THE CORRESPONDENCE PRINCIPLE

AND

THEORY CHOICE IN PHYSICS

BY

YOUSSEF SAMADI ALIABADI

London School of Economics

Submitted in Fulfilment of the Requirement for the

University of London Degree of Ph.D

June 1996

RESUME

A conception of the Correspondence Principle which Bohr deployed implicitly in developing a new theory of atomic constitution in 1913, is made explicit through an extensive examination of his classic paper of that year. Arguments are considered which purport to show that the application of the principle must be restricted to few isolated cases. These arguments are either defused or rejected. In particular an extensive review of issues concerning the interpretation of Quantum Mechanics is made to counter the claims that an insurmountable conceptual gap exits between the tenets of this theory and those of Classical Mechanics which makes it logically impossible for the latter to be regarded as the 'limiting case' of the former. In the light of a particular interpretation adopted and defended, a proposal is made that suggests that the Hamilton-Jacobi formulation of Classical Mechanics, as well as Maxwell's electromagnetic theory, can be viewed as 'limiting cases' of Quantum Mechanics. Having established a case for the global validity of the requirement imposed on physics by the Correspondence Principle, it is then argued that this requirement is indispensable if a particular brand of realism is adopted for the interpretation of theories in physics. Taking on board the assumption that an ultimate theory exists which mirrors the underlying physical constitution of the world, it is subsequently argued that the intertheory order established by the global imposition of the principle in physics, can be used to solve the problem of rational theory choice for this brand of realism.

CONTENTS

CONTENTS / continued ...

CHAPTER 5

Is Quantum Mechanics Inconsistent With the Demands of the Generalized Correspondence Principle?

CHAPTER 1

INTRODUCTION

The question of what counts as the right approach towards the solution of the problem of theory choice in science, depends very much on how this problem is formulated. If the problem is viewed as how best to select theories which *are not* suited to be believed (so far as rationality of choice is concerned), then Popper's falsificationist approach may be regarded as an adequate step in the right direction. But if the problem is considered to be how to select theories which *are* worth believing, this approach, on its own, can no longer serve as adequate. It must either be augmented with additional considerations or else replaced with an altogether different approach. Alternatives in both these categories already exist: Watkins [1984] offers an approach of the first kind, and the varieties of Baysianism may be considered as species of the second.

With perhaps the possible exception of the objective brand of Baysianism, these approaches share in keeping the notion of truth basically out of their prescription for theory choice. The main reason for this may be cited as difficulties, which following the failure of Popper's account of verisimilitude, were perceived by many authors to stand in the way of formally defining the notion of proximity of a false theory to the truth. Many people, however, feel that truth is too attractive a notion to be given up in accounting for rational theory choice in science. This feeling is particularly shared by those who adopt a strong realist outlook on scientific theories. If science is a genuine *cognitive* enterprise, and if scientific theories purport to describe the world as it really is, the feeling is that, in comparison to its rivals, the merit of a theory for being believed *ought to* reside in the greater similarity of what it says to the truth. In the absence of a knock out argument that shows this feeling to be based on a mere illusion, the search for an approach to the problem of rational theory choice which incorporates the notion of truth is, therefore not an unrewarding exercise.

The motivation for this thesis is the exploration of one possibility towards this end. The idea is the following: Suppose in a branch of science an ordering can be established between theories which somehow ranks each according to its explanatory, as well as predictive capability with sufficient clarity. Then the assumption that this ordering converges to a theory that surpasses all others in these respects, would help to set up a progression which may be exploited to indicate the relative place of a theory vis à vis the one which forms the ideal in that branch. With a strong realist gloss, this ideal may be taken to represent that theory which mirrors the underlying constitution of the world in the relevant field. Now, such a theory may be taken as a surrogate for the truth in the corresponding domain, and consequently the progression in question may be taken as representing a verisimilitude ordering in the given branch. In this way, a reasonably clear representation for the *relative proximity* of theories in a branch of science may be obtained while circumventing the difficulties associated with the formal definition of this notion.

In 1913 Niels Bohr suggested an original idea which he exploited ingeniously to develop a radically new theory about the constitution of atoms. The idea may be roughly stated as follows: Suppose in a branch of physics a theory exists which has proved spectacularly successful in accounting for and predicting a wide range of phenomena. Should, as a result of experimental research into the further depths of the field, circumstances arise which call for the emergence of a new theory, of all the theories that can possibly succeed where the old one had failed, only those should be considered as worthy of an established place within the branch which can at least reproduce its successes in the appropriate domain. This demand, which essentially amounts to a call for the preservation of *continuity* in a branch of physics, is known as the Correspondence Principle. Two questions immediately arise here: (1) Precisely what sort of inter-theory relation must be envisaged between successive successful theories in a branch of physics if this demand for continuity is to be realistically fulfilled? (2) Can this principle be generalized to hold in all branches of physics? Issues surrounding these questions are contentious and there is by no means a consensus among authors who have taken an interest in them.

The aim in this thesis is not to settle accounts on these issues. They are nevertheless raised and dealt with for the purposes of *exploring* the potentials of the Generalized Correspondence Principle in establishing a progression between several major theories from various branches of physics, which can then be used (along with additional assumptions) to represent a quasi-verisimilitude ordering. Towards this end a conception of the Correspondence Principle is needed which can not only hold its ground in particular cases of interest, but also lend itself to generalization. To find this conception the strategy will be to go back to Bohr and start by examining his intuitions on the subject. The reason for this choice is not only that Bohr was the first physicist who introduced the Principle and exploited it in the practice of developing a series of theories which culminated in the birth of Quantum Mechanics, but also he envisaged it as holding generally in any case where a new theory is needed due to the failure of attempts to extend the resources of an otherwise successful theory to new domains.

The thesis gets under way in chapter 2 with an investigation of Bohr's deployment of the Correspondence Principle in his [1913]. The term 'Korrespondenzprinzip', according to van der Waerden [1967], p. 7, first appeared in the literature in a paper published by Bohr in 1920 (Z. Phys. 2, p. 423). The concept, however, was introduced and heuristically exploited by him seven years earlier in his classic 1913 paper. During the period when Quantum Mechanics was being developed, there are notable references to, and further articulations of, the Principle in several papers which Bohr published either by himself or jointly with other authors. Nevertheless, throughout this period not only did the concept remain essentially unchanged, but also his deployment of it in successive stages of working out a new theory of atomic constitution paralleled that inaugurated in his [1913]. The choice of this paper is therefore justified by the fact that it contains all the ingredients which underwrite Bohr's intuitions about the application of the Correspondence Principle in physics. There is also the additional advantage that the heuristic use of these intuitions is conspicuously displayed in this paper thus throwing further light on how the Principle was conceived by its author.

In Bohr [1913], the Principle appears as an assumption (without yet receiving its name) according to which for a certain range of values of a variable parameter, the predictions of the new theory 'coincide' with those of the theory to be replaced. This 'coincidence' is then invoked to legitimize the use of a hypothesis in the old theory in order to find the value of an unknown coefficient in the new theory. In Bohr [1918], however, this idea is generalized, appearing first as an expectation and later as a requirement that *any* new theory must satisfy if it is to be an acceptable replacement for the overthrown theory. Having introduced the idea, radically at odds with the classical notions, that electromagnetic radiation is emitted from or absorbed by atoms only during a discontinuous jump from one stationary state to another, Bohr writes: 'If we next consider a transition between two stationary states, it is obvious at once from the essential discontinuity... that in general it is impossible even approximately to describe this phenomenon by means of ordinary mechanics or to calculate the frequency of the radiation absorbed or emitted by such a process by means of ordinary electrodynamics. On the other hand, from the fact that it has been possible by means of ordinary mechanics and electrodynamics to account for the phenomenon of temperature-radiation in the limiting region of slow vibrations, we may expect that any theory capable of describing this phenomenon in accordance with observations will form some sort of natural generalisation of the ordinary theory of radiation.' (Bohr [1918], reprinted in van der Waerden [1967], p. 99). Soon after in the paper, what in the quoted passage appeared as an expectation is proposed as 'the necessary relation' (Ibid, p. 101), and 'the necessary connection' (Ibid, p. 110), that should hold between the new and the old theories.

In the specific context of working out a viable theory of atomic structure, Bohr's intuitions about a new theory being a 'natural generalization' of the classical electrodynamics is premised on the type of success which the latter enjoyed in adequately accounting for a wide range of phenomena. It is, therefore, natural to expect that the 'necessary relation' he has in mind should obtain in a wider context whenever a theory to be replaced has, *at least*, been established as enjoying similar successes. Feyerabend and Hanson have launched some rather influential arguments which conclude by suggesting that this expectation is unwarranted. These arguments will be presented in chapter 3. Feyerabend's argument will be examined in this chapter and found to attack only an assumption which Bohr's conception of the Correspondence Principle does not contain.

Hanson's case, on the other hand, is premised on the claim that the machinery deployed by the formalism of Quantum Mechanics is so alien to that deployed by the Classical Mechanics that a convergence, in any sense, between the concepts of the two theories would be impossible. This he presents as an insurmountable obstacle which prevents any 'correspondence' to hold between the two theories. It will be claimed that Hanson's case is based on an interpretation of the Quantum formalism which is not the best available. In particular, it suffers from the defect of making psychology an integral part of physics. It will be argued that there is nothing in the machinery deployed by the Quantum formalism to make this position an unavoidable consequence. On the contrary, a family of interpretations for this formalism exits which are superior to the former in that they do not share *this particular* problem. This is not to say that they are without problems of their own. Rather, seeking possible solutions to these problems does not require hiring help from disciplines other than physics itself. The case for these claims will be fully developed in chapter 5, at the end of this thesis, because it involves a wide range of issues concerning the interpretation of Quantum Mechanics. The aim of that rather long chapter, however, will remain basically to present a counter-example to Hanson's argument.

The conception of the Correspondence Principle which is found not only to best capture Bohr's intuitions and practice, but also suited to the aim pursued in this thesis, turns out to be unlike any extant ones in the literature on the subject. Depending on the kind of entities theories are taken to be, attempts at articulating the Generalized Correspondence Principle may be broadly divided into two groups. One group views theories roughly as sets of *propositions* which describe events and processes in physical space-time together with laws that govern them. It is fair to say that this view, which agrees well with common-sense, enjoys a wide-spread support among philosophers of science and for this reason it may be branded the 'standard' view. The rival group utilises model theoretic techniques and regards theories as sets of mathematical *models* (intended or potential) in which the formal equations proposed by a theory hold. We may call this the 'non-standard' view. These views will be briefly discussed in chapter 3 where the standard view be found to suit the purposes of this thesis better than its rival.

Within the standard tradition, there are basically two strands which nevertheless share the view that in the transition between theories related by the Correspondence Principle, something must remain somehow unchanged. According to one strand the 'correspondence' holds as a relation solely between the mathematical equations of the superseded and superseding theories. As some parameters in some equations of the superseding theory are allowed to tend towards an extreme value (either zero or infinity), these equations would tend increasingly to look like their counterparts in the superseded theory. In the limit where these extreme values are assumed, it is claimed that the respective equations would actually become *identical*. We may call this strand the 'structural' conception of the generalized Correspondence Principle. According to the other strand, the Correspondence Principle demands that part of the *ontology* of the superseded theory carries over, unchanged, to that of the superseding one. We may call this the 'ontological' conception. Representative proponents of both these conceptions will be presented in chapter 3 and their accounts will be tested against the paradigm case contained in Bohr [1913]. Both will be found wanting.

The alternative version of the Generalized Correspondence Principle conjectured in this thesis rests within the standard tradition. Unlike the extant strands within this tradition, however, this conception does not *require* that anything should remain unchanged in the transition from the superseded to the superseding theories. There will be a common range of values for the parameters of both theories for which their respective equations will have, on the whole, nearly the same empirical consequences. In general, however, it is neither the case that some of these equations can be 'turned into' the others, nor is it the case that there will be significant portions in the ontologies proposed by the theories involved which escape change in the transition. All that is required is that the superseded theory should, in a sense that I will try to specify, be a *good approximation* to the superseding one. This conception of the Generalized Correspondence Principle, which we may call *'approximational'* will be presented in chapter 3. It should be emphasized, however, that this presentation will be exploratory rather than critical; it is intended to serve the purpose of seeing how much mileage may be got from it in setting up the desired progression for some successive major theories in physics. In this connection, the case involving Quantum and Classical Mechanics, for the reasons already mentioned, will be dealt with separately in the last part of this thesis, chapter 5.

The motivation for choosing the approximational conception is twofold. First, the cases in which either the structural or the ontological correspondence hold, appear to fall out naturally as its special cases. Second, - and this is more important for the purposes of this thesis - it appears that the approximational conception, if tenable, can lend support to a particular brand of realism for handling the problem of rational theory choice in physics. The brand of realism in question is one among several positions of varying strengths which constitute the general realist outlook about scientific theories. What all these positions have in common can, perhaps, be best highlighted in a comparison with the instrumentalist outlook about such theories. According to the latter, the merits of scientific theories is exhausted by their success in predicting observable phenomena. Thus construed, anything these theories appear to say about the underlying constitution of the world is either vacuous or else irrelevant to their actual epistemic value. Realists, on the other hand, share instrumentalists' high regard for empirical success, but consider it only as *one of the merits* of a scientific theory. To them, what a theory says is neither vacuous nor irrelevant to its cognitive role; a scientific theory is comprised of genuine propositions which are either true or false depending on how the world is independently of human cognition and purposes.

The brand of realism I will concentrate on goes even further. It incorporates, in addition, the view that scientific inquiry is capable of providing, not just a series of 'educated' conjectures, but genuine information about the deep structure of the world. According to this position, a theory which satisfies all the standard methodological desiderata and has accumulated an impressive record of empirical success, contributes to *our knowledge* of the world even though it may turn out to be false. An adherent to this brand of realism is thus not only curious to find out what underlies the observable features of the world, but would turn partly to physics in order to satisfy this curiosity. Because this position rests, in addition to the *ontological commitment* common to all brands of realism, on an additional *epistemological commitment*, I shall call it *'two-tier'* realism. Arguments for or against this position will not be considered in this thesis. It will be presented in chapter 3 where a case is going to be made for why two-tier realism requires a principle that can safeguard the achievements of an outstanding but overthrown theory. It will be found that one purpose which the approximational conception of the

Correspondence Principle can serve appears to be the fulfilment of this requirement.

Two-tier realism, in addition to being attractive to common sense, is not without a pedigree in contemporary philosophy of science. Karl Popper, at least at one stage in his philosophical career, was certainly committed to it. This was the period when he introduced and tried, without success, to articulate the idea of *verisimilitude*. This idea, however, remains of vital relevance to a two-tier realist. In the absence of any direct cognitive access to the underlying constitution of the world, he must be content to turn to scientific inquiry in order to gain what knowledge he can about it. This, in turn, will bring him face to face with the problem of rational theory choice in science.

Here, there are certain (essentially Popperian) methodological constraints that would have to be met by any theory which is to be acceptable in science. These include internal consistency, unity of axioms, symmetry of laws, exclusion of ad-hoc stratagems, and last but not least, testability. If in addition to satisfying these constraints a theory also meets with spectacular empirical success under stringent testing conditions, a two-tier realist may be tempted, but cannot rationally conclude that it is true. The reason, which we owe to Popper, is that we know (a) verification of scientific theories is logically impossible, and (b) there are theories of this kind which had to be rejected subsequently as false. So, the only option that is left open to a two-tier realist is to regard a historical sequence of such theories as *progressing towards the ultimate truth* about the underlying constitution of the world. This option leaves science a promising kind of inquiry: although it cannot offer a guarantee to deliver the truth at any time, it churns out theories which proceed in the direction towards it.

After failing to find a precise and viable definition for 'verisimilitude', Popper gave up attempts at bringing this notion to bear on the problem of rational theory choice in science. As a consequence, his account of theory choice remains, so far as the aspirations of a two-tier realist are concerned, essentially negative: it can tell him which theories not to believe, but it cannot *recommend* the one which contributes the most to our knowledge of the world. A brief discussion of some issues concerning the notion of verisimilitude and Popper's account of it is presented in chapter 4. However, the bulk of that chapter will be devoted to an exploration of the following question: how can the approximational conception of the Generalized Correspondence Principle be exploited by two-tier realism to fill the gap left by the absence of a formally precise and viable definition for the concept of verisimilitude? This is not to say that this concept remains undefined; indeed there are more than one attempts at formal definitions for it in the literature. The point is rather to explore what needs to be done, with the approximational conception at hand, towards the fulfilment of the aspirations of a two-tier realist concerning the question of rational theory choice in physics, without having to formally define the notion of verisimilitude.

On the whole, the purpose followed in this thesis is to take up some ideas in order to see how they can be brought to fit together to form, rather like a jigsaw, an account of theory choice in physics. Such an account is desirable given the point of view of two-tier realism, but is missing in the literature. Wherever I have found it necessary, I have tried to make the ideas not only as clear as I can, but also plausible. I recognize, however, that this is no guarantee that they are either free of flaws, or indeed even tenable. Delving into questions pertinent to these issues can no doubt form the purpose of further study.

CHAPTER 2

BOHR'S CONCEPTION OF THE CORRESPONDENCE PRINCIPLE

1. General Features of Bohr's 1913 Theory

The theory of atomic constitution proposed in Bohr [1913], despite sharing many of its features with other existing theories at the time, has a radically new feature. It boldly postulates a process which is in breach of all the familiar mechanical explanations, and is not itself explained in any way. One could argue that the same could perhaps be said of Planck's theory of black body radiation which proposes that electromagnetic energy is absorbed and emitted by charged oscillators in discontinuous lumps only. It must be noted, however, that even though this discontinuity is indeed alien to Classical theories, the process of radiation itself is explained purely in mechanical terms. In Planck's theory, just as in Maxwell's electrodynamics, electromagnetic radiation is the result of disturbances created by an accelerating charge in the electromagnetic field.

Bohr had to somehow put a stop to an accelerating charge radiating energy in all circumstances. The reason for this was the following: Earlier, the results of experiments on the scattering of α particles from a few atoms, had forced Rutherford to propose the planetary model of the atom in 1911. This model consisted of a positively charged massive nucleus with a sufficient number of negatively charged electrons in circular orbits around it, such that the atom remains electrically neutral. Since the orbiting electron is constantly accelerating, this model along with the Classical theory of radiation entail that the electron should radiate itself out of energy. As a result, it should soon come to rest by collapsing into the nucleus. In order, therefore, for the Rutherford model of the atom to be a stable structure, it is necessary that the electron should be prevented from radiating energy while it remains undisturbed in an orbit. Bohr achieves this by fiat.

Bohr's electron is similar to Planck's oscillator in that *when it does* create disturbances in the electromagnetic field, the latter propagate as discontinuous lumps of energy. Whereas Planck's oscillator, however, is Classical in that its acceleration invariably disturbs the field (albeit discontinuously), Bohr's electron in an atom exchanges energy with the field only when it goes from one nuclear orbit to another. This exchange always involves discontinuous lumps of energy. The electron goes into an orbit farther away from the nucleus when it absorbs energy from the field, while it emits energy by going into one with a smaller radius.

There is another important dissimilarity between Bohr's electron and Planck's oscillator. According to Planck, in every exchange of energy between the oscillator and the electromagnetic field, the frequency of radiation is always equal to the frequency of the oscillator's periodic motion. This, in effect, is how Planck's theory can *mechanically* account for radiation from accelerating charges. Bohr, as we shall see shortly, sees no alternative but to break this equivalence. Since he offers no explanation for the relationship that he proposes between the two, his theory ends up with a distinctly non-mechanical character.

The similarities and dissimilarities noted are not, however, haphazard. Bohr [1913] cites the results obtained by J. W. Nicholson in a series of publications in which he was working with a model of atomic structure which appears to be a cross between those of J. J. Thompson's and Rutherford's. According to this model, the electrons are held, by an inverse-square force, on etherial rings that surround a positively charged nucleus. The electrons are allowed to oscillate around fixed points either horizontally on the plane of the rings, or vertically perpendicular to this plane. Using Planck's theory of radiation verbatim, Nicholson was able to derive ratios between the wavelengths of different sets of *tightly packed* lines in the spectrum of the solar corona, which showed excellent agreement with measured results (Bohr [1913], p. 6]).

Nicholson's theory, however, is incapable of accounting for unevenly dispersed lines in the emission spectrum of the hydrogen atom, such as the Balmer lines, a phenomenon which was known at the time. The reason for this is that as long as oscillators whose exchange of energy with the field is Planckian, accelerate, they would have to lose energy by emitting quanta of radiation. This emission, in turn, should change the frequency of the electron's periodic motion because the latter is tied up with the frequency of radiation. Consequently, a single orbital frequency should not be sustained for a prolonged period of time in the emission spectrum of the atom (Bohr [1913], p. 7]). The lesson Bohr draws here is that a more realistic model of the atomic structure should have the electrons behave as Planckian oscillators *only* in the regions where the frequencies of radiation correspond to tightly packed spectral lines.

A look at the energies and the frequencies associated with not only Balmer's, but also the Paschen and Lyman lines in the emission spectrum of the hydrogen atom (the latter was not yet available at the time), shows that the tighter packed lines correspond to lower energies and frequencies. Adopting Rutherford's model of the atom, then, Bohr is left with the thought that in the higher frequency areas of its orbit, the electron must behave in a non-Planckian way, *approaching* Planckian behaviour only as its orbital frequency gets lower and lower.

Now, the difference between the regions of high and low frequencies cannot be attributed to the orbital motion of the electron, since *that* remains the same in all the orbits. However, if the *separation between the orbits* is different in these regions, transitions between orbits could account for the difference. It turns out that as the orbital radius increases, the separation between the orbits gets smaller and smaller. At large distances from the nucleus, therefore, the electron could jump from one orbit to another without *appearing* to have changed it at all. It would *look as if* the electron were radiating in a single orbit, whereas in fact it is radiating by jumping between orbits.

2. The Average Rule

Severing the mechanical relationship between frequencies of the periodic motion and radiation, can secure stability for the electron's orbits in Rutherford's model, but it creates a problem. As will become clear later, Bohr needs a numerical relation between the two frequencies in order to calculate 'characteristic' values for the energy of the hydrogen atom, as well as the size of the electron's orbit in its most stable state.

In Rutherford's model of the hydrogen atom, the electron is held in orbit by a mutually attractive Coulomb force between it and the positively charged nucleus. If we denote the charges on the nucleus and the electron by e and -e respectively, this force is $F_c = -ke^2/r^2$, where r is the radius of orbit and k a constant. When the atom is in equilibrium, this force must be equivalent to a centripetal force F_p due to the revolution of the electron. If v is the velocity of the electron and m_e its mass, $F_p = -m_e v^2/r$. v^2/r is the radial acceleration of the electron, and the minus sign indicates the inward direction of the force. Equivalence of the two forces yields $m_e v^2/r = ke^2/r^2$, or $m_e v^2 = ke^2/r$.

The left hand side of the latter equivalence is twice the kinetic energy E_k of the electron. This entails $E_k = 1/2$

(ke^2/r). The electric field due to the positive nucleus gives rise to a potential V_r everywhere on the surface of a spherical shell with a radius r from the nucleus. Since the electron carries a charge -e, its potential energy E_p at a distance r from the nucleus is given by $E_p = -V_r e$. The potential V_r, in turn, is given by the work necessary for bringing a positive charge e from a point infinitely away from the nucleus to a point at a distance r from it. So $V_r = \int_{\infty}^{r} ke/r^2 dr$, which yields $V_r = ke/r$. Since $E_p = -V_r e$, we obtain $E_p = -ke^2/r$.

We are now in a position to find the total energy E_t of the electron: $E_t = E_k + E_p = 1/2\ (ke^2/r) - ke^2/r = -1/2\ (ke^2/r)$. The minus sign indicates that a total energy of $1/2\ (ke^2/r)$ must be pumped into the system in order to break the electron loose from the nuclear hold. Let us call this the 'characteristic' energy of the electron in its stable orbits, and denote it by W. We have then, $W = 1/2\ (ke^2/r)$, or $r = 1/2\ (ke^2/W)$.

We note that W has the same value as the kinetic energy of the electron in its revolutions at a distance r around the nucleus. Assuming the electron travels with a constant speed, its speed v in a circular orbit of radius r is related to its orbital frequency w by $v = 2\pi rw$. Since the kinetic energy of the electron is $E_k = 1/2\ (m_e v^2)$, substituting we get $E_k = 1/2\ m_e(4\pi^2 r^2 w^2) = 2\pi^2 r^2 w^2 m_e = W$. From this we get an expression relating the characteristic energy of the electron to its orbital frequency: $w = \sqrt{W} / (\sqrt{2}\sqrt{m_e}\pi r)$. Substituting the values of r in this relation we obtain: $w = 2W\sqrt{W} / (\sqrt{2}\sqrt{m_e}\pi ke^2) = \sqrt{2}W^{3/2} / (\sqrt{m_e}\pi ke^2)$. Now the problem is to get expressions for w and W which would enable us to calculate results that can be tested against the available experimental data.

According to the postulate laid down by Bohr, the electron should emit radiation only when it changes orbits to one closer to the nucleus. Suppose initially we start out with an electron which is infinitely far from a hydrogen ion, and at rest relative to its nucleus. Suppose further that as a result of some interaction, the electron gets trapped in the mutual nuclear hold, finally settling down at a stable orbit of radius r around the nucleus. During this transition, radiation is given off whose energy should be, according to Planck's theory, $nh\mu$, where h is the Planck's constant, μ the frequency of radiation and n a variable ranging over integers.

This energy must be equivalent to W, since, as we saw, W is just the amount of energy that the electron needs (in reverse order) to break loose from the binding nuclear force, thus turning the hydrogen atom into an ion. So $W = nh\mu$. If we had Planck's theory of radiation in its entirety available at this point, we could sail home, because according to this theory $\mu = w$. In the absence of this luxury, Bohr comes up with the suggestion that the frequency of the radiation should be equivalent to the *average* of the orbital frequencies before and after the binding takes place (Bohr [1913], p. 4f). Call this proposal, the 'average rule'.

Since the electron was initially at a relative rest, its orbital frequency w_i at this stage must be zero. At the final stage, the electron would be orbiting the nucleus with the frequency w_f. The average $1/2\ (w_i + w_f)$ of these frequencies would be just $1/2\ (w_f)$. If by w we now understand any orbital frequency to which the electron has settled after initially starting at rest relative to the nucleus, and substitute this for μ in Planck's expression for the radiated energy, we get $W = nhw/2$. Substituting this for W in the expression we derived for the orbital frequency, we obtain: $w = \sqrt{2}(nhw/2)^{3/2}/\ \sqrt{m_e}(\pi ke^2)$, or $w = 4m_e\pi^2k^2e^4/n^3h^3$. We can see that this astonishing result gives the orbital frequencies of the electron in the hydrogen atom, mostly in terms of constants whose values were all available at the time. The only variable in this expression is n which ranges over integers. With this expression at hand, we can not only determine the energies of the electron, but the radii for its stable orbits as well. They come out respectively as $W = 2m_e\pi^2k^2e^4/n^2h^2$, and $r = n^2h^2/4m_e\pi^2ke^2$.

We must, at this point, note a change in the interpretation of the variable n. In Planck's theory of radiation the values of this variable are *cardinal numbers*, each determining a 'size' for the amount of the radiated energy. In Bohr's theory, the values of n are *ordinal numbers* each labelling an orbit of the electron in the order of their closeness to the nucleus. Thus for $n = 1$ we get the orbit with the smallest possible radius; for $n = 2$, we get the radius for the next possible orbit further away, and so on without any orbits being allowed in between. So, although nothing is changed so far as the numerical results are concerned, where in Planck's theory $nh\mu$ signifies n lumps of $h\mu$ energy which are simultaneously radiated out, in Bohr's reinterpretation it signifies a single lump of energy that is radiated out with the *frequency* $n\mu$. This frequency, in turn, is characteristic for a jump from the state in which the electron is at rest infinitely far from the nucleus, to the state in which it is orbiting the nucleus with a radius determined by n.

Bohr writes in this connection: 'We are thus led to assume that the interpretation of the equation [$W = nhw/2$] is not that the different stationary states correspond to an emission of different numbers of energy-quanta, but that the frequency of the energy emitted during the passing of the system from a state in which no energy is yet radiated out to one of the different stationary states, is equal to different multiples of $w/2$, where w is the frequency of revolution of the electron in the state considered.' (Ibid, p. 14).

The value of W is greatest for $n = 1$ which means that it must be the ionization potential for the hydrogen atom. The value of this potential was experimentally known in 1913, and the agreement between it and the value calculated from his theory is claimed by Bohr as its first confirmation. The same goes for the value of r for $n = 1$, which must fix the dimensions of the hydrogen atom at its most stable state. Further confirmations are claimed for the agreement between the calculated energies and frequencies for transitions between orbits corresponding to $n \geq 2$ and those measured for Balmer's lines in the emission spectrum of the hydrogen atom. Rydberg's empirically determined relation for the frequencies of the line spectra, is also seen to follow directly from this theory. All this and more seems to indicate that Bohr had struck gold with his average rule. The question is how did he come upon that rule?

3. The Informal Justification of the Average Rule

Let us first examine some answers to this question in the popularized literature. Holton and Brush [1973] say that it was a bold but informed conjecture which happened to meet with success. 'Bohr', they write, '... takes a bold step on the basis of little more than a hunch; he assumes that if the electron is initially at rest at a very large distance from the nucleus... and if, when it is captured or bound by the atom, it drops into a final state, ... the frequency ν of the emitted light is simply the average of the initial and final frequencies of orbital revolution.' (Ibid, p. 482). And further, 'The critical steps - assuming that the frequency of emitted radiation is the average of the frequencies of revolution in the two orbits, and then redefining the quantum number n - were leaps in the dark which could only be justified by their subsequent success.' (Ibid, p. 486). Short of the colourful adjectives, this is pretty much how Olenick et al [1986], (pp. 448-452), also tell the story. In neither of these expositions is there any mention of the Correspondence Principle, which was, as already mentioned, only formally articulated in later works of Bohr's.

Holton and Roller [1958], (pp. 624-628), on the other hand, do not mention the average rule at all, but summon the Correspondence Principle in an attempt to account for Bohr's derivations. After they spell out their case, however, they leave the perplexed reader with such a flagrant case of inconsistency that must surely discredit the principle they correctly claim to be at work. Perhaps Lakatos [1970] was influenced by this and similar accounts when it claimed that Bohr's model of the hydrogen atom was developed on 'inconsistent foundations' (pp. 142, 144, and 153). But more on this later.

Bohr, Holton and Roller write, 'found a systematic way of deducing radii of the allowed orbits in the hydrogen atom by a consideration of the following type. When an atom is strongly excited, its electron should be moving in an orbit with a large diameter. But for a sufficiently large orbit, we must expect that the classical electrodynamic theory of Maxwell and Hertz applies; in that case the circulating current (here only one electron) should radiate light at a frequency equal to the frequency of the orbital motion. At the same time, Planck's quantum postulate should hold for this as for any system with periodic motion, so that the energy changes of the orbiting electron should proceed in smallest intervals of value $h\nu$. Here is a place where classical and quantum physics join, and we may assume that they overlap without leaving a discontinuous transition. In this region the quantum physics and classical physics merge into each other, and their predictions must correspond.' Holton and Roller [1958], (p. 625).

The fact that Maxwell and Hertz are brought in makes the inconsistency especially glaring. For according to their theory, electromagnetic radiation is incessantly linked to the instantaneous acceleration of electric charges. If there is no pause in this acceleration, there would be no gap whatsoever in the release of electromagnetic energy into the ether. Bohr's theory, in sharp contrast, would allow electrons to accelerate around an atomic obit as long as it can be sustained, *without radiating any energy whatsoever*. In order for these contradictory accounts to '*merge*' at any level, one of them must give way, somehow exempting the electrons from behaving in their otherwise natural way. Neither, however, appears to allow any such exemption.

According to a serious and detailed study made in Kuhn and Heilbron [1969], the average rule was needed to enable Bohr to derive Balmer's formula for the spectral lines. However, Bohr had no justification to sanction its use. 'In order', they write, 'to *derive* the Balmer formula, Bohr needed a quantum condition to determine energy levels. The required condition proves to be equation (15), $W_\tau = \tau h w_\tau /2$ [where W_τ is the binding energy of the electron at the τth energy-level, and w_τ is its orbital frequency at this level], and this equation gains whatever plausibility it possesses from its resemblance, emphasized by Bohr, to the quantum condition governing Planck's oscillator. Such an oscillator can emit several quanta at a time but only at a single frequency determined by its mechanical structure. Equation (15) must, by analogy, govern a process in which τ quanta are emitted, each of frequency $w_\tau /2$. To reach the Balmer formula Bohr will ultimately change this interpretation, saying instead that equation (15) represents the emission of a single quantum with frequency $\tau w_\tau /2$. That interpretation, however, at once destroys the analogy between the Bohr and Planck radiators, and some other justification for (15) is therefore required. In fact, Bohr found none.' (Ibid, p. 270).

A sketch of the correct answer has already been presented, and it involves Bohr in neither a leap in the dark, nor proceeding from inconsistent assumptions or through unjustified moves. Informally speaking, the success of Nicholson's application of Planck's theory of radiation to the densely packed spectral lines, shows that in the domain of low frequencies the behaviour of the electrons is observationally indistinguishable from that of Planck's oscillators. This domain corresponds to large values for n which, in turn, determine large orbital radii for the electron. Let us take two adjacent orbits in this domain, such that, compared with the dimensions of an unexcited atom the radii of these orbits are huge. Since the orbits in this domain are closely packed, the difference between the two adjacent radii is negligible. Therefore, an electron which changes its orbit from the outer to the inner one, would *appear* to be revolving with still the same radius. Assuming constant speeds in each orbit, the frequencies of the electron's revolution would thus *appear* unchanged.

If w_i is the frequency in the outer orbit and w_f that in the inner one, the average rule dictates that the frequency μ of the radiation, emitted as a result of the jump, should be $\mu = 1/2\ (w_i + w_f)$. Since $w_i \approx w_f$ is an excellent approximation, we would have $\mu = 1/2\ (2w_f) = w_f$, or $\mu/w_f = 1$. That is, the electron jumping between these orbits, would appear to be radiating with its orbital frequency, just as Planck's theory says it should. This informal argument shows that the average rule is indeed promising in securing convergence, in an appropriate domain, between predictions from the picture of an electron in a Rutherford atom, and that of a Planck's oscillator.

4. The Formal Justification of the Average Rule

In the fully-fledged argument that Bohr launches in [1913] to formally account for this convergence, he reveals a clear grasp of the heuristic potential of the Correspondence Principle (without yet giving it this name or elucidating its content). Moreover, he goes on to a brilliant exploitation of this potential in order to derive the average rule. As we shall see, nowhere in this exploitation is there any hint that predictions from two inconsistent theories overlap *because they 'merge'* at some stage.

Bohr starts by noting that his theory of the electron is, in some respects (given the change in the interpretation of the integral variable), faithful to Planck's picture of a radiating oscillator. 'Firstly', he writes, 'it will be observed that it has not been necessary, in order to account for the law of the spectra by help of the expressions... for the stationary states, to assume that in any case a radiation is sent out corresponding to more than a single energy-quantum, hv.' Bohr [1913], (p. 12). After all, elsewhere in the same paper he writes, 'the essential point in Planck's theory of radiation is that the energy radiated from an atomic system does not take place in the continuous way assumed in the ordinary electrodynamics, but that it, on the contrary, takes place in distinctly separated emissions, the amount of energy radiated out from an atomic vibrator of frequency v in a single emission being equal to τhv, where τ is an entire number, and h is a universal constant.' (Ibid, p. 4).

On the crucial question of the relation between the frequencies of the periodic motion and radiation, he goes on to say: 'Further information on the frequency of the radiation may be obtained by comparing calculations of the energy radiation in the region of slow vibrations based on the above assumptions with calculations based on the ordinary mechanics. As is known, calculations on the latter basis are in agreement with experiments on the energy radiation in the named region.' (Ibid, p. 12). This passage encapsulates the heuristic deployment of the Correspondence Principle. Since the Classical, but refuted, theory is empirically adequate in the domain where vibrations are slow, the relation it assumes between the frequencies can be exploited to guide the search for the one that is to supersede it.

As was noted earlier, the domain of 'slow vibrations' for the electron in the hydrogen atom, corresponds to orbits with relatively huge radii. Since the latter are determined by large values for the integral variable n, let us choose two adjacent radii corresponding to two such values of n which we name 'N' and 'N-1' respectively. According to Bohr's postulate for the emission of radiation, a single quantum of energy $h\mu$ should be released as a result of the electron jumping from the orbit determined by N to that determined by N-1. This quantum of energy is, in turn, according to a rule postulated by Einstein in 1905 in connection with the photoelectric effect, equivalent to the difference of the energies possessed by the electron in each of the respective orbits.

If we recall, the average rule dictated that the relation between the energy W radiated from the hydrogen atom, and the orbital frequency w of its electron in a stationary state, should be $W = (n/2)\ hw$. Now, however, Bohr treats the multiplier of hw in the last expression, as an unknown, and goes on to launch the following argument (Ibid, pp. 12-13):

We assume only that the multiplier in question should be a function $f(n)$ of the integral variable n. We will then have $W = f(n)hw$. Substituting this for the expression we obtained linking W and w viz., $w = \sqrt{2}\ W^{3/2} / \sqrt{m_e}(\pi ke^2)$, and squaring both sides, we get $w^2 = 2f(n)^3h^3w^3 / m_e\pi^2k^2e^4$. If we abbreviate the denominator to B and solve for w, we get $w = B / 2f(n)^3h^3$, which, in turn, yields $W = B / 2f(n)^2h^2$.

