

THE
INTERNATIONAL SERIES
OF
MONOGRAPHS ON PHYSICS

GENERAL EDITORS

R. H. FOWLER AND P. KAPITZA

THE
PRINCIPLES
OF
QUANTUM MECHANICS

BY
P. A. M. DIRAC
FELLOW OF ST. JOHN'S COLLEGE
CAMBRIDGE

OXFORD
AT THE CLARENDON PRESS
1930

OXFORD UNIVERSITY PRESS

AMEN HOUSE, E.C. 4

LONDON EDINBURGH GLASGOW

LEIPZIG NEW YORK TORONTO

MELBOURNE CAPE TOWN BOMBAY

CALCUTTA MADRAS SHANGHAI

HUMPHREY MILFORD

PUBLISHER TO THE

UNIVERSITY

P R E F A C E

THE methods of progress in theoretical physics have undergone a vast change during the present century. The classical tradition has been to consider the world to be an association of observable objects (particles, fluids, fields, &c.) moving about according to definite laws of force, so that one could form a mental picture in space and time of the whole scheme. This led to a physics whose aim was to make assumptions about the mechanism and forces connecting these observable objects, to account for their behaviour in the simplest possible way. It has become increasingly evident in recent times, however, that nature works on a different plan. Her fundamental laws do not govern the world as it appears in our mental picture in any very direct way, but instead they control a substratum of which we cannot form a mental picture without introducing irrelevancies. The formulation of these laws requires the use of the mathematics of transformations. The important things in the world appear as the invariants (or more generally the nearly invariants, or quantities with simple transformation properties) of these transformations. The things we are immediately aware of are the relations of these nearly invariants to a certain frame of reference, usually one chosen so as to introduce special simplifying features which are unimportant from the point of view of general theory.

The growth of the use of transformation theory, as applied first to relativity and later to the quantum theory, is the essence of the new method in theoretical physics. Further progress lies in the direction of making our equations invariant under wider and still wider transformations. This state of affairs is very satisfactory from a philosophical point of view, as implying an increasing recognition of the part played by the observer in himself introducing the regularities that appear in his observations, and a lack of arbitrariness in the ways of nature, but it makes things less easy for the learner of physics. The new theories, if one looks apart from their mathematical setting, are built up from physical concepts which cannot be explained in terms of things previously known to the student, which cannot even be explained adequately in words at all. Like the fundamental concepts (e.g. proximity, identity) which every one must learn on his arrival into the world, the newer concepts of physics can be mastered only by long familiarity with their properties and uses.

From the mathematical side the approach to the new theories presents no difficulties, as the mathematics required (at any rate that which is required for the developments of physics up to the present) is not essentially different from what has been current for a considerable time. Mathematics is the tool specially suited for dealing with abstract concepts of any kind and there is no limit to its power in this field. For this reason a book on the new physics, if not purely descriptive of experimental work, must be essentially mathematical. All the same the mathematics is only a tool and one should learn to hold the physical ideas in one's mind without reference to the mathematical form. In this book I have tried to keep the physics to the forefront, by beginning with an entirely physical chapter and in the later work examining the physical meaning underlying the formalism wherever possible. The amount of theoretical ground one has to cover before being able to solve problems of real practical value is rather large, but this circumstance is an inevitable consequence of the fundamental part played by transformation theory and is likely to become more pronounced in the theoretical physics of the future.

With regard to the mathematical form in which the theory can be presented, an author must decide at the outset between two methods. There is the symbolic method, which deals directly in an abstract way with the quantities of fundamental importance (the invariants, &c., of the transformations) and there is the method of co-ordinates or representations, which deals with sets of numbers corresponding to these quantities. The second of these has usually been used for the presentation of quantum mechanics (in fact it has been used practically exclusively with the exception of Weyl's book *Gruppentheorie und Quantenmechanik*.) It is known under one or other of the two names 'Wave Mechanics' and 'Matrix Mechanics' according to which physical things receive emphasis in the treatment, the states of a system or its dynamical variables. It has the advantage that the kind of mathematics required is more familiar to the average student, and also it is the historical method.

The symbolic method, however, seems to go more deeply into the nature of things. It enables one to express the physical laws in a neat and concise way, and will probably be increasingly used in the future as it becomes better understood and its own special mathematics gets developed. For this reason I have chosen the symbolic method, introducing the representatives later merely as an aid to practical

calculation. This has necessitated a complete break from the historical line of development, but this break is an advantage through enabling the approach to the new ideas to be made as direct as possible. I have given the connexion between the new theory and Bohr's orbit theory, because the latter is likely to be useful in an elementary way for a long time to come.

The second half of the book contains applications to all the main fields in which quantum mechanics has been found useful. These applications all follow strictly from the general assumptions of the first half, with the exception of those of the last chapter, which gives a further theoretical development.

P. A. M. D.

ST. JOHN'S COLLEGE, CAMBRIDGE.

29 May 1930.

... +  ...

CONTENTS

I. THE PRINCIPLE OF SUPERPOSITION	1
1. Waves and Particles. 2. The Polarization of Photons. 3. Superposition and Indeterminacy. 4. Compatibility of Observations. 5. Further Discussion on Photons. 6. Definition of Superposition.	
II. SYMBOLIC ALGEBRA OF STATES AND OBSERVABLES	18
7. Addition of States. 8. Multiplication of States. 9. Algebra of Observables. 10. Conjugate Complex Observables. 11. Physical Interpretation of Algebra of Observables. 12. Example of Algebra of Observables.	
III. EIGENVALUES AND EIGENSTATES	35
13. Definitions and Elementary Properties. 14. The Expansion Theorem. 15. Functions of an Observable. 16. Examples of Functions of Observables. 17. Simultaneous Eigenstates. 18. Some Probability Theorems. 19. Contact Transformations.	
IV. REPRESENTATIONS OF STATES AND OBSERVABLES	55
20. General Properties. 21. Orthogonal Representations. 22. The δ Function. 23. Case of a Continuous Range of Fundamental States. 24. The Weight Function. 25. General Case of Representation.	
V. TRANSFORMATION THEORY	73
26. Eigenstates as Fundamental States of a Representation. 27. Canonical Transformations. 28. Probability Amplitudes. 29. Example. 30. Second Example.	
VI. EQUATIONS OF MOTION AND QUANTUM CONDITIONS	92
31. General Remarks. 32. Poisson Brackets. 33. Equations of Motion and Quantum Conditions obtained from Analogy with the Classical Theory. 34. Schrödinger's Form for the Quantum Conditions. 35. The Transformation Function $(q' p')$. 36. The Space-displacement Operator. 37. The Time-displacement Operator. 38. Heisenberg's Matrices.	
VII. ELEMENTARY APPLICATIONS	117
39. The Free Particle. 40. Wave Packets. 41. The Harmonic Oscillator in One Dimension. 42. The Harmonic Oscillator in Two Dimensions. 43. The Spin of the Electron	
VIII. MOTION IN A CENTRAL FIELD OF FORCE	137
44. Properties of the Angular Momentum. 45. Transition to Polar Co-ordinates. 46. Energy-levels of the Hydrogen Atom. 47. Selection Rules. 48. The Zeeman Effect for the Hydrogen Atom. 49. Combination of Angular Momenta.	

