim for science, since different philosophers and methodologists mean quite different things by the one term 'knowledge'. According to one view (Ayer [1]) to say that we know p is to say (a) p is true, (b) we believe p is true, (c) we have good reasons for believing p is true. But as many writers have in effect pointed out, we can hardly suppose that science seeks knowledge in *this* sense. We have good reasons for supposing that many of our best scientific theories are in fact strictly *false*. All three conditions are violated. Some might try to get around this by holding that in science we accept a theory as belonging to our knowledge if it is more likely to be closer to the truth than any of its rivals, in the light of the available evidence. Popper, however, holds that even our best corroborated theories remain utterly improbable and conjectural. For Popper, to say that a theory "belongs to our knowledge" is to say roughly that it is highly falsifiable, and has withstood all our attempts to falsify it. Any trace of a justificationist or verificationist element is expunged from the Popperian concept of knowledge. Again, conventionalists or instrumentalists may uphold a different concept of knowledge. According to conventionalism, to say that a *theory* is a part of our knowledge is to say that it systematizes established facts in a neater, more convenient, orderly fashion than other rival theories. And finally, in this essay, I put forward the view that to say that a theory is a part of our knowledge is to say that it helps us towards discovering more about an underlying intelligibility which we hope really is inherent in the universe.

Confronted by this dismaying multiplicity of concepts of knowledge we clearly have the problem of specifying rather general requirements that any concept of knowledge must satisfy if it is to qualify as a *legitimate* concept of knowledge. Or, to put the thing slightly differently, we need to specify rather general requirements that any proposed fundamental *aim* for science must satisfy if it is to constitute a legitimate aim for science. The point here is that we may hold it to be a *tautology* that the fundamental aim of science is to seek knowledge (in *some* appropriate sense of 'knowledge'). To specify a precise fundamental aim for science is just to give a precise explication of the notion of "knowledge," and *vice versa*. Furthermore, to lay down general adequacy conditions for proposed fundamental aims of science is just to lay down adequacy conditions for possible concepts of knowledge.

What, then, lies behind all the very different ideas about knowledge, about the aims of science, indicated above? There is, I suggest, one root idea in common: namely that theories which deserve to be called a part of our knowledge are theories which deserve to be especially *prized* or *valued* in some way or other. We may value our best scientific theories for their supposed relative certainty, or for their ability to survive refutation despite extreme vulnerability to refutation, or for their amazing systematizing power, or for their staggering ability to explain and render intelligibile. There are, in other words, any number of quite different ideas as to what it is that is of such value about our best scientific theories. There is, however, general agreement that theories, insofar as they belong to our knowledge, do deserve to be especially prized, especially valued, for some reason or other. In saying that science leads to authentic knowledge whereas pseudoscientific disciplines lead to nothing of the kind, we are at the very least asserting that the best theories of science deserve to be taken much more seriously, deserve to be held to be much more important, than the theories of those other disciplines.

I propose, then, that one condition a concept of knowledge must satisfy in order to be acceptable is this: the concept must be such that in saying a theory belongs to our knowledge we are at least saying that the theory deserves to be especially valued for some reason or other. Again, a proposed fundamental aim for science, in order to be acceptable, must be such that it is clear that theories which genuinely help us towards realizing the aim deserve to be especially valued, for some reason or other. Values are in short inherent in the very meaning of 'knowledge'.

Can we, without becoming controversial, say anything more specific about the kind of values that ought to be inherent in any concept of "knowledge" if it is to be legitimate and acceptable? Can we specify slightly more precise value conditions a proposed aim for science must satisfy if it is to constitute an acceptable possible aim for science?

There are, I suggest, inherent in science two very different kinds of values and aims. On the one hand there are the aims and values of pure science; on the other, the aims and values of technological science. It is almost as if "scientific knowledge" is compounded of two distinct notions of "knowledge."

Much of our scientific knowledge has arisen as a consequence of men pursuing the aims of pure science (whatever precisely these may be) in sublime indifference to the question of whether this pursuit will lead to discoveries of any practical or technological value. Men have pursued scientific inquiry in order to increase our understanding of the world, in order to discover explanations for puzzling phenomena. Put roughly, science is pursued for its own sake, and not for any technological spin off it may lead to. And in this context of pure science, theories are judged to be contributions to "knowledge" insofar as they help us realize the objectives of pure science. Cosmological theories, or a theory such as general relativity, may be judged to be contributions to knowledge of the very highest order, even though from a technological standpoint their value may be nil. There can be no doubt that pure science has its own values, quite distinct from any question of practical or technological value.

This said, it must be acknowledged that, where relevant, theories especially

valued from the standpoint of pure science, are also especially valued from the standpoint of practical or technological use. Even though new scientific theories may not have been developed with the explicit aim in mind of being technologically useful, nevertheless, where relevant, we judge it rational to choose accepted scientific theories for technological use rather than rival nonscientific theories.

Corresponding to the two *kinds* of aims and values inherent in science there will be two different rationality problems, which can be formulated like this:

> (a) *The pure rationality problem*: To specify a definite worthwhile aim for pure science, and then show that our present scientific theories are rationally, sensibly chosen given this aim, whereas other rival theories such as empirically successful aberrant theories, would be irrationally chosen, given the specified aim.
>
> (b) *The pragmatic rationality problem*: To show that theories rationally chosen in the context of pure science are the rational, sensible theories to choose (where relevant) granted that our aim is practical or technological, it being irrational to choose other rival theories such as empirically successful aberrant theories.

An acceptable concept of "knowledge," an acceptable proposed fundamental aim for science must, I suggest, ultimately do justice to both of the above values and aims of science, the pure and applied. In order to be fully satisfactory, a theory of scientific inquiry must be capable of resolving *both* of the above two rationality problems.

It is, however, I suggest, important that we begin by clearly distinguishing the aims and values of pure and applied science, and the two rationality problems that this distinction gives rise to. For it is a very striking fact about science, *which we should seek to explain*, that theories chosen in order to realize the aims of pure science, are also of great value from the point of view of technological use. If we conflate the two sets of values and the two problems, we prejudge, in an *a priori* way, that pure science has valuable practical applications, and we thus preempt the possibility of *explaining* this striking fact about science.