If we now denote the energy of the electron while in the orbit determined by N, W_i and that determined by N-1, W_f, Einstein's rule tells us $W_f - W_i = h\mu$. Substituting the expressions we derived for the energies and rearranging, we obtain $W_f - W_i = (B / 2h^2)\ (1/f(N-1)^2 - 1/f(N)^2) = h\mu$. Solving for μ, we get $\mu = (B / 2h^3)\ (1/f(N-1)^2 - 1/f(N)^2)$. This expression has the same form as Rydberg's empirical formula for the frequencies of spectral lines (viz., $\mu = R\ (1/(n_f)^2 - 1/(n_i)^2)$, where R is Rydberg's constant and n_i and n_f are the initial and final values of the integral variable respectively).

In Rydberg's formula, if we put $n_f = 2$ and let n_i vary, we get the frequencies for the Balmer lines. Comparison of the two expressions for the frequencies in question suggests that in order for Bohr's result to be empirically adequate with regard to the Balmer series, $f(n)$ should be the product of a constant with the integral variable. This means $f(n) = cn$, where c is a constant. For the choice of $n_i = N$ and $n_f = N-1$, we will have $f(n_i) = cN$ and $f(n_f) = c(N-1)$, respectively.

Substituting these in the expression for the frequencies of the radiation, we will get $\mu = (B / 2h^3)\ (1/c^2(N-1)^2) - 1/c^2N^2)$, or simply $\mu = (B / 2h^3c^2)\ (2N-1 / N^2(N-1)^2)$.

If w_N and w_{N-1} are the orbital frequencies of the electron before and after the radiation from the atom, we have $w_N = B / 2c^3N^3h^3$, and $w_{N-1} = B / 2c^3(N-1)^3h^3$ respectively. As N gets larger and larger, w_{N-1}/w_N gets closer and closer to 1. We would, therefore, not be too far off the mark if we treated the orbital frequencies in the domain of large N as approximately identical. The fact that Planck's theory of radiation has established itself with spectacular success in this domain, would then lead us to expect that $\mu = w_{N-1}$ is an excellent approximation. This expectation yields $\mu/w_{N-1} = 1$. Substituting the expressions for the frequencies, we get $c(N-1)^3(2N-1) / N^2(N-1)^2 = 1$. Rearrangement of this yields $c = N^2 / (2N^2 - 3N + 1)$, or $c = N^2 / 2N^2(1 - 3/2N + 1/2N^2) = 1 / 2(1 - 3/2N + 1/2N^2)$.

As N tends towards infinity, we can see that the bracket in the denominator of the last expression tends towards 1. Therefore, only for $c = 1/2$ would the numerical equivalence of the frequencies be an excellent approximation in the domain of large N. Inserting this into the expression for the radiated energy, we get $W = 1/2\ (hw_f)$, which is the same as what we obtained using the average rule. Q.E.D.

5. Methodological Review of Bohr's Argument

The problem facing Bohr [1913] is that Rutherford's planetary model of the atomic structure is incompatible with Planck's theory of radiation, given the stability of atoms and the experimental evidence on the line spectra. Rutherford's model established a clear superiority over its rival in dealing with the scattering behaviour of α and β particles. The latter, on the other hand, scored major successes with black body radiation as well as with densely packed spectral lines associated with low frequencies of radiation from atoms.

Since there was, at the time, no independent evidence for questioning the validity of Rutherford's model, the failure of Planck's theory in dealing with unevenly dispersed lines in the spectrum of the hydrogen atom, suggests that it is perhaps in Planck's theory that the source of the incompatibility lies. Bohr chose to question the mechanistic assumption in Planck's theory which causally links radiation to the acceleration of charged particles in the electromagnetic field. This choice can be justified by the fact that its benefit outweighs its cost. The cost is that a pictorially neat and familiar story for how radiation is produced must be given up. The benefit is that not only can stability for a viable model of atomic structure be secured, but also empirical features of the radiation from the deeper levels of this structure can be formally accounted for.

This choice entails that the frequency of the electron's orbit in the atom and that of the radiated energy from it are not numerically equivalent. Since, however, a numerical relation between the two is needed for calculations of interest, the problem becomes more sharply focused as follows. What numerical relation should hold between the frequencies so that the new theory of radiation, along with the Rutherford model, can generate the radiation frequencies associated with the line spectra from the hydrogen atom?

That some numerical relation should indeed exist between the frequencies, is strongly hinted by the empirical adequacy of the Classical theory in 'the area of slow vibrations'. Although the numerical equivalence assumed by this theory cannot be sustained for all frequencies, its success in this area signals the clue that a relation *which converges towards equivalence* under suitable conditions, is there to be found.

There are, therefore, two boundary conditions for the problem at hand. At one end, where we empirically probe into the finer structure of the atomic constitution, we find the unevenly dispersed spectral lines that appear to defy the assumption that the frequencies of radiation and periodic motion of the electron are numerically equivalent. At the opposite end, there is the empirical evidence that in the grosser regions of the scale, the ratio of these frequencies converges to unity. Bohr uses the first condition to infer *the form* that the unknown relation should have. He then works backwards, assuming the other condition for the appropriate values of the variable parameter, to deduce the relation that he has been looking for.

Bohr makes this point in no uncertain terms: '... taking the starting-point in the form of the law of the hydrogen spectrum and assuming that the different lines correspond to a homogeneous radiation emitted during the passing between different stationary states, we shall arrive at exactly the same expression for the constant [$B / 2h^3c^2$]... as that given by [the use of the average rule]..., if we only assume (1) that the radiation is sent out in quanta $h\nu$, and (2) that the frequency of the radiation emitted during the passing of the system between successive stationary states will coincide with the frequency of revolution of the electron in the region of slow vibrations.' (Ibid, p. 14).

The second assumption in the above passage is dictated by the view that Planck's theory is an excellent approximation in the domain of low frequencies of radiation from the atoms. The numerical coincidence of the frequencies follows from a theoretical assumption which is incorporated in Planck's theory. Even though this theory has proved unsatisfactory in accounting for the results from deeper empirical probes into the phenomenon in question, it has achieved a spectacular run of success in dealing with a range of results from shallower layers of empirical investigation into the same phenomenon.

The licence that Bohr takes in assuming some proposals of a falsified, but outstanding, theory as a guide in the search for the new theory he is developing, is premised on the adequacy of the former over a limited domain of empirical facts. This guiding role, in the case we have been considering, can be described in general terms as follows. Let T_1 be the false theory that has proved empirically adequate over the domain D_1. Suppose an element of T_1 has been identified as the source of its shortcoming when this domain is extended to cover deeper levels. The task is to rectify this shortcoming in a new theory T_2 in a way that preserves the adequacy of T_1 over D_1.

The quantitative expression whose introduction in T_2 is supposed to do the trick may, at first, be represented as a qualitative expression, some of whose formal features can be conjectured from the available information. To preserve the adequacy of T_1 over D_1, however, T_2 must recognize that the defective element in T_1 is valid (in the sense of being a good approximation) for those values of its parameters that coincide with values contained in D_1. If now an equation can be set up assuming this proposal of T_1, such that all the parameters in it range over D_1 and the only unknown is the quantitative expression which is being sought, it may be solved and the unknown found.

CHAPTER 3

GENERALIZED CORRESPONDENCE PRINCIPLE
AND TWO-TIER REALISM

1. Conceptions of The Generalized Correspondence Principle

Once a new theory T_2 has been put together, heuristically exploiting the successes of T_1 in a domain D_1, T_2 entails that T_1 is a good approximation to it. That is to say, T_2 will contain parameters such that when they are given values approaching those contained in D_1, conditional predictions of T_2 will become virtually indistinguishable from those yielded by T_1. The question is whether in physics a theory that is to supersede one which has been established as outstanding over a certain domain, *should* invariably entail that the latter is a good approximation in that domain. The conception of the Correspondence Principle which is strong enough to serve the purposes of this thesis must include the imperative that it should. The family of conceptions which share this imperative, because of their generality, is usually referred to as the *Generalized Correspondence Principle*. Any conception of the Correspondence Principle which is generalized, imposes a constraint on a superseding theory which otherwise does not have to be there. It is possible that T_2 is a theory which has nothing to say about the phenomena contained in D_1, but is empirically adequate over the domain where T_1 has failed. What we want to know, therefore, is whether the imposition of such a constraint can be *justified* across the board in physics. Before the search for such a justification is undertaken, however, we need a clear and workable conception of the Correspondence Principle.

Depending on the kind of entities theories are taken to be, attempts at articulating the Generalized Correspondence Principle may be broadly divided into two groups. One group views theories roughly as sets of propositions which purport to describe laws and processes which hold between events in the world. It is fair to say that this view enjoys the most widespread support among the philosophers of science and for this reason it may be branded the 'standard' view. The rival group utilises model theoretic techniques and regards theories as sets of mathematical *models* (intended or potential) in which the descriptions and laws proposed by a theory hold. The proponents of this view are still in minority among the philosophers of science and their approach may be branded 'non-standard' (it is sometimes called the 'semantic' view of theories, but this name is in some way unsatisfactory).

Within the non-standard tradition, Moulines [1984] and Pearce and Rantala [1984] offer representative formulations of the Correspondence Principle as a general intertheory relation. Redhead [1975] introduces the novelty of viewing a theory as 'a unary relation on a generalized function space' (Ibid, p. 90), but his approach remains within the non-standard tradition in that it seeks to 'effectively reduce the problem of intertheory relations to one of relations between mathematical structures' (Ibid, p. 89).

The main advantage claimed for this approach appears to be its ability to capture intuitions about the relation between theories which satisfy the Correspondence Principle. According to these intuitions (which underlie approaches within the standard tradition as well), the theory that is superseded must somehow 'approximate' what is proposed to be the case by the new theory. It was once customary to understand this property in terms of the relation of *reduction* between the superseded and superseding theories. The idea of the reduction of one theory to another was originally introduced within the standard tradition. It was first explicated in Nagel [1961] in terms of a strictly deductive relation between the superseding theory plus some additional statements (called 'bridge laws'), and the superseded theory. In Causey [1977] the idea was further articulated in considerable detail, and the bridge laws replaced by 'non causal and synthetic identity statements'. The function which these additional statements are required to perform is to somehow bridge over the conceptual gap which must exist between two theories that make inconsistent claims about the underlying constitution of the world.

One problem that emerged from the discussion of this type of intertheory relations was the possibility of some loss in content in the transition from one theory to another. This problem was raised in Kuhn [1970] and has, subsequently, been known as the problem of 'Kuhn-loss'. The alleged difficulty with regard to this possibility is that in case the superseded theory has a greater content than the superseding one, the strictly deductive relation between the two would be untenable even with the introduction of bridging statements (Kuhn [1970] goes even further and claims this possibility as a reason for the theories involved being incommensurate). Detailed case studies (e.g. in Koertge [1969], Worrall [1989a] and Worrall [1989b]) suggest that there is no loss of *empirical* content in transition from one theory to another. Assuming that there is no Kuhn-loss *in empirical content* (or if there is, there is always a common part in which the claims of both theories 'intersect with regard to their reference'), Post [1971], operating within the standard tradition, introduces the idea of 'slicing' a theory into various explanatory levels. (Redhead [1975] endorses this idea, but claims that it is also best articulated within the non-standard approach). The 'General Correspondence Principle', as Post calls it, is then formulated as a relation between those parts of the relevant theories that share common empirical contents.

According to Post, if S is the successful but superseded theory, and S^* a 'slice' of S which has proved empirically adequate over a range of phenomena, 'the new theory must correspond to the old theory, in the sense of coinciding with S^*, the working part of the old theory in the range in which S has been found to work. According to the General Correspondence Principle, this is a necessary requirement.' (Post [1971], p. 233). To achieve correspondence in this sense between theories that make radically different claims about the constitution of the world, 'we need ... a system of translation *T*, from the language of L [the superseding theory] into that of S, with the necessary restriction on *T* that it should carry the statement of a given event in L into the statement of the same event ... in S ...' (Ibid, p. 230).

The picture that emerges from this account is that successive successful theories in the history of science which satisfy the Correspondence Principle, '*build on*' the successes of previous theories by somehow taking over slices from them. 'Quite generally, the thesis may be put this way: no theory that ever 'worked' adequately turned out to be a blind alley... The most radical revolutions have destroyed the top levels and give a somewhat different interpretation to the lower levels, but they have not destroyed the whole of the lower-level structure.' (Post [1971], p. 237).

Within the same tradition, there exists a rival view about how the successes of a superseded theory are *preserved* by the superseding theory. This view introduces the idea of 'structure' in the world and attributes the successes of outstanding theories in the history of physics to the discovery of some of these. The theme of one theory preserving the discovery of some 'structure' by the successful, but falsified, theory it replaces, is adopted in Worrall [1989], and its pedigree traced back to Poincaré. Like Post [1971], the task of safeguarding this preservation is given to the Correspondence Principle. It appears, however, that unlike Post [1971], the preservation Worrall has in mind occurs, not between whole theories (or parts of their ontology), but some of their *formal* laws only. Worrall is not alone in holding this view; Zahar [1983] and [1989], Krajewski [1977], Yoshida [1977], Fadner [1985] and Radder [1991] all concur in conceiving the Correspondence Principle as holding between formal laws rather than theories as wholes.

The reason for adopting this view is that the only part of a superseded theory that appears salvageable by being deducible from the superseding theory, is its *mathematical* equations (which are equated with the formal laws laid down by a theory). In other words, the mainstream conviction that cumulative progress in science can only be safeguarded if there is some *logical* bond between successive successful theories, forces the selection of mathematical equations as *units of correspondence*. The ontological interpretations of the equations, according to this view, changes too radically in some transitions to make it possible to secure a deductive relation between theories as wholes. Uninterpreted equations, however, can survive these radical changes. If one set of equations belonging to a falsified but successful theory can somehow be derived from its counterpart in the superseding theory (sometimes with the use of conditions imposed on some of the latter's parameters), the feeling is that the desideratum of cumulative progress in science is essentially achieved. What matters most in the old theory, namely its main equations, is thus preserved by the new.

Post [1971], Koertge [1969] and Worrall [1989a] share the belief that some sort of a term-by-term (or phrase by phrase) correlation or mapping (or as Post calls it, a 'translation') between the languages of the appropriate theories is necessary in a correspondence (those who take formal laws as units of correspondence appear not to). Koertge operates with Post's idea of *slicing*, whereas Worrall does not. This enables Koertge to incorporate non-mathematical scientific theories, such as Stahl's and Lavoisier's theories of chemical reactions, into a correspondence scheme. The advantage gained in Worrall's position, however, is that the notion of *structure invariance* which never received sufficient clarification in Post [1971] and Koertge [1969], can be clearly unpacked in terms of the sameness of mathematical equations.

In Worrall's account 'structure' is preserved in the transition between two theories if the following condition is satisfied: after establishing a mapping between the referring terms of the theories, some of their mathematical equations either remain identical, or else those of the superseding theory *become* identical with those of the superseded one when certain limiting conditions are imposed on some of the former's parameters. This enables Worrall to say that irrespective of whatever entities or attributes each theory postulates at the ontological level, what they, jointly get right are the *relations* that prevail between *whatever* are the real constituents of the world. Ontological claims may come and go, but the formal feature of the relations between the entities that actually inhabit the world, remains preserved through the succession of successful theories. 'The rule', Worrall [1989a] claims, 'in the history of physics seems to be that, whenever a theory replaces a predecessor, which has however itself enjoyed genuine predictive success, the 'correspondence principle' applies... But the principle applies *purely* at the mathematical level and hence is quite compatible with the new theory's basic theoretical assumptions (which *interpret* the terms in the equations) being entirely at odds with those of the old. I can see no clear sense in which an action-at-a-distance force of gravity is a 'limiting case' of or 'approximates' a space-time curvature... Yet Einstein's equations undeniably go over to Newton's in certain limiting special cases. In this sense, there is 'approximate continuity' of *structure* in this case.' (Ibid, p. 120).

The trouble with this account is that from the *realist* point of view equations in physics are not merely a set of uninterpreted mathematical symbols 'hanging in a limbo'. If one adopted an instrumentalist outlook on theories, there would of course be no problem here. Mathematical equations in this outlook are *merely* formal tools for the prediction of future events from a set of available data *irrespective of what actually is the case in the real world*. From the realist point of view, on the other hand, all the parameters occurring in a mathematical equation in physics *stand as interpreted signs*. If an equation from Einstein's theory actually does turn into *the equation* that states Newton's second law when certain limiting conditions are imposed on it, then the former must have turned into a formal relation that *holds between masses, their accelerations and the forces acting on them in a three dimensional Euclidean space in which the passage of time is everywhere invariant*. In the realist's book, a mathematical statement of Newton's second law is never identical to a mere formal relation. The formalist procedure of separating the formalism from its interpretation, so fruitful in the foundational studies of mathematics, does not carry over to physics so far as a realist is concerned. For the two-tier realist, in particular, the story a theory tells in physics *together with* the equations it proposes, form a single net, as it were, that is cast to capture as many states of the observed physical systems as is possible.

2. The Approximational Conception of the Generalized Correspondence Principle

The stumbling block which prevents authors in the field from making any sense of the idea that one theory can be an 'approximation' to another while they clash on *both* the ontological (clearly stated in the quoted passage from Worrall [1989a]), as well as the formal level, is that for this relation to hold they require some *significant part* of the superseded theory to somehow carry over to the superseding one. Only then, they feel, from the standpoint of the superseding theory can one claim that the superseded theory achieved something which would warrant regarding it as an approximation to a better theory. The question is, if this requirement is to be given up, what alternative conception of the Correspondence Principle can be stated?

The study of Bohr [1913] in chapter 2 revealed a case involving theories between which the Correspondence Principle *has to* hold (otherwise the value of an important parameter in the new theory could not be justified). This study also suggests *the possibility* that the Correspondence Principle may hold between two successive theories where no significant part of the ontology or formalism of the superseded theory can remain preserved in the transition. What in Bohr [1913] is taken as a good approximation to a better future theory, is a theory which the Cambridge astronomer J. N. Nicholson had proposed in a series of publications in 1911 and 1912. In Nicholson [1911], a model of the atom is proposed in which the positive charge is concentrated in a massive nucleus and the electrons are embedded in fixed positions on an aetherial ring which rotates slowly around the nucleus. The potential responsible for holding this system together is worked out using Thompson's calculations. The ring is set to rotate with a velocity sufficiently small for no detectable radiation to be observed. This is the first major difference between the ontologies of Bohr's theory of atomic constitution and that which it supersedes.

The second difference in this category, has to do with the *mechanism for radiation* from the atom. In Nicholson [1911] detectible radiation is emitted from the atom as a result of the electrons either flying off the ring (due to acquiring excess energy), or else vibrating horizontally or vertically (relative to the plane of the ring) round their fixed positions in the ring. Here, no mention of Planck's quantization of the field is made. Postulating a hypothetical element whose ring contains four electrons (named 'Nebulium'), and making all the calculations based on Thompson's theory of the atom, Nicholson [1911] manages to derive the main lines in the spectrum of Orion and three other nebulae with an error of '3 or 4 tenth-meters ' (ibid, p. 52), which translates into an accuracy of '3.6 in about 4400' (ibid, p. 55). Later in Nicholson [1912], another hypothetical element (named 'Protofluorine') is introduced whose ring is postulated to hold five electrons. Here the idea of Planck's oscillator is applied to these atoms in order to account for the lines in the spectrum of the solar corona. Each line in this spectrum is produced by oscillations performed by the atom as a whole as a result of its electrons vibrating within the ring. When one electron is ejected from or gained by the atom, the angular momentum of the oscillator changes abruptly. The new mode of vibration by the oscillator differs from the previous one discretely, thus giving rise to the appearance of a new separate line in the spectrum. Assuming that the energy given off by such an oscillator is always Planckian (i.e. of the order $nh\mu$, where n is an integer, h the Planck's constant and μ the frequency of the oscillator), Nicholson accounted for 14 lines in the spectrum of the solar corona with an accuracy of 3 or 4 parts in 1000. This theory also managed to predict, with an accuracy of 1 in 10,000, a previously undetected line in this spectrum.

The characteristic (ontological) features of Nicholson's ringed model of the atom are the following: (1) it obeys the Classical law of 'no acceleration by charges in the electromagnetic field without radiation'; (2) the relation between the frequency of its oscillations and the optical frequency of the resulting radiation has a mechanical explanation; and (3) it releases energy into the field always as a 'bundle' of n individual quanta (n being an integer). Bohr's theory differs from Nicholson's on every significant count. First, it is based on a *planetary* model of the atom in which the electrons are moving round the nucleus due to the action of central forces (as opposed to being carried by a vehicle). Second, by introducing the idea of *stationary states* it not only violates the Classical law governing the acceleration of charged particles in the field, but also provides no explanation for the transitions between these states. Third, it postulates the release of energy from the atom always as a single quantum whose *optical frequency* in different emissions changes by an integral multiple. Fourth, the optical frequency of the radiation is linked to the orbital frequency of the electrons without there being any *mechanical* explanation for this link (the only explanation Bohr provides is based on the use of the Correspondence Principle).

These differences at the ontological level lead to differences at the level of formalism which cannot be bridged over. First, in Nicholson [1912], the general equation linking the mechanical energy of the system to that of the radiation from it is of the form $2\pi mna^2\mu^2 = \tau h\phi$, where m is the mass of an electron, n their number in the ring, a the radius of the ring, μ the frequency of oscillations, τ an integer, h Planck's constant and ϕ the frequency of radiation. The integer τ in this equation, just as in Planck's theory, is a *cardinal* number since it indicates *how many* quanta of energy are being simultaneously released. The same integer which also appears in Bohr's theory, however, is an *ordinal* number indicating the order of the final orbit to which the electron has settled in the atom after the exchange of energy between the system and the field has taken place. Apart from the fact that cardinals are very different species of number than ordinals, there is no way that one can be *turned into* the other under certain circumstances.

Second, the expressions for the total energy of *the atom* are clearly different in the two theories. In Nicholson's theory no quantum number appears in this expression, while in Bohr's theory such numbers are always present (when the atom is not radiating there is only one of these present, indicating its stationary state, and when it is radiating, there are two of these present, each indicating the states between which a transition has occurred). According to the *structural* approach (within the standard tradition), these expressions have to become identical under some limiting conditions because the Correspondence Principle clearly holds between the two theories. Zahar [1989] expresses this point of view by stating the *structural* conception of the Correspondence Principle more precisely than other members of this approach as follows: '... [W]e should note that the correspondence principle can be given the form of a meta-statement... [L]et $\phi = 0$ denote a known law which is to be modified; the correspondence principle could read as follows: "The new law is of the form $\Phi = 0$, where Φ is a function of the quantity Γ such that $\Phi \rightarrow \phi$ as $\Gamma \rightarrow 0$" [the arrow reads 'tends towards']'. (Zahar [1989], p. 22). We are no longer asked to deduce the old law from the new, but instead require that the equation yielded by the new law should be of such a form that as a parameter in terms of which it is formulated tends towards zero, it gradually turns into that yielded by the old law.

Let us take the parameter introduced in the equation for Bohr's law governing the frequency of radiation from the hydrogen atom, say, to be the inverse of the integral variable n_j (so that as the values for the latter get larger, the parameter in question would tend towards zero). Even in the domain of extremely slow vibrations, *the equation* expressing Bohr's law would contain terms referring to *both* an initial n_i and a final n_f. This is because Bohr's atom radiates *only when* its electron jumps from an orbit further away from the nucleus to one nearer to it and never when it is on the same orbit. The sole condition under which the reference in Bohr's equation to two values for the integral variable could be eliminated, is when $n_i = n_f$ and there is no change in the electron's orbit. In that case, Bohr's theory allows no radiation whatsoever from the atom.

According to Planck's theory of radiation, on the other hand, the atom should be radiating so long as its bound electron is in accelerated motion. Suppose we somehow incorporate Bohr's interpretation of the integral variable within Planck's theory. A single value of this variable, within certain limits, determines an orbit for the electron. Since in Planck's theory the atom should radiate energy even when its electron stays on the same orbit, the equation offered by Planck's theory for the frequency of radiation from the atom, contains a single occurrence of this variable. Bohr's and Planck's equations, therefore, cannot possibly *turn into* one another under any circumstances. True, in the event that the integral variable takes infinity as its value (i.e. when the atom's electron is at a distance infinitely away from its nucleus), both equations predict zero radiation *from the atom*. But they do so for entirely different reasons, i.e. by containing expressions which remain different under all conditions.

This situation invites an alternative conjecture about what is demanded by the Generalized Correspondence Principle. One conjecture, which was named *approximational* in the Introduction, if tenable, appears to avoid these difficulties. For the purposes of this thesis it is preferable that this conjecture should rest within the *standard* tradition of viewing theories. The two-tier realist's interest in scientific theories is primarily geared to what they have to *say* about the actual world. The standard view of theories serves this interest quite naturally. Sets of possible models as suggested by the non-standard tradition, on the other hand, is non-committal as to which model reflects the actual world. For the two-tier realist this disadvantage offsets any formal advantages this tradition may offer for the articulation of intertheory relations.

The approximational conception may be stated as follows: A theory T_1 is a *good approximation* to a theory T_2 just in case (1) T_1 is an empirically successful but falsified theory, and (2) relative to the standards of experimental accuracy which were operative in the confirmation of T_1, T_2 at least matches all the successes of T_1 and, moreover, succeeds where T_1 has failed (perhaps it even scores successes where T_1 has nothing to say). Now, T_1 and T_2 may say things about the underlying physical constitution of the world which are logically incompatible with one another. But so long as their conditional predictions (i.e. 'if such and such is the case here and now, so and so would be the case there and then') over a common domain differ at most within the relative margin of experimental error, T_1 may be regarded as a good approximation to T_2. I take the standards of experimental accuracy, and the margin of experimental error, to be determined by the scientific practices of the day. In this conception, one theory may be a good approximation to another, while no explanatory 'slices' or mathematical equations in the two necessarily remain *the same*. The essential feature that such theories must share is rather that over a common domain, they should *perform* similarly (in the sense just articulated).

With regard to this notion of similarity of performance, several points need to be made clear. First, two theories may make observationally indistinguishable conditional predictions while the theoretical (metaphysical) attributes which each ascribes to observable entities differ hugely. In astronomy, for example, the Copernican and Tychonic theories of the solar system yield observationally indistinguishable predictions for the behaviour of the planets as well as the sun from the vantage point of the earth. But in one theory the earth is not a planet and must rest stationary at the centre of the universe while the sun would be rotating around it dragging the planets (which are in orbit around it) along. According to the other theory, the earth does not occupy a privileged position in the universe; it is a planet and together with the rest of the planets it orbits the sun which lies immobile at the centre of the universe.

Secondly, this notion of similarity is general enough to allow as special cases the identity of either some equations or ontological portions. Two theories which perform similarly with regard to some observable entities, may share some common ontology or equations, or indeed one can have some equations which are convertible to those of the other under appropriate limiting conditions. Lastly, this notion, if tenable, ties up naturally with the idea of progress, and in particular the related idea of verisimilitude in physics. If T_1 is a good approximation to T_2 in the sense just articulated, then there is a clear sense in which T_1 may be said to bear a *resemblance* to T_2; T_1 resembles T_2 not in what it says, but in what it *does*. The transition from one to the other may be regarded as progress because while T_2 outperforms T_1, it nonetheless safeguards the achievements of the latter through this resemblance. This idea of *similarity of performance* appears to capture Bohr's intuition that the superseding theory should be a 'natural generalization' of the superseded one while the two remain incompatible both on the ontological as well as the formal levels. Moreover, if there is, in a branch of physics, an ultimate theory towards which a succession of outstanding but falsified theories converge, and which manages to capture the truth of the matter in the corresponding field, then all its successful predecessors would bear some resemblance to it to the extent determined by the position which they occupy within this succession. The closer a theory lies in this succession to the ultimate theory, the greater its resemblance to the truth. More on this point later.

It is possible that there may be cases where the statement of the approximational conception, as presented above, fails to effect a discrimination as to which superseded theory is a good approximation to a superseding one. I shall present a pseudo-historical case exemplifying this possibility later in an argument concerning the two-tier realist assessment of a theory's success. Suppose T_1 and T_2 are rival theories which despite being incompatible at the ontolological (metaphysical) level, are nevertheless equally empirically adequate. Suppose T_3 is a theory which (1) supersedes both T_1 and T_2, (2) is ontologically different from both to different extents, and (3) is distinguishable from both at the observational level to exactly the same extent. In this case if the approximational conception is to help us single out either T_1 or T_2 (but not both) as a good approximation to T_3, it will have to be augmented with an additional proviso. We may find this proviso by considering the nature of the differences between the theories involved at the ontological level. That theory would be a good approximation to T_3 which exhibits, in its ontological claims, a closer continuity to those of the latter than its rival. Admittedly, this is a rather vague stipulation. Since, however, we are here dealing with a hypothetical eventuality, we may postpone further discussion until we come to present the promised pseudo-historical case which involves Kepler's theory on the one side and the Tychonic as well as a version of the Copernican theory on the other.

The approximational conception, I submit, is not without a respectable pedigree, but is rather suggested by the general intertheory relation which Bohr saw as holding whenever the Correspondence Principle is satisfied in physics. Bohr is the first author to notice not only the existence of this principle in physics, but its heuristic potential. He also inaugurated attempts to articulate it. It is, therefore, surprising to find that very little attention has been paid by philosophers interested in the subject to his work (in Post [1971], for example, which is regarded by many as a classic in this subject, reference to Bohr [1913] appears in a footnote and only in passing).

3. Objections to Generalization of the Correspondence Principle

Since the heuristic potential of an outstanding theory that is to be superseded would not be available for use in the context of discovery without the constraint imposed on the superseding theory by the Correspondence Principle, the availability of this resource can very well justify the admission of the Principle. It may be felt, however, that to get *just* this benefit out of a theory that is to be superseded, the imposition of such a far reaching constraint is not necessary. The empirical adequacy of a theory over a certain domain can be exploited in one particular circumstance for the purposes of discovery, without requiring that superseding theories in general and in all circumstances should live up to the demands of the Correspondence Principle.

Indeed this appears to be the position argued for in Feyerabend [1962] and [1981]. Here, however, the motivation for taking this position stems from a particular reading of the Correspondence Principle according to which it obliges the superseding theory either to *'contain'* the theory it supersedes or be *consistent* with it. 'The argument runs as follows: (A) a good theory is a summary of facts; (B) the predictive success of T'... has shown T' to be a good theory inside [the domain] D'; hence (Γ) if T, too, is to be successful inside D', then it must either give us all the facts contained in T', i.e. it must give us T', or at least it must be compatible with T'.' (Feyerabend [1962], in Feyerabend [1981], p. 69). This constraint is then claimed to impose a straight-jacket stifling the creativity of scientists in the field who must inevitably be working not only in different sociological and cultural environments, but also under a variety of aesthetic influences as well. '... [I]t is to be expected that theoreticians working in different countries, will arrive at theories which, although in agreement with all the known facts, are mutually inconsistent. Indeed, any consistency over a long period of time would have to be regarded not... as a methodological virtue, but as an alarming sign that no new ideas are being produced and that the activity of theorizing has come to an end.' (Ibid, p. 60).

As for Bohr's heuristic deployment of the Principle, Feyerabend [1981] insists that it should be viewed as being restricted locally only to the calculation results in the particular problem he was trying to solve. '...Bohr has.. repeatedly emphasized that "the principle of correspondence must be regarded as a purely quantum-theoretical law which cannot in any way diminish the contrast between the postulates... and electrodynamic theory."' (Ibid, p. 253n). And, 'Thus Bohr warns us in 1922... "that the asymptotic connexion" between the quantum theory and classical physics "as it is assumed in the principle of correspondence... does not at all entail a gradual disappearance of the difference between the quantum theoretical treatment of radiation phenomena and the ideas of classical electrodynamics; all that is asserted is an asymptotic agreement of the numerical statistical results". In other words, the principle of correspondence asserts an agreement of numbers, not of concepts.' (Ibid, p. 253).

Two things must be said about Feyerabend's reservations. Firstly, the conception of the Correspondence Principle under exploration in this thesis does in no way require the obliteration of fundamental differences between two inconsistent theories, one of which is to supersede the other, at any level. On the contrary, an attempt has been made to accommodate the intuitions which Feyerabend rightly attributes to Bohr. Thus, this concern of Feyerabend's relates to a non-existent feature in the conception in question, and he has no other reasons that would warrant barring its applicability across the board in physics. This seems to be especially the case since he appears to be quite happy regarding it as a *law* in Quantum theory. In the absence, therefore, of the feature that worries him, it seems that no cogent reasons remain within his position against regarding the Principle as a law *in the rest of physics as well*. Secondly, on my reading, the distinction that he, and not Bohr, introduces between the agreement of numbers and 'concepts' does not appear to make much sense. If Bohr's theory along with the right initial conditions entails that the numerical equivalence of the frequencies is a good approximation over the domain of slow vibrations, this also means that the Classical idea of the mechanical relation between the frequencies works quite well *in that domain*. Even though that idea does not reflect the truth anywhere, the phenomena included in this domain appear to conform to it, and in the light of the new theory we can, as we have seen, explain exactly why they do.

Hanson [1961], reflects very much the same sentiments although he appears not to go too far against the current of mainstream physics, so far as some sort of global validity for a version of the Correspondence Principle is concerned. His main reservation about the constraint imposed by the Principle is that, unless it is restricted merely to calculation results, it would force a continuity in the languages of a superseding and superseded theories which he feels is not maintained everywhere in physics. 'Theoretical works... appeal to the correspondence principle in at least two ways, both legitimate and clear. (1) In the formative days, namely 1913-27, quantum physicists often used classical mechanics as a criterion for the correctness of their calculations, and as a storehouse of suggestions about research and development within the theory... (2) It is a standard in all science that whatever the other merits of a new theory, unless it explains everything that could be explained by the theory it purports to replace, it is a non-starter.' (Hanson [1961], p. 226).

One claim the second clause in this passage is making in particular, is that whenever in science a new theory is to be developed with the aim of accounting for phenomena which proved the downfall of an otherwise successful extant theory, the transition from the latter to the new theory must be continuous. This is just the basic requirement of all the conceptions belonging to the Generalized Correspondence Principle. The first clause, on the other hand, by appealing to a historical fact, appears to express the same sentiment that was expressed by Feyerabend on the basis of distinguishing 'agreement between numbers' from that between 'concepts'. This is confirmed in the following passages from Hanson [1961]. 'That $(4\pi^2e^4m/h^3n_i^3)(n_i - n_f)$ gives a 'classical' frequency for the transition $\Delta n = 1$ proves at most that there are formal analogies between certain reaches of quantum theory and certain reaches of classical theory.' (Ibid, p. 154f). And, 'As an indication of how the mathematics of elementary particle physics can be managed, the correspondence principle is clear and useful. But when spoken of in more spectacular ways... the nature of intelligibility in physics hangs in the balance.' (Ibid, p. 157).

What, according to Hanson [1961], puts 'the nature of intelligibility' in the balance, is reading 'the preservation of continuity' as the preservation of the *language* of the old theory by the new one. To see the kind of breakdown in intelligibility which is allegedly overlooked by some conceptions of the Correspondence Principle, Hanson [1961] discusses the relation between the Quantum Theory and Classical Mechanics in some detail. 'In classical mechanics uncertainties in state determination are in principle eradicable. Expositions refer to punctiform masses, the paradigm examples of mechanical behaviour. Point-particles are distinct possibilities within classical particle physics... Elementary particle physics presents a different logical situation..., the discoveries of 1900-30... forced physicists to combine concepts in unprecedented ways... A direct consequence of these combinations of concepts is expressed in $\triangle p \,.\triangle v = h/m$, where $\triangle p$ and $\triangle v$ are the limits of uncertainty in a particle's co-determined position and velocity. Within quantum theory, to speak of the co-ordinates and momentum of an elementary particle at time t is to make no intelligible assertion at all...' (Ibid, p. 150).

In Classical Mechanics the state of a particle at any one time is fully determined by single values of at least two appropriate dynamical parameters. In any fully developed versions of Quantum Mechanics, on the other hand, such pairs of values do not, in general, determine a state for the particle. Instead, a new parameter is introduced and its distribution over all the points in the space of either one of the dynamical parameters at any one time, is taken to determine the state of the particle at that time. This distribution, in turn, determines *probability distributions* for possible values that the corresponding dynamical parameters are allowed to possess at the time. The Uncertainty Principle is the statement of a relation that holds between the range of possible values that each member of the appropriate pair of the dynamical parameters of the particle in a given state are allowed to possess with varying probabilities.

'Here the perplexity arises. A certain cluster of symbols S expresses an intelligible assertion in classical mechanics, yet that same S may not be so regarded in quantum mechanics... If S can be used to express an intelligible statement in one context, but not in another, it would be natural to conclude that the languages involved in these different contexts were different and discontinuous... Ordinarily this would be conclusive evidence that the languages are... *logically discontinuous*. But the correspondence principle apparently instructs us to regard them as continuous; quantum theory embraces the old classical laws as a limiting case... Either the uncertainty principle holds: that is, the S of classical physics makes no assertion in quantum physics, or the correspondence principle holds:

that is, the S of classical physics is a limiting case of quantum physics; but not both.' (Ibid, pp. 151-152).