CONTENTS

IX. PERTURBATION THEORY	157
50. General Remarks. 51. The Change in the Energy-levels caused by a Perturbation. 52. The Perturbation considered as causing Transitions. 53. Application to Radiation. 54. Transitions caused by a Perturbation Independent of the Time. 55. The Anomalous Zeeman Effect.	
X. COLLISION PROBLEMS	176
56. General Remarks. 57. The Scattering Coefficient. 58. Solution with the p -Representation. 59. Dispersive Scattering. 60. Resonance Scattering. 61. Emission and Absorption.	
XI. SYSTEMS CONTAINING SEVERAL SIMILAR PARTICLES	198
62. Symmetrical and Antisymmetrical States. 63. Permutations as Observables. 64. Permutations as Constants of the Motion. 65. Determination of Energy-levels. 66. Application to Electrons.	
XII. THEORY OF RADIATION	218
67. Theory of Einstein-Bose Assemblies. 68. Discussion of Einstein-Bose Assemblies. 69. Application to Photons. 70. Determination of the Interaction Energy between a Photon and Atom. 71. Emission, Absorption, and Scattering of Radiation. 72. Einstein's Laws of Radiation.	
XIII. RELATIVITY THEORY OF THE ELECTRON	238
73. Relativity Treatment of a Single Particle. 74. The Wave Equation for the Electron. 75. Invariance under a Lorenz transformation. 76. Existence of the Spin. 77. Transition to Polar Variables. 78. The Fine-Structure of the Energy-levels of Hydrogen. 79. Physical Meaning of the Negative-Energy Solutions.	

I

THE PRINCIPLE OF SUPERPOSITION

§ 1. Waves and Particles

In the application of classical electrodynamics to atomic phenomena one meets with difficulties of a very fundamental nature, which show that the classical theory is irreconcilable with the facts. For instance, it is quite hopeless on the basis of classical ideas to try to account for the remarkable stability of atoms and molecules that is required in order that substances may have definite physical and chemical properties. These difficulties have necessitated a modification of some of the most fundamental laws of nature and have led to a new system of mechanics, called quantum mechanics, since its most striking (although not its most important) differences from the old mechanics apparently show a discontinuity in certain physical processes and a discreteness in certain dynamical variables.

Classical electrodynamics forms a self-consistent and very elegant theory, and one might be inclined to think that no modification of it would be possible which did not introduce arbitrary features and completely spoil its beauty. This is not so, however, since quantum mechanics, after passing through many stages and having its fundamental concepts changed more than once, has now reached a form in which it can be based on general laws and is, although not yet quite complete, even more elegant and pleasing than the classical theory in those problems with which it deals. This is brought about by the fact that the changes made in the classical theory are very few in number, although they are of a fundamental nature and involve the introduction of entirely new concepts, and are such that practically all the features of the classical theory to which it owes its attractiveness can be taken over unchanged into the new theory.

The necessity for a fundamental departure from the laws and concepts of classical mechanics is seen most clearly by a consideration of experimentally established facts on the nature of light. On the one hand the phenomena of interference and diffraction can be explained only on the basis of a wave theory of light; on the other, phenomena such as photo-electric emission and scattering by free electrons show that light is composed of small particles, which are called photons, each having a definite energy and momentum de-

pending on the frequency of the light. These photons appear to have just as real an existence as electrons, or any other particles known in physics. A fraction of a photon is never observed, so that we may safely assume it cannot exist.

To obtain a consistent theory of light which shall include interference and diffraction phenomena, we must consider the photons as being controlled by waves, in some way which cannot be understood from the point of view of ordinary mechanics. This intimate connexion between waves and particles is of very great generality in the new quantum mechanics. It occurs not only in the case of light. All particles are connected in this way with waves, which control them and give rise to interference and diffraction phenomena under suitable conditions. The influence of the waves on the motion of the particles is less noticeable the more massive the particles and only in the case of photons, the lightest of all particles, is it easily demonstrated.

The waves and particles should be regarded as two abstractions which are useful for describing the same physical reality. One must not picture this reality as containing both the waves and particles together and try to construct a mechanism, acting according to classical laws, which shall correctly describe their connexion and account for the motion of the particles. Any such attempt would be quite opposed to the principles by which modern physics advances. What quantum mechanics does is to try to formulate the underlying laws in such a way that one can determine from them without ambiguity what will happen under any given experimental conditions. It would be useless and meaningless to attempt to go more deeply into the relations between waves and particles than is required for this purpose.

§ 2. The Polarization of Photons

Although the idea of a physical reality being describable by both particles and waves, which are connected in some curious manner, is of far-reaching importance and wide applications, yet it is only a special case of a much more general principle, the *Principle of Superposition*. This principle forms the fundamental new idea of quantum mechanics and the basis of the departure from the classical theory.

In order to lead up to an explanation of this principle, we shall first take a very simple special case of it, which is provided by a consideration of the polarization of light. It is known experimentally

that when plane-polarized light is used for ejecting photo-electrons, there is a preferential direction for the electron emission. Thus the polarization properties of light are closely connected with its corpuscular properties and one must ascribe a polarization to the photons. One must consider, for instance, a beam of light plane polarized in a certain direction as consisting of photons each of which is plane polarized in that direction and a beam of circularly polarized light as consisting of photons each circularly polarized. Every photon is in a certain state of polarization, as we shall say. The difficulty is now how we are to fit in these ideas with the known facts about the resolution of light into polarized components and the recombination of these components.

Suppose, for instance, that we have a beam of plane-polarized light passing through a polariscope and getting resolved into two components polarized at angles of α and $\alpha + \frac{1}{2}\pi$ with the direction of polarization of the incident beam. The intensities of the two components will be, according to classical optics, respectively $\cos^2\alpha$ and $\sin^2\alpha$ times that of the original beam. Let us say that a photon of the original beam is in the state of polarization 0 and a photon in one or other of the two components is in the state α or $\alpha + \frac{1}{2}\pi$ respectively. The question that now arises is: What must we consider happens to each individual photon when it reaches the polariscope? How do the photons in the state 0 change into photons in the states α and $\alpha + \frac{1}{2}\pi$?

This question cannot be answered without the help of an entirely new concept which is quite foreign to classical ideas. We shall therefore first consider another question of a different type, namely, what will be the result of any particular experiment which one may perform to try to determine what happens to an individual photon when it reaches the polariscope. It is only questions of this type that are really important, and quantum mechanics always gives a definite answer to them. Any answer that may be given to our first question, *i.e.* any description of the whole course of a photon during the experiment, would be simply a device to help us to remember the results of the experiments. We ought not to be surprised if no such description based on classical ideas is possible.

The most direct experiment of this kind would be to use an incident beam consisting of only a single photon and then to measure the energy in each of the two components. The result predicted by

quantum mechanics is that sometimes one would find the whole of the energy in one component and other times one would find the whole in the other component. One would never find part of the energy in one and part in the other. Experiment can never reveal a fraction of a photon. If one did the experiment a large number of times, one would find in a fraction $\cos^2\alpha$ of the total number of times that the whole of the energy is in the α -component and in a fraction $\sin^2\alpha$ that the whole of the energy is in the $(\alpha + \frac{1}{2}\pi)$ -component. One may thus say that a photon has a probability $\cos^2\alpha$ of appearing in the α -component and a probability $\sin^2\alpha$ of appearing in the $(\alpha + \frac{1}{2}\pi)$ -component. These values for the probabilities lead to the correct classical distribution of energy between the two components when the number of photons in the incident beam is large.