Furthermore, there is, I suggest, a sense in which the pure rationality problem has *priority* over the pragmatic rationality problem. It is only when we have succeeded in characterizing pure science as the rational pursuit of a rational, worthwhile aim that we can be in a position even to *formulate* the pragmatic rationality problem properly, let alone solve it. It is only when we know what rationale lies behind the choice of theory in pure science that we can go on to ask sensibly why theories chosen in terms of this rationale should also be theories to choose in a technological situation.

In this essay, therefore, I concentrate on putting forward a theory of scientific inquiry which succeeds in resolving the above pure rationality problem. Elsewhere, I show how this theory can be extended to solve the pragmatic rationality problem as well (Maxwell [20]).

It might be thought that if we ignore the pragmatic rationality problem, then the pure rationality problem is easily solved. But this is far from being the case. As we in effect saw in the last section, no theory of scientific inquiry so far put forward

succeeds in solving this problem. Neither Popper nor Lakatos nor Duhem is primarily concerned to solve the pragmatic rationality problem. They are all, however, concerned to show pure science as an internally rational enterprise, as the rational pursuit of a rational, worthwhile aim: they are, in other words, concerned to solve the pure rationality problem. And yet, as we have seen, they fail. No acceptable rationale is given for excluding empirically successful aberrant theories, even though the technological reliability question is not at issue. Thus the pure rationality problem is not the light matter it might at first appear to be.

Let me conclude this section by spelling out a little more explicitly conditions that a theory of scientific inquiry will need to satisfy if it is to solve the pure rationality problem as conceived here.

First, the theory will need to specify:

(a) An unambiguous, fundamental aim A (or group of aims A) for pure science, and a set of methodological rules R, which specify how scientific theories are to be chosen between, accepted and rejected, in the light of evidence.

The specified aim A, and rules R will then need to meet the following conditions:

(b) The aim A must not be such that it is known in advance to be unrealizable.

(c) The aim A must be sufficiently worthwhile for theories which genuinely take us towards realizing A to have sufficient *intellectual value* to deserve the honorific title of 'knowledge' (within the context of pure science).

(d) The aim A and rules R must be such that we have sufficiently good reasons for holding that theories accepted in accordance with the rules R really do help us towards realizing the aim A. We must be able, for example, to give some kind of reason for holding that the rules R give us a better hope of realizing A than alternative sets of rules R', R''.

(e) The aim A and rules R must reflect reasonably well the realities, the practice of pure scientific inquiry.

Clearly a theory of scientific inquiry which fails to satisfy (e) completely could hardly be held to be satisfactory. Such a theory might specify *some* kind of rational enterprise, but it could hardly succeed in exhibiting *science* as a rational enterprise.

We should note, however, that if we are to be really ambitious we do not want (e) satisfied *too* well. For if our theory does indeed succeed in satisfactorily exhibiting science as a rational enterprise, we may well hope that it will enable us to correct present scientific practice—enable us, that is, to do science even more successfully than we have been able to do up to now. We may then add an additional, extremely demanding condition:

(f) The specified aim A and rules R ought to enable us to correct and improve present scientific practice.

The above metamethodological thory embodied in (a) to (f) above (put forward briefly in [19], without argument) is advocated here essentially as a reasonable *conjecture*, as specifying reasonable conditions of adequacy for a theory of scientific inquiry which constitutes a solution to the pure rationality problem. It is put forward as a theory which is *itself* open to criticism. And in fact, in [20], I shall in

effect argue that the theory in the above form is inadequate, and needs to be quite substantially changed and improved. However, before we can be in a position to see these inadequacies in the above metamethodological theory, we need first to develop a theory of scientific inquiry which succeeds in satisfying the criteria spelled out in (a) to (e) at least (disregarding (f) for the time being).

6. Aim Oriented Empiricism. A fundamental thesis of this essay is that all versions of standard empiricism fail to resolve the Humean rationality problem because these views all profoundly misrepresent the basic *aims* of scientific inquiry. The argument of section **3** in effect shows that the aim of discovering more and more about the phenomena whatever the phenomena may turn out to be like is an *irrational* aim to pursue, since given this aim, there must always be a potentially infinite number of easily formulated theories which are, from a rational standpoint, just as good as each other. Instead of attempting to patch up standard empiricism in the face of Hume's onslaught, we need, I suggest, rather to accept that Humean considerations show decisively that standard empiricism fails as an ideal of rationality and scientific propriety. We need to go to the heart of the trouble, and completely *reject* the basic standard empiricist thesis that the fundamental aim of science is simply to discover more and more about the phenomena *as such.*

According to the view to be advocated here—a view which may be called 'aim oriented empiricism'—the fundamental aim of pure science is to discover more and more about an underlying simplicity, coherence, unity, harmony, order, beauty, or *intelligibility*[3] which we conjecture to be inherent in the universe (or inherent in the

[3] I use the terms simple, intelligible, non-*ad hoc*, coherent, harmonious, unified, explanatory, beautiful more or less interchangeably, the core idea here being perhaps intelligibility. It might be argued, however, that there is not even an intuitive connection between "explanatoriness" on the one hand and, say, "beauty" on the other hand. So let me indicate how these ideas link up in an intuitive way.

To begin with, it might be argued that a theory T explains an experimental result or law, L, if (roughly speaking) T entails L and is independently testable so that T has excess empirical content over L. (This is perhaps the standard conception of explanation in the philosophy of science: I ignore refinements irrelevant to the point at issue here.) Nothing is said here about T being *simple*, or non-*ad hoc* etc. Hence, simplicity etc. have nothing to do with explanatoriness. The obvious objection to this is that if we collect together n different experimental laws $L_1 \ldots L_n$, and let T_1 be the conjunction of these laws, then, according to the above account, this theory T_1 "explains" each of L_1, L_2 etc. And this is clearly absurd. Such an entirely *ad hoc* theory as T_1 would not be considered explanatory for a moment. This shows that non-*ad hocness* is a crucial component of the notion of explanatoriness.