These remarks suggest, not that in Quantum Mechanics single values for a particle's position and velocity are *inadequate* for the determination of its state at any one time, but rather that the sentence which says they possess such values, is devoid of meaning. That is, if one were literate only in the language(s) of Quantum Mechanics, one would fail to understand that any proposition is expressed by the sentence in question. That this suggestion is in fact false, requires a closer look at Quantum Mechanics and in a considerably more detailed manner. This will be done in chapter 5. It will be shown there not only that in Quantum Mechanics to speak of a particle's exact position and velocity at a certain time makes as precise a sense as it does in Classical Mechanics, but that to speak in any other terms would render the assertions of the former, if not entirely unintelligible, obscure beyond reasonable possibilities of comprehension. The chapter will also suggest that, contrary to the claim just quoted, the Uncertainty Principle may hold *together* with the view that Classical Mechanics (minus a theory of gravitation) is a limiting case of Quantum Mechanics. In the meanwhile I shall assume that the reservations expressed in Hanson [1961] do not threaten the justification of the global validity of the Correspondence Principle, as I read it, in physics.

4. The Charge of Inconsistency

In Lakatos [1970] concerns of an entirely different kind are expressed about the role of the Correspondence Principle. Although the heuristic efficacy of this principle, as deployed in Bohr [1913], is acknowledged, the Principle is nevertheless described as an '*ad hoc*' stratagem, 'the only purpose of which is to hide' a 'deficiency' (p. 142). The deficiency in question refers to an alleged inconsistency which supposedly lies at the heart of Bohr's theory of electrons in atomic orbit. 'The background problem was the riddle of how Rutherford atoms... can remain stable; for, according to the well-corroborated Maxwell-Lorentz theory of electromagnetism they should collapse. But Rutherford's theory was well-corroborated too. Bohr's suggestion was to ignore for the time being the inconsistency and consciously develop a research programme whose 'refutable' versions were inconsistent with the Maxwell-Lorentz theory.' (Lakatos [1970], p. 141).

If there has been any suggestion by Bohr to *ignore* the inconsistency between his theory and the Classical electrodynamics, it is not to be found in his [1913]. There is, and we have noted this, an element of fiat in Bohr's proposal for overcoming the difficulties he had faced. But postulating changes by fiat, unsatisfactory as it may be, is not to countenance inconsistencies. Lakatos's description of the role of the Correspondence Principle in Bohr [1913] is, therefore, incorrect and misleading. Incorrect, because there is no inconsistency which Bohr *intended to hide*. Misleading, because it presents a legitimate principle which performs an effective function in the context of discovery as a mere trick towards obtaining a desired result. This function, although unnoticed until Bohr, was always available for exploitation by physicists throughout the ages and may also have been used implicitly by some of them.

From the examination of the heuristic potentials of the Correspondence Principle in the particular case of Bohr [1913], we notice another function that it performs. Besides guiding the search for a new relation between the frequencies, it enables the theory that finally incorporates this relation, *to explain why* the old theory is successful in the domain of low frequencies. In the light of the new theory, this domain is seen to correspond to larger and larger values of the integral variable which determines various orbits of the electron. From the new theory it can, therefore, be *demonstrated* that treating the frequency of the radiation from the atom as numerically equivalent to that of the periodic motion of its electron, is a perfectly adequate approximation for large values of the integral variable.

Radiation from the atom in the new theory is attributed to the electron jumping from a higher to a lower nuclear orbit. But the new theory also correlates increases in the radius of the electron's orbit to larger and larger values for the integral variable. When this radius becomes so large that separations between adjacent orbits can hardly be detected at all, the atom *appears* to be radiating *as though* the electron remains on a stationary orbit. The relation that the new theory postulates between the frequencies of orbit and radiation, then yields results that for jumps between such orbits are, for all practical purposes, virtually indistinguishable from those yielded by the old relation.

Going back to our T_1 and T_2, we can generalize this point. If T_2, along with some initial conditions, entails that T_1 is a good approximation in a certain domain of phenomena, then T_2 explains why T_1 is successful over that domain. Since the Correspondence Principle requires T_2 to entail (with the right initial conditions) that T_1 is a good approximation in the appropriate domain, it requires that T_2 should explain the successes of T_1. Indeed, the heuristic efficacy of the Correspondence Principle in the context of discovery stems from this very requirement. For since this efficacy consists in allowing the use of some of T_1's assumptions in the development of T_2, even though they are false, this use can be justified only if T_2 explains why those assumptions have a limited validity.

5. The Two-tier Realist Conception of Theories

The extra explanatory burden that the Correspondence Principle puts on a superseding theory, assumes a particular urgency for the two-tier realist view of theories. According to this view it is desirable to assess the empirical successes of a theory that is methodologically sound, in terms of a truth-related property. The best candidate for such a property in physics appears to be something like progress in the direction of capturing the underlying physical constitution of the world. However, mere combination of empirical success and methodological soundness is not sufficient for such an assessment. Without an additional requirement that would preserve some continuity in the transition between theories, the two-tier realist assessment of outstanding theories could face the danger of running into inconsistency.

Two circumstances can be envisaged in which basing the two-tier realist assessment of successive theories on methodological soundness and empirical success alone, is vitiated. Let us, to begin with, assume that this combination suffices for the assessment in question: (*I*) Suppose T_1 is a methodologically sound theory with an impressive record of empirical success. According to our assumption, T_1 must be regarded as having progressed in the right direction towards the truth. Suppose now that T_1 is falsified, and replaced with a theory T_2 which is just as methodologically sound, but with a more impressive record of empirical success. The inclination would be to say that on these scores T_2 has progressed further than T_1 in the direction of the truth. But suppose that T_2 clashes with T_1 on most fundamental issues in their respective descriptions of the underlying constitution of the world without, in the meanwhile, offering any recognition of T_1 as a *legitimate* theory over a certain domain (i.e. as a step in the right direction towards the truth). In the light of the assessment of T_2 as progress towards the truth, it would be awkward to maintain that T_1 is also progress *along the same direction*, but only to a lesser extent.

(*II*) This time suppose that T_1 and T_2 are both methodologically sound theories which have an *equally impressive* record of empirical success. According to our assumption, they must both be assessed as having progressed to the same extent in the direction of capturing the underlying physical constitution of the world. However, let the two theories, as before, clash on most fundamental issues in their description of this constitution. The assessment of the two theories in this case would then amount to regarding radically different descriptions of the world as enjoying the same proximity to the truth.

In both (*I*) and (*II*) it appears that the two-tier realist assessment that is accorded to one theory cannot hold when it is made, *on the same grounds*, of its rival. In (*I*), T_1 is assessed as having progressed towards the truth on the grounds of its methodological soundness as well as empirical successes. But when T_2 is assessed on the same grounds, it becomes necessary to revise the original assessment of T_1, and perhaps even to reverse it completely. In (*II*), assessing T_1 and T_2 on the same grounds would have to accord to both the same proximity to the truth. This assessment, however, cannot be consistently maintained since the two offer radically different descriptions of the underlying constitution of the world. The two-tier realist assessment of theories, therefore, needs to be modified if it is to handle the relation between superseding and superseded theories in general.

6. Some Concrete Examples

Let me illustrate the point with the help of some examples that I take from the history of science. In order to accommodate these examples to my purposes, I shall have to make some contrary-to-fact assumptions which are otherwise innocuous. Consider the Ptolemaic and the Copernican theories realistically interpreted. For the Copernican theory, however, let us not take the product completed by Copernicus himself by the time of his death in 1543 (presented in the *Commentariolus*, and *De Revolutionibus Orbium Coelestium*). Instead, let us take the version that incorporates all the modifications introduced by Kepler, not only in his *Mysterium Cosmographicum* of 1596 (Dreyer [1953], pp. 376-378), but also by the end of the year 1604 (Ibid, p. 382).

Kepler's motivation in choosing the Copernican theory against its chief rivals was not only that it is more fertile in the explanation of the known facts and the prediction of new ones, but also that it allows for greater unity and symmetry. He expresses this preference in no ambiguous terms at the beginning of *Mysterium Cosmographicum*: 'My confidence was first established by the magnificent agreement of everything that is observed in the heavens with Copernicus's theories;

since he not only derived the past motions which have been recapitulated from the remotest antiquity, but also predicted future motions, not indeed with great certainty, but far more certainly than Ptolemy, Alfonso, and the rest.' Kepler [1596], p. 75. And shortly afterwards, 'For, to turn from astronomy to physics or cosmography, these hypotheses of Copernicus not only do not offend against the Nature of things, but do much more to assist her. She loves simplicity, she loves unity. Nothing ever exists in her which is useless or superfluous, but more often she uses one cause for many effects.' (Ibid, p. 77). Let us look at these points one by one.

I. *Explanatory improvements*

(*a*) In the Ptolemaic systems, the observed stations and retrogressions of the planets (i.e. the reversal of the direction of their motion after a short period during which they appear to stand stationary) has to be accounted for, in each individual case, by adjusting the size of the epicycles which are pegged on the principal circles which are, in turn, assigned to take them round the centre of the universe (i.e. the earth). In the Copernican system, these fall out (albeit not with great accuracy) naturally from the geometry of the sun-centred orbits without requiring such ad-hoc adjustments.

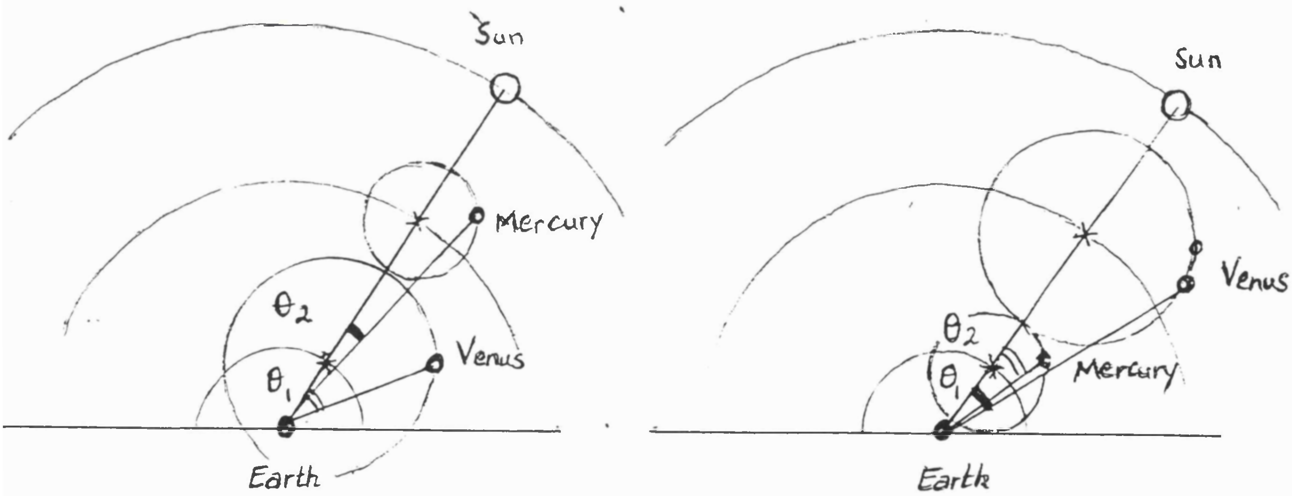
(*b*) The observed positions of the inferior planets (those whose orbits lie between the earth and the sun), namely Venus and Mercury, never exceeds a fixed distance from the observed position of the sun. This fact could only be accounted for in the Ptolemaic systems by means of an ad-hoc requirement according to which the centres of each planet's epicycle must always lie on a straight line connecting the earth and the sun. In the Copernican system, on the other hand, because the earth orbits the sun with the orbits of these planets lying in-between, their positions, as viewed from the earth, can never deviate from that of the sun by more than a fixed amount. The facts in this case, therefore, follow from the sun-centred arrangement of the orbits in the Copernican system.

(*c*) The frequencies of the retrogressions of the inferior planets are hugely different: That of Venus is once in 584 days, and that of Mercury is once in 116 days. In the Ptolemaic systems, this fact could only be accounted for by requiring the epicycle of Mercury to rotate with a speed five times greater than that of Venus. This requirement, however, must be imposed for no other reason than saving the appearances. In the Copernican system, on the other hand, the difference in these frequencies follows naturally from the fact that the orbit of Mercury is smaller than that of Venus, and the latter smaller than that of the earth.

II. *Predictive improvements*

(*a*) For the inferior planets, the angle of elongation (i.e. the angle by which the position of a planet, as viewed on earth, deviates from that of the sun) is always acute. The maximum elongation of Venus, however, is greater than that of Mercury. Since in the Ptolemaic systems it is necessary for

Venus and Mercury to be carried on epicycles, the relative order of their deferents (i.e. the principal circle around the centre of the planetary system on the circumference of which epicycles are centred) must be determined arbitrarily. These maximums can be reproduced by either putting Mercury's deferent closer to the sun than that of Venus, or vice versa, adjusting the sizes of their respective epicycles accordingly (as in Figure 1).



θ_1 = maximum elongation of Venus
θ_2 = maximum elongation of Mercury

Figure 1

In the Copernican system, on the other hand, the relative sizes of the orbits of the inferior planets can be determined from the geometry of the sun-centred orbits plus the values of their maximum elongations. The greater the maximum of elongation of an inferior planet, the larger the radius of its orbit from the sun (as can be seen from Figure 2).

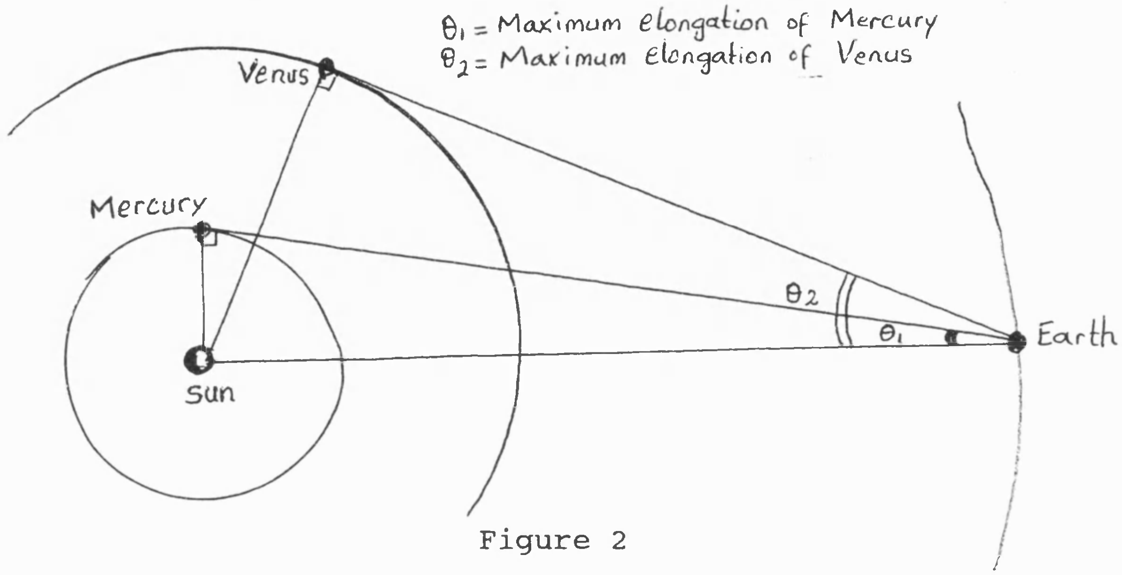


Figure 2

(*b*) Sizes of the orbits of all the planets can be determined using the Copernican system plus some initial conditions. The same, however, cannot be said of the Ptolemaic systems. For the inferior planets, their maximum elongation is used for this purpose. At this maximum, the distance between the earth and the planet lies along the tangent to the planet's orbit (see Figure 2). It, therefore, forms an angle of 90° with the radius of this orbit. Together with the radius of the earth's orbit, these form a right-angled triangle, one acute angle of which is known. The value of one acute angle of a right-angled triangle fixes the ratios of its sides. From these ratios, in turn, the relative sizes of the inferior planets' orbits can be deduced. Similar, but more complicated, calculations could be used to derive the relative sizes of the other planets (Kuhn [1957], pp. 175-176).

(*c*) The periods of all the planets can be calculated using the Copernican system and some direct measurements. The same cannot be said of the Ptolemaic systems. Let F be the frequency of the event that a planet and the earth lie on a straight line to the sun. From the geometry of the sun-centred orbits, F is determined by the difference of the frequencies of the earth's and the planet's revolutions. If we represent the period of the planet's revolution by T, the frequency of its revolution is 1/T. We then have:

$$F = 1/T - 1/365, \text{ for the inferior planets, and}$$

$$F = 1/365 - 1/T, \text{ for the others.}$$

F can be measured by taking the time interval between two successive retrogressions. T can, therefore, be calculated. It is interesting to note that with this calculation we can test the order of the inferior planets' orbits which were independently derived using their maximum elongations. Assuming a direct proportionality between the period of a planet and the size of its orbit, it follows that a planet with a longer period must have a greater orbit around the sun. We can, therefore, compare the order of the orbits dictated by the maximum elongations, with that derived from the frequency calculation.

(III) *Improvements of unity*

(*a*) A common requirement in both the Ptolemaic and Copernican systems is that the planets should move along circular paths with uniform speeds. In the Ptolemaic systems, the deferents of the sun and the planets are centred on the earth. The speed of their rotation, however, can only be viewed as uniform, not with respect to this centre, but with respect to a geometrical point, called the 'equant', some distance away from the centre. This necessity introduces an element of asymmetry in the cosmological presuppositions underlying the Ptolemaic systems. Realistically interpreted, the deferents are, in accordance with the physical principles of Aristotle, geometrical representations of corporeal spheres. These are all centred around the earth because the latter has a privileged position in the universe. That the speed of their rotation should be uniform not with respect to this centre, therefore, begs an explanation which is not provided.

In the Copernican system, all points of equant are done away with. This was, in fact, claimed by Copernicus as the major advantage of the sun-centred cosmology over its rivals. The new system, however, had problems of its own in this respect. Although the sun was now claimed to be the centre of the universe, Copernicus never managed to centre all the deferents on the position of the sun. So, although the speed of the deferents could be viewed as uniform with respect to *their* centres, these did not coincide with the centre of the universe. Nevertheless, if a way could be found in which the centres of the deferents are brought to coincide with the position of the sun, a major advantage, so far as the unity of the system is concerned, could be claimed. This must have provided Kepler with a strong motivation for seeking to improve the Copernican system.

(*b*) In the Ptolemaic systems, the sizes of the orbits (and not all of them at that, as we have seen) could only be determined by taking on board an additional assumption (besides that of the proportionality of the periods to the distances), commonly known as the 'principle of plenitude'. According to this principle, physical space contains no empty places and is filled by spheres made of an ephemeral material that are tightly packed one inside the other in a cogwheel-type arrangement. This principle also plays a dynamical role in that it allows the motion of the stellar sphere - the outermost sphere in the hierarchy of spheres - to be imparted to all other spheres that must rotate around the earth. But this role is secured only at the cost of banishing the primary cause of all motions (which is necessary to set the stellar sphere in rotation), namely the 'prime mover', to outside the realm of physics. So, there is an additional source of asymmetry in the cosmology underlying the Ptolemaic systems: physical *as well as* non-physical causes are required to account for the motions of corporeal entities.

In the Copernican system, the principle of plenitude is, as we have seen, not needed for calculating the sizes of the orbits. In addition, because the stellar sphere is set at rest at a vast distance from the last rotating sphere (in order to avoid parallax), its violation is strongly suggested. This affords the chance for seeking only physical causes from within the universe in order to account for the motion of the planets. Kepler sought these causes first in terms of forces emanating from the sun alone. This appears to tie up nicely with the assumption of the direct proportionality between the length of the periods and the size of the orbits. If, therefore, progress could be made in the direction of securing centrality for the sun, a more unified cosmological account would, in its light, appear as a promising prospect.

The best Kepler could achieve by proceeding in this direction, while preserving the assumptions of the circularity of the orbits and the uniformity of the planets' speeds, was to introduce the following modifications to the Copernican system: (*i*) Copernicus had conceived the planes of the planetary orbits as intersecting in the centre of the earth's orbit. Not being able to adopt the position of the sun as the centre of this orbit, the best he could do was to set it on an epicyclic rotation around a circle centred on the sun (see Figure 3). The circle which carries the centre of a planet's orbit is called an 'eccentric'. The effect produced by this construction is to approximate, by means of circles, the eccentricity in a planet's orbit which we now know (and Kepler soon afterwards discovered) is due to its elliptical shape.

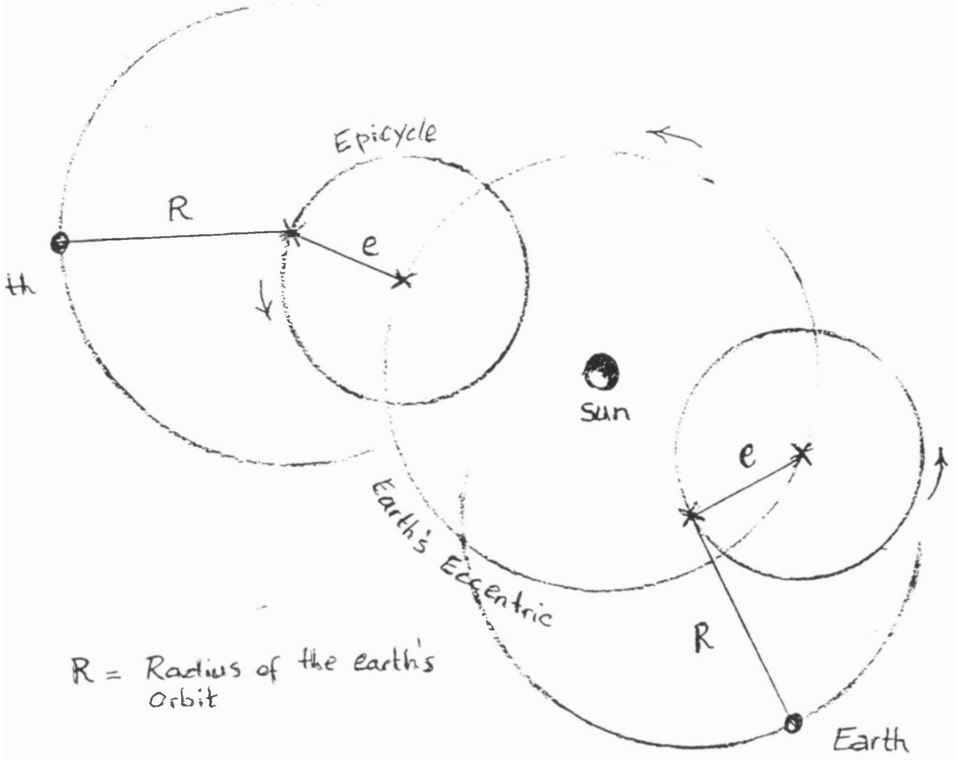


Figure 3

By making the planes of planetary orbits intersect in the position of the sun itself, Kepler observed that considerable improvements result for the latitudes (deviations north and south from the plane of the ecliptic) of all the planets (Kuhn [1957], p. 210, and Dreyer [1953], pp. 383, 393). (*ii*) Copernicus had used two different origins for the eccentricities of the planetary orbits. The eccentricity of the earth's was measured from the sun (as in Figure 3), while those of every other planets' were measured from the centre of the earth's orbit. The latter was motivated by the wish to account for his belief, based partly on inaccuracies in the data available to him, that the eccentricities of Mars and Venus had undergone some change since the time of Ptolemy (Kuhn [1957], pp. 210-211, Dreyer [1953], p. 378. Note that Kuhn names Mercury rather than Mars in this regard. A comparison with Kepler [1596], pp. 161, 181, and 183 shows that this is a misidentification). By measuring all eccentricities from the sun, Kepler observed that these variations disappear and a better fit to more accurate data is achieved.

The theory which emerges from Kepler's gloss on the Copernican system (let us, for brevity's sake, call it 'Kop'), appears to hold encouraging promises for a fully fledged sun-centred cosmology. Moreover, in comparison with the best available version of the Ptolemaic system (let us call it 'Ptol'), Kop proves superior on methodological, as well as empirical, grounds. From the realist point of view, therefore, it merits the property of being closer to the truth than its rival.

7. The 'Paradox' of the Two-tier Realist Assessment of Success

As has been noted, Ptol suffers from serious methodological shortcomings. However, for the sake of the argument, let us overlook them and assume that during the period in which it faced no serious rival, it could be assessed, from the same realist point of view, as progress in the direction of capturing the cosmological order in the universe. This assumption, however, does not appear capable of holding its ground when the same realist considerations are applied to the superior Kop, which must also be assessed as progress in this direction. For not only on almost all the points at issue between them Kop and Ptol offer diametrically opposed conceptions, but also there is nothing in Kop in the light of which Ptol could be considered as a fairly good approximation *if the cosmological story that Kop has to tell is assumed to be true*.

First, there is a clash between the two on the conception of the physical space. Ptol says 'no place without a substance', whereas Kop asserts the existence of vast regions of empty places. Moreover, physical space in Ptol is highly regimented: (*i*) The centre is the place for substances whose 'natural state' is that of rest. Everywhere else is reserved for substances whose natural state is that of uniform motion. (*ii*) The central region is inhabitable only by substances which have weight and are subject to change and decay. The rest can be inhabited only by substances which are weightless and unperishable. Physical space in Kop, on the other hand, is not at all sensitive to the kind of entities that can or cannot occupy it. The centre allows motion like everywhere else; and any region can be occupied by perishable substances.

Second, the earth in Ptol is regarded as the stationary centre of the universe, whereas in Kop it appears as a wanderer like all the other planets. Moreover, dynamical states of rest or motions in Kop are not 'natural' to particular kinds of substances. A celestial object such as the stellar sphere is allowed to stay at rest, while the earth is set in perpetual motion (of three different kinds, at that).

Third, the moon appears in Ptol as a planet going around the centre of the universe like all the other planets. In Kop it is portrayed as a satellite orbiting the planet Earth, just as the moons of Jupiter, discovered earlier by Galileo, are satellites orbiting that planet. Finally, the cause of planetary motions is conceived in Kop as residing in a physical object, the sun, and placed within the universe. In Ptol it is viewed as originating in a supernatural entity which resides outside the physical universe.

The descriptions of the physical universe offered by the two theories are too far apart to make it easy to consider them in the same league so far as progress towards the truth is concerned. The new theory, furthermore, makes no allowance whatsoever for sustaining the assumption of progress that we attributed to the old theory. Consequently, there is no continuity in the transition from the old to the new theory; no meeting point on the basis of which the progress of the new theory could be viewed as somehow having been *built on* the progress of the old. Any merit, therefore, that is due to the account Kop offers, on the grounds of its superiority, must be claimed at the expense of whatever merit that we were previously led to ascribe to Ptol. If Kop is to be considered as having made some progress towards capturing the truth of the cosmic constitution, Ptol must, in the light of the changed circumstances, be considered as having been on the wrong track all along.

Historically, in fact, we do not face the problem of such radical reversal of decision in the transition from Ptol to Kop. Ptol actually suffers enough methodological shortcomings for the two-tier realist to consider it out of serious contention for his truth-related property of merit. It lacked predictive fertility and achieved most of its empirical 'successes' by making ad hoc adjustments to the incoming, as well as improved, data. But if, contrary to facts, Ptol were not to suffer from all these shortcomings, and there were no requirements on the theory succeeding it to entail that it was a good approximation, its two-tier realist assessment could endow it with the merit of progress towards the truth at one time, only to strip it of this merit altogether when it is superseded.

If such a situation were to prevail for theories which not only are satisfactorily adequate over an empirical domain, but pass the methodological test of soundness as well, the two-tier realist conception of theories could be faced with something like a paradox. In the absence of continuity in the transition from one theory to another, this outlook could appear to allocate contradictory assessments on a refuted, but otherwise outstanding, theory. On the one hand, limited but hard won success is interpreted as indicating partial progress towards capturing the truth. On the other, it would appear to amount to no such thing at all when the theory that once achieved it is superseded by one which offers a radically different account.

The problem appears more pronounced when we consider the possibility of rivalry between two theories that are equally outstanding, but radically different in accounting for the same facts. Once again, let me take an example from the history of science and tailor it for my purposes by making an innocuous counterfactual assumption. In 1588 Tycho Brahe published his *De Mundi aetherei recentioribus phaenomenis liber secundus*, in which a new system of planetary arrangements was presented. The novelty of this system consisted in grafting portions of the Copernican system on to the geocentric view of the universe in such a way that all the explanatory, as well as predictive, advantages of the former over the Ptolemaic systems are preserved.

Tycho was an astronomer *par excellence*, in the sixteenth century sense of the word, and hence with his attention firmly focused on the agreement of calculations, based on geometrical constructions, with observational data. Thus the fact that his proposed system was, from the view point of physical cosmology, very awkward, to say the least, appeared not to trouble him at all. The rough outline of his system has the earth positioned at the centre of the universe, with the moon and the sun in orbits around it. The rest of the planets (Mercury, Venus, Mars, Jupiter and Saturn) are set on orbits centred at the sun (in that order).

Tycho, himself, refused to believe in the reality of celestial spheres (Letter to Kepler of 9 December 1599; in Koyré [1973], pp. 161-162), which makes it difficult to accommodate a causal account of the motions of the celestial objects within his system. But even if these spheres are to be introduced to his system, there is no physical mechanism whereby the motion of the stellar sphere could be imparted to the spheres of the moon, the sun, and the inferior planets. The reason is that the stellar sphere, along with the spheres of Saturn, Jupiter, and Mars would have to rotate in the opposite direction to those of the rest (Kuhn [1957], p. 202).

Let us overlook these shortcomings and assume, again for the sake of the argument, that the Tychonic system is, from the methodological point of view, as good as Kop. Since empirically the two systems are equally adequate, in the light of our assumption they must both be regarded by the realist as having equal merits. The Tychonic system, however, is deeply entrenched in the same cosmological principles that underlie the Ptolemaic systems. In fact, it was in an attempt to uphold these very principles in the face of serious challenges that were raised against them that this system was produced. In comparing Kop and Ptol, we saw that these principles pull in opposite directions. The Tychonic system and Kop, therefore, cannot both be regarded as progress to the same extent towards the truth. We seem to face an inconsistency in our assessment of these theories.

Although the impasse we face in both this example and the one we contrived involving Ptol and Kop, stems from the same problem (viz. the insufficiency of methodological soundness and empirical adequacy, on their own, for the assessment of the proximity of a theory to the truth), the situation is different in the two cases. In the latter, we are forced to assess Kop as closer to the truth than Ptol. This assessment, however, also forces us to reverse the decision we made earlier regarding Ptol when it was unrivalled. But *that* decision was made on exactly the same grounds as the one which forced us to assess Kop as a closer theory to the truth. One and the same criterion, therefore, appears to produce contradictory assessments of a theory depending on whether or not a suitable rival is on the scene. This makes the truth-related property that the two-tier realist wishes to detect in theories, rather too shaky for comfort. After we decide that it is there in a theory, it can evaporate merely as a result of the appearance of an eligible rival on the scene. On the other hand, in cases exemplified by the rivalry which we set up between the Tychonic system and Kop, the assessment that seems to be equally deserved by the protagonists cannot be accorded to them both. Assessing the two theories by the same criterion, we are forced to regard them as equally close to the truth. This assessment, however, cannot be maintained because the two are diametrically opposed in their descriptions of the underlying constitution of the universe.

8. Resolution of the 'Paradox'

Obviously the two-tier realist wishes his belief in the relative proximity of a theory to the truth to have a sound objective basis. It is, therefore, desirable that if this property is detected in a theory, it is consistently upheld vis à vis the kinds of challenges we have been considering. The explanatory burden that the Correspondence Principle imposes in physics, appears to meet this desideratum admirably. Let us see how this is done in the particular cases we have been considering.

In the case of an outstanding theory T_1 that has run into trouble, the requirement that the Correspondence Principle imposes amounts to the following: a theory T_2 is a suitable replacement for T_1 if it is not only methodologically sound and empirically superior, but also entails that T_1 is a good approximation over the domain where it proves empirically adequate. By meeting this requirement, T_2 ensures that the merit achieved by T_1 in virtue of its successes is preserved, no matter how deeply the accounts offered by the two theories clash. The assessment of T_2 as progress towards the truth, far from necessitating the reversal of the earlier decision about the proximity of T_1 to the truth, secures for the latter the assessment that it is also progress, though to a lesser extent, in the same direction. By taking on board the additional requirement demanded by the Correspondence Principle, therefore, the problem of reversing the decision about the merits of a successful theory is resolved.

In the case we discussed involving Ptol and Kop, the latter proved a superior theory to the former without, in fact, offering any explanation of its successes. This can be accounted for by observing that Ptol, as we saw, suffered from serious methodological shortcomings. But since Kop not only does not share these defects, but proved to be a promising theory, one should expect that an outstanding theory that would supersede it fulfils the requirement demanded by the Correspondence Principle. Such a theory was produced by Kepler and was partly presented in his *Astronomia Nova* of 1609, and partly in his *Epitome Astronomiae Copernicanae* of 1618, 1620, and 1621 (Dreyer [1953], p. 395).

Apart from a few remnants from the Copernican system, Kepler's theory is radically different from extant cosmological theories at the time. It sets the boundary of the universe in the stationary stellar sphere; it has the sun firmly positioned at the centre of this sphere; and it treats the earth as a planet. However, it jettisons the laws of circular motions and uniform speeds, and assigns to each planet a simple elliptical orbit, fixed around the sun at one of the foci, as well as a varying (though law-governed) speed. It banishes standard devices such as eccentrics, deferents and epicycles from astronomy once and for all, and introduces new laws and dynamical causes to cosmology.

One of the laws governs the speed of the planets. According to this law each planet moves with such a speed that during equal time-intervals, the straight line connecting the planet to the sun sweeps out equal areas. Consequently, the speed with which a planet moves at a given time is inversely proportional to its distance from the sun at that time. When the planet is farther away from the sun, it moves slower than when it is nearer the sun. The other law connects the ratio of the periods of any two planets to that of their average distances from the sun, and was presented separately in *Harmonice Mundi* of 1619.

In Kepler's theory there are two essentially physical causes responsible for planetary behaviour. One originates from the sun alone and is the moving force which perpetually pushes the planets along their orbits. This may be visualized as lines of force that emanate from the sun, like spokes of a bicycle wheel, across the plane of planetary orbits around the sun. The sun spins round a fixed axis with uniform speed, turning, as a result, the lines of force that extend from it to all the planets. In positions nearer the sun these lines are more densely packed. Therefore, planets in orbits closer to the sun are, at any one time, driven by more lines of force than those with orbits further apart. This accounts for the increase in the length of a planet's period as the size of its orbit increases.

The shape of the planetary orbits and variations in the orbital speeds of the planets are accounted for by the action of magnetic forces which Kepler ascribes to the sun as well as to each planet. This idea was reinforced, if not motivated, by the fact that William Gilbert had established, in his *On the Magnet*, that the earth was itself 'a huge round magnet' (Dryer [1953], pp. 394-395). The magnetic property which Kepler attributes to the sun turns out to be very peculiar indeed. Its south pole is buried deep in the centre of the sun, so that it can exert no influence at all; and its north pole is distributed uniformly all over the sun's surface. The planets, in turn, have their magnetic poles, very much like those of the earth's, at the opposite ends of an axis that remains parallel to itself throughout their orbits (Kuhn [1957], p. 246).

The upshot of this arrangement is that a planet throughout its orbit is always under the influence of the north pole of the sun. There are, however, two positions at the opposite ends of its journey where both magnetic poles of a planet are equally distanced from the sun's north pole. The net magnetic influence of the sun on the planet's motion is zero when the latter occupies these positions. In between these positions, as the planet is gradually pushed along its orbit, the distance between one of the planet's poles and the sun's north pole decreases towards a minimum. When this happens to the south pole of the planet, the latter is attracted towards the sun. When it happens to the north pole of the planet, the latter is repelled from the sun. As this minimum is passed for one pole, the planets' poles approach the position of equidistance, after which the distance involving the opposite pole begins to decrease, and so on.

The net effect of a planet's periodic attraction towards, and repulsion away from, the sun, while constantly being pushed around by the turning lines of force, is a motion round an elliptic path. As the planet gets closer to the sun, it meets a greater number of lines of force (because they are more densely packed in such regions). It, therefore, experiences a greater pushing force, and thus will move with a greater speed than when it is farther away. The elliptical orbit, as well as variations in the speed, of the planets are thus accounted for by the action of two forces.

As the magnetic interaction that Kepler postulated to hold between the sun and the planets, is allowed to diminish in strength, the periodic swing that the planets make towards and away from the sun, would become less and less pronounced (in fact it *is* less pronounced already in the case of the outermost planets). In the limit, where this interaction vanishes altogether, one would expect the planets to move along circular orbits centred at the sun. Under such circumstances, the number of lines of force that extend from the sun to each planet would remain constant at all times. Since the sun is spinning uniformly, it follows that the planets should move with constant speeds on their orbits.

The system of planets which orbit the sun in circles with uniform speeds was the very system that Copernicus presented in his *Commentariolus* as an ideal towards which astronomy should strive. He never managed to attain to this ideal himself, but Kop went as far as it is possible to achieve it. It managed to position the sun not only at the centre of the stellar sphere, but at the centre of the planes of all planetary orbits as well. In addition, it endowed the sun with moving forces which alone were responsible for all planetary motions.

Now, Kop is crowded with eccentrics, deferents and epicycles. But if we cut it down to the same rough form as the system presented in Copernicus's *Commentariolus*, we will have a heliocentric system of planets that move in circular orbits, with uniform speeds and on account of a dynamical influence from the sun. This system differs from Kepler's own in empirical adequacy. But it does so only in so far as it ignores the magnetic interaction between the sun and the planets. In fact if we just introduce this interaction into this rough form of Kop, we will get precisely the system that Kepler finally arrived at through a much more arduous route. From the vantage point of Kepler's theory, therefore, it is seen that Kop is the best one can get if the magnetic interaction between the sun and the planets is not taken into account.