Thus the individuality of the photon is preserved in all cases, but only at the expense of determinacy. The result of an experiment is not determined, as it would be according to the classical theory, by the conditions under the control of the experimenter. The most that can be predicted is the probability of occurrence of each of the possible results. This lack of determinacy, which runs through the whole of quantum mechanics and is in sharp contradiction to the classical theory, may at first sight appear to be unsatisfactory, as implying a departure from the law of causality. It should be remarked, though, that if one makes any experimental arrangement to observe the energy of one of the components (*e.g.* by reflection by a movable mirror and measurement of the recoil momentum communicated to the mirror), it will always be impossible subsequently to recombine the two components to produce interference effects. The observation must inevitably produce, as we shall see from the general laws of quantum mechanics, a change in phase of uncertain and unpredictable amount. One may therefore, as has been pointed out by Bohr,* ascribe the lack of determinacy in the result to the uncertainty in the disturbance which the observation necessarily makes, although one cannot inquire closely into how it comes about. The apparent failure of causality is from this point of view due to a theoretically necessary clumsiness in the means of observation.

We must now consider the answer to our first question and give a description of the photon throughout the course of the experiment. A description consisting of a continuous picture in the classical sense

* See the article by N. Bohr in *Nature*, p. 580, 1928.

is not possible. The description which quantum mechanics allows us to give is merely a manner of speaking which is of value in helping us to deduce and to remember the results of experiments and which never leads to wrong conclusions. One should not try to give too much meaning to it.

It is necessary to suppose a peculiar relationship to exist between the different states of polarization, which is such that when, for instance, a photon is in the state 0, it may be considered as being partly in the state α and partly in the state $\alpha + \frac{1}{2}\pi$. Similarly it could be considered as partly in state β and partly in state $\beta + \frac{1}{2}\pi$, where β is any other angle of polarization, or as partly in the state of left-circular polarization and partly that of right-circular polarization. More generally, one could consider it partly in each of two states plane polarized in two directions that are not at right angles, though this is seldom convenient, or one could consider it partly in each of more than two states. There are thus many ways of describing the photon, which are all always permissible and equally good theoretically, although, of course, the one that says the photon is entirely in state 0 is simpler than those that say it is 'distributed' over two or more states. When we say that the photon is distributed over two or more given states the description is, of course, only qualitative, but in the mathematical theory it is made exact by the introduction of numbers to specify the distribution, which determine the *weights* with which the different states occur in it.

One cannot picture in detail a photon being partly in each of two states; still less can one see how this can be equivalent to its being partly in each of two other different states or wholly in a single state. We must, however, get used to the new relationships between the states which are implied by this manner of speaking and must build up a consistent mathematical theory governing them.

In our polarizing experiment, if we choose to consider the incident photon as being partly in state α and partly in state $\alpha + \frac{1}{2}\pi$, the action of the polariscope is then quite simple. It separates the two components α and $\alpha + \frac{1}{2}\pi$ into two distinct beams, so that after the photon has passed through we must say that it is partly in one beam with the polarization α and partly in the other with the polarization $\alpha + \frac{1}{2}\pi$. There is now no way of saying the photon is wholly in one state, without a generalization of the meaning of a state, which will be made later. The simplest description is the one just given, in

which the photon is distributed over two states. Other possible descriptions would require the photon to be distributed over three or more states; e.g. one could say it is partly in the first beam with the polarization α , partly in the second beam with the polarization β (arbitrary), and partly in the second beam with the polarization $\beta + \frac{1}{2}\pi$. Such descriptions would not, however, be of value unless the beams were subsequently passed through other polarizing instruments.

Let us consider now what happens when we determine the energy in one of the components. The result of such a determination must be either the whole photon or nothing at all. Thus the photon must change suddenly from being partly in one beam and partly in the other to being entirely in one of the beams. This sudden change may be counted as due to the disturbance of the photon which the observation necessarily makes. It is impossible to predict in which of the two beams the photon will be found. Only the probability of either result can be calculated from the previous distribution of the photon over the two beams.

This way of describing the photon during the course of the experiment leads to one important conclusion, namely, the above-mentioned circumstance that when once the energy in one of the components has been determined, it will be impossible subsequently to bring about interference between the two components. When the photon is partly in one beam and partly in the other, if the two beams are superposed interference can take place, as the mathematical theory will show. This possibility disappears when the photon is forced entirely into one of the beams by the energy observation. The other beam then no longer enters into the description of the photon, so that if any experiment is subsequently performed on the same photon it will count as being entirely in the one beam in the ordinary way.

We have obtained a description of the photon throughout the experiment, which rests on a new rather vague idea of a photon being partly in one state and partly in another. The reader may, perhaps, feel that we have not really solved the difficulty of the conflict between the waves and the corpuscles, but have merely talked about it in a certain way and, by using some of the concepts of waves and some of corpuscles, have arrived at a formal account of the phenomena, which does not really tell us anything that we did not

know before. The difficulty of the conflict between the waves and corpuscles is, however, actually solved as soon as one can give an unambiguous answer to any experimental question. *The only object of theoretical physics is to calculate results that can be compared with experiment*, and it is quite unnecessary that any satisfying description of the whole course of the phenomena should be given.

With regard to the objection that the present description does not seem to take us any farther than we could, perhaps, have gone with very hazy notions of the relations between photons and electromagnetic waves, such as, for instance, those one had before the discovery of quantum mechanics, it should be remarked that the conclusion obtained above, that when once the energy of one of the beams has been measured subsequent interference between the beams would be impossible, could not have been drawn from *very* hazy notions, and also that the present discussion is really too qualitative for the advantages of the new theory to show up clearly. In § 5 the discussion on the nature of light will be renewed on a slightly more quantitative basis, which will bring out definitely the difference between the present theory and the previous hazy notions. For many elementary optical experiments, moreover, the hazy notions would suffice to give answers to questions concerning the results of observations and in such cases quantum mechanics would not give any further information. The object of quantum mechanics is to extend the domain of questions that can be answered and not to give more detailed answers than can be experimentally verified.

§ 3. Superposition and Indeterminacy

The new ideas that we have introduced in our description of the photon must be extended and applied to any atomic system, *i.e.* to any set of electrons and atomic nuclei interacting with each other and perhaps also with photons. We must first generalize the meaning of a 'state' so that it can apply to any atomic system. Corresponding to the case of the photon, which we say is in a given state of polarization when it has been passed through suitable polarizing apparatus, we say that any atomic system is in a given state when it has been prepared in a given way, which may be repeated arbitrarily at will. The method of preparation may then be taken as the specification of the state. The state of a system in the general case includes any information that may be known about its position in space from the

way in which it was prepared, as well as any information about its internal condition.