But we can go further. Non-*ad hocness* or simplicity is a crucial component even of the notion of scientific *law* (itself an essential component of the notion of "explanation"). For if we allow laws to be just strictly universal statements which can be as aberrant, as *ad hoc* as we please, then the notion of a *lawful* sequence of events becomes completely empty of content. This may be shown as follows. Consider any completely chaotic, "unlawful" universe, whose successive instantaneous states are $S_1, S_2, \ldots S_r \ldots$ We may suppose the universe never gets into precisely the same state twice. We now construct universal theories $T_1, T_2, \ldots T_r \ldots$, where T_r asserts: If a state of type S_r occurs then the next state is of type S_{r+1}. (Here S_r and S_{r+1} are described in strictly universal terms.) Let T be the conjunction of $T_1, T_2, \ldots T_r \ldots$, T is a strictly universal theory, which predicts precisely the course of events in our chaotic universe. Thus if we make strict universality a sufficient condition for lawfulness, we would be obliged to say that any chaotic universe whatsoever is "lawful." Clearly we have to rule out theories such as T as not specifying laws because of a grotesque degree of *ad hocness* or aberrance.

phenomena we are investigating). According to aim oriented empiricism there is, in other words, inherent in the basic aims of science a wild metaphysical *conjecture*—namely, that the world (or the domain of phenomena under investigation) is intelligible. At the very least, there is the metaphysical conjecture that the phenomena are such that it is in principle possible to develop theories of increasing simplicity, unity, coherence or intelligibility which *also* meet with increasing empirical success. The fundamental aim of science is, in other words, to develop successive theories which progressively articulate and make precise more and more of a metaphysical conjecture M (which asserts, roughly, that the phenomena are intelligible) in such a way that these theories meet with more and more empirical success. To assert M (i.e. that the phenomena are intelligible) is just to assert that there is some theory T, compatible with M, which successfully predicts the phenomena of the domain under consideration. According to this view, to assert that a theory belongs to our knowledge is just to assert that the theory helps us towards realizing the basic aim of articulating M in a precise, testable form. According to this view, a theory only belongs to our knowledge insofar as it is more or less compatible with M. The methodological rules of scientific inquiry are, then, in the first instance, rules which favor theories that are: (a) compatible with M; (b) empirically successful. On this view, theories that are incompatible with M will quite rationally be disregarded, however "empirically successful" they may be, for the simple reason that considering such theories cannot help us to articulate more and more of M.

It may be granted, provisionally, that as long as M, the metaphysical "blueprint" (as I shall call it), can be made to be sufficiently precise to be incompatible with all aberrant theories, then this view resolves the traditional Humean rationality

---

Some degree of non-*ad hocness*, simplicity, coherence, unity, harmony, is thus an essential ingredient of our idea of explanatoriness and even of our idea of scientific *law*. To see an intuitive link-up between *explanatory* and beautiful, we can argue as follows. To *explain* is to show why the thing to be explained *has* to be the way it is, why it could not be different. Thus we may "explain" a particular event E or law L by showing that E or L is part of an overall lawful pattern described by a theory T—a pattern which would be seriously disrupted if the event E did not occur, or if L did not hold. Clearly, this requires that the pattern described by T is reasonably nonarbitrary, is such that any modification to this pattern, such as the cancellation of E or L, more or less destroys the whole effect. In other words, the more amenable a theory is to modification and tinkering without essential loss of simplicity, the less explanatory that theory is. The more a theory resists tinkering, so that even the slightest change introduces a drastic loss of simplicity and coherence, the more explanatory that theory is. This fundamental feature of explanatoriness is however also a feature of beauty. A central doctrine of Western art, emerging from ancient Greece and from Renaissance Italy, is that the artist seeks perfection; he seeks to produce a work whose every detail embodies such perfect harmony and unity that even the slightest change would introduce a drastic discordant note. According to this classic conception of beauty, there is nothing arbitrary about a great work of art; there is nothing that could have been otherwise. The whole thing has an air of inevitability, of coherence, balance and unity. Thus we see that in terms of this classical conception of beauty, 'beauty', as applied to theories, can be more or less used interchangeably with 'explanatoriness'. It was, I believe, to this classical conception of beauty that Einstein was alluding when he stressed, as he repeatedly did, that a new acceptable theory must be beautiful.

The whole problem of simplicity, intelligibility, beauty will be raised again in **Part II**, where I shall consider in particular the problem of how simplicity assessments can be *objective* and *rational*.

problem of pure science *as long as the basic aim for science advocated by this view is rational.* But it is just here, it may be objected, that aim oriented empiricism fails. How, it may be asked, can it conceivably be rational to try to articulate, in a testable form, more and more of the metaphysical blueprint M when we can have no reason whatsoever for supposing that M is true? How can theories, accepted because they are compatible with M, be held to constitute *knowledge* when M can be no more than an entirely unfounded conjecture? Does not aim oriented empiricism transgress the most basic requirement of all for a discipline to be scientific, namely that nothing will be assumed permanently about the world *as an article of mere faith*, independently of all empirical considerations? In short, does not aim oriented empiricism characterize science as irredeemably *irrational*, in that it brands science as a kind of theology?

The problem before us is in essence this: How can it conceivably be rational to search for *intelligibility* inherent in the universe when we can have no reason whatsoever for supposing that intelligibility exists to be found? In order to answer this question we need to consider, in general, under what circumstances it can be rational to search for a thing, X, when we do not know that X exists. The answer I suggest is that the more *important* it is to us to discover X, so the more rational it will be for us to search for X, given that we do not know X exists.

Ordinarily we would not hold it to be very rational to search for water in a desert, just because water is very unlikely to exist in a desert. But if we are unfortunate enough to be lost in a desert, and our supply of water is running low, then it becomes entirely rational to search for water, for an oasis, just because discovering water has become so supremely important—our very lives depending on it. Our actions, in a sense, quite rationally presuppose the existence of an oasis somewhere nearby, not because we think it very likely that such an oasis exists, but rather because we have no alternative but to hope that such an oasis exists. In such a situation it becomes pointless to question that a nearby oasis exists, not because we are sure of the existence of the oasis, but rather because the nonexistence of the oasis leaves us no alternative but to give up and die.