Kepler's theory thus ensures that the merit of Kop as a step in the right direction towards the truth is preserved no matter how radically the two theories differ from one another. Therefore, the assessment of Kepler's theory as progress towards the truth also secures for Kop the assessment that it is also progress (though to a lesser extent) in the same direction. The same, however, cannot be said of Ptol. There is nothing *in Kepler's theory* which can shed any light on the *reasons for* the empirical successes achieved by Ptol; no parameter whose behaviour in certain regions of empirical phenomena may suggest that Ptol is a good approximation to the underlying structure of the universe. As it happens, we know that Ptol is methodologically unsound. May be *that* has something to do with its being off course in the direction of the truth. Intriguing as this question may be, I cannot go into it here. What I wanted to show is how the requirement imposed by the Correspondence Principle ensures that a two-tier realist assessment of methodologically sound and empirically adequate theories which differ fundamentally from one another in their description of the world can be consistently maintained.

The Generalized Correspondence Principle is equally efficacious in deciding which of the two radically different theories which enjoy *the same* degree of empirical adequacy is the one that is directed towards the truth. The case in point is that which I contrived between Kop and the Tychonic system. Here the problem arose when both theories were assumed to demand the assessment of being equally close to the truth, but could not be so assessed because the stories they offered clashed radically with one another. In cases such as this if the constraint imposed by the Correspondence Principle is respected generally, the question can be decided by the appearance of the theory that supersedes both protagonists. *That* theory has the merit of being progress towards the truth which is singled out as a good approximation by the superseding theory.

In my example, the theory which superseded both Kop and the Tychonic system was Kepler's theory. Its methodological soundness, as well as empirical superiority, suggest that it has the merit of being closer to the truth than either of the theories which it supersedes. Kepler's theory, in turn, meets the requirement demanded by the Correspondence Principle by entailing that Kop, and not the Tychonic system, is a good approximation to what it describes as the truth of the matter. Therefore, the assessment of Kepler's theory as progress towards the truth singles out Kop as progress, though to a lesser extent, in the same direction. This assessment leaves the Tychonic system out of the contention for this merit altogether. The question, 'which of the two theories, Kop or the Tychonic system, while equally adequate in dealing with the empirical phenomena, enjoys proximity to the truth?' is thus decided.

Like Ptol, the Tychonic system too is, in fact, unsatisfactory on methodological grounds. The traditional explanation for why the celestial objects revolve around the earth was that it is stationary and occupies a privileged position at the centre of the universe. Neither of these properties is, in the Tychonic system, possessed by the sun; yet all the planets are set to revolve around it. Moreover, there does not seem to be a physical mechanism which is capable of sustaining this system. The point is, however, that even if it did not suffer from any such defects, the Tychonic system would lack the merit of being progress towards the truth because the superior theory that supersedes it and its rival Kop, picks the latter as a good approximation to the truth. Again, a problem which the two-tier realist may face in assessing two equally empirically adequate theories which clash fundamentally in their descriptions of the world, is resolved if the satisfaction of the requirement demanded by the Correspondence Principle is deemed compulsory in physics.

What emerges from this discussion is that if theories are conceived from the two-tier realist point of view, then the Correspondence Principle must be an essential requirement in physics. Non-realist conceptions of theories have no need for such a *principle*, because they are not interested in any truth-related (in the realist sense of 'truth') properties of theories. Their primary interest lies generally in finding expedient schemes for making predictions which by fitting the widest available data would provide reliable estimates for what is to be expected in the future. Discontinuity in the change between theories, therefore, is not something which would pose a problem for them.

It is, of course, very much a matter of debate how much of the business of doing physics relies on the two-tier realist conception of theories. Certainly in the early days when Quantum Mechanics was in the making and physicists working in this area were very much groping in the dark, non-realist sentiments were quite often aired by some prominent contributors to this development. Nor is this the only period in which sentiments of this kind are expressed by prominent physicists. More often than not, when the adequacy of a particular account offered by a spectacularly successful theory is in question because it faces a serious crisis, such sentiments appear to gain force and popularity. To what extent the two-tier realist interpretation of theories is essential in physics is a question which does not concern me here. What does concern me here is (1) which conception of the Correspondence Principle can have the chance of holding generally in physics, and (2) to what extent this conception may be viewed as entrenched within physics. The conception I have proposed in this chapter, if at all tenable, appears to be general not only in the sense of holding in more than one branch of physics, but also in the sense of having the *reductionist*, as well as the *ontological* and the *structural* conceptions as its special cases. It also appears to account for cases of correspondence in which they fail to hold. I have also tried to show that heuristic considerations apart, the requirement demanded by the Generalized Correspondence Principle must be maintained if the success of theories in physics is to be consistently assessed from the two-tier realist point of view.

CHAPTER 4

VERISIMILITUDE AND RATIONAL THEORY CHOICE IN PHYSICS

1. The Correspondence Relation

The empirical adequacy of Kepler's theory does not extend beyond the domain of phenomena over which Kop may be considered as a good approximation. It is a better theory over exactly the same domain. Newton's theory, however, not only improves on Kepler's over this very domain, but is also adequate over a larger domain. It covers terrestrial as well as celestial motions. This large difference apart, the relation between Newton's theory and Kepler's is essentially the same as that between the latter and Kop. The two theories are radically different, and Newton's theory explains why Kepler's was successful in accounting for planetary motions. The moving force postulated by Kepler's theory to emanate from the sun is done away with in Newton's theory, and the magnetic properties of the planets or the sun play no role in the motions in the solar system. Instead, the latter postulates a single attractive force, *gravitation*, which acts between massive bodies in proportion to their masses, and in inverse proportion to the square of the distance between them.

The law of gravitation proposed by Newton's theory, when applied to the mass possessed by each planet and the sun, given the distances between them, predicts roughly elliptical orbits for the former with the sun *oscillating* around one of the foci. The reason for this effect is that just as the sun exerts an attractive force on a planet, the latter also exerts an attractive force on the former. When the planet involved has a mass much smaller than that of the sun, this effect is negligible. But when the planet has a considerable mass (as is the case, for example, with Jupiter) then the effect becomes detectable at times. Also, according to this law, there is an influence on the motion of one planet by the attraction exerted on it from its neighbours. Depending on the masses and the distances between them, this effect can become pronounced thereby producing perturbations in the otherwise smooth path of a given planet.

When all these effects are taken into account, a better fit between calculations and data is observed than was the case using Kepler's theory. However, from the vantage point of Newton's theory it can be seen why Kepler's theory was successful. If the masses of those planets whose orbits lie closer to the sun are negligible compared to that of the sun, then they can exercise very little effect on the sun itself. Likewise, if the more massive planets have orbits far away from the sun, their effect on the sun would be very little too. Under these circumstances the suggestion that the position of the sun remains stationary approximates the real situation very well. In addition, if the masses of the planets are very small compared to that of the sun, and the orbits are well spread out, the effect each planet would experience from its neighbours would be negligible compared to that it experiences from the sun. Given these facts, the suggestion that each planet orbits the sun in such a way that it is only under the influence of the latter, would not be far from what is the case according to Newton's theory.

As the gravitational pull which a given planet exerts on the sun and on the other planets is allowed to diminish, the perturbations in the orbits of the planets, as well as the position of the sun, decrease accordingly. For those values of the mass of the planets and distances between them that this pull is so small as to be negligible, the displacements of the sun from the foci of planetary orbits, as well as perturbations in the planets's orbits, would fall within the limits of experimental accuracy which were operative in the confirmation of Kepler's theory. Under these conditions the planets would appear to behave as they would if Kepler's theory were true, namely turning around the sun while *concurrently being pulled in and pushed out* periodically by the latter. The description of their behaviour in terms of Kepler's laws (simple elliptical orbits; equal areas in equal times; and the periods law) would thus be a good approximation to the description offered by Newton's theory. Newton's theory, therefore, entails that *as* the gravitational effect that each planet exerts on all other planets and the sun is diminished, *so* the solar system would tend to become Keplerian. Since this pull effect throughout the solar system is minute compared to the gravitational pull of the sun, Kepler's theory works as a good approximation, in the sense outlined in the previous chapter, to what Newton's theory describes to be the underlying constitution of the solar system.

We have seen that Kop can be singled out as the very first cosmological theory which not only was methodologically sound, but showed better empirical adequacy in comparison to its extant rivals as well. We have also seen that, unlike its extant rivals, assessment of it as progress towards the truth survived its replacement by a superior theory which was radically at odds with its description of the underlying constitution of the cosmos. The reason this assessment survived the transition to Kepler's theory, was seen to be the fact that the latter complied with the requirement demanded by the Correspondence Principle with respect to it rather than to its extant rivals. Similarly, the assessment of Kepler's theory as better progress in the same direction was, in turn, seen to survive its replacement with Newton's theory. The reason for the survival of this latter assessment is the same: Newton's theory satisfies the demand imposed by the Correspondence Principle on any theory that is to supersede Kepler's.

Newton's theory, in turn, is not only sound so far as our methodological requirements are concerned, but empirically more adequate than any theory which had been proposed prior to it. The Correspondence Principle, therefore, requires that any theory T_m which in the future is to replace it should, with appropriate initial conditions, entail that it is a good approximation to what T_m describes to be the case. Taking Kop as the starting point and moving up to Newton's theory, we see a succession of theories which satisfy the following conditions: (*i*) all are methodologically sound, (*ii*) each is relatively more adequate over a domain of empirical phenomena than those which come before it in the succession, and (*iii*) should the need arise, each is required to be superseded by a theory which, together with the right initial conditions, entails that it is a good approximation to the new description of the underlying physical constitution of world.

A succession of theories which satisfy all these conditions exhibits an order that may be called, for obvious reasons, the *'Correspondence Relation'*. Any theory T_i that is a member of the sequence ordered by the Correspondence Relation, stands in a definite temporal relationship to all other members of the sequence. If T_i occupies a position in the sequence which lies between one theory which it entails to be a good approximation, and another theory which entails that *it* is a good approximation, then T_i is proposed after the former and before the latter. If T_i does not have any successors, it is proposed after all other members of the sequence. That theory which precedes T_i in the sequence and which T_i entails to be a good approximation, I call a *'limiting case'* of T_i.

The Correspondence Relation has the following properties which Krajewski [1977], pp. 51-53, also points out. It is *irreflexive*, because a theory cannot be a limiting case of itself. It is, moreover, *asymmetric*, because if T_1 is a limiting case of T_2, then T_2 cannot be a limiting case of T_1. Finally, it is *transitive*. Let T_1 be a limiting case of T_2, and the latter a limiting case of T_3. Since T_3, along with the appropriate initial conditions, entails that T_2 is a good approximation over the domain of empirical facts D_2, and T_2

likewise entails that T_1 is a good approximation over the domain D_1, then provided D_1 is contained in D_2, T_3 also entails that T_1 is a good approximation over D_1.

It must be pointed out, however, that the definition offered in Krajewski [1977] for the Correspondence Relation is not only at odds with that offered here, but it is unsatisfactory as well. There, the relation (abbreviated as CR) is defined in the following terms: '... a new theory T_2 is in CR with an old one T_1 when there is a CR between basic laws of T_1 and T_2.' (Ibid, p. 6). And, 'A law L_1 is in a CR with a law L_2 if the equation $F_2(x) = 0$ of L_2 passes asymptotically into the equation $F_1(x) = 0$ of L_1 when some characteristic for L_2 parameters p_i tend to zero (p_i -> 0). If we assume that they reach the limit ($p_i = 0$), we may deduce the equation of L_1 from the equation of L_2: $F_2(x) = 0$ & $p_i(x) = 0 \Rightarrow F_1(x) = 0$.' (Ibid, p. 42).

This definition is, for most cases at any rate, satisfied only vacuously. For, presumably the new theory works better than the old *because of* the essential role that parameter p_i plays in the equations prompted by its 'basic laws'. This parameter, in turn, is assigned, by T_2, either a constant value, in which case it cannot vanishes altogether. Or, it is allowed to vary, in which case so long as the physical effect that T_1 and T_2 are attempting to account for is present, p_i cannot quite assume zero as a value (even though it may approach values that, compared to those assumed by the other parameters, become less and less significant). The equation prompted by the basic law L_2 of the theory T_2, therefore, presupposes that p_i does, under no circumstances of interest, vanish altogether. In most cases, therefore, the conjunction of this equation with the condition that the value of p_i is zero, is inconsistent. Since from inconsistent premises any proposition follows, we can deduce any equation we like from Krajewski's condition. This makes not only T_1, but any other theory at all a limiting case of T_2.

Once a theory is the only one in a branch of physics which satisfies all the methodological conditions required by the two-tier realist conception of theories, and establishes itself as empirically adequate over a domain not smaller than that of any of its predecessors, the Generalized Correspondence Principle ensures that it *should be* a limiting case of any theory that in future may supersede it. This status, in turn, would guarantee that in case the theory runs into refuting circumstances, any theory that would replace it in the future must stand in the Correspondence Relation to it. The Generalized Correspondence Principle, therefore, ensures that such a theory is guaranteed a place in the sequence ordered by the Correspondence Relation.

In the branch of physical astronomy or cosmology, the first three members of this sequence, as we have seen, displayed a smooth progression of the domains of empirical phenomena which they cover. After Newtonian mechanics, however, the great theoretical advances that were made in physics do not always preserve the same smooth extension of empirical domains. For example, Maxwell's theory of electromagnetism, though an outstanding achievement in every respect, covers a domain of empirical phenomena which has no overlap with that covered by Newton's theory. Einstein's theory of relativity extends Lorentz transformations, presupposed by Maxwell's equations, to the domain covered by Newton's theory. It thus proposes a radically new theory of gravitation. It nevertheless fails to account for such quantum phenomena as radiation from the atoms, the photo-electric effect, as well as a variety of phenomena involving the so-called 'weak' and 'strong' nuclear forces. Quantum Mechanics, on the other hand, can deal with these and many more like them adequately, and (as will become apparent in the next chapter) may, with some difficulties outstanding, be regarded as a natural extension of not only the Hamilton-Jacobi theory of Classical Mechanics, but Maxwell's electromagnetism as well. It has, nonetheless, not been able, so far, to come to terms with the phenomenon of gravitation satisfactorily.

2. The Relative Proximity of a Theory to the Truth

Glossing over these problems (any one of which may prove fatal for the overall success of the project under exploration in this thesis), the progression of theories, wherever it is ordered by the Correspondence Relation in physics, at least so far as the two-tier realist view of it is concerned, aims to discover the underlying physical constitution of the world in the appropriate domain. Since this constitution determines the truth of the matter in that domain, the place a theory occupies in this progression may be taken as an *indication* of the *advance it has made towards the truth*. Popper once thought that the idea of advancing towards the truth can be salvaged, on purely logical grounds, from the kind of transitions considered. His intuition was that the content of a theory can always be divided into classes of its true and false consequences. The better of the two theories, according to this intuition, must have (1) a greater overall content, (2) a larger class of true consequences and (3) a class of false consequences which is smaller or at least not larger. These classes, the hope was, can thus be compared, pair-wise, for size between the theories in question, and a decision about which has advanced more towards the truth can be made on the outcome.

Popper [1972] introduces these ideas as follows: '... I have introduced a logical notion of *verisimilitude* by combining two notions, both originally introduced by Tarski: (a) the notion of *truth*, and (b) the notion of the (logical) *content* of a statement; that is, the class of all statements logically entailed by it... The class of all the *true* statements which follow from a given statement... and which are not tautological can be called its *truth content*... The class of false statements entailed by a statement - the subclass of its content which consists of exactly those statements which are false - might be called... its 'falsity content'...' (Popper [1972], pp. 47-48).

After making these notions more precise he concludes: 'With the help of these ideas we can now explain more clearly what we intuitively mean by truthlikeness or *verisimilitude*. Intuitively speaking, a theory T_1 has less verisimilitude than a theory T_2 if and only if (a) their truth contents and falsity contents... are comparable, and either (b) the truth content, but not the falsity content, of T_1 is smaller than that of T_2, or else (c) the truth content of T_1 is not greater than that of T_2, but its falsity content is greater. In brief, we say that T_2 is nearer to the truth... than is T_1, if and only if more true statements follow from it, but not more false statements, or at least equally many true statements but fewer false statements.' (Ibid, p. 52)

A collection of papers by David Miller, John Harris, and Pavel Tichy which were simultaneously published in *The British Journal for the Philosophy of Science*, 25, (1974), demonstrated that conditions (a) - (c) in the above passage can be satisfied only if T_2 is true. All the outstanding theories in physics which we would like to regard as having made some advances in the direction of the truth, have either been decided to be false, or else cannot be decided to be true. It follows, therefore, that comparative assessment of verisimilitude, on *such* logical grounds, between the theories of interest in physics is untenable.

The argument can be presented as follows: Since the classes of true and false consequences of theories of interest in physics are infinite, any comparison of their sizes will have to be made in terms of the set-theoretic relation of inclusion. Couched in these terms, if T_i(T) and T_i(F) are respectively classes of the true and false consequences of T_i, (i=1,2,...), according to conditions (b) - (c), T_2 would have greater verisimilitude than T_1 just in case either (*i*) $T_1(T) \subset T_2(T)$, and $T_2(F) \subseteq T_1(F)$; or (*ii*) $T_1(T) \subseteq T_2(T)$, and $T_2(F) \subset T_1(F)$.

If (*i*) is to be the case, there will be at least one true consequence of T_2 which is not a consequence of T_1. Let us call it 't_j'. Assuming that T_2 is false, we can take any one of its false consequences, say f_j, and conjoin it with t_j. The resulting statement t_j & f_j will be a false consequence of T_2 which is not a consequence of T_1. If, on the other hand, (*ii*) were to be the case, assuming that T_1 is false, it would have at least one false consequence which is not a consequence of T_2. Let us call it 'f_k'. If now we take the disjunction $f_k \vee \neg f_j$, we would have a true consequence of T_1 which is not a consequence of T_2. It follows that neither (*i*) nor (*ii*) could hold if both T_1 and T_2 are false; hence T_1 and T_2 are incomparable as to their closeness to the truth in terms we set out to compare them.

As an alternative to the determination of the relative verisimilitude of theories by means of the comparison of their contents, it may be thought that perhaps the relative *accuracy of predictions* from them ought to be taken as the criterion. The theory which yields better predictions for the observable quantities of interest is, according to this alternative, the one which could be regarded as closer to the truth than its rivals. General Relativity, for example, may be regarded as closer to the truth than Newtonian Mechanics on the grounds that there are domains of empirical data over which predictions from the two theories are in conflict, and those from the former are invariably more accurate than those from the latter.

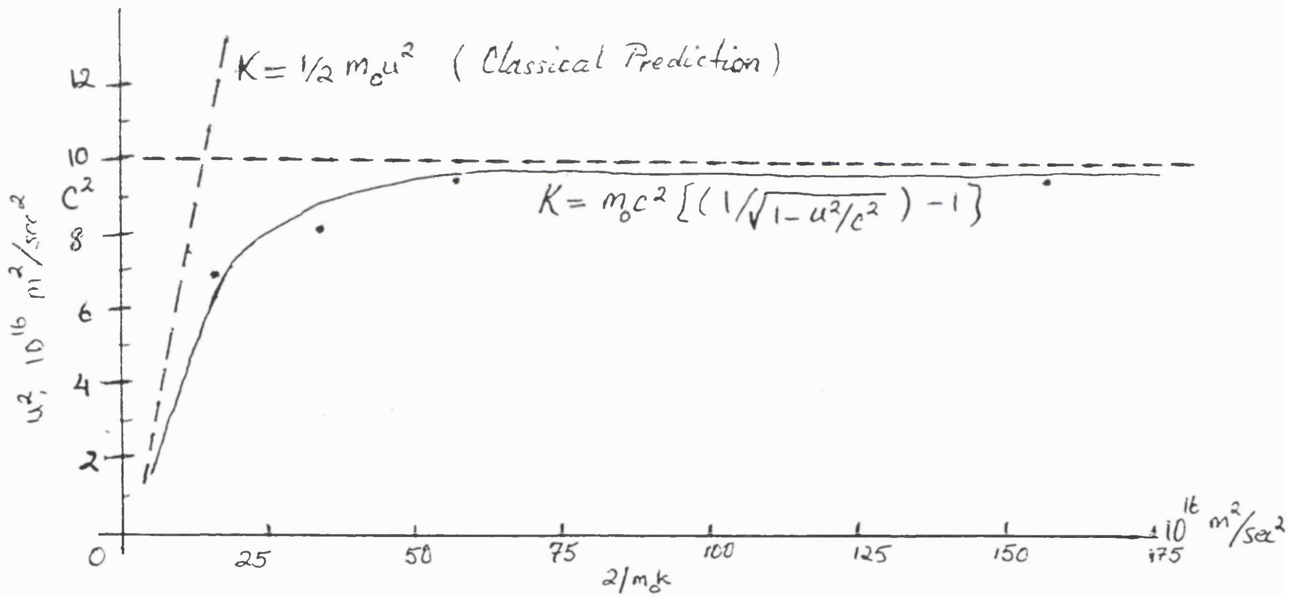
Miller [1975] produces an argument whose stated target is to demolish this alternative as well. Despite formal complexities present in the original, the gist of Miller's argument has been presented, in rough but very simple terms, in Oddie [1986], (pp. 156 - 157). Let us take theories under comparison to be T_1 and T_2 again. Let us assume that predictions from T_2 for every measured value of a standard quantity **Q** are, on the whole, more accurate than those obtained from T_1 (but not always the same as the measured ones). The main move involves the introduction of a transformation by means of which a non-standard quantity is defined such that (*a*) it appears that the new quantity is indistinguishable, from the logical point of view, from **Q**, and (*b*) the accuracy of predictions for the latter from the two theories, is exactly reversed. The aim is to show that if T_2 is claimed to be closer to the truth on the grounds that its predictions for **Q** are more accurate than those from T_1, the latter can equally be claimed to be closer to the truth than T_2 on the grounds that *it* predicts more accurate values for the contrived quantity than T_2.

This can be done as follows: Let the true values of **Q** be determined in terms of the variable X by the function F(X). Let $F_1(X)$ and $F_2(X)$ be the functions by means of which T_1 and T_2 predict values for **Q** respectively. We assume that $F_2(X)$, on the whole, predicts values for **Q** that are better approximations to those determined by F(X) than values predicted by $F_1(X)$. We can always define a transformation *f* which for every X takes the value determined by $F_1(X)$ and replaces it with that determined by $F_2(X)$, and vice versa; i.e.

$$f(X,\mathbf{Q}(X)) = \begin{cases} (X,\ F_1(X)) & \text{if } \mathbf{Q}(X) = F_2(X), \\ (X,\ F_2(X)) & \text{if } \mathbf{Q}(X) = F_1(X), \\ (X,\ \mathbf{Q}(X) & \text{otherwise.} \end{cases}$$

Under this transformation a new quantity **K** emerges whose values are determined by a new variable X', i.e. (X', **K**(X')) = *f*(X, **Q**(X)). The true values of **K** are determined by **K**(X') = F(X'), and theories T_1 and T_2 predict values for **K** according to **K**(x') = $F_2(X')$ and **K**(X') = $F_1(X')$ respectively. The truth in terms of **K** is the same as the truth in terms of **Q**; and just as **K** is defined in terms of **Q** by means of the function *f*, the latter is definable in terms of the former by exactly the same function. So, the argument goes, there is no reason, from the logical point of view, to regard one as somehow meriting a privileged status vis à vis the other. Rather, just as T_2 'outsmarts' (to use a phrase from Miller [1975], p. 170, T_1 with respect to **Q**, T_1 outsmarts T_2, with respect to **K** in exactly the same way. It is finally concluded that as long as none of the theories being compared get all their predictions for a standard quantity right (which is typically relatively better results cannot justifiably be taken as an indication of its closer proximity to the truth.

That this result, if correct, goes far beyond the issue of verisimilitude is effectively highlighted in Oddie [1986]. '... Miller-argument does not just hit truthlikeness - it will also hit confirmation and refutation. For suppose that we have two curves expressed by two theories, and a collection of data $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$. Typically the data will not fit either curve *exactly*, but will fit one curve better than the other.



'Consider, for example,... a graph of the classical and relativistic predictions for K, kinetic energy, against velocity (u) found in standard physics textbooks... The data is taken to confirm the relativistic prediction... But we now have a recipe for reversing this by means of a transformation. We have:

Classical relation: $K = 1/2\ m_o u^2$,

Relativistic relation: $K = m_o c^2 [(1/(1 - u^2/c^2)^{½}) - 1]$.

The appropriate transformation is the following:

$$t(u, v) = \begin{cases} (u, 1/2\, m_o u^2) & \text{if } v = m_o c^2 [(1/(1 - u^2/c^2)^{½}) - 1] \\ (u, m_o c^2 [(1/(1 - u^2/c^2)^{½}) - 1] & \text{if } v = 1/2\, m_o u^2 \\ (u, v) & \text{otherwise.} \end{cases}$$

'This transformation exactly reverses the relation of the classical and relativistic theories of kinetic energy to the data... If Miller is correct then...[t]he data confirms the relativistic formula no better than it confirms the classical formula.' (Ibid, pp. 157-158)

Indeed, for this very reason, Miller's argument, as hinted in Oddie [1986], p. 158, may be taken as a *reductio* against the premise that the quantities in question *are* logically indistinguishable from one another. To shed more light on this point, let me offer the following variation on the same theme: Suppose there are two points A and B which are exactly 10 meters apart on a flat stretch of land. We are given two straight sticks S_1 and S_2 of fixed lengths, to use as whole units in terms of which this distance should best be divided into segments of roughly equal lengths. S_1 is about 1.249 meters long and S_2, about 1.66 meters long. We are then asked to judge which stick is relatively more suitable to do the job. Using S_1 eight times we can divide up the distance nicely except for a residue of less than 1 centimetre after the final segment. On the other hand, the best we can do with S_2 is to use it six times and end up with a left over of about 4 centimetres.

If we represent the distance in question by D, the truth of the matter is D = 10 meters. A rough division of this distance in terms of S_1 would produce the estimate $D = 8S_1$, which evidently fits the bill better than the rival estimate $D = 6S_2$. So if S_1 and S_2 were all there was to divide the distance roughly into segments of equal lengths, S_1 would be more suitable for the job. Now a Miller type conversion can be produced for this case with exactly the same results as in the theoretical cases we have considered. It is *possible* to define a new quantity C by the following transformation *f* on D:

$$C = f(D) = \begin{cases} 6S_2 & \text{if } D = 8S_1, \\ 8S_1 & \text{if } D = 6S_2, \\ D & \text{otherwise.} \end{cases}$$

Dividing up C six times with S_2 leaves a residue of less than one centimetre, while doing the same eight times with S_1 leaves a residue of about four centimetres. C has exactly the same dimension as D; the truth of the matter is the same for both D and C; and D can be defined in terms of C by means of the same transformation *f*. However, to divide up C according to the available scales S_1 and S_2 forces a complete reversal of the previous judgement. S_2 is now seen to be more accurate than its rival, and, therefore, a better choice for dividing up a quantity that appears to be logically indistinguishable from D.

This argument, naturally, fails to cast any doubt on the wisdom of choosing S_1 as a better alternative for dividing up the distance between A and B. One suspects that it is for very much the same reasons that such contrived predicates as Goodman's *grue* and *bleen*, as well as Miller's *Minnesotan* and *Arizonan*, have failed to make their way into talks about diamonds and the weather. That there is a connection between these predicates and the concocted quantities we have been considering is hinted in Miller [1975], p. 183, and discussed in Oddie [1986], chapter six, in considerable detail. The point is not that logic alone can not be expected to decide which quantities we chose to measure, or which predicates are more suitable for describing the world. Nobody in their right mind would turn to logic for making this decision. The fact remains, however, that logically indistinguishable quantities or predicates should be interchangeable, every thing else remaining the same.

Oddie [1986] chooses to question, with apparent success, the legitimacy of claims that such contrived constructions are in fact logically indistinguishable from predicates and quantities that enjoy widespread circulation. His motivation for choosing this line of attack, as stated in Oddie [1986], p. 158, is the observation that arguments based on these concoctions do not *just* target isolated notions such as 'accuracy' and 'truthlikeness'. They go much further and threaten to render a whole array of notions such as 'structure, change, sameness of state, confirmation and disconfirmation' as 'spurious'.

Arguments and counter-arguments such as these have prompted some authors of realist persuasion to steer clear of the notion of verisimilitude in connection with the problem of rational theory choice in science altogether. A paradigm case in point is presented in Watkins [1984]. The project pursued here is to articulate a set of methodological desiderata in terms of which an '*optimum aim for science*' may be defined. A theory which best satisfies these against its rivals at any one time, can rationally be chosen on the grounds that it best fulfils this aim and thus is the best theory available. It is clear that from the view-point of the two-tier realism this prescription leaves something to be desired. A two-tier realist in his choice of theories is not motivated by abiding by rules which do not clearly link up to *the truth*. He wants to know which theory offers a description of the underlying constitution of the world that is closest to the truth than any of its rivals at any one time. Better accuracy of predictions, or the superiority of the size of a theory's content *on their own (or even jointly together)*, may not count as sufficient reasons for the judgement that a theory has made more advances in the direction of the truth than its rivals. As was argued in the previous chapter, the two-tier realist supposition, prompted by the successes achieved by a methodologically sound theory at the empirical level, that it has latched on to a truth-related property, needs to be strengthened by the additional constraint imposed by the Generalized Correspondence Principle.

3. The Ultimate Theory

In attempting to circumvent the problems associated with the notion of advancement towards the truth or verisimilitude while fulfilling the aspirations of the two-tier realist, two points beg immediate clarification. The first is *what* exactly should be taken to represent the truth, and the second is how the proximity of a theory to it is to be *indicated*. Since the Correspondence Relation is an intertheory relation, if it is to have a bearing on the second question, it would be natural to expect that what represents the truth should either be a single true theory, or a package of several true theories. If there exists a theory (or a package of theories), let us call it '***T***', that mirrors the underlying physical constitution of the world (either in some or all the branches of physics), then whatever it says about the world would be true. In that case, ***T*** may be identified as the *ultimate goal* towards the discovery of which physics aims. In this capacity, it could then serve as the upper bound in the sequence of theories ordered by the Correspondence Relation. However, ***T*** must, in line with what we took for granted at the outset, satisfy all the required methodological conditions. These conditions, in turn, push successive theories in the direction of having greater explanatory, as well as predictive, fertility while utilizing fewer number of independent axioms. ***T*** (or each single theory in it if it is a package of true theories), therefore, would have to possess these attributes more than any other theory in physics.

A question immediately arises: In assuming that ***T*** exists, do we not commit ourselves to more metaphysics than we can possibly justify? The assumption that there are a few axioms and laws which capture all the complexities of the physical world down to their most intricate details, imputes to the world a specific kind of underlying structure. If our experience with the kinds of theories that have been proposed up to now in physics is anything to go by, indications are that this structure would involve something like a few basic entities and a handful of basic relations between them. But we do not *know* that the world is like that; and to extrapolate, *on the basis of the successes that more and more unified theories have achieved so far*, that it ought to be like that, is to commit an unjustifiable inductive leap.

Watkins [1984] suggests that the assumption of the existence of theories such as ***T***, amounts to taking a side in a metaphysical dispute, and for this reason, should not be countenanced by the aim for science. 'The issue is whether or not there is, in the constitution of the physical world, a bottom layer consisting of irreducible entities (simples, elements, atoms, or whatever) such that: (*i*) science could, in principle, explain the nature of all things at all layers above this layer in terms of the intrinsic properties of, and mutual relations between, such ultimate entities; (*ii*) science could not explain their nature in terms of the properties of further entities because there are no further entities. If there is such a layer, then there is the possibility of ultimate explanations; otherwise not'. (Ibid, p. 131).

The issue, Watkins [1984] goes on, is the subject of an undecidable metaphysical dispute. As an example of an extant metaphysical position Watkins [1984], p. 131, cites Leibniz's theory of the infinity of structures in the world. According to this theory, 'If we could penetrate ever deeper into the microstructure of things, we would always discover worlds within worlds, each as richly complex as the last.' (Ibid). 'So', he concludes, if the aim for science '...is to be impartial between this metaphysical thesis and the thesis of classical atomism that there are ultimate bits, we should discard the ideal of ultimate explanations and retain that of *ever deeper explanations*.' (Ibid).

Several things may be said about this reservation. First, if physics is to be regarded as a *truth-directed* inquiry (and being firmly grounded in the realist position, this is granted in Watkins [1984]: 'Science aspires after truth' (Ibid, p. 155), and again, '... what science aspires after is *truth*' (Ibid, p. 280), then the idea that it should aim at 'ever deeper explanations' presupposes *some kind of structure* in the world (without picking any particular one). Surely, the latter is a position very much in dispute amongst metaphysicians. There are extant metaphysical theories (any brand of idealism would do) that deny any structure at all in the world.

If, however, it is unobjectionable to take *this much* metaphysics on board, why should it be objectionable to add a bit more? I realize that it sometimes is more prudent to be as economical as possible. But then again, how much it is wise to carry depends very much on what one can get in return. If there are no standard 'weight limits' on how much metaphysics one is allowed to carry, then if one can solve a few problems by taking one more bit of 'baggage' along, one may justifiably be excused for not being able to produce justifications for it. This strategy has paid well in science itself, and there is no reason why it should not be put to good use in its methodology as well.

Second, aiming to get greater fertility from fewer axioms is not a dictate which has been imposed on physics from the outside. Successive generations of physicists have produced theories that, on the whole, appear to have manifested this property to greater and greater extents. So, for someone who is interested in describing developments in theoretical physics, the suggestion that it progresses in the direction of more unified theories is, by far, the best account of the facts. The question is, whether it is better to view this progression as open-ended or as converging to something like the above T?

Frankly, I cannot see any particular vice, *per se*, in viewing this progression as having an end; nor any particular virtue inherent in the belief that it is open-ended. Popper [1972] suggests that there is a vice associated with the conjecture that the former is the case. This vice appears to be a violation of the aim of science, and is alleged to commit us to the existence of essences: 'If it is the aim of science to explain, then it will also be its aim to explain what so far has been accepted as an *explican;* for example, a law of nature. Thus the task of science constantly renews itself.' (Ibid, p. 194). But surely, one can retort, if it so happens that a theory such as ***T*** exists - and we have seen no reason, so far, to rule out this possibility - then if physics ever attained it, there would be no more explanation to be had. It may be deemed desirable to seek *metaphysical* explanations for the laws proposed by ***T***. But physics has, in the shape of ***T*** all the explanations it has been aiming for. The mere fact, then, that physics aims for explanations does not, by itself, establish that it is not possible for an ultimate explanation of *the physical constitution of the world* to exist. We may never be able to *decide*, even when we happen to stumble on ***T***, that we have at last reached the end of the line in physics. But what reason is there for believing that *there is no theory such as* **T** for us to stumble upon?

Popper [1972], p. 195, answers this question as follows: 'There can be no explanation which is not in need of a further explanation, for none can be a self-explanatory description of an essence...' Granting, with Popper, the obscurantism of the doctrine of essences, I do not see why ***T*** should contain anything even remotely resembling essences. Like most extant outstanding theories in physics, ***T*** presumably contains a few principles which together stipulate certain laws of nature. *T* differs from the rest, however, only in that its proposed laws are the most comprehensive there is, and they happen to coincide with the actual laws that govern the physical world.

So far as *laws of nature* are concerned, they may be viewed in the very sense that Popper takes them: 'Laws of nature are conceived... as (conjectural) descriptions of the structural properties of nature...' (Ibid, p. 196). The laws that *T* presumably proposes can easily fit this description. If they are ever proposed, they could not possibly be proposed as anything but conjectures; and they would describe the structural, as opposed to the 'essential', properties of the physical world just as our extant theories in physics attempt to do. Popper [1972], p. 195) regards as 'essential' those properties which are 'inherent *in each individual or singular thing*', and 'which may be appealed to as the explanation of this thing's behaviour.' The laws that *T* presumably puts forward are no more descriptions of 'essential' properties in this sense, than the laws proposed by other extant outstanding theories in physics aspire to be.

Popper's case for denying the existence of *T*, in the end, becomes rather confusing, to say the least. He starts by appealing to the role of explanations in the aim of science. We saw, however, that just because science aims at explanations, it does not follow that there could not be an ultimate explanation which itself can have no *scientific explanation*. Then the argument shifts to essentialism. The belief in the existence of an ultimate explanation in physics is presented as amounting to the belief in the existence of essences. We saw that nothing of the sort is entailed by the belief in the existence of a theory such as ***T***. So, the pitfalls that Popper [1972] wished to warn us against in connection with the belief in the existence of ***T***, do not appear to be there to threaten us.

The fact of the matter is that the issue concerning the existence or otherwise of ***T*** is not only undecidable, but, at the moment at least, there is no a priori consideration that would tilt the balance either way. Therefore, if we opt for the conjecture that ***T*** exists, we would be no more in danger of committing a sin against intellectual propriety than if we had sided with its rival. True, the first conjecture is stronger than the second. But as long as this extra strength can be employed to do some good work, our indulgence in adopting it would be well rewarded, and therefore, excused.

4. The Truth-likeness Sequence

With ***T*** in place, we proceed to deal with the second question, viz. how is the relative proximity of a theory to the truth to be *indicated*? As a first step towards answering this question I observe that the Correspondence Relation is transitive. Suppose theories T_1, T_2, and T_3 are ordered by the Correspondence Relation such that T_1 is a limiting case of T_2, and the latter, in turn, that of T_3. Suppose, further, that D_1, D_2, and D_3 are domains of empirical facts over which T_1,

T_2, and T_3 are, respectively adequate. If D_1 is contained in D_2, and D_2 in D_3, then by safeguarding the successes of T_2 over D_2, T_3 also safeguards the successes of T_1 over D_1.