We must now imagine the states of any system to be related in such a way that whenever the system is definitely in one state, we can equally well consider it as being partly in each of two or more other states. The original state must be regarded as the result of a kind of *superposition* of the two or more new states, in a way that cannot be conceived on classical ideas. Any state may be considered as the result of a superposition of two or more other states, and indeed in an infinite number of ways. Conversely any two or more states may be superposed to give a new state, even also when they refer to different positions of the system in space. Thus in our previous example of the polarization experiment, when the photon is partly in the one beam with the polarization α and partly in the other with the polarization $\alpha + \frac{1}{2}\pi$, we may still count it as being entirely in a certain single state. In fact it still satisfies the definition of having been prepared in a definite way which may be repeated at will.

When a state is formed by the superposition of two other states, it will have properties that are in a certain way intermediate between those of the two original states and that approach more or less closely to those of either of them according to the greater or less ‘weight’ attached to this state in the superposition process. The new state is completely defined by the two original states when their relative weights in the superposition process are known, together with a certain phase difference, the exact meaning of weights and phases being provided in the general case by the mathematical theory of the next chapter. In the case of the polarization of a photon their meaning is that provided by classical optics, *e.g.* when two perpendicularly plane polarized states are superposed with equal weights, the new state may be circularly polarized in either direction, or linearly polarized at an angle $\frac{1}{4}\pi$, or else elliptically polarized, according to the phase difference. This, of course, is true only provided the two states that are superposed refer to the same beam of light, *i.e.* all that is known about the position and momentum of a photon in either of these states must be the same for each.

It is convenient at this stage to modify slightly the meaning of the word ‘state’ and to make it more precise. We must regard the state of a system as referring to its condition throughout an indefinite period of time and not to its condition at a particular time, which

would make the state a function of the time. Thus a state refers to a region of 4-dimensional space-time and not to a region of 3-dimensional space. A system, when once prepared in a given state, remains in that state so long as it remains undisturbed. This does not, of course, imply that it is not undergoing changes which could be revealed by experiment. In general it will be following out a definite course of changes, predictable by the quantum theory, belonging to that state. It is sometimes purely a matter of convenience whether we are to regard a system as being disturbed by a certain outside influence, so that its state gets changed, or whether we are to regard the outside influence as forming part of and coming in the definition of the system, so that with the inclusion of the effects of this influence it is still merely running through its course in one particular state. An illustration of this is our previous example of a photon being passed through a polariscope and becoming partly in each of two beams. Either we may consider the polariscope as disturbing the photon, so that after it has passed through it is in a different state; or else we may consider the polariscope as forming part of the 'field' in which the photon is moving, so that it is in the same state when it is in the incident beam as later when it is partly in each of the two component beams, and it is just following out its course in that state. The general laws of quantum mechanics apply equally well for either of these meanings of the state. There are, however, two cases when we are in general obliged to consider the disturbance as causing a change in state of the system, namely, when the disturbance is an observation and when it consists in preparing the system so as to be in a given state.

With the new space-time meaning of a state we need a corresponding space-time meaning of an *observation*. This requires that the specification of an observation shall include a definite time at which the observation is to be made, or at which the apparatus used in making the observation is to be set in motion, relatively to the time when the system was prepared. It should be noticed that it has a meaning to consider an observation being made on a system in a given state before this state is prepared. If the system is prepared at time t_0 , so that after time t_0 it is in a given state, we can imagine what it would have to be like before time t_0 in order that, if left undisturbed, it may become in the given state after time t_0 . Thus we can imagine the given state being produced backwards in time

and can give a meaning to an observation being made before time t_0 on the system in this state.

The introduction of indeterminacy into the results of observations, which we had to make in our discussion of the photon, must now be extended to the general case. When an observation is made on any atomic system that has been prepared in a given way and is thus in a given state, the result will not in general be determinate, *i.e.* if the experiment is repeated several times under identical conditions several different results may be obtained. If the experiment is repeated a large number of times it will be found that each particular result will be obtained a definite fraction of the total number of times, so that one can say there is a definite probability of its being obtained any time the experiment is performed. This probability the theory enables one to calculate. In special cases this probability may be unity and the result of the experiment is then quite determinate.

The indeterminacy in the results of observations is a necessary consequence of the superposition relationships that quantum mechanics requires to exist between the states. Suppose that we have two states *A* and *B* such that there exists an observation which, when made on the system in state *A*, is certain to lead to one particular result, and when made on the system in state *B*, is certain not to lead to this result. Two such states we call *orthogonal*. Suppose now that this observation is made on the system in a state formed by superposition of *A* and *B*. It is impossible for the result still to be determinate (except in the special case when the weight of *A* or *B* in the superposition process is zero). There must be a finite probability p that the result, that was certain for state *A*, will now be obtained and a finite probability $1-p$ that it will not be obtained. By continuously varying the relative weights in the superposition process we can get a continuous range of states, extending from pure *A* to pure *B*, for which the probability of the result, that was certain for state *A*, being obtained varies continuously from unity to zero.

It was mentioned above that an observation is not specified unless the time when it is made is given. In special cases it may so happen that the result of the observation, or the probability of any particular result being obtained, is independent of this time. If the state of the system is such that this is so for every observation that could be made on the system, then the state is said to be a *stationary* state and we should picture it as one in which the conditions are not varying.

The possibility in quantum mechanics of superposing states to get new states is connected with the fact that in the mathematical theory the equations that define a state are linear in the unknowns. It is not unnatural that one should try to establish analogies with systems in classical mechanics (such as vibrating strings or membranes), which are governed by linear equations and for which, consequently, a superposition principle holds. Such analogies have led to the name ‘Wave Mechanics’ being sometimes given to quantum mechanics. It must be emphasized, however, that *the superposition that occurs in quantum mechanics is of an essentially different nature from that occurring in the classical theory.* The analogies are therefore very misleading. Their inadequacy may be seen from the following special case. Suppose one compares the states of an atomic system with the states of vibration of a membrane. If one superposes any state of the vibrating membrane with itself, the result is a new state of double the amplitude. On the other hand, if one superposes an atomic state with itself according to quantum mechanics, the resulting state will be precisely the same as the original one. There is nothing in the atomic case that is analogous to the absolute value of the amplitude, as distinct from the relative amplitudes of different points, of the vibrating membrane.

§ 4. Compatibility of Observations

In general a system is disturbed when an observation is made on it, so that after the observation it is no longer in the same state as before. Only when the initial state and the observation are such that there is a probability unity, *i.e.* a certainty, for one particular result is it possible that the observation may produce no change of state. The necessity for this conclusion may be seen from the following argument.

Suppose that there is a probability p for a given result being obtained from the observation. Consider one occasion on which this result was actually found and suppose the observation was repeated immediately afterwards on the system in the state in which it was left by the first observation. There must have been a probability unity for the given result being obtained a second time, since we may assume the system could not have changed in the infinitely short time between the two observations. Thus while the first state is such that there is a probability p for a given result from a certain

observation, the second state (*i.e.* the one in which the system was left by the first observation) is such that there is a probability unity for this same result from a practically equivalent observation. Hence the second state must differ from the first when p differs from unity, since the probability of a result is quite definite for each state. It must be understood that the second state here considered is the one that arose on that particular occasion referred to above when the first observation was found to give the particular result desired. There will be a different second state corresponding to each different result for this observation. They must all be different from the initial state when p differs from unity.