I suggest that the situation is somewhat analogous to this as far as the search for intelligibility in science is concerned. From an intellectual standpoint (and ultimately also, I would wish to argue, from a practical, technological standpoint) our need for there to be intelligibility inherent in the universe is so enormous, so utterly irreplaceable, that we have no alternative but to take for granted that intelligibility does exist even though we can have not the slightest reason for supposing this assumption to be *true.* To cast doubt on the existence of intelligibility is idle, not because we have such wholly convincing reasons for holding it to be *true* that intelligibility does exist (quite the contrary!), but rather because if intelligibility does not exist at all, then our case is hopeless, and both science and, ultimately, life become impossible.

The suggestion I wish to put forward, is that as far as pure science is concerned, and in particular as far as physics and cosmology are concerned, the supreme value, the supreme desideratum and objective, is the discovery of harmony, order, simplicity, beauty or intelligibility inherent in the world. As Einstein once put it:

> . . . the cosmic religious feeling is the strongest and noblest motive for scientific research. Only those who realize the immense efforts and, above all, the devotion without which pioneer work in theoretical science cannot be achieved are able to grasp the strength of the emotion out of which alone such work, remote as it is from the immediate realities of life, can issue. What a deep conviction of the rationality of the universe and what a yearning to understand, were it but a feeble reflection of the mind revealed in this world, Kepler and Newton must have had to enable them to spend years of solitary labor in disentangling the principles of celestial mechanics! Those whose acquaintance with scientific research is derived chiefly from its practical results easily develop a completely false notion of the mentality of the men who, surrounded by a skeptical world, have shown the way to kindred spirits scattered wide through the world and the centuries. Only one who has devoted his life to similar ends can have a vivid realization of what has inspired these men and given them the strength to remain true to their purpose in spite of countless failures. ([4], pp. 39–40)

Again, Poincaré has put the point very forcefully like this:

> The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful, it would not be worth knowing, and if nature were not worth knowing, life would not be worth living. Of course, I do not here speak of that beauty which strikes the senses, the beauty of qualities and appearances; not that I undervalue such beauty, far from it, but it has nothing to do with science; I mean that profounder beauty which comes from the harmonious order of the parts and which a pure intelligence can grasp. This it is which gives body, a structure so to speak, to the irridescent appearances which flatter our senses, and without this support the beauty of these fugitive dreams would be only imperfect, because it would be vague and always fleeting. On the contrary, intellectual beauty is sufficient unto itself, and it is for its sake, more perhaps than for the future good of humanity, that the scientist devotes himself to long and difficult labors. ([22], p. 89)

The view that I wish to advocate can be seen as a slight modification of what Einstein and Poincaré here express. According to this view, the search for beauty, for harmony, for intelligibility, inherent in the world ought not to be seen merely as the purely private, personal motivation of the individual scientist of genius, having little to do with the public, overall aims and objectives of scientific inquiry (as the passage from Einstein might suggest): on the contrary, the quest for intelligibility should be seen as the fundamental public objective of pure science itself. We should see the whole *raison d'être* of pure science to be the discovery of beauty or intelligibility in the universe. And indeed as pure scientists we should hold the discovery of beauty, harmony, simplicity, intelligibility inherent in the world to be of such supreme importance, to be a wonder, a joy, of such outstanding value, that quite rationally we are willing to undertake the "long and difficult labors" of

scientific research merely on the *off chance* that what we hope to find really does exist. Empirically successful aberrant theories are ignored, not because we have good reasons for holding such theories to be *false*, but rather because a consideration of such theories cannot help us to realize our basic aim, even if some such theory is *true*. Aberrant theories, even if true, are dismissed because they can have no intellectual value. In short, if nature is not (sufficiently) beautiful then knowledge is *impossible*, not merely *worthless* (as Poincaré says), since according to aim oriented empiricism a theory only deserves to be considered a part of our "knowledge" to the extent that it helps us discover beauty in the universe.

Again and again, throughout the history of the natural sciences, at least from Kepler and Galileo onwards, creative scientists have spoken of their passionate desire to discover beauty, and of the vital role that this desire has played in leading them to the discoveries that they have made. And invariably these remarks have been interpreted as having relevance only for the *psychology* of discovery, having no bearing on the logic, the rationality, of scientific method and scientific discovery. I believe just the opposite of this. Unless we appreciate the enormous intrinsic value that natural science places on beauty, harmony and intelligibility, we cannot hope to understand the rationality of pure science. It is only if we appreciate the immense value placed on beauty or intelligibility, for its own sake as it were, that we will be able to grasp the simple point that scientists are prepared to persist in the hunt for beauty or intelligibility, despite the most appalling setbacks, and even though they cannot know that what they look for does really exist. It is not just that it would be psychologically impossible for Kepler, let us say, to have discovered his laws if he had not been fired by the conviction that there is an inherent harmony in the motions of the planets. Much more important, it would have been *irrational* for him to have investigated the motions of the planets in the way that he did if he had not held, at least as a possibility he passionately hoped to realize, that the planets move in accordance with precise, simple mathematical laws. And furthermore, without the aim of discovering mathematical harmony and simplicity in the motions of the planets, there could have been no *verification* of Kepler's laws. The actual, observed motions of the planets "verify" any number of easily invented, ugly, aberrant laws just as much as Kepler's simple, harmonious laws. It is only rational to *accept* Kepler's laws, given the observations, as being of sufficient intellectual value to constitute a genuine contribution to knowledge, to pure science, if we hold, with Kepler, that the supreme value, the supreme objective, of pure science is to discover a hoped-for harmony, beauty, intelligibility in the universe.

In short, the aim of seeking intelligibility in the universe is rational not because we have good reasons for supposing that this aim can be realized (in that intelligibility "really does exist") but rather because, as far as science is concerned, we place such a supreme value on the discovery of intelligibility that we are willing, quite rationally, to hunt for intelligibility even though we cannot know that what we seek really does exist.

There is now one crucial problem that must be solved if I am to make plausible my thesis that aim oriented empiricism succeeds in resolving the Humean pure rationality problem, in that it rebuts Humean considerations which seem to under-

mine the very *possibility* of pure science being rational. The problem is this: How is it possible to make a rational choice between rival blueprints, rival ideas as to how the world is intelligible? We cannot assess the truth or falsity of such rival metaphysical blueprints, since these theories will of course be empirically untestable. How, then, can it be possible to make a rational choice between rival specific aims for science, given that the general overall aim is to discover a hoped-for intelligibility in some sense or other?