Next, I claim that the condition for the membership of a theory in the sequence of theories ordered by the Correspondence Relation, is that it should be a limiting case of any future theory that may supersede it. I shall attempt to specify this condition. It will then be observed that, given a finite period of time, the question whether a theory satisfies this condition is decidable.

With the help of two assumptions, these observations will be employed to realize the two-tier realist ambition of grounding the choice of a theory in physics on its relative proximity to the truth. One assumption is that in whichever branch of physics where the Correspondence Relation is defined, it is bounded on both ends; and the other is that its upper bound is ***T*** (or in ***T*** if it is a package of theories). The first assumption is natural to make. The history of physics in any of its many branches, starts somewhere, and the very first theory in a given field which satisfied the necessary methodological conditions, and managed to achieve an acceptable degree of empirical adequacy over a certain domain of empirical phenomena, qualifies as the lower bound of the Relation in that field. In physical astronomy, for instance, this theory may be identified as the physical version of the Copernican theory which was worked out by Kepler by the end of 1604.

The proposition that $\boldsymbol{T}$ is the upper bound of the Correspondence Relation follows from the assumption of its existence and the indispensability of the Generalized Correspondence Principle. If $\boldsymbol{T}$ exists, it would be both methodologically and empirically adequate. The Generalized Correspondence Principle would, therefore, require that if it were to be superseded, it should be a good approximation to the theory that replaces it. Thus, $\boldsymbol{T}$ would belong to the sequence of theories ordered by the Correspondence Relation. Since there are no theories superseding $\boldsymbol{T}$, it follows that it would be the upper bound of the Relation where it is defined.

With $\boldsymbol{T}$ in place as the upper bound of the Correspondence Relation, I have all the pieces I need to put together my argument. Given that this relation is transitive, it follows that every member of the sequence S of theories ordered by the Relation, is a limiting case of $\boldsymbol{T}$. Since $\boldsymbol{T}$ captures the entire truth that is of interest to physics, every member of S, compared with those theories that fail membership in it, could be considered as having gone some ways towards capturing this truth. Moreover, if T_i is a member of S, then compared to all theories that are its limiting cases, T_i could be considered as having gone the farthest in this direction. If, therefore, it has been decided that a theory qualifies for membership in S, and there is no theory available that has it as a limiting case, this theory can be chosen as one that has gone the farthest in the direction of capturing the truth.

Granted these assumptions, the succession of theories wherever ordered by the Correspondence Relation would converge towards ***T***. Because ***T*** is the best theory (or package of theories) in physics, it must be methodologically sound as well as empirically adequate. Since, moreover, the requirement demanded by the Correspondence Principle is globally valid in physics, ***T*** is also obliged to satisfy it. Therefore, ***T*** is a member of the sequence ordered by the Correspondence Relation. Since there is no other alternative which would surpass ***T*** in capturing the underlying physical constitution of the world, ***T*** must be the last member of this succession. Hence, the succession ordered by the Correspondence Relation converges to ***T***.

Because the Correspondence Relation is transitive, every member of the succession ordered by it would be a limiting case of ***T***. Moreover, because ***T*** is true, each member of this succession can be regarded as a step closer to the truth than the theory which immediately precedes it in the succession. For these reasons, this succession of theories may be called the *truth-likeness sequence*. A member of the truth-likeness sequence may be said to resemble the truth to a greater extent than its predecessors on the grounds that it goes further towards ***T*** than any of them.

5. The Problem of Rational Theory Choice

With the notion of the truth-likeness sequence, we appear to be well poised to solve the problem of rational theory choice within the context originally envisaged by Popper. In this context what was needed, but never attained by Popper, was a *clear* and *objective* account that would satisfactorily capture the two-tier realist intuition about hard won successes of a theory. According to this intuition, spectacular successes achieved by a theory plus methodological soundness, add up to indicate that it has made some progress towards the truth. The constraint that this account has to satisfy in order to entail that this intuition is rational, is that it should rely, either tacitly or explicitly, on no inference which is in violation of the requirements of validity in ordinary logic.

Although membership in the truth-likeness sequence is performance-related, it requires no invalid inference for its justification. Whether a theory is methodologically sound is a question which can be effectively decided by carrying out a specific number of checks. Empirical adequacy, in the sense required for a theory *to qualify as a limiting case of any future theory that may supersede it*, too, is an effectively decidable characteristic. A theory in physics is customarily deemed empirically adequate *in this sense*, if a sufficient number of measurements, from a sufficient variety of experiments, bear out its predictions in the following manner.

Let $\mathbf{Q}$ be an observable quantity, measurements on which under conditions $\mathbf{C}_i$, within a certain period up to the time t, have been observed to fall within a finite set of values whose average we denote by $\mathbf{Q}_i$. Let T_i be a theory whose prediction for $\mathbf{Q}$ with $\mathbf{C}_i$ as initial conditions, is $\mathbf{q}_i$. There is always an interval of tolerance which we may designate by a positive number ϵ (and which includes experimental errors), such that in order for the predictions of T_i to be acceptable at all, $|\mathbf{q}_i - \mathbf{Q}_i| \leq \epsilon$. Let us call the quantity $|\mathbf{q}_i - \mathbf{Q}_i|$, the *deviation from the mean* of the value $\mathbf{q}_i$.

If at time t it has just been observed that all established measurements of all the observable quantities for which T_i has predictions, yield deviations from the mean which are better than those obtained for any other *available* theory, T_i may be considered promising, so far as empirical adequacy is concerned. If a promising theory sustains its run of success for a variety of additional measurements over a sufficiently long (but always finite) period of time, it will, as a rule, be established as empirically adequate for the purpose of being a limiting case of any theory that in the future may supersede it. Establishment of the empirical adequacy of a theory, in this sense, does not appear to require a logically inadmissible inference. Consequently, the membership of a theory in the truth-likeness sequence can be decided in a finite number of steps, without violating any rules for logical validity.

By the time the membership of an unfalsified theory T_i in the truth-likeness sequence is confirmed, that is, once it becomes incumbent on any future theory that may supersede T_i to entail that it is a good approximation over a certain domain, T_i would be established as a step in the right direction towards the truth. In so far as T_i occupies a position in this sequence which is nearer than all other theories already in it to $\boldsymbol{T}$, it is a theory which among all the available ones is the closest to the truth. The proximity of T_i to the truth in this sense can be decided on the basis of the number of theories which are its limiting cases. At any given time, the theory with the greatest number of limiting cases is the closest theory to the truth. Choosing it as the best available theory would, therefore, be a rational choice based on its proximity to the truth.

It is possible, however, that at a certain time t the decision about the proximity of T_i cannot be made even though it does satisfy all the requirements we have been demanding. Suppose that concurrently with T_i there is also a theory T'_i on the scene which is not only methodologically as sound as T_i, but also equally adequate empirically. Suppose that neither T_i nor T'_i have been falsified up to t, and that both have the last falsified member of the truth-likeness sequence as their limiting case. If T'_i seriously clashes with T_i on its description of the underlying physical constitution of the world, then we would be unable to decide at t, which is a closer theory to the truth.

Here, again, the Generalized Correspondence Principle comes to our aid. Even though both T_i and T'_i qualify as limiting cases of the theory T_{i+1} which in the future may supersede them both, only one of them can be entailed by T_{i+1} as a good approximation. Whichever, therefore, is singled out as its limiting case by T_{i+1} when the latter is discovered, *it* would be found to be a member of the truth-likeness sequence. Assuming that a finite period of time is required for the discovery of T_{i+1}, it follows that the question of the proximity of T_i or T'_i to the truth can be decided in a finite period of time. Note that the same cannot be said of the methodologies which recommend choice solely on the basis of the empirical performance of outstanding theories. As long as the theories remain unfalsified, the question of which one is better is undecided. But in case they are falsified simultaneously, while remaining equally empirically adequate, this question must remain undecided for ever.

6. Convergent Realism and Verisimilitude

The account that emerges from these considerations falls under a family of positions which goes under the generic name 'convergent realism'. It must, however, be distinguished from a version first proposed in Putnam [1978], and subsequently attacked in Laudan [1981]. This version is a variation on the traditional theme of cumulative success in terms of the preservation of some ingredients of a superseded theory in the superseding one. Only, what here is preserved is neither 'explanatory structure' nor mathematical equations, but the reference of terms. The superseding theory is claimed to preserve the reference of the terms employed by the successful but false theory it replaces, thereby rendering the latter as 'approximately true'. This idea is then claimed to be justified because it provides the best explanation for how scientists behave and why science is successful.

'To begin with', we read in Putnam [1978], pp. 20-21, 'let me say that I think there *is* something to the idea of *convergence* in scientific knowledge. *What* there is is best explained... by means of two principles: (1) Terms in a mature science typically *refer*. (2) The laws of a theory belonging to a mature science are typically approximately true... [S]cientists act as they do because they *believe* (1) and (2), and their strategy works because (1) and (2) are *true*. One of the most interesting things about this argument is that, if it is correct, the notions of 'truth' and 'reference' have a causal-explanatory role in epistemology. (1) and (2) are premises in an *explanation* of the behaviour of scientists and the success of science - and they essentially contain concepts from referential semantics.'

The reason that this is not a cogent argument is twofold: On the one hand the strategy that is ascribed to scientists could work even if (1) and (2) were false. On the other, we would like to be able to regard a theory as a limiting case of another even if from the vantage-point of the latter some of the former's terms are found not to refer at all. That Kepler's theory of planetary motion is a limiting case of the Newtonian theory of gravitation may be taken as uncontroversial. However, the dynamical behaviour of the planets in Newton's theory is accounted for by a mutual gravitational attraction acting between them and the sun, thereby scrapping Kepler's *vis* (or *anima*) *motrix* as a mechanical attribute of the sun alone.

I am not at all certain how strongly scientists feel about preserving the reference of terms deployed in outstanding, but falsified, theories. I do not even know how one may proceed to ascertain either the strength of this feeling or the claim that it is shared by all scientists throughout the ages. For my part, where I do have to bring in scientists' attitudes, it has to do with a respect for consistency. Since everywhere else in the practice of science staying within the bounds of consistency is of paramount importance, it is not too much to require that in the particular case of assessing the successes of one and the same theory which has established itself as an outstanding contribution in a field, those of a two-tier realist persuasion will want to stay within the same bounds as well.

The resources which I have employed to draw up my account of convergent realism are either drawn from the practice of physics or else are consistent with it. The Generalized Correspondence Principle, which is my main resource, has an indispensable heuristic role to play in the development of new theories. Another resource, the assumption that $\boldsymbol{T}$ exists, though perhaps not shared by all physicists or methodologists, has no cogent arguments against it and is consistent with the practice of physics. With the aid of these resources, I have drawn up a procedure for gauging the comparative proximity of a theory to the truth. This procedure, I hope, will live up to the aspirations of a two-tier realist who turns to physics in order to satisfy his curiosity about what the world is like.

The account of the progression of successful theories towards the truth which emerges in this thesis is not generally conservative. No part, either in the ontological claims or the mathematical equations, of the superseded theory is required to be preserved in the superseding theory. It may be felt that this is an unacceptable feature of my account. For in the conservative accounts, two inconsistent theories may be regarded as approximations to the truth by virtue of sharing some parts with the ultimate truth. Successful theories, in these accounts, are successful in virtue of *having discovered something about the world* which will naturally be incorporated in any future theory that may supersede them.

I find that this conception of progress towards the truth, though attractive, is not supported by all cases in which, in particular, the Correspondence Principle applies. What emerges from my study of Bohr's deployment of this principle in his [1913] is that it applies between theories which have nothing in common save the ability to ascribe and forecast similar behaviour for observed physical systems in a restricted domain. That this ability in a superseded theory is indicative that it has proceeded a step in the right direction towards the truth, is grounded, not in the fact that some bits of it will be preserved by any future theory that may supersede it, but by the fact that *it thus qualifies as a member of the Correspondence Relation.*

In my account, attempts at the description of the physical constitution of the world fall basically into two groups. One containing theories which proceed in the general direction towards the truth (even if they are false), and the other, those that lie off this course altogether. The property of proceeding in the general direction of the truth is possessed by a theory which qualifies as a member of the truth-likeness sequence. This sequence may not hold as a linear ordering in the whole of physics, one theory following another smoothly towards ***T***. There may be branches of physics that despite the hope and endeavours of some physicists may never be brought in line with others. But so long as all branches ultimately lead to ***T*** (in the words of Post [1971], so long as there are no 'blind alleys' culminating in a methodologically sound, empirically adequate and *false* theory), my purposes are satisfied. Moreover, the proximity of a given member of this sequence to the truth, relative to all the extant members in its branch, can be specified on the basis of the position it occupies relative to them. The greater the number of theories that are limiting cases of a chosen theory, the closer it is to the truth compared to the rest. Finally, at any given time, the choice of a theory which qualifies as a member of this sequence, and has the greatest number of theories in its branch as its limiting cases, not only best satisfies the aspirations of a two-tier realist, but is rational in the Popperian sense of not involving any inductive leaps as well.

CHAPTER 5

IS QUANTUM MECHANICS INCONSISTENT WITH THE DEMAND OF THE GENERALIZED CORRESPONDENCE PRINCIPLE?

1. Introduction

The question whether Classical Mechanics can, in some sense, be conceived as a 'limiting case' of Quantum Mechanics has been a subject of a long controversy among interested philosophers. Opinions on the issue range from denying that it can, through proposing a piecemeal 'correspondence' between the two, to insisting that the equations of one theory are reducible, wholesale, to those of the other. Representatives of the 'denial school' have been discussed when we considered the arguments of Hanson and Feyerabend (Post [1971] is also a member of this school, but for different reasons).

At the opposite end of the spectrum, a concise statement of the wholesale reductionist position is offered in Yoshida [1977]: 'In the standard discussion of the classical limit of quantum mechanics one considers a particle in a potential $V(\vec{r})$ with the modulus and phase in its wave function separated, i.e., $\psi(\vec{r}) = A(\vec{r}) \exp (i/\hbar) S(\vec{r})$; one substitutes the wave function in the Schrödinger equation; then by separating the real and imaginary parts one obtains two equations:

$$(\partial S/\partial t) + (\nabla S^2/2m) + V = (\hbar^2/2m)(\Delta A/A), \qquad (1)$$

and $$m(\partial A/\partial t) + (\nabla A \cdot \nabla S) + (A/2)\,\Delta S = 0\ ;$$

then by setting $\hbar = 0$, the right side of equation (1) vanishes and the result is taken to be the Hamilton-Jacobi equation.' (Ibid, p. 51).

Apart from general considerations, already discussed, that speak against this reductionist view, there are serious technical problems with this particular proposal. One, highlighted in Redhead [1993], (p. 332), is that for the value zero of $\hbar$, treated as a variable, the state function in question would turn into a singularity. Thus whatever is gained by the imposition of this limiting condition on the Schrödinger equation, is obliterated by losing the state description altogether. The second problem is that there is a clear sense in which Maxwell's theory of electromagnetism can be conceived as a limiting case of Quantum Mechanics. The principal equations of Maxwell's theory (the celebrated Maxwell equations), however, contrary to the Hamilton-Jacobi equations, are linear. If a simple substitution of $\hbar = 0$ into the Schrödinger equation turns it into the non-linear Hamilton-Jacobi equation, it surely cannot deliver Maxwell's equations as well. Yet Schrödinger's equation is the equation of motion governing the behaviour of photons as well as other sub-atomic particles.

Lastly, the idea of 'piecemeal correspondence' is premised on a belief in the existence of cumulative progress in physics on the one hand, and the denial of the tenability of wholesale reduction, in general, on the other. Representatives of this school of thought may be found in Radder [1991] and Redhead [1993]. According to them, for cumulative progress to obtain in physics it is enough that some of the equations (in the case of Radder, for instance), or solutions to equations (in the case of Redhead), of a superseded theory, under appropriate limiting conditions, become identical to those of the superseded theory. As I have argued earlier, there appear to be serious problems in upholding this position as being *generally* the case; there are some important instances where one wants to speak of cumulative progress where neither equations nor their solutions remain (or become) identical. If, however, in an intertheory transition some equations or their solutions do remain (or become) identical, we may have a special case of the approximational conception of Correspondence, provided the other conditions for one theory being a good approximation to another (as spelled out in chapter 3) are also satisfied.

The question whether Quantum Mechanics stands in the Correspondence Relation to Classical Mechanics presents additional problems. If, for the purposes of this thesis, we take Schrödinger's mathematical formalism in its non-relativistic form, then, unlike other extant cases in physics, there are not one but several competing interpretations for this formalism (as well as the 'Orthodox' interpretation, one may mention de Broglie's 'Pilot-Wave', Bohm's 'Ontological', DeWitt and Graham's 'Many-World', Nicholas Maxwell's 'Propensiton', 'GRW' and 'Modal' interpretations). The idea of Correspondence Relation put forward in this thesis, rests on the approximational conception of the Correspondence Principle. This conception, in turn, is premised on viewing theories in physics as whole packages containing mathematical equations together with their interpretations. To proceed, therefore, with a discussion of the above question it is necessary that an interpretation is adopted for the non-relativistic Schrödinger equation. To justify the choice of one interpretation from the class of extant ones, however, requires a thoroughgoing critical examination of all of them which is clearly beyond the scope of this chapter. My main purpose in this chapter is to dispel claims, such as made in Hanson [1961], to the effect that (*i*) Schrödinger's mathematical formalism *requires* for its interpretation a language which is untranslatable to that of Classical Mechanics, and *as a consequence*, (*ii*) the uncertainty principle (an indispensable part of Quantum Mechanics) is incompatible with the Generalized Correspondence Principle (which is required for the conception of theory choice being conjectured in this thesis).

Claims that have been made on the basis of (*i*) and (*ii*) stem from a reading of the Quantum formalism which is commonly referred to as the 'orthodox' interpretation. According to this reading, *it is necessary* to postulate, at the sub-atomic level, entities which must be considered, in some sense, *both* as particles and as waves. The credibility of this reading will be examined in this chapter in some detail and it will be concluded that far from facilitating a comprehension of the formalism *it* stretches intelligibility to its limit. Upon further scrutiny, I shall subsequently argue, the orthodox interpretation is found to suffer from more serious defects than this. These will be identified as having to extend an unjustifiable invitation to psychology *and* epistemology to intervene for the task of providing a *physical* interpretation of Schrödinger's mathematical formalism.

These conclusions, however, have only a negative character; they would suggest that such claims as launched in Hanson [1961] are based on an unsatisfactory interpretation of Schrödinger's formalism. The question whether a viable interpretation of this formalism is possible which does not suffer from these defects and which can lend itself to the demands of the Generalized Correspondence Principle vis à vis Classical Mechanics, still remains unanswered. To settle this question, in turn, it suffices to find an appropriate interpretation that appears promising in delivering the desired goods. This, however, has proved to be no easy task. It may be the case that there is no interpretation of Quantum Mechanics which has Classical Mechanics as its 'limiting case', even in the weak sense specified in chapter 3. If this proves to be the case, and if Quantum Mechanics remains a deeply entrenched theory in physics, then the idea being explored in this thesis, namely that the Correspondence Principle should govern theory choice, would have to be restricted to those branches in physics where a truth-like sequence can be defined.

In this chapter, I will outline an interpretation which, *prima facie*, appears to me to be promising in rendering Classical Mechanics as a limiting case of Quantum Mechanics (in the sense of chapter 3). The interpretation in question is based on regarding the wave function in Schrödinger's equation as representing a *real physical entity*. The first attempt to view the wave function in this light is due to de Broglie. He, however, felt obliged to abandon it for a while due to attacks from Bohr and his followers. The idea was revived in a somewhat different form in the early 1950's by Bohm and has been gaining increasing strength and popularity in recent years. Regarding the wave function as a real physical entity, makes the interpretation to be outlined in this chapter particularly akin to the spirit of two-tier realism which lies at the heart of the conception of theory choice being conjectured in this thesis. I will also attempt to show in this chapter that the interpretation to be proposed has the advantage of not suffering from the two major defects specified above which beset the Orthodox interpretation. To set the background to the discussion, I will start from a description of general features of Quantum Mechanics in comparison to Classical Mechanics. Because the issues involved pertain to questions of the interpretation of Quantum formalism, I shall try to keep the discussion as non-technical as I can.

Before the emergence of Quantum Mechanics, the fact that an event was observed to occur with a certain frequency was explained without introducing an extra physical reality, besides those already postulated by Classical Mechanics. Apart from masses, forces and a degree of randomness in the interactions between the systems, which typically arise from a degree of uncertainty in the initial conditions, there was nothing else that was thought to play a causal role in determining how an event is to occur.

Take the case of Classical particles passing through a barrier with two slits suitably separated from one another, for example. If a large number of particles could be released from their source one by one, so that they would only pass through the centre of either slits (without encountering either the gap between the slits or their edges), Classical theory does not allow for probabilities to enter into the outcome. However, if at the point of release the degree of freedom of the particles is increased, so that directions of the outgoing particles now span a small angle (allowing them to arrive at the barrier on either sides of the slits, as well as on the gap in between), an element of randomness will thus be introduced in the behaviour of those that emerge on the other side. According to the Classical theory, this is entirely due to the collision of the passing particles with the edges of the slits in a manner that is not controlled by the initial conditions, and their subsequent scattering in directions which are not controlled as a result.

When the emerging particles are captured on a screen some distance away from the barrier, based on these considerations, one would expect the frequencies with which each spot on the screen receives particles from the barrier, to be smoothly distributed over a circular area in the centre of the screen. On the centre region in this area, the frequency with which particles arrive should be the highest, gradually decreasing towards the circumference, and diminishing beyond . If we take a long exposure photograph of all these particles, travelling one after the other from the source to the screen, we would expect to have a picture that resembles Figure 1.

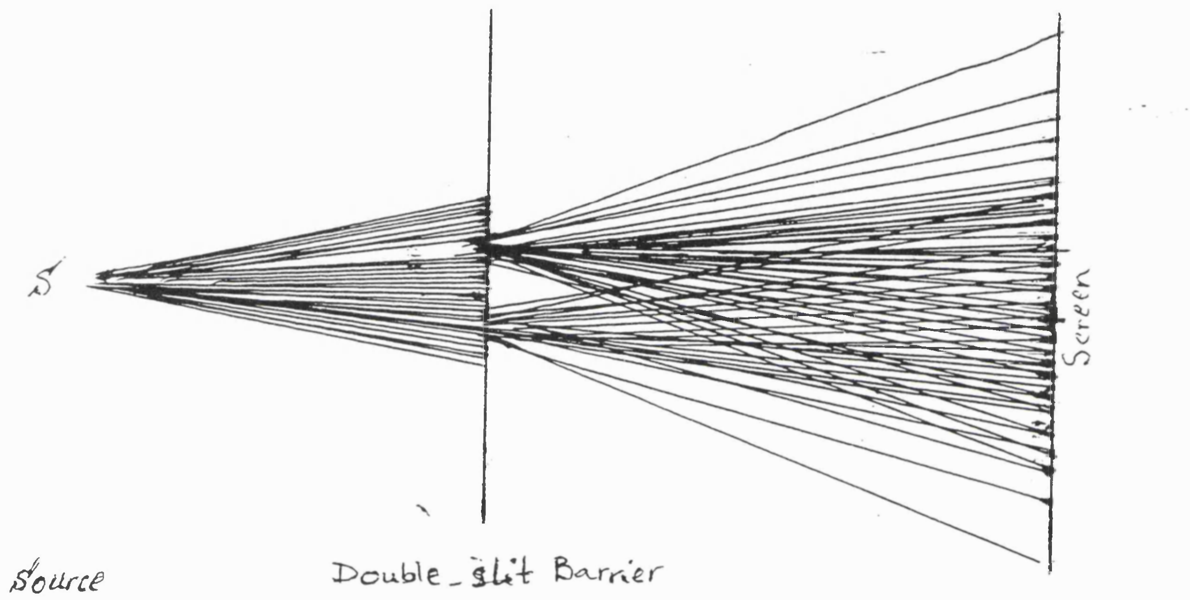


Figure 1

This distribution of frequencies can be represented by a bell-shaped solid, the cross-section of which on a plane perpendicular to the plane of the screen would look like Figure 2.

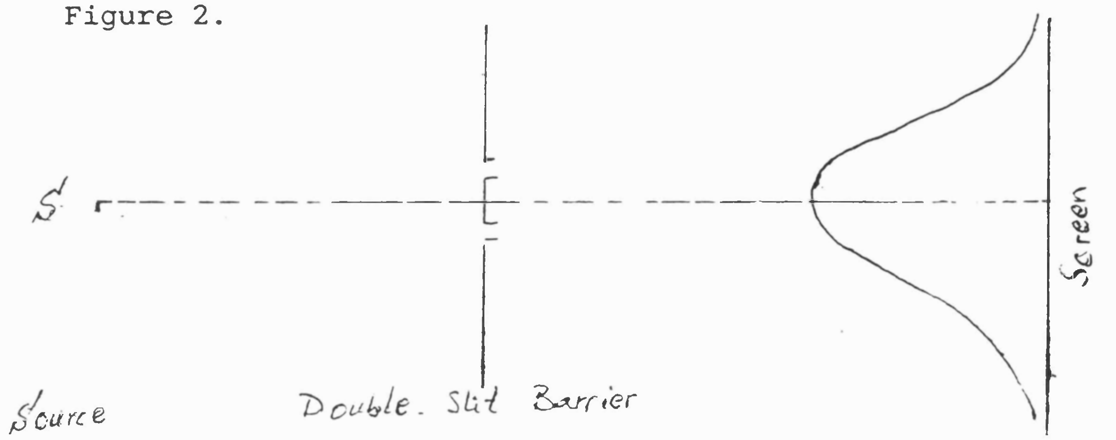
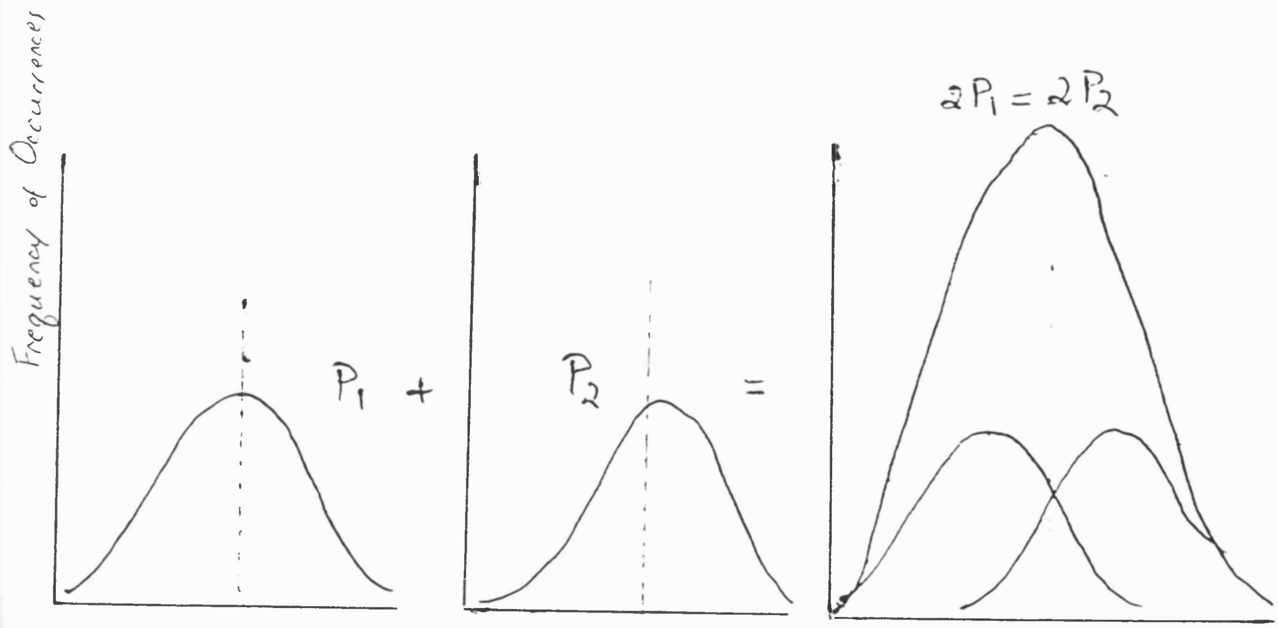


Figure 2

The outcome should remain the same if, instead, all the particles are released together within the same angular span as before, and a snapshot of their travel from the source to the screen is taken. Repeating this experiment with only a single slit open should, on the basis of the same theory, lead us to expect similarly shaped distributions of the frequencies. The only difference occurs on the sizes of the peaks, as well as the area over which the particles are spread. The number of particles that arrive on the central region of the distribution on the screen when only a single slit is open, should be half that of the same when both slits are open. This suggests that the frequency or probability (more on this later) of particles arriving on the screen from the two slits, is a simple sum of the frequencies or probabilities from each of the slits on their own.



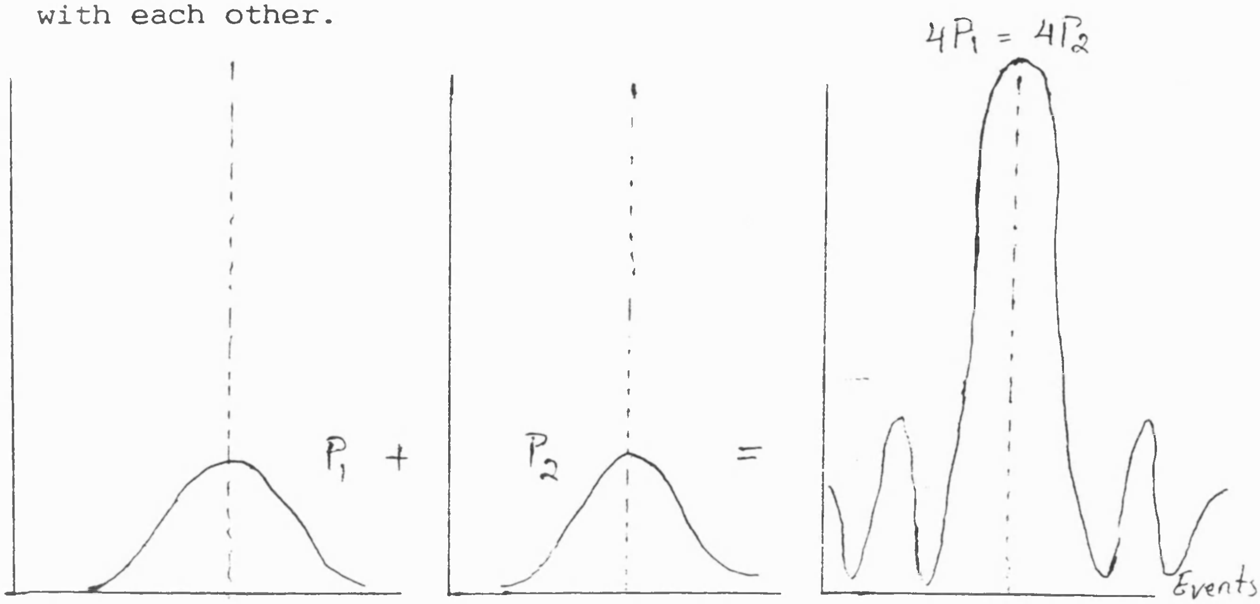
Probability distribution from slit no. 1 | Prob. dist. from slit no. 2 | Prob. dist. from both slits

Figure 3

2. Anomalous Results from Slit-Experiments

This Classical picture breaks down when we start dealing with particles whose parameters range over values on a scale close to Planck's constant (i.e. in the region of 10^{-34} Joules). The same initial and boundary conditions together with the Classical theory, lead us to distribution patterns which simply are never observed. In what follows, we are going to continue talking about experiments involving such particles passing through barriers with slits, or mirrors, or other gadgets that detect their presence in a region. This is more a matter of convenience than, in some cases, strict accuracy. Some such experiments may never have been performed, or indeed not even possible to perform, in the manner that we are going to describe here. But our description is consistent with all the relevant results that have actually been accumulated through a variety of experiments, and manage to show in simple, schematic ways, the essential features that characterize the peculiar behaviour of the systems we wish to consider.

Although passage of the particles in question through a single slit does produce, under appropriate conditions, similar bell-shaped curves to the Classical case, when two slits are opened at an appropriate distance from one another, the resulting distribution is never observed to be a simple sum of such curves. Instead of an enlarged bell-shaped curve, what is actually obtained is a pattern resembling a diffraction pattern typical of ordinary waves interfering with each other.



Probability distribution from slit no. 1

Prob. dist. from slit no. 2

Prob. dist. from both slits

Figure 4

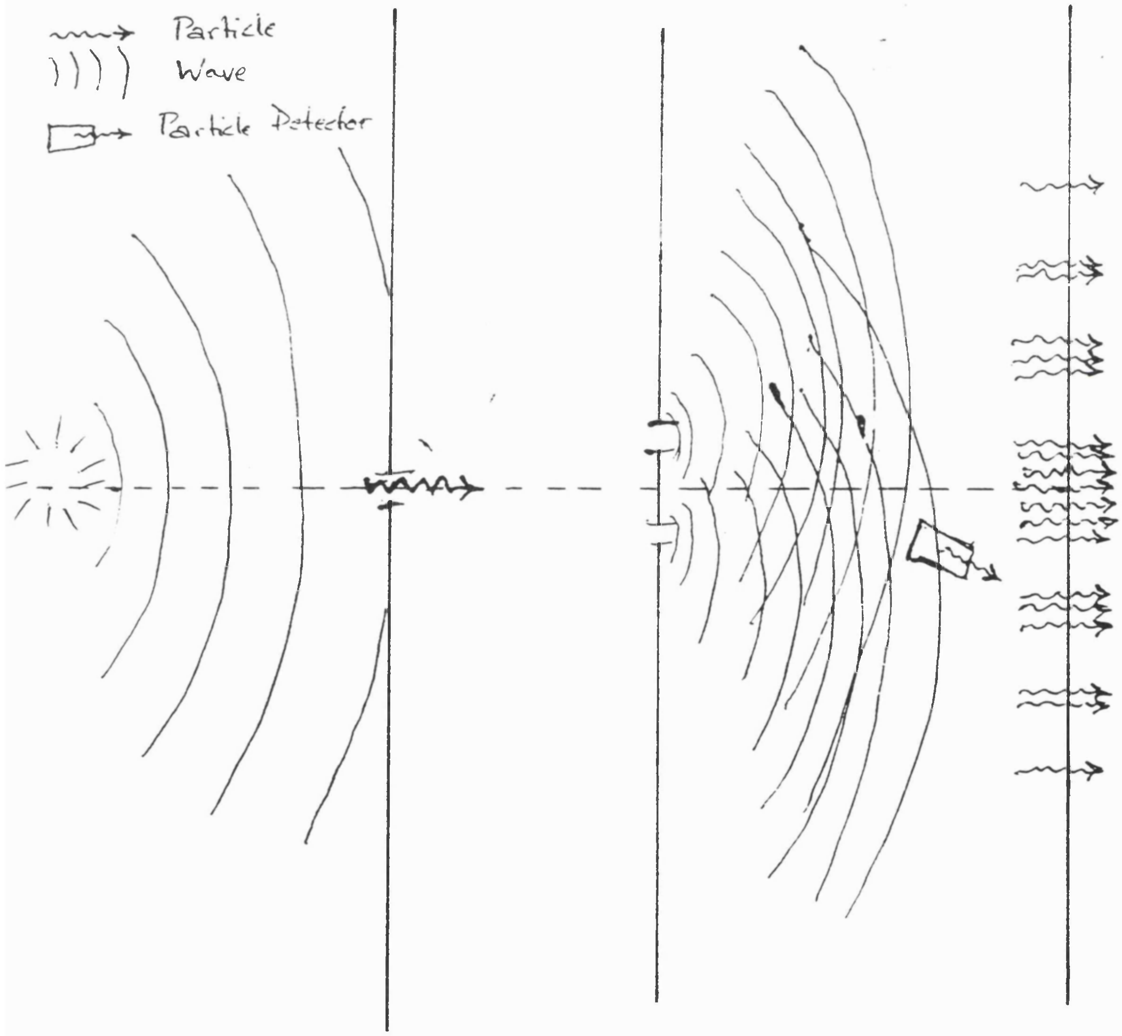
The resemblance of the distribution of our particles from the double-slits to a typical diffraction pattern, is unaltered either when it is obtained from a swarm of identically prepared particles falling on the screen at one go, or from the cumulative traces left by them arriving there in succession, over a period of time. Obviously a simple scattering explanation along with the Classical picture of the initial and the boundary conditions does not suffice to account for this outcome. In the new distribution pattern, there are areas on the screen where no particle passing through the double-slits is ever observed to land. However, the very same areas appear quite permissible when the same sort of particles are all made to go through only one slit on its own. So, instead of increasing the chances of arrival on those areas, when two slits are left open simultaneously, this option actually prevents them from ever arriving there.

3. 'Wave-Particle Duality'

Because of a striking resemblance to the diffraction pattern produced by undulatory phenomena under similar circumstances, the natural reaction is to bring some kind of wave behaviour to bear on this situation. The problem that would have to be faced as a consequence, is how to reconcile this behaviour with entities which otherwise display quite distinct corpuscular constitution.