Hence when once an observation of a system in a given state has been made, one cannot in general make a second observation and suppose it to apply to the same state. The first observation spoils the state of the system, which must then be prepared again before one can make the second. The two observations may, however, be such that, although the first one alters the state of the system, yet it does so in such a way as not to make any difference to the probability of any given result being obtained with the second. By the probability of a given result being obtained with the second is here meant its probability at the beginning of the experiment, before one knows what the result of the first observation is, and not its probability after a particular result has been obtained with the first observation. Two observations for which this is so when they are made (or at least when the first is made) with the minimum of disturbance allowed by theory, which can be attained in practice only under the most favourable conditions, are called *compatible*. Three or more observations are called compatible when any two are compatible. Two or more observations may be compatible only with respect to one particular state as initial state before any of the observations, or they may be compatible with respect to all initial states. In future when it is said that two or more observations are compatible, the second alternative is to be understood unless the contrary is stated.

The condition for the compatibility of two observations is, according to the laws of quantum mechanics, a symmetrical condition between them. If one of two compatible observations, α_1 say, is made at the time t_1 and the other, α_2 say, at the time t_2 which is later than t_1 , then, according to the definition given above, the probability of a given

result being obtained for α_2 must be the same whether this observation is made on the system in the initial state or in the state ensuing after observation α_1 . The symmetry condition now requires that the probability of a given result being obtained for α_1 must be the same whether this observation α_1 is made on the system in the initial state or in the state ensuing after observation α_2 , it being necessary to suppose this latter state, which is prepared at time t_2 , to be produced backwards in time, in the way mentioned in the preceding section, in order that the observation α_1 at time t_1 may be made on it. By the probability of a result for the state ensuing after a certain observation, is meant in each case the average probability for each state that can ensue after this observation, each of these states being weighted in the averaging process with the probability that it does ensue after this observation.

It has been pointed out that the state of a system after any observation has been made on it is such that this observation, if made on the system in this final state, would for a certainty give one particular result. Suppose now that a number of compatible observations $\alpha_1, \alpha_2, \dots$ are made on the system. Then the final state must be such that, if any of the observations α_r is made on the system in this final state, there will be a certainty for one particular result, since there was a certainty for one particular result as soon as the observation α_r was made in the preparation of the final state, and this will not be affected by the subsequent observations $\alpha_{r+1}, \alpha_{r+2}, \dots$, owing to the compatibility condition. The existence of states for which the result of any of the observations is a certainty forms one of the main properties of compatible observations. The order of the observations need not, of course, be their order in time, since we are allowed to consider an observation being made on a state before it is prepared.

The case of greatest interest of the compatibility of two observations is when they both refer to the same instant of time. The compatibility condition is now that if either is made a very short time before the other, the probability of any given result being obtained with the second shall be the same as if the first had not been made.

It is often convenient to count two or more compatible observations, particularly when they are simultaneous, as a single observation, the result of such an observation being expressible by two or more numbers. We shall frequently have to consider the greatest

possible number of independent compatible simultaneous observations being made on a system and shall, for brevity, call such a set of observations a maximum observation. *When a maximum observation is made on a system, its subsequent state is completely determined by the result of the observation and is independent of its previous state.* This may be considered as an axiom, or as a more precise definition of a state.

The state of a system after a maximum observation has been made on it is such that there exists a maximum observation (namely, an immediate repetition of the maximum observation already made) which, when made on the system in this state, will for a certainty lead to one particular result (namely, the previous result over again). Any state can be specified only as the state ensuing after a given maximum observation has been made for which a given result was obtained, or in some equivalent way. We can therefore draw the conclusion that for any state there must exist one maximum observation which will for a certainty lead to one particular result, and conversely, if we consider any possible result of a maximum observation, there must exist a state of the system for which this result for the observation will be obtained with certainty.

§ 5. Further Discussion on Photons

When quantum mechanics is applied to a system composed of simply a freely moving corpuscle, the equations that define a state of the system are, as we shall find from the mathematical theory, the ordinary equations for wave motion. It is this circumstance that gives to the corpuscle many of the properties of waves and allows us to consider a corpuscle in a given state as associated with, or controlled by, a given wave. In order to show more definitely the nature of the relations between the waves and the corpuscle, a typical example will be given of the conflict between the wave and the corpuscular theories of light and of the solution which quantum mechanics provides.

Consider a beam of light to be split into two components of equal intensity, which are made to interfere. According to the old corpuscular theory we would say that each of the two components contains an equal number of photons and we should then require that a photon in one component could interfere with one in the other. Under certain conditions they would have to annihilate one another, and under others to produce four photons. This contradicts the idea of photons

being discrete particles and is, besides, in disagreement with the conservation of energy, which should hold for each process in detail and not be merely statistically true.

The answer that quantum mechanics gives to the difficulty is that one should consider each photon to go partly into each of the two components, in the way allowed by the idea of the superposition of states. Each photon then interferes only with itself. Interference between two different photons can never occur. The solution of Maxwell's equations that forms the wave picture of the phenomenon represents *one* of the photons and not the whole assembly of photons. The relative intensities that this solution gives for the light at different points determine the relative probabilities of that photon being found at these points when an experiment is made to find its position. Only the relative intensities at different points are of importance; *the absolute intensity has no interpretation*. One must not try to establish any connexion between the absolute intensity of the waves and the total number of particles, which is in sharp distinction to the older ideas of the relations between waves and particles.

The quantum-mechanical views do not, of course, get over the difficulty of enabling us to picture something having properties between those of waves and corpuscles, but they serve to remind us, by their way of saying a photon is partly in one component and partly in the other, of the close connexion between the components and so prevent us from intuitively drawing wrong conclusions, as we do on the older views when we picture each component as having its own photons. For instance, we are reminded, by the requirement that the total probability of a photon being anywhere must be and must remain unity, that in whatever way the two component beams interfere, if they neutralize each other in one place they must reinforce each other in another so that conservation of energy is preserved. We thus get into no difficulty with the detailed conservation of energy.

§ 6. Definition of Superposition

A definition of the superposition of states will now be given. *We say that a state A may be formed by a superposition of states B and C when, if any observation is made on the system in state A leading to any result, there is a finite probability for the same result being obtained when the same observation is made on the system in one (at least) of the two states*

B and C. The Principle of Superposition says that any two states *B* and *C* may be superposed in accordance with this definition to form a state *A* and indeed an infinite number of different states *A* may be formed by superposing *B* and *C* in different ways. This principle forms the foundation of quantum mechanics. It is completely opposed to classical ideas, according to which the result of any observation is certain and for any two states there exists an observation that will certainly lead to two different results.

From our definition of superposition some elementary theorems follow immediately. For example, the states *B* and *C* themselves are particular cases of states formed by superposition of *B* and *C*. Again, if we superpose two states *A* and *B* obtaining a state *P*, which is then superposed on another state *C*, the resulting state *Q* will have the property that, if any observation is made on the system in this state leading to any result, there will be a finite probability of this same result being obtained when the observation is made on the system in one of the two states *P* and *C*, and hence there must be a finite probability of this result being obtained when the observation is made on the system in one of the three states *A*, *B*, and *C*. Thus the property possessed by the state *Q* is symmetrical in the three states *A*, *B*, and *C*, so that when superpositions are made successively their order is unimportant. This, of course, is necessary for the word ‘superposition’ to be suitable for describing the relations between the states.