Let us suppose that we have before us n rival proposed aims or metaphysical blueprints for science, $M_1, M_2, \ldots M_n$, each of which in effect attributes a different kind of intelligibility to the world (or to that part or aspect of the world which is the concern of our science). M_1, for example, might be the kind of view, held by many primitive people, that the world is full of demons and spirits, which activate the different things in the world, "natural phenomena" thus being explicable in terms of the aims and actions of these spirits. M_2 might be a rather more unified version of this view, according to which the whole cosmos is activated by one being, namely God. M_3 might be a somewhat depersonalized version of this view, according to which there is some overall cosmic purpose, all natural phenomena occurring in order to realize this purpose, and being explicable as tending to bring about the purpose. M_4 might be the kind of metaphysical system advocated by Plato, according to which material objects are imperfect copies of the ideal forms. M_5 might be Aristotle's teleological metaphysical system. M_6 might be the kind of corpuscular view held by seventeenth century natural philosophers. M_7 might be the point-atomistic view advocated by Boscovich. M_8, the Einsteinian view that the universe is made up entirely of one unified field. M_9 the rather more general Galilean view that the book of nature is written in the language of mathematics, or, a little more specifically, the view that the world is made up entirely of a few different sorts of fundamental physical entities which interact in accordance with a few relatively simple mathematical laws. M_{10} might be the vaguer, more instrumentalistic view that the world is such that the search for simple universal mathematical laws can meet with considerable empirical success. M_{11}, M_{12} to M_n includes, we may imagine, all the main wildly imaginative ideas about the nature of the universe that men have dreamed up during the course of recorded history, including many relatively *aberrant, ad hoc* blueprints, for good measure. The question before us is simply: How can we possibly make a rational choice between $M_1, \ldots M_n$, granted that we have no way of assessing the truth or falsity of the *theories* which these blueprints represent?

The crucial point to recognize here is that the task before us is *not* to assess rationally rival metaphysical thories about the world *whatever the world may be like*: rather the task is to assess rival metaphysical theories about a world *presupposed to be intelligible*. In other words, the task before us is not to assess rival metaphysical theories given that our basic aim is to discover (or choose) *truth as such*; rather the task is to assess rival metaphysical theories granted that our basic aim is to discover *intelligible truth*, to discover intelligibility we take for granted to be inherent in the universe (for reasons given above). We need, in other words, to choose between rival specific *aims* for science, rival statements $P_1, \ldots P_n$ of the

form: "We should aim at developing testable scientific theories compatible with the metaphysical theory M_r (granted that our overall aim is to discover intelligibility in the world)," where $r = 1, \ldots n$.

At this point it may be objected: but all this conflates two quite distinct things, namely, on the one hand rival *theories* $M_1 \ldots M_n$, and on the other hand rival *aims* for science, rival *proposals* for developing new scientific theories, $P_1, \ldots P_n$. Ought not these things to be sharply distinguished? Surely we cannot assume that choosing rationally between $M_1, \ldots M_n$ is the same problem as choosing rationally between $P_1, \ldots P_n$?

The fundamental point of aim oriented empiricism is just that within the context of scientific inquiry these two problems are precisely the same. According to aim oriented empiricism, in choosing between rival theories we are choosing between rival attempts to articulate intelligibility conjectured to exist in the world. Thus, according to aim oriented empiricism, if M_r is the most scientifically acceptable theory about how the world is intelligible from $M_1, \ldots M_n$, then P_r is inevitably the most rational choice of aim for science from $P_1, \ldots P_n$; and *vice versa*. Thus, as far as aim oriented empiricism is concerned, we do not really need to bother about the distinction between the two sets of statements $M_1, \ldots M_n$ and $P_1, \ldots P_n$. The term 'blueprint' is used here precisely as a technical term of aim oriented empiricism to unite (or rather stand indifferently for) the two ideas "most scientifically acceptable metaphysical theory about how the world (or the relevant domain of phenomena) is intelligible" and "best aim for science"—two ideas that may of course be distinct for standard empiricism.

There are, I suggest, essentially *two* ways in which we can choose rationally between the blueprints $M_1, \ldots M_n$, within the context of scientific inquiry, given the overall *aim* of scientific inquiry postulated by aim oriented empiricism.

To begin with, we can assess rival blueprints in terms of the relative *empirical success* of attempts to articulate precisely, in a testable form, these rival blueprints. In other words, we can assess the relative empirical success of the rival research programs to which the blueprints give rise, in the kind of way described by Lakatos [15]. Some blueprints (such as M_9 and M_{10} for example) can be construed as having led to enormous, ostensible empirical success, recognized by all, whereas other blueprints have at best only led to some empirical success acknowledged only by adherents of the blueprint in question. For example, miracles fit intelligibility criteria of specific religious blueprints, and hence can be recognized as constituting empirical success by those who uphold the relevant metaphysical blueprints: miracles do not, however, fit the intelligibility criteria of science, as it exists today, and hence are not recognized as constituting empirical success which undermines scientific blueprints. The reverse situation does not hold, however: those who uphold religious blueprints *are* obliged to recognize the empirical success of science.

It must, of course, be acknowledged that however much empirical success a research program based on some blueprint M_r may meet with, there will always remain the possibility that M_r is completely *false* (some aberrant blueprint perhaps being true). Ostensible success in attempting to realize a blueprint does not necessarily mean *real* success. We can, however, say this. If our attempt to articulate

in a testable form some blueprint M_r meets with relatively little empirical success, then we can at least be sure that we really are failing to make progress. Ostensible failure means *real* failure. In this way we can assess rationally the relative success and failure of rival attempts to discover intelligibility even though we cannot assess rationally the truth or falsity of the blueprints on which these rival attempts are based.