One popular solution to this problem has been to suggest that unlike the ordinary phenomena, which is divided by Classical Mechanics into mutually exclusive categories of particles and waves, the phenomena whose parameters take on only exceedingly small values, are neither waves nor particles on their own. However, *the observable traces* that the latter leave on ordinary objects with which they interact are, under certain circumstances, corpuscular in character, and undulatory under others. This, in general, is neither ad hoc nor untenable. For there are extant theories which by filling the physical space with an all-pervasive field manage, with considerable success in some cases, to reconstruct corpuscular features (by concentrating the intensity of the field in a confined region of space), as well as undulatory ones (by propagating disturbances in the field in various directions).

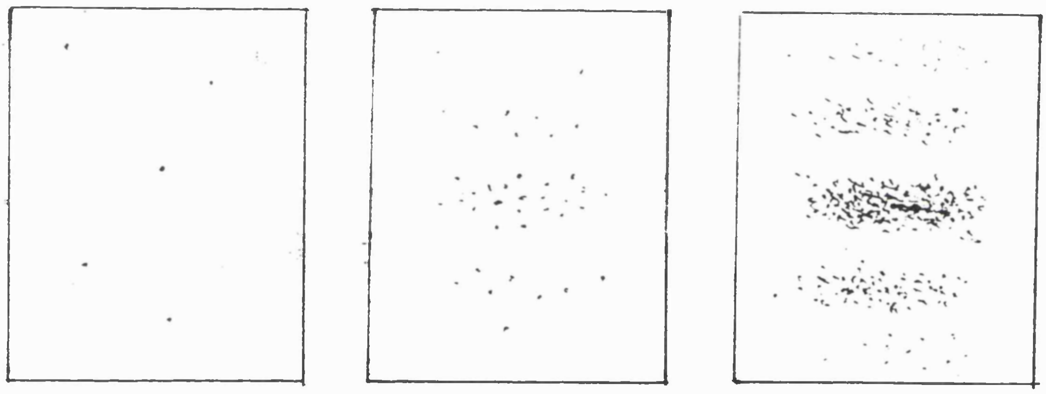
Take a disturbance in the electromagnetic field, for example, such as a flash of light with an exceeding low intensity. When passing through a slit with suitable aperture, the emergent radiation appears to behave in a particle-like manner. When passing through a double- slit barrier with suitable apertures and separation, the radiation that emerges appears to possess continuous wave-like characteristics.



Weak intensity light | Single-slit barrier | Double-slit barrier | Photographic plate

Figure 5

The problem arises when the intensity of the electromagnetic radiation from a source is so weakened that to observe its effect on a photographic film installed on the screen in a double-slit experiment, long exposures of a highly sensitive film in a tightly insulated environment is required. The first result would appear as a few dots scattered randomly over the film area. As the time of exposure is increased, more and more dots would appear, gradually arranging themselves in a pattern converging on a diffraction pattern produced by interfering waves.



Development of the pattern on the photographic plate as the time of exposure is prolonged.

Figure 6
(Reproduced from Rae [1986], p. 7)

The dots on the film clearly suggest that what has been arriving there from the slits are indeed a succession of individual particles rather than a continuous wave. The strange thing is that individual particles should, by the mere fact of going through two rather than one slit, not only arrive on the film only at those regions where a constructive interference between two waves from the slits would otherwise take place, but also in numbers that is directly proportional to the intensity of the resultant wave at those locations.

In line with the sentiments of some of the early contributors to the formulation of the Quantum theory, Rae [1986] suggests that as a rule, the character of physical systems should be viewed as determined by their interaction with particular types of apparatus deployed to observe them. That is, we should revise the attitude shaped by common sense, and legitimized by Classical Mechanics, whereby physical systems are pictured as having definite characteristics which they retain independently of whether or not they are interacting with measuring devices on which they leave observable traces. He writes: 'The fact that processes like two-slit interference require light to exhibit both particle and wave properties is known as *wave-particle duality*. It illustrates a general property of quantum physics which is that the nature of the model required to describe a system depends on the nature of the apparatus it is interacting with: light is a wave when passing through a pair of slits, but it is a stream of photons when it strikes a detector or a photographic film.' (Rae, [1986], p. 9). When two different interactions are combined, as in a double-slit experiment, according to this view, it should come as no surprise that the exposed parts of the photographic film show up as lots of grains arranged in a diffraction pattern.

4. Some Critical Remarks

The trouble with suggestions such as this is that they are easier said than understood in connection with the particular cases we have been considering . For suppose that the film in a double-slit set up is exposed to a very weak source of light, only so long that just a single dot appears on it. This suggests that the field has been absorbed by the film at a finely localized spot on its surface. Yet the same field is alleged to have spread out all over the space between the double-slit barrier and the film. How a wave-form, which is supposed to permeate through the whole volume confined by the set-up, suddenly contracts at the instant of making contact with the film, to leave its mark at a spot on the latter's surface, is a mystery for which no physical explanation is offered.

Let us look at this mystery from a slightly different angle. Since light passing through some single-slit barriers (those just large enough as not to give rise to a diffraction phenomenon) allegedly behaves in a corpuscular manner, it should, in principle, be possible to isolate a single photon on the other side, if light from a sufficiently weak source is directed at it. If, now, the same photon is brought to pass through a double-slit barrier, according to the suggestion under review, it should somehow emerge on the other side as a continuous wave emanating from both slits. The introduction of an additional slit, suitably separated from the first, is supposed to transform a highly localized field intensity into a wave-form which would now burst open all over the space beyond.

A question that immediately comes to mind is exactly when this transformation is supposed to take place. Is it immediately before the photon makes contact with the barrier? If this were the case, how would it recognize that there are two, rather than one openings waiting for it to go through so as to adjust its posture afterwards accordingly?

Perhaps, one might say, it is at the very instant that the contact is actually made that the necessary adjustments are executed. If this were the case, then surely just prior to the time of contact the field must still be in its localized form. At this instant, one possibility is that at the next instant it would enter the opening provided by either one of the slits, i.e. when it does make contact with the barrier, it would find itself within the opening provided by one of the slits. But then at the time when the contact has actually been made, the field, so far as *it* is concerned, is in the situation of interacting with a single slit and should remain as localized as before. How does the mere presence of another opening some distance away disturb its localized posture, spread it out, making it somehow emanate from the other side of the barrier as two wave-forms?

None of this makes any sense. Yet the view that leads to these mysteries by no means suffers from lack of endorsement. Authors of considerable weight and authority persist in writing as though such mysteries must be tolerated, perhaps as a necessary price to pay when we attempt to come to terms with the intricacies of nature at levels to which we have no direct access. Penrose, for instance, writes: '...each individual photon behaves like a wave entirely on its own! In some sense, each particle travels through *both slits at once* and it interferes with *itself*!' (Penrose [1989], p. 304).

The standard retort to the difficulties raised here is that they highlight the futility of trying to advance beyond an essential limitation when description of small-scale physical systems is being sought. We meet with nonsense, it is claimed, only when we attempt to describe systems in states that remain unobserved. The state of a field when it is in-between interactions where it has actually left detectible traces, remains unobserved. When attempts are made to extrapolate a particular picture which is suggested by an observed trace a field has left on another system, to states from which no such traces are at hand, difficulties ensue. In 1927 Heisenberg even went so far as to suggest that the very meaning of the observable terms in our language, such as 'position', should be strictly restricted only to the observed states of a system. 'If one wants to clarify', he wrote, 'what is meant by "position of an object",... for example of an electron,... one has to describe an experiment by which the "position of an electron" can be measured; otherwise, this term has no meaning at all.' (Jammer [1966], p. 328).

In whichever guise it is stated, the limitation prescribed here results in the abandonment of an ambition which underlined the tradition of seeking descriptions of systems and their time-development in physics. This ambition consists of seeking a coherent view of physical systems which is capable of providing the information we require about their behaviour at all states, observed or unobserved.

Resorting to such limitations is not, in itself, an illegitimate move which ought to invite resistance at all cost; mere entrenchment of an ambition within a tradition is, after all, not a sufficient justification for believing in its merits. Equally, giving up the ambition in question, unless it is shown to be inevitable by the physics of the situation, is bound to raise legitimate questions about its wisdom. So far, we have not mentioned anything about the physics of the situation apart from schematically describing some experimental results which defy satisfactory explanation from Classical Mechanics.

5. The Conception of State in Classical Mechanics

Historically, proposals such as mentioned and briefly examined here were advanced subsequent to the birth of Quantum Mechanics. The mathematical formalism of this theory has proved amply adequate to deal with the numerical data as well as the distribution patterns obtained, not only from the experiments schematized above, but from numerous others besides. This highlights the fact that the difficulties touched upon relate to conceptual matters that arise when the formalism is subjected to physical interpretation. Before proceeding any further, however, we must have a brief look at what this formalism has to deal with, and in what respects the latter demands departures from the formal devices employed by the Classical theory.

Let 'C' denote the class of physical systems whose behaviour is satisfactorily described by Classical Mechanics (CM for short). Descriptions in CM proceed by attributing to each member of C a set of parameters S. For each system, some of these parameters, such as electric charge or mass (in non-relativistic descriptions), take on values which remain constant throughout its behaviour, while others, such as location in physical space or velocity, are allowed to vary as the states of the system evolve. We must pause to point out here, that all our discussion is going to be restricted to non-relativistic descriptions, since it is on this territory that all the major battles over interpretation are fought.

In each case, a set of constraints determined by the physical conditions under which the system is to evolve, is specified which limit the values its parameters can take. Within these limitations, the members of S take, as their values, any real number in a continuous range. At any instant t, each parameter takes on one and only one value. Let us call a set of values for S 'complete', if every number in this set is a value of a distinct parameter, and for every member of S there is one and only one value in this set. A complete set of values for S at t, identifies a unique state of the system at t.

CM postulates a set of rules R, which uniquely determines a complete set of values for S at time t, given a complete set of values for a proper subset of S at t. Let us call the latter, the set of 'D-parameters' of the system. At any given time, therefore, a complete set of values for the D-parameters of a system uniquely determines the state of the system at that time. At least two parameters are needed in order to define a set of D-parameters for any system, and they usually are taken to be its location (if the system is a single particle), or its configuration (if a system is a combination of particles) in the physical space, and its velocity (or generalized momenta, if the system is complex). Because a complete set of values for a pair of D-parameters of a system exists at any one time, the parameters may be regarded as compatible.

It is implicitly assumed in CM that a complete set of values for the D-parameters of a system can be measured at any time. A set of laws, L, are further provided which, given the state of a system at any one time, uniquely determine all its subsequent states at every later times. Together with R and L, a complete set of values for the D-parameters of a C-system suffice for the comprehensive and unqualified description of the evolution of its states through time.

6. The Quantum Mechanical Situation

Quantum Mechanics (QM for short), may be described as a theory which grew out of attempts to pursue the aspirations of CM for the description of a class Q of physical systems whose behaviour indicates violation of some of the former's precepts. First, there was the observation that in certain circumstances, some parameters of Q-systems take only on values which fall within a discrete set of real numbers, instead of ranging over a continuum. For certain parameters, such as the total energy or the angular momentum of some Q-systems, this set is observed always to contain values that are only integral multiples of a certain constant, irrespective of the initial conditions or the frame of reference used to measure them.

Then came the discovery that not only (*i*) a complete set of values for S at a given time t, does not always determine a unique state for a Q-system (meaning that one such set may correspond to different states), but also (*ii*) a complete set of values for what in CM is singled out as the D-parameters of a Q-system at t, together with some modifications of R and L, do not always determine its states subsequent to t.

Against this background, however, came the evidence of remarkable statistical regularities among the values obtained from repeated measurements of the parameters of identically prepared Q-systems. When individual parameters are singled out, and their values measured by identical procedures for a large number of Q-systems, all of which start initially under identical conditions (we will clarify in more detail what this involves later; for the moment, let us take it as sufficiently understood), distributions are obtained which display invariant features. Furthermore, when such distributions are compared with those obtained for other parameters under the same conditions (using, in each case, appropriate but invariant measuring procedures), invariant relations appear to be exhibited between them.

7. The Uncertainty Principle

The quantitative expression of the latter fact was formulated by Heisenberg, and is known as the Uncertainty Principle. Suppose we take a large number of a certain kind of Q-systems (electrons, for example), and prepare them initially under identical conditions. Suppose, further, that we conduct an experiment with these (like, for instance, a double-slit experiment) in the outcome of which we hope to measure their positions. It does not matter if we do this for each individual one by one, or for all of them together in one go (although the latter will prove very difficult with electrons).

Let Δq represent the interval in the scale we use to measure these positions, within which the values we obtain are observed to fall with varying frequencies. Next, suppose we take the same number of the same kind of systems, initially prepare them in the same manner as before, and conduct an experiment (not necessarily the same one as before), in the outcome of which we hope to measure their momenta, that is, roughly, their velocities raised throughout by a constant factor equivalent to the value of their mass. Let Δp represent the interval in the scale used for this purpose, within which the values obtained are observed to occur with different frequencies. The Uncertainty Principle legislates that values thus obtained for the quantities Δq and Δp should always obey the relation $\Delta q \cdot \Delta p \geq \frac{1}{2} \hbar$, where $\hbar$ is the Planck's constant divided by 2π.

Starting from these facts, several schemes were proposed to capture the behaviour of Q-systems, which although they shared more or less the same physical content, differed by employing different mathematical techniques. One, due to Schrödinger, sets out to describe the evolution of Q-systems by employing the customary phase space. Phase space is a mathematical construction, developed in the later part of the evolution of CM, which is made up of a set of D-parameters of a physical system, namely its configuration in the physical space, as well as its momenta in an abstract momentum space.

8. Phase Space and the States of Physical Systems

Suppose a system consists of N number of particles. Each particle, at an instant, is located somewhere in physical space which itself is three dimensional. So, the instantaneous location of each particle in this space can be specified by three coordinates in a certain frame of reference. The rate at which this location changes, that is, the instantaneous velocity of each particle, can also be projected onto three components, each one of which corresponds to a principal direction in the physical space. An abstract three dimensional space can, therefore, be constructed in which an instantaneous velocity of a particle can be represented as a point specified by three coordinates. Each of these coordinates would then represent the component of the instantaneous velocity of the particle along an appropriate principal direction in physical space.

If these two spaces are now put together such that all the axes form a linearly independent set (which, for our purposes at this point, roughly means that no one axis can be reconstructed by adding up different multiples of the others), a six dimensional space is formed in which the location of a particle in the physical space at a given instant, as well as its velocity at that same instant, can be represented by a single point. This newly constructed space is called the 'phase space' for a single particle. Since the location and momentum of a particle jointly determine a set of D-parameters for it in CM, each point in the phase space represents a unique instantaneous state of the particle. As the number of particles which make up a physical system increases, so does the number of dimensions needed to represent its states, accordingly. To represent the states of a system which comprises a total of N particles, a space of 6N (linearly independent) dimensions would be required.

For a system of N particles, a complete set of values for its D-parameters consists of 6N real numbers. Customarily, physicists are interested in the representation of the changes of states in the phase space. Therefore, it is the rate at which states change, in various directions in this space, which is of primary concern to them. Since this rate of change is different in different directions, to each point a vector can be assigned whose direction indicates the direction of the greatest change from that point, and whose magnitude shows the 'size' of that change. Larger vectors would thus represent a more rapid change in a certain direction, and smaller ones a slower one in certain others.

As was noted earlier, this project for the description of the behaviour of physical systems, was frustrated by the discovery of Q-systems. This meant that the representation of the states of a system as points in an appropriate phase space lacks generality, and is untenable if the domain of physical systems is to be truly universal. There are, however, new adequacy requirements which any theory purporting to describe the behaviour of physical systems must satisfy. One is that it must entail, with appropriate initial conditions, the statistical regularities observed for the corresponding values of each parameter. The other is the Uncertainty relations between the values of the appropriate pairs of parameters.

9. Schrödinger's Scheme for the Description of States

One scheme for the description of the states of a system is Schrödinger's. It starts out by selecting either one of the parameters that make up the phase space, and introduces a new parameter over all its values, such that the values of the latter everywhere depend on the values of the former. This means that the values of the new parameter are assigned by means of a function whose domain ranges over the values of the parameter selected from the phase space. Either of the selected parameters specify a space 'half the size' of the original phase space. So, for a system consisting of N particles, the space defined by its configurations, as well as that by the momenta of its constituent particles, is 3N dimensional. The first is appropriately called the 'configuration space' of the system, and the second, its 'Momentum space'.

Let us take the configuration space of a system and see how the new parameter is defined over it. Each point u_i in this space, as we have seen, corresponds to a particular arrangement of N particles in the physical space. To it, a complex number is assigned such that its absolute square, i.e. the product of the number with its complex conjugate (which always comes out a real number) represents, with proper adjustments in some cases, the probability of finding the system in the close vicinity of u_i.

This one-one pairing of a real number u_i with a complex number, finds its geometrical representation in the correspondence between u_i and a unique point on the so-called 'Argand plane'. Roughly, this plane is set up by introducing two additional (linearly independent) dimensions to the space of the chosen parameter. The plane defined by these new dimensions at u_i serves as the complex axis associated with that point. Each point in this Argand plane, specifies one value for the new parameter at u_i, and is called an 'amplitude' for the particular configuration of the N particles which is represented by u_i.

At any one time t, the assignment of amplitudes to the points in the space of a selected parameter must be complete, in the sense that every point in this space should have a unique amplitude corresponding to it at t (in regions where, for example, the spatial arrangement of the particles is never found, the corresponding amplitudes assume zero for their values). The function which for every t determines a particular distribution of the amplitudes over the space of a selected parameter, is obtained by solving a first order differential equation. This equation was first discovered by Erwin Schrödinger in 1926 and is known as the time-dependent Schrödinger equation.

We briefly note here another peculiar feature of QM in comparison to CM. When this picture is applied to determine possible values for the position of an electron in a Hydrogen atom, which is itself in a state of fixed energy, positions are allowed (with a non-vanishing probability) that according to CM are impossible for the electron ever to be found at. Needless to say that the experimental confirmation of this consequence of the procedure described was another major triumph for QM.

Now it can be seen how the Uncertainty Principle is brought to bear on determining the function which assigns amplitudes to the values of one parameter in the phase space, once a function for the other has been found. Suppose we have a function that determines the distribution of the amplitudes over the configuration space of a system. At every time, the spread of the values of this parameter, i.e. the spread over which the amplitudes are distributed, represents the uncertainty in the configuration of the system. The function that is to assign amplitudes over the corresponding momentum space, must, at each time, determine a spread of momentum values which is related to that of the configurations at that time, through the Uncertainty Principle. Mathematically a device called 'Fourier transformation' is available, by the application of which on a function with the mentioned characteristics over one half of the phase space, one with precisely the right sort of features over the other half may be produced.

10. The Conception of State in Quantum Mechanics

In what must be acknowledged as a radical break with the Classical tradition, it is now proposed that a complete set of values for the amplitudes over the space of a selected parameter of a physical system at any one time, picks out a unique state of the system at that time. The function that assigns values to the amplitudes at each point in the space of a selected parameter of a system in a particular physical situation, is, as a result, called, appropriately, its 'state function'.

This view of the state of physical systems differs sharply from its counterpart in CM in two respects: (*i*) the set of numbers that singles out a state in QM, are values of a parameter that simply does not exist in CM, and (*ii*) the minimal set of parameters whose values determine states in CM, are no longer D-parameters in QM.

In general, any set of numbers may specify a vector, provided a suitable structure exists into which it can be embedded. Usually, a set of objects which includes zero, is closed under addition and scalar multiplication, and whose members satisfy few additional requirements, provides such a structure and is, what we have already met, a vector space.

It so happens that there exists a special vector space, called a 'Hilbert space', in which a complete set of values for the amplitudes of any one parameter at any one time can be embedded. Each number, therefore, in this set would specify one component, for a unique vector, along one principal direction in the corresponding Hilbert space. Hilbert spaces, in general, have an infinite number of such directions. Each can be marked off at the origin of the space, much in the same manner as in the case of the three dimensional Cartesian space, by a vector of unit 'length' which is called a 'basis'. The set of all the bases of a Hilbert space is linearly independent, which roughly means that none can be reconstructed by linearly combining the others, and, every vector in the space which is not a basis, can be expressed as a sum of all the bases, each multiplied by a suitable factor (called a 'coefficient').

We now have a picture of the components of a vector in a Hilbert space which, in many ways, is similar to the familiar components of a vector in the three dimensional Cartesian space. Each component is a unit vector multiplied by a [complex] number, and together they add up to define a unique vector in the appropriate Hilbert space. It must be kept in mind that such a component represents a single amplitude associated with a particular value of a selected parameter at a given time, and thus determines the probability, for the system involved, of actually having that value for the parameter in question at that time.

Each vector in the Hilbert space appropriate for a selected parameter of a given physical system, would now represent a unique state of that system, and is fittingly called its 'state vector'. The scheme for the vectorial representation of Quantum Mechanical states of physical systems was developed by Von Neumann and Dirac. It is more general than Schrödinger's, in the sense that it can accommodate, in a straightforward manner, handling of those dynamical parameters that have no Classical counterparts (such as a particle's spin, it's parity, etc.).

In both these schemes, one and the same set of amplitudes for all the values that a selected parameter of a system can possibly have in a given situation at any one time, corresponds to a unique state of the system. Amplitudes are not themselves observable quantities, but the probabilities they determine, are. An examination of how these probabilities are to be measured would shed some further light on what the Quantum Mechanical states look like.

11. Credible Measurement Results

In order to observe a particular value for a system's parameter, calibrated instruments must be used which correlate a particular feature in the system's behaviour with a number on the scale built into the instrument. The simplest examples of such instruments are measuring rods, clocks or weighing scales.

C-systems behave very conveniently when various instruments are used, in laboratory conditions, to measure their parameters. This convenience stems entirely from the fact that CM is perfectly adequate for the description of their states, and single measurement results on just two parameters of the system suffice to determine each state according to this theory. This, of course, is not to say that any measurement result is just as good as others. The possibility of sloppy readings, instrument malfunction, as well as breakdowns in what ought to be tightly controlled conditions in the laboratory, always exist and must, therefore, be vigilantly guarded against. The point is, once due care is exercised and errors eliminated, a single credible result from the measurement of two parameters counts as adequate data for the CM to determine states of an observed C-system.

We have seen that Q-systems do not generally lend themselves to a similar convenience. Not that these systems are so fragile that the slightest contact with a measuring instrument destroys any chances of obtaining a credible reading for the value of their parameters. The suggestion, often cited in writings on Quantum theory, that it is impossible to measure values for 'canonically conjugate' variables (i.e., roughly, variables that *in CM* would either be themselves the D-parameters of a system, or else would be a pair of their components) simultaneously, is a red herring. If, indeed, it were impossible to obtain credible measurement results that somehow tie up with the values of a parameter, the testability of the theories whose equations range over their mathematical surrogates would be seriously impaired. On the other hand, such results invariably come in the shape of particular readings from a certain scale.

As we have mentioned, a single measurement result on the members of what in CM amounts to a set of D-parameters for a Q-system, does not constitute adequate data for deciding the state of that system. Suppose we carry out a controlled measurement on the position of a Q-particle at time t, and manage to obtain a credible reading. Since this result is a particular number, there is no dispersion in the values of this parameter, and hence, if we let q represent the latter, Δq is zero.

Substituting this in the Uncertainty relation, we get a corresponding dispersion for the values of the particle's momentum which is indeterminately large. There is no obligation to take this to mean, as some writers suggest, that the particle has no particular momentum at time t. Stretching the meaning of a particle's 'momentum' this far, risks its comprehension. Rather, it indicates that any particular value for the particle's momentum, is compatible with the particular position measurement obtained at t. One consequence of this is that, no matter how vigilant we may try to be in guarding against errors creeping into our momentum measurement at t, we have no way of distinguishing a credible result from an error-ridden one, once a credible result has been obtained for the particle's position at t. In this sense, we can say that the particle's momentum is, *indeterminate* at the time.

All this should come as no surprise if it is to QM that we turn for deciding the states of Q-systems. What we need here are probability distributions for possible values of certain parameters, and single credible results from particular measurements are far from providing them. But never too far.

For it is from the accumulation of these very results, in the repetition of the very same measurements, that such distributions emerge. That is why we have been insisting on a stock of a large number of systems of the same kind, each of which must be initially prepared in the same manner as all the others.

If after carrying out a particular measurement on a parameter of a Q-particle, we could regain it intact and bring it back to exactly the same state as we started with initially, we would not need to start with a large number originally. As it happens, Q-particles are generally not easily retrievable after they have left a trace on a measuring instrument; they are either absorbed by the latter, annihilated after the process, scattered in unpredictable directions, or else next to impossible to capture and bring back to the starting line-up.

To repeat the same measurement on a parameter, then, we start out with a large number of Q-systems that are all of the same type (electrons, for example, or photons, etc.), and prepare them to be in identical situations. This means that their constant parameters, such as mass, size, charge, and the like, must all have exactly the same values. Then, the same set of constraints is imposed on all of them initially. That is, for example, if they are to start in an electric field, the potential difference should be the same for all of them. Or, if they are to be in an electromagnetic field, the field's intensity should be the same for all of them. Or, if they are to get on to some target, they should all start from the same 'launching pad'. Finally, one and the same procedure must be used in each trial the outcome of which is to be a single measurement on a selected parameter.

12. Dispersion of Measured Values and the State of Q-systems

As was noted before, in contrast to the case of C-systems in a given state, where repetitions of the same measurement establishes a single value for a selected parameter, from the repetition of the same measurement on a large number of identically prepared Q-systems of a given type, a dispersion emerges over which values occur with varying frequencies. This dispersion is further cleansed of experimental errors when a considerably large series of repetitions of the same measurement are carried out. The size of a thus established dispersion remains unchanged, as long as the trials involve systems initially prepared under the same conditions. It begins to vary accordingly, as variations are introduced in the initial preparation of the systems of an identical type.

The frequency to which a particular value for a selected parameter converges over an indefinitely large number of identical trials is, customarily in physics, taken to be the probability of the event that the parameter in question assumes that value. Since a unique distribution of probabilities for the values of a selected parameter is always obtained from credible measurements on numerically different, but identically prepared, systems of the same type, it may be taken as the 'identity sign' for a particular state. Whenever such a distribution is obtained from error-cleansed measurements, it can be used to test the accuracy of a prediction from QM about that state. So in the end, even though the luxury of D-parameters is denied the Q-systems, an economy of a sort is introduced in deciding their states, in so far as only one parameter is required for the job.

Suppose the parameter selected for deciding the state of a stock of identically prepared Q-particles of the same type is their location in the physical space, designated by 'q'. For ease in picturing the situation, let us restrict q to one dimension only. A particular distribution P_1 of probabilities obtained for a range of q's values, singles out a unique state of the system. Let this state be designated by 'Z_1'. We have, therefore, a one-one correspondence between Z_1 and P_1. The latter, moreover, rises to, and falls from, a peak over an interval of length Δq_1. This interval indicates the range of values for q which are observed to occur with significant frequencies. This is schematically represented in Figure 7.

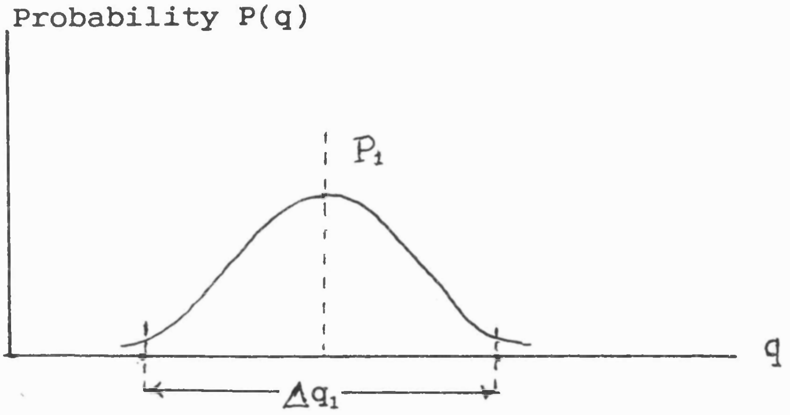


Figure 7

Suppose now, another stock of the same type of Q-systems is taken and are all prepared under exactly the same conditions as those under which Δq_1 was measured. It would be natural to say that all the members of this sample are also in the state Z_1. This time, however, let measurements be made, with an appropriate and invariant measuring procedure, on the momentum of the systems which we designate by 'p'. A dispersion of length Δp_1 would be obtained for the frequencies of observed values of p. The fact that all the systems involved in these trials are supposed to be in state Z_1, establishes a relation between Δq_1 and Δp_1. This relation is specified by the Uncertainty Principle to be $\Delta q_1 . \Delta p_1 \geq \frac{1}{2} \hbar$.

As it happens, q and p are only one pair of parameters which in CM define a set of D-parameters for a system. The Uncertainty Principle actually extends the same relation to any other pair of parameters which in CM may define a set of D-parameters. Each parameter in such a pair is called the 'canonical conjugate' of the other. The Uncertainty Principle entails that from the size of the dispersion in the values of one parameter of a system in a particular state, the lower bound for that in the values of it's canonical conjugate is determined. For Q-systems in a certain state, this is as close as we can come to the Classical case of having complete sets of values for canonically conjugate parameters. Because the values of these parameters always display dispersions when Q-systems are in a certain state, the parameters may be said to be 'non-compatible'. We may have to be content with probabilities and dispersions, in place of single sharp values, but with only one measuring procedure we can do what would otherwise require at least two if we were to operate with CM.

13. Instantaneous State of a Single Q-Particle

There is a twist in the tale we have told, that can be the cause of bewilderment as well as perhaps some confusion. We have seen that the observation of a Quantum Mechanical state, which is done by measuring a probability distribution for a selected parameter, requires a certain collection of Q-systems. A question can obviously be asked about the state of a single member of such a collection at any one time. If we refer back to the experiment, the outcome of which was depicted in Figure 6, we can see that the question of the state of each particle involved may be decided only after the result of the measurement for all of them has come through. If, at the very early stages when the positions of only a handful of particles have been detected on the photographic film, we ask 'what state are these particles in?', the answer must be 'any one which is compatible with the distribution pattern that has emerged up to that stage'. And clearly there are many such states.

This calls for introducing a distinction. In the case of the C-systems, their state can, in principle, be decided at any given time, by the results of single credible measurements on relevant parameters at that time. In the case of a Q-system, although it is in a particular state at any given time, this state cannot always be decided at that time. One must go on repeating the measurements *ab ovo*, to use Schrödinger's phrase (Schrödinger, [1935]), until a particular distribution pattern emerges and is stabilized. So, in general, being in a certain state at a particular time does not entail that it can be decided at that time. Conversely, that the state of a system cannot be decided at a given time, does not entail either that it has no state at all at that time, or that its state is somehow inconceivable.

Still, however, a nagging question hangs over the state of a particle, for a parameter of which a particular value has been credibly measured, say, at time t. Even if we manage, somehow, to obtain a particular value, at t, for the parameter which is the canonical conjugate of the first, we saw that its state at this time, remains undecided. But in that case, what is it exactly that we are lacking at t for deciding the state of the particle? Have we not got at hand all that is required for fixing a state of the particle?

The answer ought to be obvious. What we need, but have not yet managed to obtain, are amplitudes. Remember, it is the set of all the amplitudes over the possible values of a parameter that singles out a state. A single value for the parameter in question is just one, among perhaps as much as infinitely many others, of its possible values. It is not even a measure of a single amplitude, let alone all the others that are required. To measure the amplitude associated with this particular value alone, we need to find the frequency with which this value recurs over a sufficiently large number of trials. The numbers we need in order to decide a state of the particle at t, can only be dug up from an enormous amount of readings from the instruments used for the measurement in question.

The element of unease that was felt in connection with this particular question, therefore, appears to be traceable to mixing two radically different conceptions. On the one hand, the Quantum Mechanical conception of a state requires amplitudes and these cannot be decided by single credible measurement results. On the other hand, the Classical conception of a state is complete with just such results and requires no further information. The two conceptions are inconsistent, thus one must be adopted with the total exclusion of the other. We cannot chose to pursue the procedure prescribed by one conception for deciding a state, only to stop in the middle of it, ask questions that presume a switch to the rival conception, and expect to get away without any problems.

14. Evolution of Quantum Mechanical States

Quantum Mechanical states, very much like their Classical counterparts, are determined by a set of numbers. The major difference, however, between the two conceptions of states is that in the latter these numbers represent values that the dynamical parameters of a system must have at a certain time, to the exclusion of all the others in the spaces of those parameters. Whereas in the former, they represent something, as yet we know not what, which somehow fixes the probabilities for an array of possible alternatives in the space of the dynamical parameters of the system. To highlight this difference, we may call, for lack of better words, the Classical state of a system a 'definitive' collection, and that of its Quantum Mechanical counterpart a 'statistical' one over the spaces of the canonically conjugate parameters.

The progression of successive states in time, from the point of view of CM is a sequence of definitive collections, and from that of QM, a sequence of statistical ones. Both progressions are governed by strict laws, provided the system involved is isolated and its states not subject to any interferences from outside the constraints already specified. That is to say, the state of a system at any one time, together with the laws, uniquely determine its state at any other time, only if the system remains undisturbed. In CM this law is expressed by the Hamilton-Jacobi equations which apply to points in the phase space. In QM it is expressed by Schrödinger's equation which applies to state functions in the space of a single parameter ('half the size' of the phase space).

The progression of the state function in time, as described by the Schrödinger equation, displays a periodic character, and defines a plane wave propagating over the space of a selected parameter. The frequency of this wave (and hence its wave-length) determines the total energy of the system involved up to a factor which is simply Planck's constant. This wave behaves in ways which in some respects is typical of Classical waves: it spreads continuously over the space in which it is travelling, and in cases where there is more than one disturbance propagating so that their paths overlap, they interfere constructively or destructively in the appropriate regions of that space. For this reason the state function is also referred to as a wave function.

Now we seem to have the wave we need in order to explain the diffraction pattern observed on the photographic film in the double-slit experiment. From the source, plane waves are set in motion, along with the particles and in the direction of their travel, falling on the surface of the barrier as the latter arrive there. On the barrier, everywhere except at the openings, the incident wave is partly absorbed and partly reflected in accordance with Classical laws. Through each of the openings, the incident wave passes, emerging on the other side as two waves which propagate simultaneously towards the screen with the photographic film. These waves, in turn, interfere as soon as they meet each other. The result, again in accordance with the Classical picture, is that they add up everywhere to reinforce one another if they are both rising together, or to cancel if one is rising while the other is subsiding.

15. The Reality of the State Function

Waves that produce visible diffraction patterns must be physically real. Yet the state function whose evolution in time happens to share some characteristics with ordinary waves, describes a mysterious phenomenon whose values cannot be directly measured at any time. The nearest this function comes to being observable, is by somehow determining probability distributions for the possible values of a given parameter. Are we to believe, then, that there is a physically efficacious phenomenon whose propagation in the real space fixes probabilities for where a particle may be located?

Many people regard this a very unpalatable suggestion. In the early days of the development of Quantum theory, de Broglie proposed the idea of 'pilot-waves' which was something along this line. According to this idea, a particle travels at the centre of a wave whose intensity at every point in space is proportional to the probability of the particle's presence there. Based on this imagery, he produced a relation between the wave-length of this 'pilot-wave' and the momentum of the associated particle that remains a permanent fixture in QM. That this imagery was indeed a guiding principle in his early work, is attested in de Broglie [1956],(p. 89): 'At the time when I conceived the first ideas of Wave Mechanics, I was convinced that it was imperative to accomplish a fusion of the physical notions of waves and particles... So I sought to represent the wave-particle dualism to myself by a picture in which the particle would be the centre of an extended phenomenon. This idea is found again and again in my early works'

Most other founding fathers of the theory neither found this idea acceptable (although some held similar views for a while), nor were necessarily led to their contributions by a particular imagery. Their major concern lay in finding a mathematical algorithm from which the observed statistical regularities in the behaviour of Q-systems could be produced. Indeed, some of them were not at all prepared to attribute the emergence of the statistical collections to any reason other than irremovable uncertainties in our knowledge of the initial conditions. Very much influenced by Bohr's ideas on the subject, Heisenberg wrote in 1927 : '...in the strong formulation of the causal law "if we know exactly the present, we can predict the future", it is not the conclusion but rather the premise which is false. We cannot know as a matter of principle, the present in all its details.' (Quoted from Jammer [1966], p. 330).

As a result, a view emerged and soon became dominant, according to which the significance of the state function must be restricted only to measurement results obtained in making observations on Q-systems. The numbers that make up the value of a state function at a given time, are, according to this view, to be construed, not as values that a physically significant parameter may have in reality, but as a catalogue of possible results to be expected if some measurements were to be carried out on the system. When Schrödinger, in his celebrated [1935], described the state function as 'the means for predicting probability of measurement results', and called it an 'expectation-catalog' (In Wheeler and Zurek [1983], p. 158), he was very much expressing this sentiment.

This view, which at one period enjoyed the status of an orthodoxy, confers (as we shall see later in more detail) on observation, and thus the observer, an unprecedented commanding position in the determination of the states of physical systems. So far, we have not seen anything in the physics of the Q-systems that would compel submission to this position apart from a conceptual difficulty in connection with the existence of 'probability waves'. This difficulty, which is really about how physical reality *ought to be viewed*, must, however, be balanced against the fundamental problems which are raised by the orthodox view about how the reality of states in physics is to be conceived.

In more recent times the recommendation that the conception of the reality of states in physics should be fundamentally revised has come to be increasingly questioned. However, not all the alternative proposals appear to converge on a single picture. The main reason for this seems to be that none has actually managed to avoid one startling consequence or another, and there is a lack of consensus on which of these is more tolerable than the others.