Another example of a deduction from the definition of superposition is the following: If an observation of the system in a state *A* is certain to lead to one particular result and if this observation for another state *B* is certain to lead to the same result, then the observation is also certain to lead to this result for any state obtained by superposition of *A* and *B*. This is because it cannot lead to any other result, as the probability of this other result for both the states *A* and *B* is zero.

One could proceed to build up the theory of quantum mechanics on the basis of these ideas of superposition with the introduction of the minimum number of new assumptions necessary. Although this would be the logical line of development, it does not appear to be the most convenient one, as the laws of quantum mechanics are so closely interconnected that it would not be easy, and would in any case be somewhat artificial, to separate out the barest minimum of

assumptions from which the rest could be deduced. The method that will be here followed will therefore be first to give all the simple general laws in the form in which they are most easily expressed and remembered, and then to work out their consequences. This will mean that we shall continually be deducing results that are obviously necessary for the physical meaning of the theory to be tenable, or that follow from the foregoing ideas of superposition. Such deductions will then merely show the reasonableness and self-consistency of our fundamental assumptions.

II

SYMBOLIC ALGEBRA OF STATES AND OBSERVABLES

§ 7. Addition of States

We introduce certain symbols which we say denote physical things such as states of a system or dynamical variables. These symbols we shall use in algebraic analysis in accordance with certain axioms which will be laid down. To complete the theory we require laws by which any physical conditions may be expressed by equations between the symbols and by which, conversely, physical results may be inferred from equations between the symbols. A typical calculation in quantum mechanics will now run as follows: One is given that a system is in a certain state in which certain dynamical variables have certain values. This information is expressed by equations involving the symbols that denote the state and the dynamical variables. From these equations other equations are then deduced in accordance with the axioms governing the symbols and from the new equations physical conclusions are drawn. One does not anywhere specify the exact nature of the symbols employed, nor is such specification at all necessary. They are used all the time in an abstract way, the algebraic axioms that they satisfy and the connexion between equations involving them and physical conditions being all that is required. The axioms, together with this connexion, contain a number of physical laws, which cannot conveniently be analysed or even stated in any other way.

We denote each state of a dynamical system by a symbol ψ . Different states may be distinguished by suffixes, e.g. ψ_1 , ψ_2 , ψ_r . If a state ψ_0 may be formed by superposition of the states ψ_1 and ψ_2 , we express this relation between the states by an equation of the type

$$\psi_0 = c_1 \psi_1 + c_2 \psi_2, \quad (1)$$

where c_1 and c_2 are numbers, which may be imaginary or complex. The different states that may be formed by the superposition of ψ_1 and ψ_2 are given by different coefficients c_1 , c_2 . *Any two ψ -symbols denoting any two states may be added in this way with arbitrary coefficients c_1 and c_2* and the sum will always be another ψ -symbol denoting a state that can be formed by superposition of these two states, except in the special case when this sum is zero. The usual algebraic

axioms of addition are assumed to hold, *i.e.* the commutative axiom

$$c_1\psi_1 + c_2\psi_2 = c_2\psi_2 + c_1\psi_1$$

and the associative axiom

$$(c_1\psi_1 + c_2\psi_2) + c_3\psi_3 = c_1\psi_1 + (c_2\psi_2 + c_3\psi_3).$$

The first of these axioms implies that superposition of two states is a symmetrical process between them, which is obvious from the definition of § 6, while the second implies the theorem, which was proved in § 6, that in successive superpositions the order is unimportant.

Our assumptions so far are thus consistent with the definition of superposition. They do, however, go farther than this definition and contain new physical laws. For example, we can infer that if the state ψ_0 may be formed by superposition of ψ_1 and ψ_2 so that equation (1) holds, then (provided $c_1 \neq 0$) ψ_1 may be formed by superposition of ψ_0 and ψ_2 . The condition of superposition (1) is, in fact, symmetrical between ψ_0 , ψ_1 , and ψ_2 . This could not have been deduced from the definition of superposition in § 6. When three states are symmetrically related in this way, we say that they are *dependent*. We can extend the definition and say that any number of states $\psi_1, \psi_2, \dots, \psi_n$ are dependent or independent according to whether there is or is not a relation between them of the type

$$c_1\psi_1 + c_2\psi_2 + \dots + c_n\psi_n = 0. \quad (2)$$

It has been mentioned that when a state is superposed on itself, the resulting state is the same as the original one. Thus our symbolic scheme should be such that $\psi_1 + \psi_1$ or $2\psi_1$ denotes the same state as ψ_1 . Actually we make a more general assumption than this, namely, that $c\psi_1$ denotes the same state as ψ_1 where c is any number, not zero, and can be imaginary or complex. The nature of the connexion between the states and the symbols ψ required by this assumption may perhaps be more easily understood if one pictures the ψ 's as vectors in some space with a sufficiently large number of dimensions. The number of dimensions required is equal to the number of independent states that the system has, which is in general infinite. An equation of the type (1) or (2) can now be regarded as a vector equation. The vectors are, of course, in general complex. A state must now be considered as completely specified by *the direction of a vector*. Vectors of different lengths and the same direction specify the same state.

We now introduce another set of symbols ϕ_1, ϕ_2, \dots also denoting states. Any state denoted by a ψ -symbol ψ_r can be equally well

denoted by a ϕ -symbol ϕ_r having the same suffix. When the ψ 's that denote three states satisfy (1), the ϕ 's that denote these states are assumed to satisfy

$$\phi_0 = \bar{c}_1 \phi_1 + \bar{c}_2 \phi_2, \quad (3)$$

where the bar over a number denotes its conjugate complex. The ϕ 's are also assumed to satisfy the commutative and associative laws of addition and to have all the other properties that the ψ 's have, e.g. $c\phi_1$ denotes the same state as ϕ_1 , and we may define a number of states denoted by $\phi_1, \phi_2, \dots, \phi_n$ to be independent when there is no relation between them of the type

$$c_1 \phi_1 + c_2 \phi_2 + \dots + c_n \phi_n = 0.$$

The theory will throughout be symmetrical between the ϕ 's and ψ 's. The sum of a ϕ and a ψ has no meaning and will never appear in the analysis.

The introduction of a second set of symbols to denote the states may appear to be superfluous, but actually it is necessary when one allows complex coefficients c_r in order to preserve the symmetry between the two roots of -1 . A superposition process such as (1), which is specified by the two complex numbers c_1 and c_2 , must be equally well specifiable by the conjugate complex numbers \bar{c}_1 and \bar{c}_2 so that we are obliged to introduce equation (3) and treat it on the same footing as (1).