This kind of Lakatosian empirical assessment of rival blueprints, although of course of decisive importance for science, nevertheless cannot *on its own* take us very far. For given that one research program, based on M_9 let us say, has led to far greater empirical success than any rival program, we can nevertheless easily accommodate rival blueprints to this success and of course we can always argue that *aberrant* versions of M_9 can be construed as having just as much empirical support as M_9 itself. We may try to favor M_9 on the grounds that M_9 will not have been deliberately modified to accommodate the empirical success whereas rival blueprints will have been so modified; however we will not be able to give a *rationale* for favoring M_9 on these grounds as far as future developments are concerned. As we saw in section **4**, the attempt to provide a rationale for our preference for non-aberrant empirically successful research programs in terms of Lakatosian criteria of empirical progressiveness alone, must fail.

According to aim oriented empiricism, however, we have an entirely rational and wholly *a priori* way of appraising rival blueprints that is entirely *in addition* to all empirical considerations. Since the fundamental aim of science is to discover intelligibility inherent in the universe, since our science places an intrinsic premium on intelligibility, we will always be rationally entitled, other things being equal, to choose the most intelligible blueprint.

Whereas *empirical* considerations, on their own, do not suffice to pick out, in a rational fashion, just one preferred blueprint from our list, and whereas *a priori* intelligibility considerations on their own do not suffice to do this either, the two types of considerations working in conjunction will, I suggest, enable us to make a rational choice of just one type of blueprint from the list. Although science does, according to aim oriented empiricism, have some features in common with theology, nevertheless science is rationally distinguishable from theology, and deserves to be valued as a more successful attempt to discover intelligibility in the universe than any rival discipline, theological or otherwise.

7. Aim Oriented Empiricism provides a Rational Method of Discovery. We saw, at the end of section **2**, that to have a rational method of discovery in science is to have a rational method for choosing between embryonic ideas for future scientific theories, a rational, purely *a priori* procedure for choosing between rival metaphysical blueprints. We saw that as long as standard empiricism is accepted, no such rational method of discovery is possible.

The moment aim oriented empiricism is accepted, however, rational discovery does become a possibility. Indeed, rational discovery becomes the very heart of the rationality of science! As we have seen, rational *acceptance* of testable theories, in terms of their empirical success, is only possible if we can rule out empirically

successful aberrant or *ad hoc* theories. But we can only do this if we can make a rational choice of blueprint in a wholly *a priori* fashion. Thus, according to the view advocated here, rational acceptance of theories depends on there being a rational *a priori* procedure for choosing between rival blueprints, which in turn provides the machinery for a rational method for developing new hypotheses. Thus, according to aim oriented empiricism, there is only rational acceptance if there is rational discovery!

Standard empiricism excludes the possibility of rational discovery in science because it excludes the possibility of there being any *a priori* knowledge about the world (in terms of which rival ideas for future research may be assessed). In terms of the aim oriented concept of knowledge, however, we do have *a priori* knowledge about the world! For according to aim oriented empiricism, only those theories which have a certain inherent intelligibility can have sufficient intellectual value to be considered a part of our "knowledge." We "know" *a priori* that the theories we accept will be intelligible. It is this *a priori* demand that acceptable theories be intelligible which makes the purely *a priori* investigation of rival possible blueprints (in terms of their intelligibility) a potentially fruitful procedure for developing valuable new scientific theories.

None of this of course means that there is an *infallible* or *mechanical* procedure for developing new theories. According to aim oriented empiricism we are rationally entitled to choose the simplest, the most intelligible embryonic idea for a future theory, and make this precise and hopefully testable in the simplest way we can think of; there is nothing mechanical about such a process, however, and aim oriented empiricism provides us with no assurance that this procedure will infallibly meet with success.

With hindsight, we can see that in order to resolve the Humean rationality problem for pure science it is necessary to develop a theory of scientific inquiry which provides something like a rational method of discovery. For the reformulated Humean argument, developed in section **3**, consisted essentially of a procedure for developing theories more empirically successful than existing scientific theories (ostensibly a foolproof method of discovery). In order to outlaw such theories, one has to outlaw certain procedures for the construction of new empirically successful theories. In other words, one has to place *a priori* (that is, nonempirical) restrictions on the field of new hypotheses worthy of consideration. But to delimit in this *a priori* fashion the field of new hypotheses in in effect to provide a more or less useful rational method of discovery. (The more restrictive such an *a priori* delimitation is, the more helpful it will be—the ideal being a delimitation that is so restrictive that only *one* new theory is left.) That such an *a priori* delimitation might have relevance to scientific practice is suggested by the fact that no precise line of demarcation can, it seems, be drawn between absurdly aberrant theories, which never get considered in science, and only moderately aberrant theories, which may well be considered by some, and dismissed by others as "*ad hoc*."

8. Concluding Remarks. In this first part of the essay, my concern has been to argue that standard empiricism profoundly misrepresents the true aims of science, and

thus constitutes a false, deplorable ideal of scientific rationality and propriety, as evinced by the fact that no version of standard empiricism is able to rebut elementary Humean considerations which appear to show that science cannot possibly be rational. I have argued that in order to overcome the Humean problem as far as pure science is concerned we need to adopt the view that the fundamental aim of science is to articulate, to make precise and testable more and more of a presupposed metaphysical conjecture—namely, that the world is simple, coherent, unified, beautiful or intelligible. In **Part II** of the essay I shall spell out in a little more detail the theory of scientific method that becomes inevitable once we accept this aim oriented empiricist view; I shall tackle the important problem of how metaphysical blueprints are to be rationally, objectively appraised with respect to their inherent intelligibility; and I shall indicate a number of important implications which this viewpoint has for scientific inquiry itself.

Before we consider further developments of the view advocated here, however, we should perhaps first consider how this view differs from views already current in the literature. I have already remarked that the view advocated here would appear to be in all essentials in agreement with the view espoused by Einstein, with the difference that Einstein appears to have held that this view constitutes something of a "miracle creed," whereas I have argued that the view in fact succeeds in representing pure science as a truly rational enterprise, whereas more orthodox and ostensibly less "miraculous" views of science fail to do this.

As far as contemporary philosophy of science is concerned, the view that appears to come the closest to the view advocated here is perhaps Lakatos' "methodology of scientific research programs" ([14], [15]). Let us then consider briefly how Lakatos' position differs from the one advocated here.