One such move away from tinkering with the very conception of the reality of the states of physical systems, and towards offering a picture for it, has been advanced in Penrose [1989], which it will prove illuminating to examine. Considering a single Q-particle he writes: 'Quantum-mechanically, *every single position* that the particle might have is an "alternative" available to it. We have seen that all alternatives must somehow be combined together, with complex-number weightings. This collection of complex weightings describes the quantum state of the particle. It is standard practice... to use the Greek letter ψ for this collection of weightings, regarded as a complex function of position - called the *wavefunction* of the particle... I am taking the view that the *physical reality* of the particle's location is, indeed, its quantum state ψ.' (Penrose [1989], p. 314).

Now, the phrase 'physical reality of the particle's location', as ordinarily understood, conjures up the image of a tiny spot in physical space which the particle *does* occupy at a given time. On the other hand, the value of the wave function that at any given time signals the particle's state, determines, in general, a statistical collection over its possible locations. From this collection a mean can be calculated which represents the average of all the locations that are possible, with varying probabilities, for the particle to occupy when it is in that state. This is called the 'expectation value' of the particle's location.

The expectation value of the particle's location in a certain state does, needless to say, represent a tiny spot in physical space that the particle *may* occupy at a certain time. But it certainly need not be identical to the physical reality of the location which the particle in that state would occupy at a given time. Conflating these notions is analogous to the case of the 1.5m tall man who did not know how to swim, but when told that the average depth of a river was only 1m, stepped confidently into it and drowned!

16. Beam-Splitting Experiments

That the wave function, on the other hand, should not be conflated with the physical reality of the particle's location, comes out clearly when we consider the particle's behaviour in the so-called 'beam-splitting' experiments. These are a variation on the theme of the double-slit experiments, with the advantage that they can actually be performed more or less along the lines that we will describe.

Ideally, when a beam of light falls on a fully silvered mirror, all of it would be reflected. We can split such a beam into a half that is reflected and a half that is transmitted through, if instead, we use a half-silvered mirror. If the intensity of the incident beam is cut down so that only a single photon falls on the dividing line in a half-silvered mirror, then it would either pass through or get reflected with an equal probability of ½.

Suppose we shoot a photon onto a half-silvered mirror that is tilted at a 45° angle to the path of incidence. There is a probability of ½ that the photon would be reflected off the mirror at 90° to its original path, and a probability of ½ that it would go through along that path. If now two fully-silvered mirrors are placed at points I and II along either routes, each tilted at 45° to the path of incidence, the photon should arrive at the point F, provided the distances travelled (which could, in principle, be made as long as one wishes) are equal. This is shown in Figure 8.

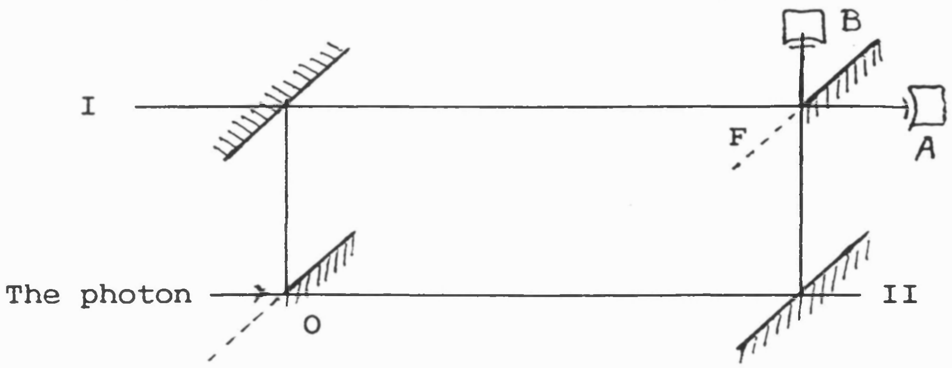


Figure 8

Suppose, further, that a half-silvered mirror , again tilted at 45° to the paths of incidence along both routes, is placed at F and photon-detectors set up at A and B to measure the probability of the photon's arrival at either point. Empirically, these probabilities are measured to be 1 at A and 0 at B, indicating that interference has taken place at F. 'What does this tell us', Penrose writes, 'about the *reality* of the photon's state of existence between its first and last encounter with a half-reflecting mirror? It seems inescapable that the photon must, in some sense, have actually *travelled* both routes at once!' (Penrose [1989], p. 330).

Whatever additional reasons there may be that add to the strangeness of the observed outcome (and there are some which we will consider), this manner of speaking appears nothing short of being inconsistent. For since we can make the distance OI shorter than OII, if the same photon is travelling along 'both routes at once', by the time it reaches mirror I, it would also be travelling unimpeded along the other route. It can then be said of the same particle that it has been simultaneously obstructed and unobstructed in its travel, which is surely absurd given our understanding of 'particles' and their properties.

It may be thought that perhaps the photon had split as a result of encountering the first half-silvered mirror. After all, a photon is just a bundle of energy and it is conceivable that it could spread just as it could appear localized. This way out, however, is blocked on empirical grounds. Whenever photon detectors are placed along both paths, only one of them is activated, and in a manner indicative of the passage of a whole photon.

17. Delayed-Choice Experiment

This way of speaking has a further embarrassing consequence. Wheeler [1981], introduces a variation in the experiment just described (In Wheeler and Zurek [1983], p. 184). Let us repeat the experiment only this time remove the final half-silvered mirror at F just before the photon is about to reach it. Now the photon will be detected at either A or B with equal probabilities. This indicates that no interference has taken place and the photon has travelled only one route: if it has gone along OIF, it will arrive at A, and if it has gone along OIIF, it will arrive at B.

In the light of the idea that the photon would have to travel both routes in the original version of the experiment, we would now face the bizarre consequence that a delayed-choice about the positioning of a half-silvered mirror at F has actually determined the events prior to when the choice was made. If we leave the mirror in place, the photon would allegedly travel both routes to get to F. But if we lead the photon to 'believe' that everything is set up so that it should travel both routes, only to 'pull the rug from under its feet', as it were, at a fraction of the millionth of a second before it is about to reach its destination by removing the half-silvered mirror at F, it would have to miraculously wipe out having travelled along one of the routes from its past altogether.

It should be noted, as Wheeler (In Wheeler and Zurek [1983], p. 192), and Bell [1987], p. 112 point out, that the orthodox view would get around this problem, by forbidding us to think about it. According to this view, the new physics does not allow any questions to be raised about what happens to the photon in-between its source and detection at either A or B. The cause of the absurdity encountered is hence diagnosed as departing from the path of physics, and embarking on that of [realist] metaphysics which pushes us in the direction of seeking realities where none is to be found. Physical reality, we are told, should be confined to measurement results alone, even if this entails that minds are ultimately essential for determining this reality. Measurement results are observable traces on measuring instruments, and observations are complete only when recorded by the mind of an observer. Since each such result, however, is entirely consistent on its own, and different results arise from different physical situations, the absurdity encountered above would not ensue.

18. The Amplitude Field

In physics, any quantity whose values everywhere depend on locations in the physical space is considered a field. The state function over the space of a particle's location, therefore, describes a field. The alternative which remains before us and which we are going to adopt, is to consider this field, which we shall call the 'amplitude field', as a real entity. This by no means is an option free of problems. Nevertheless, it has several considerable advantages over its rivals. First of all, it provides a natural explanation of all the phenomena we have encountered, without running into the absurdities just considered. Secondly, it avoids revision of the traditional conception of the reality of states, which would either suppress the freedom to be curious, or else bestow attributes on the unobserved states of physical systems, as well as on the role of minds in determining the observed ones, which appear to defy comprehension.

The main obstacle in the way of considering the amplitude field as a real entity is that it remains hidden. And that, not only in the sense of being itself directly unobservable -for there are many unobservable entities in physics whose reality is not doubted by anyone-, but also in the sense of lacking any understood physical attributes. For instance, it turns out that in order for it to be efficacious in producing some observed results in physical space, disturbances in it must be able to travel faster than the speed of light (Bell [1987], pp. 171; 106, 115, d'Espagnat [1979]). Yet, unlike other familiar fields known in physics, the amplitude field appears to lack any energy of its own (Rae [1986], p. 27).

These, however, are problems that when more sharply defined should find their solutions from within physics itself. And that is how it should be. Here, we appear to have a consistent interpretation of QM that refers all the outstanding problems which remain on the way to its adoption back to the domain of physics. This fact alone should suffice to commend it against the rivals which can do no better (as we shall see) than reverting to metaphysics or speculative psychology for offering solutions to their own. There is also an added advantage, for our purposes, that the physical picture provided by this interpretation, highlights clearly the features of the limiting case where the laws of CM meet with empirical success.

According to this interpretation, the particle, with its own position and momentum at any given time, is to be sharply distinguished from the wave function which determines its state at that time. The particle behaves in the age old Classical fashion, as all particles should, in so far as it is always localized in one and only one position, and travels with one and only one velocity at any one time. It can only pass through a single slit at a time, and travel along a single route to any target. It fails, however, to fit into the Classical picture in so far as its state at any one time is not determined by its location and velocity at that time. That is to say, its instantaneous state is compatible with a range of values for location and velocity in physical space, subject to the restriction laid down by the Uncertainty Principle.

The range of locations in physical space which a particle in a given state is allowed to occupy at any time, is determined by the size of the disturbance in the amplitude field at that time. This size , in turn, is determined by Schrödinger's equation and the initial conditions. Disturbances in the amplitude field at any one time are to be thought of as open passage-ways within physical space. Any point in this space is free for the particle to occupy to an extent proportional to the intensity of the field at that point. The greater this intensity at a given point, the more open it would be for the particle to occupy. All the points at which this field intensity is zero, are consequently closed for the particle to occupy.

Physical space according to this picture, therefore, is no longer the void that it was painted to be in CM. Everywhere within that void a particle could freely move about as long as there were no obstacles on its way. In the new picture, however, even in the absence of any obstacles whatsoever, only those locations where the intensity of the amplitude field is non-zero are accessible. The periodic features of the amplitude field are, moreover, preserved when the location space of a single particle is enlarged to the configuration space of a system of particles. They then get transformed into the corresponding momentum space in such a way that the uncertainty relation between the two parameters is always maintained.

Historically, it appears that de Broglie was not the only prominent contributor to the Quantum theory who was led, by the idea that the wave function represents an awkward field, to the discovery of an important feature of Quantum phenomena. Dugas [1955], p. 586, discussing the second paper that Max Born published in 1926 ('Quantenmechanik der Stossvorgänge', in Zeitschrift für Physik, 38), writes: 'For his part, Born suggests a new interpretation which arises out of a remark of Einstein on the relation between photons and wave fields. Einstein said that waves only served to indicate the path to the particle and, in this connection, spoke of a virtual or "phantom" field (*Gespensterfeld*). This field determines the probability that a photon, carrying energy and momentum, should take a certain path; but it does not, itself, have energy or momentum. To Born the part played by the waves of quantum mechanics would, in an analogous way, be that of a pilot (*Führungsfeld*).'

Jammer [1974], (pp. 40-41), confirms this by citing an interview with Born in October 18, 1962 which appears in Archive for the History of Quantum Physics. In addition, he also cites a lecture which Born delivered in 1955, three days before Einstein's death, and which appears under the title 'Albert Einstein und das Lichtquantum' in Die Naturwissenschaften 11, 1955. In this lecture, 'Born declared explicitly that it was fundamentally Einstein's idea which he (Born) applied in 1926 to the interpretation of Schrödinger's wave function.' (Ibid, p. 41).

Max Born's crucial contribution to the Quantum theory was the discovery that the intensity of the wave function at any point over the space of a selected parameter determines the probability that the system's parameter has that value. It is in fact through this very discovery that the wave function can be, albeit indirectly, observed at all. We seem, therefore, to have yet another instance in which the idea of amplitude field has played a powerful heuristic role in a major discovery. That this idea is not, from the point of view of physics, untenable, has been forcefully argued against all its prominent critics by David Bohm and John S. Bell in more recent times, in particular in Bohm [1952], and Bell [1987].

19. Back to the Beam-splitting Experiments

Let us set up the experiment described in **17** so that the photon will be restricted, for simplicity, to move only on the plane of this paper. We mark the directions in which the photon will be travelling by the customary x and y coordinates. The photon starts from its source S and travels along the x-axis to the first half-silvered mirror at O.

There is a probability of ½ that it will go through, as well as get reflected, at the first half-silvered mirror. If it goes through, it would continue travelling along the x-axis until it encounters the full mirror II. It would then be reflected at a 90° angle and hence travel towards F along the y-axis. If, however, it is reflected from the first mirror, it would be deflected by 90° at O and hence proceed along the y-axis until it encounters the full mirror I. Reflection from this mirror would finally send it travelling along the x-axis towards F.

Meanwhile, firing the photon at S starts a disturbance in the amplitude field. The geometrical shape of this disturbance is, as has been noted, determined by the initial as well as the boundary conditions, and will be described by an appropriate wave function ψ. It turns out (Penrose [1989], pp. 314-319), that if the initial conditions are so prepared that in all the repetitions of the experiment the photon starts out with the same momentum, ψ would describe a unique helix winding around the x-axis from S. This wave propagates towards O in such a manner that the sum of the absolute square of its amplitude over a small interval around each and every point between S and O can be normalized to 1.

Encountering the half-silvered mirror at O, the wave splits into two equal disturbances ψ_1 and ψ_2, each with an intensity which is everywhere half that of the original. That is, $\psi_1\psi_1^* = \psi_2\psi_2^* = \frac{1}{2}$, where ψ_1^* and ψ_2^* are the complex conjugates of ψ_1 and ψ_2 respectively. One passes through winding around the x-axis in phase with the original, while the other is reflected winding around the y-axis 90° out of phase with the original. This is illustrated in Figure 9, with arrows indicating the direction in which each helix is winding in a 0° to 360° dial.

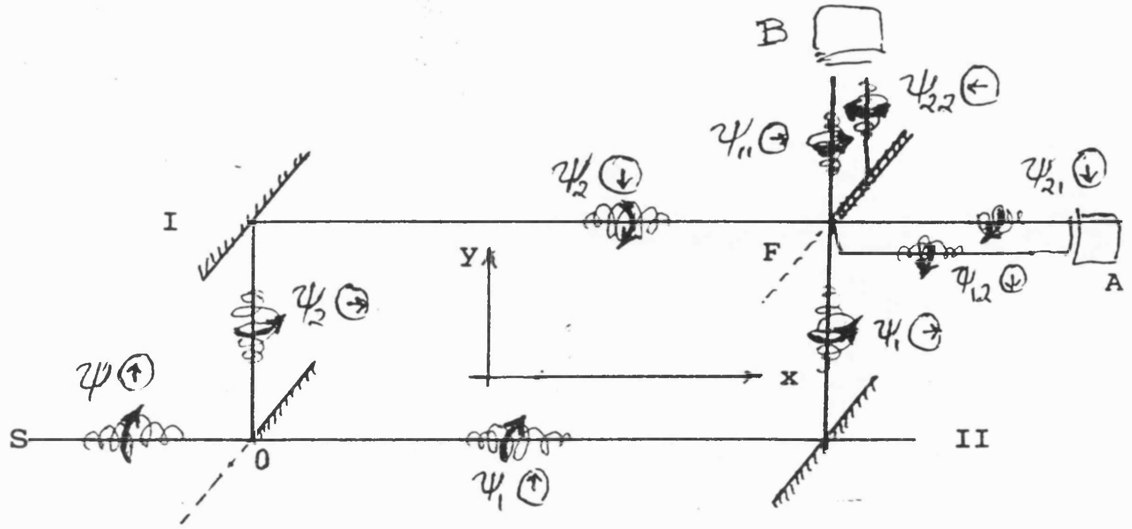


Figure 9

Assuming that nothing other than a mere change of phase happens to either waves as a result of reflection from the full mirrors, ψ_1 and ψ_2 would each undergo a 90° phase shift again, by encountering mirrors I and II respectively. They would, therefore, be approaching the half-silvered mirror at F, 90° out of phase with one another. At F each wave would split into two equal waves: ψ_1 into ψ_{11} which goes through the half-silvered mirror, and ψ_{12} which is reflected from it; and ψ_2 into ψ_{21} and ψ_{22} likewise (see Figure 9). In the same manner as the splitting of ψ at the first half-silvered mirror at O, the intensities of ψ_{11} and ψ_{12} everywhere must now be such that each is half that of ψ_1 That is, in rough notation, $\psi_{11}\psi_{11}^* = \psi_{12}\psi_{12}^* = 1/4$ The same should hold for ψ_{21} and ψ_{22}.

Each reflected wave would be 90° out of phase with the corresponding incident wave approaching F, while the transmitted ones would remain in phase with the latter. As a consequence, ψ_{12} would be propagating in phase with ψ_{21} thus producing a constructive interference at A, while ψ_{11} would be 180° out of phase with ψ_{22} thus producing a destructive interference with it at B.

Now, the probability that the photon is present in the vicinity of any point along its possible routes, is given by the intensity of the amplitude field around that point. We designate this quantity by $|\psi(x)|^2$ for the value of the wave function ψ around the point x, which is another way of writing $\psi(x)\,\psi^*(x)$. If we designate the phase differences between ψ_{12} and ψ_{21} on the one hand, and ψ_{11} and ψ_{22} on the other, by θ_1 and θ_2 respectively, we calculate the probabilities P(A) and P(B) of the photon arriving at A or B respectively, as follows:

$$P(A) = |\psi_{12}(A) + \psi_{21}(A)|^2 = |\psi_{12}(A)|^2 + |\psi_{21}(A)|^2 + 2|\psi_{12}(A)||\psi_{21}(A)| \cos\theta_1 .$$

$$\theta_1 = 0^\circ \Rightarrow \cos\theta_1 = 1 \Rightarrow P(A) = 1/4 + 1/4 + 2\,(1/2)(1/2)\ , \text{ or}$$

$$P(A) = 1/2 + 1/2 = 1 .$$

$$P(B) = |\psi_{11}(B) + \psi_{22}(B)|^2 = |\psi_{11}(B)|^2 + |\psi_{22}(B)|^2 + 2|\psi_{11}(B)||\psi_{22}(B)| \cos\theta_2 .$$

$$\theta_2 = 180^\circ \Rightarrow \cos\theta_2 = -1 \Rightarrow P(B) = 1/4 + 1/4 - 2\,(1/2)(1/2)\ , \text{ or}$$

$$P(B) = 1/2 - 1/2 = 0 .$$

By conceiving the particle and the amplitude field as separate entities, we have accounted for the observed probabilities in a straightforward manner (i.e. requiring no renormalizations), without either resorting to the absurdity of a single photon travelling both routes at once, or ad-hoc metaphysical revisions, as well as epistemological restrictions. All other empirical results obtained from introducing variations to this experiment can likewise be accounted for:

1. If either of the routes after the first half-silvered mirror is blocked by a photon-detecting device, for example, the device would be activated with a probability of ½ in a manner indicative of the passage of a whole photon. The amplitude wave along the open route would propagate unimpeded towards F, splitting into two equal waves as a result of encountering the half-silvered mirror at F. Then the photon should be detected at either A or B with equal probabilities, as observed.

2. If there is no mirror at F and both routes from O remain open, the amplitude waves from I and II would proceed intact to A and B without interfering. The photon should then be detected at either points with equal probabilities, as observed.

3. The delayed-choice about placing a half-silvered mirror at F does not give rise to any problem. Two amplitude waves would be approaching F with or without a mirror there, and the outcomes at A or B would be decided by whether or not they split at F. This, in turn, is decided by the fact that there is, or is not, a half-silvered mirror in place there.

20. Averages and Expectation Values

Although the amplitude field and the Q-system that is associated with it are separate entities, the fact that (a) the intensity of the field at each and every point in the space of a selected parameter of the system determines the probability distribution of the latter's values, and (b) this distribution is unique for all the systems that are prepared in exactly the same way, enables us to 'read off' the state of the system at a given time from what its associated field is doing at that time. Because of this association, the state of the field at a particular time can be thought of as the signature for the state of its associated system at that time. If we keep this point in mind, most of the conceptual problems arising from the current interpretations of QM will disappear, as we shall see.

To begin with, we remind ourselves that in CM the state of a C-system at a given time is uniquely determined by single values of its D-parameters at that time. Because the probability distribution of the values of a selected parameter is unique for identically prepared Q-systems, we can produce the (statistical) analogue of the Classical situation in QM by taking the average of all the values that occur in this distribution. This single number, then, would represent a unique state of a Q-system.

Let us take a large enough stock of Q-particles of the same type which are all prepared to start an experiment in the same way. We ask what the average value of their position is in a completely isolated volume of the physical space, after the lapse of a certain time from when they are released, one by one, into this volume. To measure this value, we would have to take the number of times each particular position in this volume is occupied by a particle after the time in question, multiply the two numbers, add up the results for all such locations, and finally divide by the total number of trials. This is just what we mean by 'taking the average' of all the locations that are possible for the particle to occupy, after the lapse of time in question.

If, however, we happen to have the wave function that is the signature of the state of the particles at the designated time, we can calculate this value through the following reasoning. Let us designate each obtained value of the particle's position q by q_i, the number of times it occurs by N_i, and the total number of trials by N. The average value q_{ave} of the particle's positions can be calculated from $q_{ave} = \sum_{i=1} N_i q_i / N = \sum_{i=1} (N_i/N) q_i$. The quantity N_i/N converges towards the probability P_i that the parameter q has the value q_i, as N is increased towards infinity. This means that for sufficiently large number of trials we can write $q_{ave} = \sum_{i=1} P_i q_i$. If the wave function associated with the particle has the value $\psi(q_i)$ at q_i for the designated time, we can write
$P_i = |\psi(q_i)|^2 = \psi(q_i)\,\psi^*(q_i)$.

Once we approach the limit of an infinite number of trials over a region of space, all locations within that region must be included in our calculation. This means that when we turn to probabilities in order to calculate our average, we must replace the sum with integral over the entire region of space, we are looking at. In general, if we represent the amplitude field that determines the probabilities for all the locations of a Q-particle in the physical space at all times, by a wave function $\Psi(q,t)$, the average of the particle's locations at any state can be calculated from $q_{ave} = \int q \, \Psi(q,t) \, \Psi^*(q,t) dq$, where dq is the abbreviation for the incremental volume dx.dy.dz of physical space. Because of the fact that this average is indicative of a particular state for the particle, it is also called the 'expectation value' of its location in that state.

We note that in the expression for calculating the expectation value of the particle's location, q appears as a variable that takes on all the latter's relevant values. This can be generalized to hold for all the parameters whose values can be measured. We could either switch to the space of the selected parameter, in which case we retain the simple variable over the values of that parameter, but should take care to convert the wave function that describes the amplitude field to its counterpart in that space. Or, alternatively, we could chose to work with the amplitude field itself, in which case we should take care to convert the variable that takes on the values of the selected parameter to its appropriate form in the physical (or more generally, the configuration) space.

21. Operators, Their Eigenvalues and Eigenfunctions

The form that the conversion of the variable over a selected, parameter, other than the one over which the amplitudes are described, takes, depends on the scheme that is used. In the

Schrödinger representation the converted variable becomes an algebraic operation (such as differentiation combined with multiplication) to be performed on the wave function over the physical (or the configuration) space. In the Hilbert space representation, it becomes a linear operation to be carried out on a state vector according to the rules of linear algebra.

In the simplest case of calculating the expectation value of a selected parameter, we are dealing with the multiplication of each and every relevant value of the parameter by its corresponding amplitude. On the other hand, the only measurable quantity which we can calculate using the machinery of QM, is the expectation value of a selected parameter. We can generalize these facts and say that all measurable parameters find their representation in the formalism of QM as certain operations that are to be performed on expressions that describe the behaviour of the amplitudes in a given representation. The formal expressions that specify the precise operations to be performed, are called 'operators'. If A is a measurable parameter, the operator that represents A is usually designated by '$\hat{A}$'.

Earlier on, we labelled the state of a Q-particle as a 'statistical collection' over the space of a selected parameter. This does not mean that there are no states in which certain parameters of the particle retain the same, value through time. It means, rather, that such states are more of an exception than the rule. In general, most parameters of a Q-system in a particular state display values that fluctuate around a mean. There are, however states in which the value of some parameters remains stable at a single number. Most notable among such parameters is the total energy of the system. The states in which the energy of the system retains a single constant value through time are, appropriately called the 'stationary states' of the system. The stationary states of the hydrogen atom, for example, correspond to the various energy levels that we met when discussing Bohr's theory of electrons in atomic orbits.

It follows at once that when a system is in a state in which a measurable parameter retains the same value, the expectation value of that parameter should be identical with that value. Since the expectation value of a parameter is characteristic of a particular state, the single constant value of a parameter that corresponds to one such special state is called the 'characteristic value' or the 'eigenvalue' of the operator that represents the parameter in question, associated with that particular state.

The wave functions over the space of a selected parameter, each of which in the Schrödinger representation determines one of these special states, are correspondingly called the 'eigenfunctions' of the operator representing that parameter. Because in the Hilbert space representation the equivalent of such wave functions are vectors, they constitute the 'eigenvectors' of the operator in question and are called its 'eigenstates'.

22. Eigenvalue Equation and the Superposition Principle

If a Q-system is in one of the eigenstates of a selected parameter, operating on the wave function that is the signature of that state with the operator that represents the parameter in question, results in what is called an 'eigenvalue equation'. Let the operator in question be $\hat{A}$, one of its eigenvalues A_i, and the wave function associated with A_i, ψ. The eigenvalue equation for this particular case would be $\hat{A}\psi = A_i\psi$. This is particularly convenient since it says that operating on the function ψ with the operator in question is equivalent to performing a simple multiplication on ψ with a real number.

Armed with this equation, we can easily prove the proposition that the expectation value of a parameter, when a Q-particle is in an eigenstate of its corresponding operator, is just the eigenvalue of the operator associated with that eigenstate. Let A be the selected parameter, A_i one of its eigenvalues, and $\psi(q,t)$ the eigenfunction of $\hat{A}$ associated with A_i which describes the amplitude field. The expectation value <A> of A can be calculated from

$$\langle A \rangle = \int \hat{A}\,\psi(q,t)\,\psi^*(q,t)\,dq\ ,$$

where dq is to be interpreted as before. From the eigenvalue equation we get

$$\hat{A}\,\psi(q,t) = A_i\,\psi(q,t)\ .$$

Substituting this in the expression for <A>, we get

$$\langle A \rangle = \int A_i\,\psi(q,t)\,\psi^*(q,t)\,dq\ .$$

Since A_i is a constant, it is unaffected by the integration. We can, therefore, write

$$\langle A\rangle = A_i \int \Psi(q,t)\ \Psi^*(q,t)dq\ .$$

The integral must be equal to unity since the particle's parameter q must have some value within the space covered by the integration. It follows that $\langle A\rangle = A_i$.

In the Schrödinger representation, the eigenfunctions of a selected operator over the space of its parameter describe wave motions in that space. Each of these wave motions is analogous to a characteristic mode of an undulatory phenomenon in Classical wave mechanics. Once all the characteristic modes for such a phenomenon are known, 'Any motion at all can be analyzed by assuming that it is the sum of motions of all the different modes, combined with appropriate amplitudes and phases' (Feynman et al [1963], I, 49-3). This is known as the 'principle of superposition' in wave mechanics.

The analogy enables us to carry over the superposition principle to QM. In the Schrödinger representation, this translates into the statement that any wave function at all over the space of a selected parameter can be expressed as the superposition of the eigenfunctions of the operator representing that parameter, using appropriate coefficients and phases. In the Hilbert space representation the application of the same principle amounts to the fact that once all the eigenstates of a selected operator are normalized (which involves choosing such coefficients for each that would make their 'length' equal to unity), they establish a set of bases upon which a (Hilbert) vector space can be set up.

We have already witnessed the principle of superposition at work when discussing the double-slit and beam-splitting experiments. In both these experiments, we were working with the amplitude field whose behaviour at each relevant region of the physical space determined the probabilities for the particle's location there. Here, there are two points worth making explicit. (1) Whenever there was more than one wave motion associated with a single particle, each was elementary, in the sense that it was not itself the resultant of several simpler waves being combined together. Although at every location in their joint or separate travels, the waves determined the *probability* of the particle's presence there, the particle was always found *localized* at a single spot, to the exclusion of all other locations. (2) The state of a single particle at any time must be viewed as being determined by the state of a complex wave motion that resulted, wherever applicable, from superimposing all the elementary waves associated with it.

If we take the crucial property of eigenfunctions as being that every other wave function can be analyzed in terms of their superposition, each elementary wave motion associated with a single particle would qualify for being described by an eigenfunction. To each one of these eigenfunctions there corresponds a particular eigenvalue (in non-degenerate cases these are distinct), which represents a value of the selected parameter to the exclusion of all the others. It follows, therefore, that each value of a selected parameter that is obtained by carrying out a measurement on the particle *in any state*, is an eigenvalue of that parameter (associated with one of its eigenfunctions, barring degeneracies).

23. 'Paradoxes' of the Orthodox Interpretation

In the interpretation of the formalism of QM favoured by most physicists -one which we have already referred to as being the orthodox- the distinction we have been at pains to maintain between what the wave function describes, and the state of the corresponding Q-system is not recognized. The fact that the wave function provides the maximum possible information about the system, plus the fact that the proposed amplitude field is very much unlike any familiar physical entity, have joined to force, in their minds, the only available alternative. That is, to regard the wave function as *nothing but* the maximal description of the state of the system itself.

What, however, was quite straightforward to envisage in the case of the amplitude field, namely, the superposition of waves, becomes, under the orthodox interpretation, something incomprehensible, namely, the 'superposition of states'. The difficulty is perhaps better seen in the light of the following standard distinction. A system is said to be in a 'pure state' if there is a single wave function which describes its state. The system is said to be in a 'mixture' of states if all that is known about it is that it is in one pure state with the probability P_1, another pure state with the probability P_2, and so on. (See, for example, London and Bauer [1939], §4], in Wheeler and Zurek [1983], pp. 233-235).

By superimposing the eigenfunctions of a selected operator, we end up with a single wave function. The state of a system that is defined by the superposition of the eigenfunctions of one of its parameters is, therefore, pure. But what are we to make of such a state?

Take the selected parameter to be the total energy of the system. If the system is in any of the eigenstates of energy E_i, (i = 1, 2, ...), its total energy would have the value E_i characteristic of that state. When the system is in a superposition of these eigenstates, however, we cannot describe its state as having the energy E_1 with the probability P_1, energy E_2 with the probability P_2, and so on. Each of the values E_i, (i = 1, 2, ...), being associated with a single eigenstate of the energy, *that* would amount to describing the state of the system as a mixture. And this cannot be because the system is, by definition, in a pure state.

The problem gets deeper once we note that even when ideally the most accurate measurements are carried out on the energy of the system, these probabilities, and only these probabilities, are the results observed. It looks as if the state consisting of the superposition of eigenstates cannot be identified by means of the best possible mass of observational data. Yet such states are by no means rare for Q-systems to find themselves in. Wherever there are uncertainties in the value of a parameter, and this is most generally the case, superpositions of the eigenstates of that parameter are involved.

24. The Problem of Measurement

A sure sign that a conception has outlived its usefulness is when the solutions it generates in order to overcome the problems it has run into prove even more enigmatic. This holds for the orthodox interpretation vis à vis the problem of the superposition of states. The way out, first articulated in von Neumann [1932],(chapters V and VI; also in Wheeler and Zurek [1983], pp. 549-623), and still widely held, consists of postulating that Q-systems evolve differently when left unobserved from when they are brought under observation. The first evolution is governed by a strict law, expressed by Schrödinger's equation, and proceeds from pure states to pure states; the second evolution is statistical, proceeding from either pure or mixture of states to a mixture of states.

Von Neumann writes: 'We... have two fundamentally different types of interventions which can occur in a system... First, the arbitrary changes by measurements... Second, the automatic changes which occur with passage of time...' (in

Wheeler and Zurek [1983], p. 553). And, '[the second] transforms states into states... while [the first] can transform states into mixtures. In this sense, therefore, the development of a state according to [the first] is statistical, while according to [the second] it is causal.' (Ibid, p. 559)

As a result, von Neumann concludes, 'we must always divide the world into two parts, the one being the observed system, the other the observer. In the former, we can follow up all physical processes (in principle at least) arbitrarily precisely. In the latter, this is meaningless... Now quantum mechanics describes the events which occur in the observed portions of the world, so long as they do not interact with the observing portion, with the aid of the [second] process, but as soon as such an interaction occurs, i.e., a measurement, it requires the application of [the first] process.' (Ibid, p. 622).

This, *prima facie*, does not appear to amount to much more than a paraphrase of the original problem. Indeed, similar qualitative proposals were already around, made by people like Bohr and Heisenberg. What made von Neumann's contribution particularly distinguished, was his genius to couch this idea within the same mathematical machinery that was used to formalize QM. In other words, he managed to produce a theory, using the formalism of QM, which described the process of measurement involving Q-systems, and which entailed that their state should turn into a mixture as a result of interacting with measuring devices on which they leave observable traces.

Crucial to this theory is the assumption that the wave function (or its equivalent in the Hilbert space representation) is the Quantum Mechanical version of the description of states for all physical systems. From this assumption, von Neumann proceeded along the following line of reasoning: 'In measurement we cannot observe the system S by itself, but must rather investigate the system S + M, in order to obtain (numerically) its interaction with the measuring apparatus M. The theory of the measurement is a statement concerning S + M, and should describe how the state of S is related to certain properties of the state of M (namely, the positions of a certain pointer, since the observer reads these). Moreover, it is rather arbitrary whether or not one includes the observer in M, and replaces the relation between the S state and the pointer positions in M by the relations of this state and the chemical changes in the observer's eye or even in his brain (i.e., to that which he has "seen" or "perceived")... In any case, therefore, the application of [the second process] is of importance only for S + M. Of course, we must show that this gives the same result for S as the direct application of [the first process] on S. If this is successful, then we have achieved a unified way of looking at the physical world on a quantum mechanical basis.' (Ibid, p. 554).

The success of von Neumann's efforts, which we take for granted for our purposes, means, in particular, that the idea of the superposition of states can be reconciled (to use Putnam's word in Putnam [1983], p. 250) with the fact that even ideal observations on Q-systems cannot ever reveal such states. The problem, then, of making sense of this idea appears to get defused, so far as at least Q-systems are concerned, since their time-evolution in isolation must, by necessity, be disrupted if any observation of their state is to be attempted at all. We don't *have to* worry about what their states look like when they evolve undisturbed, one could say, so long as we have a good enough unified theory which, in conjunction with appropriate initial conditions, entails the observation results actually obtained.

This, however, proves to be a short-lived respite. The reason, as we shall see, has to do with the fact that the reconciliation between the observed and unobserved states of the Q-systems has been achieved by extending the idea of superposition of states to all kinds of systems. Devices from which results of measurement on a parameter of a Q-system must ultimately be read off, would have to include C-systems. Von Neumann's theory manages to get the desired mixture for the Q-system *observed*, by viewing the combined Q and C-systems as a single system undergoing undisturbed evolution in time (according to the time-dependent Schrödinger equation).

So long as all the systems involved, the theory suggests, remain coupled in complete isolation, the states of the whole, conceived as a single unit, goes through superpositions. As soon as an observer intervenes to read off the result of the measurement on the C-part of the combined system, the coupling breaks down. Consequently, the smooth evolution of the unit as a whole comes abruptly to a halt, and the states of each de-coupled system on its own, embark on a new kind of evolution. Finally, this new evolution leaves the subsequent states of the separated systems only statistically correlated.

Although this theory ensures that the awkward states of superposition remain so well hidden that they need not bother us so long as we restrict ourselves to observables -and at least according to one school of thought, *that* is all that physics should be concerned about, it has achieved this success at the cost of contaminating C-systems with such states. The problem is that *their* state, either in complete isolation or in interaction with any other system, is successfully envisaged as a definitive collection, which although compatible with being a mixture at any one time, cannot be reconciled with being a superposition at any time. There may be uncertainties as to what the state of a C-system is at a particular time, but to ascribe superpositions to these uncertainties 'seems simply wrong', to use Schrödinger's description in Schrödinger [1935] (also in Wheeler and Zurek [1983], p. 156).

25. Schrödinger's Cat and Wigner's Friend

To highlight the gravity of the problem, Schrödinger [1935] proposed a gruesome experiment. 'A cat is penned up in a steel chamber, along with the following diabolical device (which must be secured against direct interference by the cat): in a Geiger counter there is a tiny bit of radioactive substance, *so* small, that *perhaps* in the course of one hour one of the atoms decays, but also, with equal probability, perhaps none; if it happens, the counter tube discharges and through a relay releases a hammer which shatters a small flask of hydrocyanic acid. If one has left this entire system to itself for an hour, one would say that the cat still lives *if* meanwhile no atom has decayed. The first atomic decay would have poisoned it. The ψ-function of the entire system would express this by having in it the living and the dead cat (pardon the expression) mixed or smeared out in equal parts.' (In Wheeler and Zurek [1983], p. 157)

Now, the 'smeared out' state of being alive *and* dead at the same time, is quite inconceivable for a single cat to be in, with or without being coupled to an atom. However, to those who find inconceivability suspect as sufficient grounds for questioning the merits of an idea, the follow up must surely encourage a good step in that very direction. For the cat that is *both alive and dead* 'in equal parts' while hidden from anybody's view, is instantaneously converted into a cat that is either very much alive, or else very much dead, as a result of an observer merely *casting an eye on it*! This, in the case of the cat and other as yet unseen animals, may amount to bestowing divine-like powers over life and death on mere mortal observers, but its implications go much further than this.