We have seen that a ϕ - or ψ -symbol may be multiplied by an arbitrary number and then still denotes the same state. Thus we can put

$$\psi_r = a_r \psi_r^*, \quad \phi_s = b_s \phi_s^*, \quad (4)$$

where the a 's and b 's are arbitrary numbers, not zero, and consider the ψ^* 's and ϕ^* 's as denoting the states instead of the ψ 's and ϕ 's. The a 's and b 's must, however, satisfy certain conditions in order that the connexion between equations (1) and (3) may hold also for the starred symbols. These equations give

$$\begin{aligned} \psi_0^* &= c_1 a_1/a_0 \cdot \psi_1^* + c_2 a_2/a_0 \cdot \psi_2^* \\ \phi_0^* &= \bar{c}_1 b_1/b_0 \cdot \phi_1^* + \bar{c}_2 b_2/b_0 \cdot \phi_2^*. \end{aligned}$$

In order that the coefficients in the ϕ^* equation may be conjugate complex to the coefficients in the ψ^* equation we must have

$$b_1/b_0 = \bar{a}_1/a_0 \quad b_2/b_0 = \bar{a}_2/a_0.$$

Hence

$$b_r = f \bar{a}_r \quad (5)$$

where f is a number independent of r .

The connexion between equations (1) and (3), and the condition (5) governing the most general transformation (4) that preserves this connexion, lead one to consider each ϕ_r as being proportional to the conjugate imaginary quantity of the corresponding ψ_r , the proportionality becoming an equality if a transformation of the type (4), (5) is applied with the correct value for f . Thus if we adopt the vector picture of the ψ 's we may take each ψ_r to be the conjugate imaginary vector to the corresponding ϕ_r . It should be remarked, though, that the conjugate imaginarity of the ψ 's and ϕ 's is not of quite the same nature as that of ordinary complex numbers, since we cannot give any meaning to the splitting up of a ψ into its real and pure imaginary parts. In the splitting up of an ordinary complex quantity into its real and pure imaginary parts, we obtain the real part by taking the average of the quantity itself and its conjugate imaginary, but we cannot do this for a ψ -symbol since we are not allowed to add together a ψ and a ϕ . Thus the relation between a ψ and the corresponding ϕ is not quite the same as the relation between two conjugate imaginary numbers, and in order that this difference may be remembered we shall reserve the words *conjugate imaginary* for describing relations between ψ 's and ϕ 's and use the words *conjugate complex* instead for quantities such as numbers which can be split up into real and pure imaginary parts. Ordinary vectors, of course, like numbers, can be split up into real and pure imaginary parts, so that the picturing of ψ 's and ϕ 's as vectors is not strictly correct, although it is all the same sometimes useful. We must therefore remember, when using the vector picture, that, in so far as it would allow one to add together two vectors representing a ψ and a ϕ respectively, it is imperfect and gives to the ψ 's and ϕ 's more properties than quantum mechanics requires or allows.

§ 8. Multiplication of States

Up to the present the only functions of the ψ 's and ϕ 's that we have allowed are linear functions of the ψ 's alone, or of the ϕ 's alone, with numerical coefficients. *We now suppose that any ψ and ϕ have a product, which is a number, in general complex. This product must always be written $\phi\psi$, i.e. the ϕ must be on the left-hand side and the ψ on the right.* Products such as $\psi\phi$, $\psi_1\psi_2$, $\phi_1\phi_2$, have no meaning and will never appear in the analysis.

The products $\phi\psi$ are assumed to satisfy the distributive axiom of multiplication, *i.e.*

$$\left. \begin{aligned} (\phi_1 + \phi_2)\psi &= \phi_1\psi + \phi_2\psi \\ \phi(\psi_1 + \psi_2) &= \phi\psi_1 + \phi\psi_2, \end{aligned} \right\} (6)$$

together with the axiom that

$$\phi(c\psi) = (c\phi)\psi = c(\phi\psi), \quad (7)$$

where c is any number. In the vector picture we can take the number $\phi\psi$ to be the scalar product of the two vectors ϕ and ψ . The conditions (6) and (7) are then satisfied. The vector picture, however, allows us also to form the products $\phi_1\phi_2$ and $\psi_1\psi_2$. Thus we again find the vector picture giving more properties to the ψ 's and ϕ 's than required in quantum mechanics.

In conformity with our view of regarding a ψ and the corresponding ϕ as conjugate imaginary quantities, we now make the following two assumptions:

$$\phi_r\phi_s = \overline{\phi_s\psi_r} \quad (8)$$

$$\phi_r\psi_r > 0. \quad (9)$$

From the first of these, by taking $s = r$, we can deduce that $\phi_r\psi_r$ is real. The second now states that $\phi_r\psi_r$ is positive. To examine the legitimacy of these assumptions, let us consider the effect of a transformation of the type (4), (5). Equation (8) gives

$$f\bar{a}_r a_s \phi_r^* \psi_s^* = \overline{f\bar{a}_s a_r \phi_s^* \psi_r^*}$$

and the inequality (9) gives

$$f\bar{a}_r a_r \phi_r^* \psi_r^* > 0.$$

From these relations we obtain

$$\phi_r^* \psi_s^* = \overline{\phi_s^* \psi_r^*} \quad \phi_r^* \psi_s^* > 0$$

provided f is real and positive. Thus a restriction must be imposed on the transformations (4), (5) in order that (8) and (9) may remain invariant.

In future we shall keep to the view that each ϕ is equal to, and not merely proportional to, the conjugate imaginary of the corresponding ψ , as the more general view, which is theoretically permissible, does not lead to anything of interest. This means that our equations need be invariant under transformations of the type (4) only provided $b_r = \bar{a}_r$, *i.e.* provided in (5) $f = 1$. The restriction on the transformations of the type (4) which is necessary for (8) and (9) to be invariant is included in this one.

We shall often assume that a ψ_r and the conjugate imaginary ϕ_r satisfy

$$\phi_r \psi_r = 1,$$

when they will be called *normalized to unity*, or simply *normalized*. The inequality (9) shows that it is always possible to *normalize* a ψ or a ϕ by multiplying it by a number. The modulus of this number is determined but not its argument.

A corollary of (9) is that if, for all ψ

$$\begin{aligned} \phi_r \psi &= 0, \\ \text{then } \phi_r &= 0. \end{aligned} \quad \left. \right\} (10)$$

This follows from the fact that if ϕ_r is not identically zero, its conjugate imaginary ψ_r will be a ψ that does not satisfy $\phi_r \psi = 0$. There is, of course, also the corresponding theorem with ϕ 's and ψ 's interchanged.

The theorem will now be proved that if ϕ_r and ψ_s are normalized, then

$$|\phi_r \psi_s| \leq 1, \quad (11)$$

the case of equality occurring only when ϕ_r and ψ_s denote the same state. Let a be any real number and apply the inequality (9) to the state denoted by $\psi_r - e^{ia} \psi_s$ or $\phi_r - e^{-ia} \phi_s$. This gives

$$(\phi_r - e^{-ia} \phi_s)(\psi_r - e^{ia} \psi_s) > 0$$

$$\text{or } \phi_r \psi_r - e^{ia} \phi_r \psi_s - e^{-ia} \phi_s \psi_r + \phi_s \psi_s > 0.$$

Hence, using the normalizing conditions $\phi_r \psi_r = \phi_s \psi_s = 1$, we obtain

$$e^{ia} \phi_r \psi_s + e^{-ia} \phi_s \psi_r < 2.$$

The second term on the left-hand side is just the conjugate complex of the first. Hence the real part of $e^{ia} \phi_r \psi_s$ is less than unity. Since this must hold for all values of a we must have the modulus of $\phi_r \psi_s$ less than unity. This gives the required result (11), when we take into account the fact that the inequality becomes an equality if $\psi_r - e^{ia} \psi_s = 0$ for some value of a , which means that ψ_r and ψ_s denote the same state.