According to Lakatos, we need to see science as a struggle between competing research programs. Each program consists of a "hard core," the central theory of the program, which, from a methodological point of view, is treated as metaphysical, in that from within the program it is protected from experimental refutation. The research program proceeds by building up auxiliary hypotheses, and it is only these auxiliary hypotheses which are allowed to be refuted. In order to survive, a program needs constantly to predict novel phenomena, i.e. to lead to empirical growth; and a program is eliminated if a rival program leads to a greater rate of empirical growth.

The analogy with the view advocated here is quite clear; Lakatos' methodologically metaphysical "hard core" looks very like the metaphysical "blueprint" of science, inherent in the aims of science, on which I have lain so much emphasis. Am I not simply advocating Lakatos' views with different terminology? Does not the aim oriented view advocated here amount simply to Lakatos' claim that we may see the whole of science "as a huge research program" ([15], p. 132)?

The first point to note is that Lakatos, in developing his theory, was tackling a quite different *problem* from the one considered here. Lakatos' underlying problem might be put like this. Granted Popper's fundamental point that science is rational just because it subjects its theories to ruthless criticism, to severe empirical testing, how are we to do justice to the apparent *dogmatism*, the bland *neglect* of criticism

and refutations typical of so much science, as Kuhn, in particular, has emphasized (e.g. in [12])? How can Popper's position be modified so as to accommodate both the dogmatic retention of certain key theories or "paradigms" despite ostensible refutations, and the element of continuity in scientific development that arises as a consequence?

Lakatos' solution to this problem is essentially to exploit the idea that there is a methodological *point* to dogmatism. All theories are in a sense born refuted. If we took literally Popper's edict that theories in empirical difficulties should be rejected, all progress would become impossible. In order to let criticism *encourage* progress, and not stifle it, we need to temper the severity of our criticism. In particular, powerful new theories, that have successfully predicted unexpected new phenomena, need to be treated leniently, at least for a while, so that they get a chance to reveal their latent strength. In short, we need to be *critical* in our application of criticism, if we are not to stifle all empirical progress. Thus a good theory can quite rationally be protected for a while from refutation, as long as it leads to successful empirical growth. Lakatos' methodology of competing research programs, with its *critical* application of criticism, is thus a more rational scheme than Popper's methodology of unchecked criticism, unchecked falsifiability.

It is quite clear, then, that Lakatos' methodology is a *refinement* of Popper's, which does not question Popper's fundamental point that the aim of science is to seek empirical growth *per se* by letting competing theories (or research programs) struggle for survival. And this means that Lakatos' theory cannot be interpreted as being equivalent to the aim oriented view advocated here, the view, that is, that science should be seen as one gigantic research program which takes as its aim the articulation of a presupposed metaphysical system chosen for its inherent simplicity and intelligibility. Viewing science as this kind of monolithic research program violates the crucial Popper–Lakatos requirement that there must be *competing* theories or research programs if there is to be critical, rational progress towards the truth. It violates the fundamental Popper–Lakatos thesis that the fundamental aim of science is empirical growth *per se*. And in addition it violates the Popper–Lakatos thesis that apparently *a priori* simplicity criteria in science are to be understood as criteria which favor theories of high empirical content, or programs of high empirical progressiveness.

Lakatos is himself quite clear on all these points. Thus he says only: "Even science as a whole can be regarded as a huge research program with Popper's supreme heuristic rule: 'devise conjectures which have more empirical content than their predecessors'" ([15], p. 132). Lakatos explicitly rejects the idea that the aim of science is to progress towards the "Blueprint of the Universe" ([15], pp. 188–189). And, in addition, he is emphatic that criteria of simplicity or non-*ad hocness* should not be matters of mere "taste," but should be reduced to criteria of empirical progressiveness ([15], pp. 105; 117–124; 175).

There is a further point of difference between the aim oriented view advocated here and Lakatos' methodology. As we have seen, Lakatos' basic idea is that, since our aim is to achieve empirical growth, we ought rationally to favor those theories which have led to empirical growth, even to the point of ignoring for a time some

apparent refutations. In other words, a theory, in order to become the "hard core" of a research program, and thus win the honor of achieving temporary unfalsifiability, must meet with some initial degree of empirical success. A hard core must be an empirically successful, testable theory which we then *decide* to turn into a metaphysical theory, from a methodological standpoint, for the time being. This means that Lakatos' notion of a "hard core" is very different from the idea of a "metaphysical blueprint" developed here. For the metaphysical blueprints which are, I claim, inherent in the aims of science, are irredeemably *metaphysical*; in no sense is the blueprint of a fundamental science such as physics a theory which has already had some empirical success. It is chosen, not because it has itself produced empirical growth, but because of its inherent intelligibility or simplicity; it is chosen for largely *a priori* reasons.

It may seem that all this is splitting hairs, and that Lakatos' basic position is, after all, not so very different from the aim oriented viewpoint advocated here. But from the point of view of solving our basic problem, the pure rationality problem, the differences are in fact decisive. For Lakatos' view is in the end a very sophisticated version of what I have called standard empiricism, in that Lakatos maintains that science seeks empirical growth *per se*, and does not make permanent, overall metaphysical presuppositions concerning the ultimate intelligibility and simplicity of the universe. And as a consequence of this, Lakatos' viewpoint fails to solve the pure rationality problem, as we have already seen in section **4.** Lakatos' theory provides no rationale for rejecting empirically progressive research programs based on aberrant "hard cores." In fact, even worse, Lakatos' theory provides no rational procedure for choosing between competing programs when we really need to make the choice, namely when both are more or less viable options. According to Lakatos, it is only long afterwards that such a rational choice can be made.