Let us take the view, very much uncontroversial, that observers emerged during a period in the time-evolution of the universe. Since there were no observers intervening with its evolution prior to this period, the states of the whole universe, viewed as a single complex physical system, must, according to von Neumann's theory, have been evolving undisturbed. This process of evolution must, therefore, have been of the second type, unfolding strictly according to the law laid down by the Schrödinger equation, from superposition to superposition of states.

'Events' as we understand the term, would have no place in this universe, since nothing is ever fixed one way or the other in the course of its evolution. We would, at best, have to describe the world up to this time, using Jammer's words (from a different but related context), as being 'a universe of evolving potentialities...but not of real events.' Jammer [1974], p. 474.

The emergence of observers, however, has, as its consequence, the gratuitous introduction of the first type of evolution into the world. Without any effort being required on their part, as observers merely multiply and turn their attentions around, definite events begin to take shape and proliferate.

When Putnam [1983], p. 251, wrote, 'On von Neumann's view there is a dependence of the truth upon one's perspective; there is no master truth', he was giving us only a part of the picture. One, by mere virtue of observing, would, according to this point of view, have to be engaged in *world-making*!

To complete the picture, the final master stroke is provided by Wigner [1961], in the shape of yet another ingenious experiment. The set-up includes an atom, similar to that in Schrödinger's experiment, only now its decay is made to produce an innocuous, but visible flash of light from a device that is secured to detect it.

'It is natural to inquire about the situation if one does not make the observation oneself but lets someone else carry it out. What is the wave function if my friend looked at the place where the flash might show at time t? The answer is that the information available about the *object* cannot be described by a wave function. One could attribute a wave function to the joint system: friend plus object, and this joint system would have a wave function also after the interaction, that is, after my friend has looked. I can then enter into interaction with this joint system by asking my friend whether he saw a flash.' (In Wheeler and Zurek [1983], p. 173)

Wigner deploys a reductio regarding the states of the joint system. Let us assume, to begin with, that the states of the joint system can be accounted for by QM à la von Neumann. Then, so long as the joint system remains completely isolated from the rest of the world, its states should evolve as the superposition of the following eigenstates: the device flashing and the friend seeing a flash, *and* the device not flashing and the friend seeing no flash (we assume there is no other alternative). Only when the isolation of the joint system is broken by an outside observer who enters into communication with the friend, asking him whether or not he saw a flash, is the superposition supposed to give way to either one of the definite states of the friend seeing a flash or not.

'However, if after having completed the whole experiment I ask my friend, "What did you feel about the flash before I asked you?" he will answer, "I told you already, I did [did not] see a flash", as the case may be. In other words, the question whether he did or did not see the flash was already decided in his mind, before I asked him.' (Ibid, p. 176)

This outcome is in contradiction with what the theory says should be the case. Since von Neumann's theory is quite successful in all the cases where the system involved does not *include* observers, Wigner concludes that it must stop short in cases 'where consciousness plays a role' (Ibid, p. 178). In other words, von Neumann's formulation of QM manages to unify the laws of motion governing the evolution of Q-systems' states with the measurement processes that must be employed to test them, at the cost of introducing to physics entities whose behaviour must remain outside the latter's applicability altogether.

26. Back to the Wave Function

We started off with an interpretation of the wave function which was based on the assumption that it *is* the Quantum Mechanical equivalent of the description of physical systems in CM. Since, as such, it ought to apply to all physical systems, including those that are combinations of Q and C-systems, we find that this can be secured only by dividing physical processes into two fundamentally different types.

The first (to change the order imposed by von Neumann), is allowed to enjoy the time-honoured characteristic of unfolding smoothly according to a strict law, but must pay the price of escaping comprehension by any stretch of the imagination. The reason for this is that the notion of superposition, which is perfectly understandable in the area where it originated, is forced to apply to something which appears totally alien to it. In the case of Q-systems, the sting of incomprehensibility may feel less biting at first, because of the fact that they are next to impossible to observe on their own. But the advice that it can be cured by barring curiosity from domains to which direct observations cannot be extended, proves ineffective. Superposition of states as applied to Q-systems, is soon found, by the implications of the theory which postulated it, to contaminate the states of C-systems as well, and here the application appears to verge on a category mistake.

The second type of process conforms admirably to the statistical regularities that are the best we can obtain when we make measurements on selected parameters of Q-systems. This, however, must pay the price of being introduced into the physics of the strange world that is supposedly determined by the first, as a result of perceptions that are spontaneously formed in the minds of observers. Exactly how these perceptions are formed, or through what procedures they come to transform what goes on in accordance with the first type of process into observed events, must remain a mystery that is, eventually, pushed well outside the reach of physics, and perhaps science as a whole, altogether.

What we end up with, therefore, as a result of taking on board the orthodox interpretation of the wave function, is a host of problems far more numerous, and more serious, than those for the solution of which it was conceived in the first place. It by-passes an entity with properties that do not appear to tally with our current conceptions in physics, only to run into a quagmire of entities, attributes and processes *all of which* seem to be grossly out of line with anything we know. At least with the former we stand a chance of improving our position since it remains a physical phenomenon, and no door has been shut on the possibilities of carrying out further research on it. With the latter, on the other hand, the theory ensures that no further probe into their mysteries can be made apart from speculations that cannot be independently tested.

All this adds up to indicate that the choice of this interpretation is unduly costly. One is also tempted to say that it is methodologically questionable as well. The only reason for hesitation would be the fact that untestable speculative hypotheses abound in physics. It is, we are told by methodologists, whether a theory, as a *package* of hypotheses which may or may not be individually testable, is itself testable, that matters. But it must be noted that the legitimacy of the place of untestable hypotheses in testable packages depends entirely on whether or not they have a *constructive* role to play, either in the articulation of a formalism or else in facilitating the comprehension of one.

None of these purposes are fulfilled by the untestable hypotheses which the orthodox interpretation of the wave function entail. An articulated formalism for QM was already in place *before* they appeared, and far from facilitating the former's comprehension, they have managed to shroud much of how it works in an impenetrable obscurity. They appear, therefore, as purely idle excess baggage which must be tacked on to physics solely as a result of sustaining the orthodox interpretation against *serious difficulties*. They are even barred from having any bearing at all on any area of science, and often end up more expedient as propaganda material for esoteric schools of thought.

In contrast, the alternative interpretation of the wave function that we have opted for does not appear to suffer from problems in the same scale. According to this interpretation, the wave function is *identified* neither with Q-systems themselves, nor with their state description. It is, rather, supposed to describe a separate entity, a field, whose intensities everywhere in the physical space determine distributions for the values of the dynamical parameters of the associated Q-system. Because this distribution is unique for each preparation procedure that Q-systems of a certain kind may be subjected to, the values of a wave function can be taken to single out particular states for them. This is in line with the tradition of identifying the state of a physical system by the values of its dynamical parameters, and is the reason why we have chosen to describe the wave function as the *signature* of a particular state.

The problem associated with the conception of this field, as has been noted, is that it lacks characteristics other known fields possess, and it can do things which in the light of what it lacks appear inconceivable. Even though it does not store energy itself, its ripples can move in space with phenomenal speeds. Their frequency, in turn, appears to be related to the energy of the associated particle in an unprecedented way. Even though the latter clearly lacks enough Classical energy to possibly reach certain regions in physical space, the very fact that this field extends over those regions enables the particle reach them (a phenomenon known as the 'tunnelling effect').

'Inconceivable' attributes such as these, however, are no strangers to physics. The very ideas that empty space stores energy and that ripples of energy can propagate as waves without requiring any material medium, exemplify two that were once thought to be inconceivable. Likewise, entities such as particles that have no 'mass' and travel with the same speed in all frames of reference were, in one day, considered just as inconceivable. The reason why they manage to force their way into physics, despite the fact that nobody liked them, is that *with their help* notable successes are achieved, one way or the other, by this science.

The same can be said about *our* 'inconceivable' field. We have seen that at least in the cases of two major contributions to the development of QM, de Broglie's and Born's, it had a considerable heuristic role to play. We have also seen that in the cases of the double-slit and the beam-splitting experiments (with or without delayed-choice), the conception of this field facilitates comprehension in all the areas where its rival runs either into absurdity or obscurity. Moreover, puzzles such as those of Schrödinger's cat and Wigner's friend simply do not arise, because the idea of superposition of states plays no part in this conception. In all, this conception is methodologically fertile while its rival is sterile, and it is illuminating while its rival is obscurantist.

27. Amplitude Field and Measurement

Once we have overcome the problem of conceiving the amplitude field, we can account for the process of measurement in a straightforward manner. Let us concentrate on position measurements only. This does not detract from generality since all Quantum Mechanical measurements ultimately depend on tracking traces which Q-systems leave at various locations on measuring instruments. Even a spin measurement, as in the Stern-Gerlach experiment, 'despite all the talk about 'spin', is finally about position observations' (Bell [1987], p. 162). The measuring instrument, or at any rate the device on which we finally observe traces of a Q-system, is, in turn, always a C-system.

The places where the latter comes into contact with the former, are generally determined by the behaviour of the amplitude field in the region. If the circumstances surrounding a measurement are such that from the initial to the final stages of the experiment the intensity of the amplitude field associated with a Q-system remains everywhere confined to a compact region of space, the Q-system would be in an eigenstate of its location. The progression of this intensity in time would, in turn, be determined by a wave function which, strange as it may look, is the sole solution to the Schrödinger equation appropriate for this case. This function describes a passage way opening up in physical space which looks like a narrow tube, and inside which the motion of the system is confined. Only at places where *this 'tube'* is cut or obstructed with a suitable device, would a trace of the system be observed.

If, on the other hand, the circumstances are such that there is more than one solution to the Schrödinger equation which is appropriate for the case, the intensity of the associated field everywhere would be determined by the superposition of several waves. The Q-system in motion under these circumstances would, therefore, face the passage ways that open up in the physical space, as a result of this superposition. There might be only one, as in the case of the beam-splitting experiment with the final half-silvered mirror in place. Or, there might be many with varying degrees of permeability, as in the case of the double-slit experiment. The degree of permeability of each passage way, in turn, is determined by the strength of the field which results from the interference of all the waves that are set into motion in a given circumstance.

If the instantaneous intensity of the amplitude field everywhere could be measured independently for a given situation, the state of its associated Q-system could be decided at once from this measurement. This possibility, however, must be ruled out since, so far as we know, the amplitude field cannot exchange energy with other systems, something that would be necessary if we were to obtain a measurement result on its intensity. Instead, the state of the Q-systems must be decided *indirectly*, by measuring the probabilities that the intensities of this field everywhere determine for its possible locations. This entails, as was pointed out in **13**, that the state of each Q-system, a trace for which has been observed at a particular location, is a mixture. This mixture would, in turn, be resolved only when the results from the same location measurement on an indefinitely large number of the same type of systems, all identically prepared, are obtained. *It is only then* that we would be in a position to have a picture of the intensity of the amplitude field at hand which is pertinent to the circumstances of a particular measurement situation on a Q-system.

We sum up: *States* do not superimpose, but amplitude waves, like all waves, do. In each particular circumstance, set up to obtain some traces of a Q-system, the behaviour of these waves is strictly determined by the Schrödinger equation that is written for that circumstance. Superpositions which must be allowed because of the linearity of this equation, are, therefore, quite naturally (thanks to Born) reconciled with the results obtained from measurements on Q-systems, without any need to invoke supra-physical interventions. Finally, the strange properties that must be accepted for the amplitude field itself, pose a problem which can find its resolution only from further advances in physics. Either it would turn out to be analogous to numerous other similar conceptions where conceptual adjustments have had to be made by working physicists, or else deeper principles would be discovered that can explain these properties and their effects on other things.

28. The Limiting Case Problem

Now we consider the vexed question of whether this account of QM complies with the demands of the Generalized Correspondence Principle. According to the atomic conception, Q-systems are the very building blocks with which C-systems are ultimately put together. C-systems, on the other hand, find an adequate account of their behaviour in CM. QM is a theory which has to date accounted spectacularly successfully for the behaviour of Q-systems. If it turns out that this theory is unable to meet the demand of the Generalized Correspondence Principle, then we will be faced with the following dilemma: Either QM is a transitory theory in the history of physics and will be replaced with a comprehensive theory that does meet this demand. Or else the demand made on all theories in physics by the Generalized Correspondence Principle is unwarranted. For QM is not only methodologically sound but empirically successful in the domain where CM is known to have failed, and the Generalized Correspondence Principle requires that *any* replacement of CM must have *it* as its limiting case.

Bohm and Hiley [1993], (p. 160), claim that Bohm's ontological interpretation of Quantum formalism is able to meet this demand quite naturally (they present this capability as yet another advantage of this interpretation over its rivals). In this interpretation, what we have called the *amplitude field* is regarded as somewhat analogous to a potential with non-Classical characteristics which is called the 'quantum potential'. With the aid of this potential and using a method known as WKB approximation, they rewrite Schrödinger's equation in a form which yields the Classical Hamilton-Jacobi equation for the motion of a particle when the quantum potential is negligible compared to the kinetic energy of the particle (ibid, p. 161). They rule out at the outset the viability of the idea that the Classical limit will be obtained when the Planck's constant in the Schrödinger equation is equated with zero (ibid, p. 161).

The general claim in Bohm and Hiley [1993] is that in the domains where CM is empirically adequate, the effects of this potential are not at all pronounced. They manifest themselves in the domain of the sub-atomic entities, and this is where QM can deal with them adequately. This indeed provides a suitable framework within which there is hope for finding an acceptable solution to the limiting case problem in QM. What remains to be done is to find the conditions under which the quantum potential can be justifiably neglected. The simplest and most natural expectation that suggests itself in this regard is that as the size of a collection of sub-atomic particles in a fixed region of physical space grows and hence tends to become a macro-system dealt with adequately by CM, the quantum potential would assume values tending to become so small as to be negligible. Interestingly, however, it is shown in Bohm and Hiley [1993], (p. 167), that this is not at all the case; the quantum potential is simply not a function of the number of particles that comprise a macro-entity. Bohm and Hiley are consequently forced to find the conditions under which the quantum potential may become reasonably negligible by resorting to some other devices.

They find this limiting condition, through a rather sophisticated argument, in the interaction between a macro-entity and the electromagnetic radiation to which, they contend, every entity in the physical world is actually subjected (they make a similar case for the interaction between a macro-entity and a stream of sub-atomic particles (ibid, pp. 172 & 173)). To get the desired result in this way, however, they need two essential ingredients each of which is open to challenge. First, is an assumption about the form of the wave function for the sub-atomic particles at the moment of the creation of the universe. They chose two forms for this wave function (ibid, p. 166f) and base their entire case upon them, thus exposing themselves to an obvious charge of ad hocness. Second, is an assumption about the way the supposed interaction must take place in order to render the value of the resulting quantum potential negligible. For this purpose, it is essential to their case that the source of the radiation falling on a macro-entity be so placed in relation to the latter that a distinct 'shadow' is formed beyond the object in the direction of the radiation's propagation (ibid, pp. 168-170, also pp. 267-269). Only under such conditions are they able to argue that (1) the probability of the destruction of interference between the wave functions of the particles comprising the macro-entity and those of the quanta of radiation present *inside* the shadow increases as the number of such particles increases, thus (2) diminishing the value of the quantum potential accordingly. But obviously in real life there is no guarantee that the relation between the macro-entities and the sources of electomagnetic radiation in the universe remains so fixed that in each case *the right shadow* is thereby produced. Any macro-entity may be exposed to many different sources of radiation, each placed at different and varying locations in relation to it so that indeed no shadow is produced in any direction. On this account, such an entity would cease being a 'macro-entity'.

Let us go back to my account of the amplitude field and try to explore the simple expectation, alluded to above, that as the number of sub-atomic particles increases to form larger and larger congregation in a region of space, the probability that the resulting system would tend to behave as a C-system increases. A striking feature in QM has been the fact that probabilities do not add up in the simple way in which they add in CM in any situations where probabilities arise, and CM is perfectly adequate to account for the behaviour of C-systems. A characteristic feature of the equations of motion in CM is that they are non-linear. That is to say, if there is more than one solution to these equations at any one time, their linear combination is not a solution. The Schrödinger equation (which is the Quantum Mechanical equivalent of the equations of motion in CM), on the other hand, is linear. We are, therefore, led to the following problem: how can the non-linear behaviour of the C-systems be reconciled with the linear behaviour of their ultimate constituents? This, like the case involving the problem of measurement, is the crux of the limiting case problem for QM.

According to our interpretation, the linear behaviour of Q-systems is entirely due to the fine-structure of the physical space in which they move. We have identified this fine-structure with variations in the intensities of the amplitude field associated with each Q-system in motion. Roughly, as the number of Q-systems that join together to form a larger system in a small region of the physical space increases, two events start happening: (1) the number of amplitude waves propagating in the direction of motion of the joint system increases, and (2) the volume that is occupied by the combined Q-systems in this region also increases.

From the point of view of the mathematics, the first increase translates into an increase in the number of 'interference terms' in regions relevant to the motion of the enlarged system in a given circumstance. We recall that the intensity of a wave function, which results at a given point from the combination of the intensities of two other such functions, is not a simple sum of those intensities. The term that must be added on to yield the correct result, includes the cosine of the phase difference between the two functions at that point. If I_1 is the intensity of the first wave at a given point, I_2 that of the second, and θ the phase difference between the waves, the combined intensities I_{12} at that point is, $I_{12} = I_1 + I_2 + 2(\sqrt{I_1 I_2}) \cos\theta$.

As θ varies between 0° and 180°, $\cos\theta$ varies between 1 and -1 accordingly, being 0 at $\theta = 90°$. So, the average of $\cos\theta$ over one period is zero. We note that the significant part in each interference term is the contribution made by the cosine of the phase difference between the waves. If, therefore, as the number of waves increases, the phase differences involved get nearer and nearer to covering all the values between 0° and 180° randomly, the sum of all the interference terms should tend towards zero. This would be the case only if an ever greater number of Q-particles that come together to form an enlarged system, are in an ever greater variety of different states. We have seen that if they are all prepared to be in the same state, the interference terms are very much visible in the outcome.

Generally, it is not true that all Q-particles found in nature would have to be in different states if they come together to occupy a region of the physical space. However, a generic type known as Fermi particles, have precisely this property. They obey what is known as Fermi's 'exclusion principle' according to which it is impossible for Fermi particles in the same state to cohabit in the close vicinity of one another. They can share a small region of space only by being in different states. As it happens, all Q-particles that serve as building blocks in the structure of C-systems such as billiard balls, for example, i.e. electrons, protons and neutrons, are Fermi particles. It is, in fact, by virtue of being Fermi particles that they are able to form stable bonds which are necessary to hold them together into structures with an ever increasing size. The net result of billions of Fermi particles congregating in a small region of physical space to form a single joint system is, therefore, that the interference terms in the combination of the intensities of their associated amplitude waves vanish.

Disappearance of the interference terms explains why the behaviour of C-systems, which are themselves made up of an enormous number of Fermi particles, can be accounted for adequately by the non-linear equations of motion in CM. In turn, as the dimensions of the system formed by the congregation of Fermi particles increase, the relative significance of Planck's constant tends to decrease. In the limit, this number becomes entirely negligible, allowing us to treat the location and momentum of the enlarged system as commuting parameters. As a result, these parameters appear to admit dispersion free values simultaneously, thus allowing us to regard them as D-parameters for the enlarged system.

All this ties up nicely with at least one account of how Schrödinger was able to arrive at his famous equation. According to Wigner [1983], 'Schrödinger derived the wave equation named after him by viewing the classical Hamilton-Jacobi equation as giving an incomplete and approximate description of this wave: incomplete, in the sense that it deals only with the phase... of this wave, $\psi(x, y, z, t)$; approximate, in the sense that the equation for the Hamilton-Jacobi function... is non-linear, whereas by demanding linearity Schrödinger got the right equation for ψ.' (In Wheeler and Zurek [1983], p. 290)

Electromagnetic phenomena, however, do not, in general, appear to fit the account given above. Here, systems of an enormous number of photons appear to behave as linearly as a single photon. In fact until the discovery of photons, large scale electromagnetic phenomena were adequately accounted for by Maxwell's equations which are linear. Feynman et al [1963], vol. III, p. 21-6, explains this difference as follows: 'The wave function $\psi(\mathbf{r})$ for an electron in an atom does not, then, describe a smeared-out electron with a smooth charge density. The electron is either here, or there, or somewhere else, but wherever it is, it is a point charge. On the other hand, think of a situation in which there are an enormous number of particles in exactly the same state, a very large number of them with exactly the same wave function. Then what? One of them is here and one of them is there, and the probability of finding any one of them at a given place is proportional to $\psi\psi^*$. But since there are so many particles, if I look in any volume *dx dy dz* I will generally find a number close to $\psi\psi^* dx\, dy\, dz$. So in a situation in which ψ is the wave function for each of an enormous number of particles which are all in the same state, $\psi\psi^*$, *can* be interpreted as the density of the particles...

'Something similar can happen with neutral particles. When we have the wave function for a single photon, it is the amplitude to find a photon somewhere... [T]here is an equation for the photon wave function analogous to the Schrödinger equation for the electron. The photon equation is just the same as Maxwell's equation for the electromagnetic field, and the wave function is the same as the vector potential **A**. The wave function turns out to be just the vector potential. The quantum physics is the same thing as the classical physics because photons are non-interacting Bose particles and many of them can be in the same state - as you know, they *like* to be in the same state. The moment that you have billions in the same state (that is, in the same electromagnetic wave), you can measure the wave function, which is the vector potential, directly. Of course, it worked historically the other way...'

We must note, however, that there are some states for photons such that when photons, either singly or en masse, are found in these states they behave in ways that cannot be accounted for by classical electromagnetic theories. But there are also photon-states in which the photons behave in a classically linear manner (i.e. beams of photons in such states are capable of producing such diffraction patterns as, for example, in Young's interference experiment). These states are called 'coherent' states and are distinguished by particular phase properties for the photons. All electromagnetic phenomena that are adequately accounted for by Maxwell's theory in particular, involve photons in coherent states.

Feynman's description above applies to photons in coherent states. Bose particles, which photons are a kind of, have the opposite property to their Fermi counterparts. 'It is a property of Bose particles that if there is already one particle in a condition of some kind, the *probability* of getting a second one in the same condition is twice as great as it would be if the first one were not already there. This fact is often stated in the following way: If there is already one Bose particle in a given state, the amplitude for putting an identical one on top of it is $\sqrt{2}$ greater than if it weren't there.' (Feynman et al [1963], vol. III, p. 4-6). Once there is a photon in a region of space which is in a coherent state, the probability that others in the same region would be in the same state increases considerably. A large scale congregation of photons all of which are in the same coherent state is, therefore, highly probable once there are some already in that state in the region. The behaviour of such a congregation, unlike a similar one of Fermi particles, would thus be linear.

As mentioned before, Quantum Mechanics as yet has been unable to account for physical phenomena involving gravitation. This fact has prompted some physicists (Penrose, for example) to regard it as essentially an incomplete theory and to call for its replacement with a new kind of theory which would have this capability. Whether or not this will happen is a point I do not wish to speculate on. I have argued against some positions which claim, on the basis of the existence of a conceptual gap between Quantum and Classical Mechanics, that the Correspondence Principle cannot hold between them. I have proposed a conjecture according to which QM, in its present state, appears to entail that the Hamilton-Jacobi formulation of CM, as well as Maxwell's electromagnetic theory, are good approximations over the domain of phenomena in which they both proved empirically adequate.

CONCLUSION

In conclusion I present a summary of the main points conjectured and discussed in the preceding chapters:

1. The Correspondence Principle as envisaged and deployed by Bohr in his classic paper [1913], is Generalized and states that any theory T_{i+1}, $(i = 1,2,...)$, which is to supersede an outstanding, but falsified, theory T_i, must, among other things, entail that T_i is a good approximation over a domain in which it was established as empirically adequate.

2. The conception of the Correspondence Principle which appears to be promising as a Generalized intertheory constraint is the approximational conception.

3. Apart from its heuristic efficacy in the development of a new theory, the Generalized Correspondence Principle is necessary for a consistent assessment of outstanding theories in physics in accordance with the outlook of the two-tier realism. This outlook requires the choice of outstanding theories in physics to be based on the progress they have made in the direction of capturing the underlying physical constitution of the world.

4. The constraint imposed by the Correspondence Principle on outstanding theories, defines the Correspondence Relation among successive theories in one or more branches of physics. If the theory T_i is a member of the sequence ordered by this relation which immediately precedes the theory T_{i+1}, then T_i is a limiting case of T_{i+1}.

5. The Correspondence Relation is transitive. Assuming that the sequence of theories ordered by this relation converges to a theory T which mirrors the underlying physical constitution of the world, each member of this sequence that precedes T would be its limiting case. T can be taken to represent the entire truth which is of interest to physics. Therefore, the position a theory occupies in this sequence provides an indication of its relative proximity to the truth. The greater the number of limiting cases that a theory has in this sequence, the closer it is to the truth relative to all its other members.

6. The property of being a limiting case of any future theory that may supersede a new theory T_i is decidable in a finite period of time. At a time t if T_i is the only theory for which this property has been decided, and T_i entails that the last falsified member of the sequence ordered by the Correspondence Relation is a good approximation, it is decided that T_i is the closest available theory to the truth at t.

The imposition of the constraint demanded by the Generalized Correspondence Principle in physics on theories that are methodologically sound, as well as empirically adequate, forms the truth-likeness sequence. This sequence, in turn, can serve to provide two-tier realism with an illusive and much needed criterion of theory choice. The satisfaction of a curiosity about what the world is like, is the predominant motivation for a realist of this type in turning to science.

It is, therefore, imperative that his choice of a theory should be grounded in relative advances towards the discovery of the truth.

Attempts to seek indications of truth-likeness in mere successes a theory has managed to achieve, have met with failure because they proved unable to find a link between them and the truth. Adding the requirement demanded by the Generalized Correspondence Principle to the demand for these successes, coupled with the assumption that the ultimate theory T exists, provides us with this very link. Each successive theory in the truth-likeness sequence constitutes one additional step in closing the gap to T, and therefore to the truth. It thus represents a further advance towards the discovery of the truth compared to all those that come before it.

References

Balzar, W., Pearce, D. A. and Schmidt H. (eds.) [1984] *Reduction in Science: Structure, Examples, Philosophical Problems;* D. Reidel Publishing Co., Dordrecht.

Bell, J. S. [1987] *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, Cambridge.

Bohm, D. [1952] A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables, *Physical Review*, 85, in Wheeler and Zurek [1983], pp. 369-396.

Bohm, D. and Hiley, B. J. [1993] *The Undivided Universe*, Routledge, London and New York.

Bohr, N. [1913] On the Constitution of Atoms and Molecules, *Philosophical Magazine*, July 1913, pp. 1-25.

Bohr, N. [1918] On the Quantum Theory of Line-spectra, in van der Waerden [1967], pp. 95-137.

de Broglie, L. [1956] Translated by A. J. Knodel and J. C. Miller as *Non-Linear Wave Mechanics, A Causal Interpretation*, Elsevier Publishing Co. 1960, Amsterdam.

Causey, R. L. [1977] *Unity of Science*, D. Reidel Publishing Co., Dordrecht.

Dirac, P. A. M. [1930, 1958] *The Principles of Quantum Mechanics*, Oxford University Press, Oxford.

Dreyer, J. L. E. [1953] *A History of Astronomy from Thales to Kepler*, Dover Publications, New York.

Dugas, R. [1955] Translated by J. R. Maddox as *A History of Mechanics*, Dover Publishing Co. 1988, New York.

d'Espagnat, B. [1979] The Quantum Theory and Reality, *Scientific American*, Nov. 1979, pp. 128-140.

Fadner, W. L. [1985] Theoretical Support for the Generalized Correspondence Principle, *American Journal of Physics*, 53 (1985), pp. 829-838.

Fay, J. [1991] *Niels Bohr: His Heritage and Legacy*, Kluwer Academic Publishers, Dordrecht.

Feshbach, H., Matsui, T. and Oleson, A. [1988] *Niels Bohr: Physics and the World*, Harwood Academic Publishers, Reading.

Feyerabend, P. K. [1962] Explanation, Reduction and Empiricism, in Feyerabend [1981], pp. 44-96.

Feyerabend, P. K. [1981] *Realism, Rationalism and Scientific Method, Philosophical Papers, Vol. 1,* Cambridge University Press, Cambridge.

Feynman, R. P., Leighton R. B., and Sands, M. [1963] *The Feynman Lectures on Physics*, Addison-Wesley Publishing Co., Reading, Massachusetts.

French, S. and Kamminga, H. (eds.) [1993] *Correspondence, Invariance, and Heuristics*, Kluwar Academic Publishers, Holland.

Hanson, N. R. [1961] *Patterns of Discovery*, Cambridge University Press, Cambridge.

Harris, J. H. [1974] Popper's Definitions of Verisimilitude, *British Journal for the Philosophy of Science*, 25, pp. 160-166.

Holton, G. and Brush, S. G. [1973] *Introduction to Concepts and Theories in Physical Science*, 2nd edn, Princeton University Press, Princeton, N. J., 1985.

Holton, G. and Roller, D. H. D. [1958] *Foundations of Modern Physical Science*, Addison-Wesley, Reading, Mass. U.S.A.

Honner, J. [1987] *The Description of Nature: Niels Bohr and the Philosophy of Quantum Mechanics*, Clarendon Press, Oxford.

Jammer, M. [1966] *The Conceptual Development of Quantum Mechanics*, McGraw-Hill, New York.

Jammer, M. [1974] *The Philosophy of Quantum Mechanics*, John Wiley and Sons, New York.

Kepler, J. [1596] *Mysterium Cosmographicum* Translated by A. M. Duncan as *The Secret of the Universe*, Abaris Books, New York, 1981.

Koertge, N. [1969] *A Study of Relations Between Scientific Theories: A Test of the General Correspondence Principle*, PhD thesis, University of London.

Koyré, A. [1973] Translated by R. E. W. Maddison as *The Astronomical Revolution*, Methuen, London.

Krajewski, W. [1977] *Correspondence Principle and Growth of Science*, Reidel, Dordrecht.

Kuhn, T. S. [1957] *The Copernican Revolution*, Harvard University Press, Cambridge, Mass.

Kuhn, T. S. and Heilbron, J. L. [1969] The Genesis of the Bohr Atom, *Historical Studies in the Physical Sciences*, Vol. I, pp. 211-290.

Kuhn, T. S. [1970] *The Structure of Scientific Revolution*, University of Chicago Press, Chicago.

Lakatos, I. [1970] Falsification and the Methodology of Scientific Research Programmes, in Lakatos and Musgrave [1970], pp. 91-196.

Lakatos, I. and Musgrave A. [1970] (eds.), *Criticism and the Growth of Knowledge*, Cambridge University Press, Cambridge.

Laudan, L. [1981] A Confutation of Convergent Realism, in Leplin [1984].

Leplin, J. [1984] (ed.), *Scientific Realism*, University of California Press, Berkeley.

London, F. and Bauer, E. [1939] Translated by J. A. Wheeler et al as *The Theory of Observation in Quantum Mechanics*, in Wheeler and Zurek [1983], pp. 217-259

Miller, D. W. [1974] Popper's Qualitative Theory of Verisimilitude, *British Journal for the Philosophy of Science*, 25, pp. 166-177.

Miller, D. W. [1975] The Accuracy of Predictions, *Synthese*, 30, pp. 159-191.

Moulines, C. U. [1984] Ontological Reduction in the Natural Sciences, in Balzar, W. and Pearce, D. A. (eds.) [1984], pp. 51-70.

Nagel, E. [1961] *The Structure of Science*, Harcourt, Brace and World, New York.

Nicholson, J. W. [1911] The Spectrum of Nebulium, *Monthly Notices of the Royal Astronomical Society*, Nov. 1911, LXXII, 1, pp. 49-64.

Nicholson, J. W. [1912] The Constitution of the Solar Corona II, *Monthly Notices of the Royal Astronomical Society*, June 1912, LXXII, 8, pp. 677-692.

Oddie, G. [1986] *Likeness to Truth*, Reidel, Dordrecht.

Olenick, R. P., Apostol, T. M. and Goodstein, D. L. [1986] *Beyond the Mechanical Universe*, Cambridge University Press, Cambridge.

Pais, A. [1990] *Niels Bohr's Times*, Oxford University Press, Oxford.

Pearce, D. and Rantala, V. [1984] Limiting Case Correspondence Between Physical Theories, in Balzar, W., Pearce, D. A. and Schmidt, H. (eds.) [1984].

Penrose, R. [1989] *The Emperor's New Mind*, Vintage, London.

Petruccioli, S. [1993] *Atoms, Metaphors and Paradoxes: Niels Bohr and the Construction of a New Physics,* Cambridge University Press, Cambridge.

Popper, K. R. [1972] *Objective Knowledge,* Oxford University Press, Oxford.

Post, H. R. [1971] Correspondence, Invariance and Heuristics: in Praise of Conservative Induction, *Stud. in Hist. and Phil. of Sci.,* 2, no. 3, pp. 213-255.

Putnam, H. [1978] *Meaning and the Moral Sciences,* Routledge and Kegan Paul, London.

Putnam, H. [1983] *Realism and Reason, Philosophical Papers, vol. 3,* Cambridge University Press, Cambridge.

Radder, H. [1991] Heuristics and the Generalized Correspondence Principle, *Brit. Jour. Phil. Sci.,* 42, pp. 195-226.

Rae, A. [1986] *Quantum Physics: Illusion or Reality,* Cambridge University Press, Cambridge.

Redhead, M. L. G. [1975] Symmetry in Intertheory Relations, *Synthese,* 32, pp. 77-112.

Redhead, M. L. G. [1993] Is the End of Physics in Sight? in French, S. and Kamminga, H. (eds.) [1993], pp. 327-341.

Savage, C. W. (ed.) [1990] *Scientific Theories,* Minnesota Studies in the Philosophy of Science, Vol. XIV, University of Minnesota Press, Minneapolis.

Schrödinger, E. [1935] Translated by J. D. Trimmer as The Present Situation in Quantum Mechanics, *Proceedings of the American Philosophical Society,* 1980. Also in Wheeler and Zurek [1983], pp. 152-167.

Tichy, P. [1974] On Popper's Definitions of Verisimilitude, *British Journal for the Philosophy of Science*, 25, pp. 155-160.

Van der Waerden, B. L. [1967] (ed.), *Sources of Quantum Mechanics*, Dover Publications, New York.

Von Neumann J. [1932] Translated by R. T. Beyer as *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, 1955, Princeton, N. J.

Watkins, J. W. N. [1984] *Science and Scepticism,* Princeton University Press, Princeton, N. J.

Wheeler, J. A. [1981] Delayed-Choice Experiments and the Bohr-Einstein Dialog, *The American Philosophical Society and the Royal Society: Papers read at a meeting, June 5, 1980*, American Philosophical Society, Philadelphia. Also in Wheeler and Zurek [1983], pp. 182-200.

Wheeler, J. A. and Zurek, W. H. [1983] (eds.), *Quantum Theory and Measurement*, Princeton University Press, Princeton, N. J.

Wigner, E. P. [1961] Remarks on the Mind-Body Question, in Good, I. J., ed., *The Scientist Speculates*, Heinemann, 1961, London. Also in Wheeler and Zurek [1983], pp. 168-181.

Wigner, E. P. [1983] Interpretation of Quantum Mechanics, in Wheeler and Zurek [1983], pp. 260-314.

Worrall, J. [1989a] Structural Realism: The Best of Both Worlds?, *Dialectica*, 43, nos 1-2, pp. 99-124.

Worrall, J. [1989b] Scientific Revolutions and Scientific Rationality: The Case of the 'Elderly Hold-Out', in Savage, C. W. (ed.) [1990], pp. 319-354.

Yoshida, R. M. [1977] *Reduction in the Physical Sciences*, Dalhousie University Press, Halifax, Canada.

Zahar, E. G. [1983] Logic of Discovery or Psychology of Invention, *Brit. Jour. Phil. Sci.*, 34, pp. 243-261.

Zahar, E. G. [1989] *Einstein's Revolution, a Study in Heuristics*, Open Court, La Salle, Ill.

Text Books

Aris, R. [1962] *Vectors, Tensors and the Basic Equations of Fluid Mechanics*, Dover, New York.

Bohm, D. and Hiley, B. J. [1993] *The Undivided Universe: An Ontological Interpretation of Quantum Theory*, Routledge, London.

Darrigol, O. [1992] *From C-Numbers to Q-Numbers: The Classical Analogy in the History of Quantum Theory*, University of California Press, Berkeley, U. S. A.

Flint, H. T. [1967] *Wave Mechanics*, Methuen, London.

Goldstein, H. [1950] *Classical Mechanics*, Addison-Wesley, Reading, Mass.

Hague, B. and Martin, D. [1970] *An Introduction to Vector Analysis for Physicists and Engineers*, Methuen, London.

Hill, R. O. [1991] *Elementary Linear Algebra with Applications*, 2nd edn, Harcourt Brace Jovnovich, San Diego, Calif.

Leech, J. W. [1965] *Classical Mechanics*, Methuen, London.

Leighton, R. B. [1959] *Principles of Modern Physics*, McGraw-Hill, New York.

Loudon, R. [1983] *The Quantum Theory of Light*, Clarendon Press, Oxford.

Mehra, J. and Rechenberg, H. [1982] *The Historical Development of Quantum Theory*, Vols 1 and 2, Springer-Verlag, Heidelberg.

Schiff, L. J. [1955] *Quantum Mechanics*, 3rd edn, McGraw-Hill International, Tokyo.

Simmons, G. F. [1963] *Topology and Modern Analysis*, McGraw-Hill International, Auckland.