The essential point here is that Lakatos' theory, in that it is a refinement of Popper's version of standard empiricism, can provide no rationale for choosing between rival hard cores, C_1 and C_2 let us say, in a purely *a priori*, nonempirical way, *in advance* of empirical success or failure. For according to Lakatos, the fundamental aim of science is to seek empirical growth *per se*; *a priori* we cannot decide rationally which of C_1 or C_2 is the most likely to lead to empirical progress, since we do not possess *a priori* knowledge about the world (in terms of which such a choice could be made). Thus, any rational *a priori* choice between C_1 and C_2 is impossible. This means first that there can be no rational procedure for the development or discovery of good, new, hard cores. But second, and even worse, it means that there can be no way of choosing rationally between C_1 and C_2 as far as scientific inquiry itself is concerned. If both C_1 and C_2 have been pursued for some length of time in the past we can of course judge which has met with the greatest empirical success during that period. But for Lakatos, this can provide no rational basis for deciding which hard core will be the most successful *in the future* (which is, of course, all that interests the research scientist). As Lakatos himself emphasizes, a program may suffer repeated setbacks, and then may suddenly become amazingly progressive. Nature, although uniform in time, may be so structured that it is some

program which goes through a long phase of degeneracy and stagnation (in comparison with others) which in the end will lead to enormous empirical growth. In short, in order to choose rationally between C_1 and C_2 we need to be able to *anticipate* on a rational basis, the future performance of C_1 and C_2; which in turn requires that we make a choice between hard cores independently of past empirical success. We need, in other words, to be able to make a rational choice in terms of *a priori* considerations; we need a rational method of discovery. Because Lakatos can provide no rational method of discovery for hard cores, no *a priori* way of choosing between hard cores, he can provide no rational procedure *at all* for choosing between rival, viable hard cores, except in the context of history.

Both Lakatos and Kuhn agree that in revolutionary situations in science no rational choice between rival hard cores or paradigms can be made *at the time*, but, at best, only long after the event. According to aim oriented empiricism, however, such a rational choice *is* possible, since we can assess in an *a priori* fashion, the relative simplicity or intelligibility of the rival hard cores or paradigms, the promise which they hold out of realizing the basic metaphysical blueprint M of the science. For Lakatos and Kuhn, there is no overall continuity in science discernible in successive hard cores or paradigms: according to aim oriented empiricism however, the basic blueprint, M, of a science does change and evolve in a continuous and rational fashion, ideally, as we shall see, M becoming progressively more and more intelligible as purely *a priori* (or conceptual) problems connected with M are resolved. A science which fully articulated its basic blueprint, M, and actively sought to *improve* M—something which present day scientists are reluctant to do out of a fear of being metaphysical and unscientific—would be very different from Kuhn's view of science as a succession of normal and revolutionary episodes, and also very different from Lakatos' view of science as a perpetual struggle for survival between competing fragments of normal science, competing "research programs."

REFERENCES*

[1] Ayer, A. J. *The Problem of Knowledge.* Middlesex: Penguin Books, 1965.
[2] Barker, S. F. *Induction and Hypothesis: A Study of the Logic of Confirmation.* Ithaca: Cornell University Press, 1957.
[3] Duhem, P. *The Aim and Structure of Physical Theory.* New York: Atheneum, 1962.
[4] Einstein, A. *Ideas and Opinions.* London: Alvin Redman, 1954.
[5] Goodman, N. *Fact, Fiction and Forecast.* London: Athlone Press, 1954.
[6] Goodman, N. *Problems and Projects.* New York: Bobbs-Merrill, 1972.
[7] Hempel, C. G. *Aspects of Scientific Explanation.* New York: Free Press, 1965.
[8] Hume, D. *A Treatise of Human Nature.* London: J. M. Dent and Sons, 1959.
[9] Jeffreys, H. *Scientific Inference.* Cambridge: Cambridge University Press, 1957.
[10] Kant, I. *Critique of Pure Reason.* Translated by N. K. Smith. London: Macmillan and Co., 1961.
[11] Koyré, A. *Metaphysics and Measurement.* London: Methuen, 1973.
[12] Kuhn, T. S. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 1970.
[13] Kuhn, T. S. "Logic of Discovery or Psychology of Research?" and "Reflections on my Critics." In *Criticism and the Growth of Knowledge.* Edited by I. Lakatos and A. Musgrave. Cambridge: Cambridge University Press, 1970.

* Reference citations in **Part I** are keyed to only this **Part I** list.

[14] Lakatos, I. "Criticism and the Methodology of Scientific Research Programmes." *Proceedings of the Aristotelian Society* 69 (1968): 149–186.
[15] Lakatos, I. "Falsification and the Methodology of Scientific Research Programmes." In *Criticisms and the Growth of Knowledge*. Edited by I. Lakatos and A. Musgrave. Cambridge: Cambridge University Press, 1970.
[16] Lakatos, I. "History and its Rational Reconstructions." In *Boston Studies in the Philosophy of Science*. Vol. 8. Edited by R. C. Buck and R. S. Cohen. Dordrecht: D. Reidel Publishing Co., 1971.
[17] Mach, E. "The Guiding Principles of My Scientific Theory of Knowledge and Its Reception by My Contemporaries." In *Physical Reality*. Edited by S. Toulmin. New York: Harper Torchbooks, 1970.
[18] Maxwell, N. "Can there be Necessary Connections between Successive Events?" *British Journal for the Philosophy of Science* 19 (1968): 1–25.
[19] Maxwell, N. "A Critique of Popper's Views on Scientific Method." *Philosophy of Science* 39 (1972): 131–152.
[20] Maxwell, N. *The Aims of Science*. (forthcoming)
[21] Meyerson, E. *Identity and Reality*. London: George Allen and Unwin, 1930.
[22] Poincaré, H. Quoted in J. Hildebrand, *Science in the Making*. New York: Columbia University Press, 1957.
[23] Popper, K. R. *The Logic of Scientific Discovery*. London: Hutchinson, 1959.
[24] Popper, K. R. *Conjectures and Refutations*. London: Routledge and Kegan Paul, 1963.
[25] Popper, K. R. *Objective Knowledge*. Oxford: Oxford University Press, 1972.
[26] Reichenbach, H. *The Rise of Scientific Philosophy*. Berkeley: Univeristy of California Press, 1958.
[27] Russell, B. *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster, 1948.
[28] Rudner, R. S. "An Introduction to Simplicity." *Philosophy of Science* 28 (1961): 109–119.
[29] Scheffler, I. *The Anatomy of Inquiry*. New York: Alfred A. Knopf, 1963.
[30] Swinburne, R. G. "Choosing Between Confirmation Theories." *Philosophy of Science* 37 (1970): 602–613.
[31] Swinburne, R. G. (ed.) *The Justification of Induction*. London: Oxford University Press, 1974.