Our introduction of products of ϕ 's with ψ 's has so far been entirely a mathematical question, with no physical implications. A physical meaning will now be given to the product $\phi_r \psi_s$. Consider that maximum observation of the state ϕ_r for which there is a certainty of a particular result being obtained. We have seen that such a maximum observation always exists. Suppose now this maximum observation to be made on the system in the state ψ_s . There will be a certain probability of the same result being obtained, which we call the

probability of agreement of ψ_s with ϕ_r . It is a number that depends only on the two states ψ_s and ϕ_r . In particular it is unity if ψ_s is the same state as ϕ_r . We now assume that the probability of agreement of ψ_s with ϕ_r is equal to $|\psi_s \phi_r|^2$ when ϕ_r and ψ_s are normalized. It has just been proved that this value for the probability can never exceed unity, so that the assumption is reasonable. Again, the only transformation of the type (4) that one can make on a normalized ϕ or ψ without destroying its normalization is multiplication by a number of modulus unity. This will not change the value of $|\phi_r \psi_s|^2$ which thus has the necessary invariance for its physical meaning to be permissible.

When we give this physical meaning to the product of a ϕ and a ψ the axioms and assumptions (6), (7), (8), (9) become, to a certain extent, physical laws, as physical consequences can now be deduced from them. For instance, from (8) one can deduce that the probability of agreement of ψ_s with ψ_r equals that of ψ_r with ψ_s . Again, from (6) and (7) one can calculate how the probability of agreement of a state ψ_0 with a state $c_1 \psi_1 + c_2 \psi_2$ formed by the superposition of ψ_1 and ψ_2 varies with the coefficients c_1 and c_2 . Let us take the case when ψ_1 and ψ_2 are orthogonal, *i.e.* when there exists an observation which is certain to lead to different results for the two states, so that their probability of agreement is zero. This requires that

$$\phi_1 \psi_2 = 0, \quad \phi_2 \psi_1 = 0.$$

In order that $c_1 \psi_1 + c_2 \psi_2$ may be normalized as well as ψ_1 and ψ_2 we must have

$$\begin{aligned} 1 &= (\bar{c}_1 \phi_1 + \bar{c}_2 \phi_2)(c_1 \psi_1 + c_2 \psi_2) \\ &= |c_1|^2 \phi_1 \psi_1 + |c_2|^2 \phi_2 \psi_2 \\ &= |c_1|^2 + |c_2|^2. \end{aligned}$$

If we now take ψ_0 orthogonal to ψ_2 , we find for the probability of agreement of ψ_0 with $c_1 \psi_1 + c_2 \psi_2$ the value

$$|\phi_0(c_1 \psi_1 + c_2 \psi_2)|^2 = |\phi_0 c_1 \psi_1|^2 = |c_1|^2 |\phi_0 \psi_1|^2,$$

which is $|c_1|^2$ times the probability of agreement of ψ_0 with ψ_1 . This result as it stands is not a physical one, since we have no other physical meaning for $|c_1|^2$ which we can equate to the ratio of the probability of agreement of ψ_0 with $c_1 \psi_1 + c_2 \psi_2$ to that of ψ_0 with ψ_1 . The fact that this ratio is independent of the state ψ_0 provided it is orthogonal to ψ_2 is, however, a physical result and is an example of the physical conclusions contained in the axioms (6) and (7).

We see further that these axioms give physical meanings to the coefficients occurring in a superposition process, or at least to the squares of their moduli. The simplest such physical meanings are obtained when we put ψ_0 equal to ψ_1 or ψ_2 in the above example. This gives the result that $|c_1|^2$ is the probability of agreement of $c_1\psi_1 + c_2\psi_2$ with ψ_1 and $|c_2|^2$ is that of $c_1\psi_1 + c_2\psi_2$ with ψ_2 . The sum of these two probabilities of agreement is unity, as could have been inferred from the definition of superposition of § 6. We may call $|c_1|^2$ and $|c_2|^2$ the weights with which ψ_1 and ψ_2 occur in the superposition process. The state $c_1\psi_1 + c_2\psi_2$ is not completely determined by these weights, as a phase factor, namely, the argument of c_1/c_2 is also necessary. This phase has no such simple physical meaning as the weights.

§ 9. Algebra of Observables

We must now introduce dynamical variables into the analysis. In classical mechanics a dynamical variable, for any state of the system, is given by a particular function of the time and is thus something that refers to all times. In the quantum theory a dynamical variable is no longer given by an ordinary function of the time, although it must still be something that refers to all times if it is to be the analogue of a classical dynamical variable. In quantum mechanics it is more convenient to deal with something that refers to one particular time instead of to all times, analogous to the value of a classical variable at a particular instant of time. We shall call such a quantity an *observable*. We can now say, in both classical and quantum mechanics, that any observation consists in measuring an observable and the result of such an observation is a number. The measurement of a dynamical variable for a particular state would in the classical theory give as result a function of the time and would in the quantum theory in general have no meaning.

We now denote each observable by a symbol. Thus the value of a Cartesian co-ordinate of an electron at a particular time t_1 would be an observable and could be denoted by the symbol $x(t_1)$. A dynamical variable, such as $x(t)$, may be regarded as an observable that depends on a parameter t which denotes the time. The symbols that denote observables will be used in the analysis along with the symbols that denote states, in accordance with certain rules and axioms that will now be given.

Any symbol α denoting an observable can be multiplied into any symbol ψ denoting a state, giving a product, which must be written $\alpha\psi$ with the ψ factor on the right-hand side. This product is of the nature of a ψ and thus denotes a state and can be added to other ψ 's. In the vector picture of the ψ 's we should say that an observable α is an operator which can be applied to any vector ψ to give another vector $\alpha\psi$. We assume the distributive axiom of multiplication, i.e.

$$\alpha(\psi_1 + \psi_2) = \alpha\psi_1 + \alpha\psi_2 \quad (12)$$

and we also assume

$$\alpha(c\psi) = c(\alpha\psi) \quad (13)$$

where c is any number. In the vector picture this means that the operator α is a linear operator and thus consists of rotations and uniform extensions or compressions applied to the vector field. The multiplication of the ψ 's by a number is an operation on them which satisfies these conditions, so that an ordinary number may be regarded as a special case of an observable. Its physical meaning will be discussed later (see § 11).

If an observable α is such that $\alpha\psi = 0$ for all ψ , then we assume that $\alpha = 0$. This means that an observable is completely determined when its product with an arbitrary ψ is given, since if we have two observables whose product with an arbitrary ψ is the same, their difference must vanish. We now define the sum $\alpha_1 + \alpha_2$ of two observables α_1 and α_2 by the condition

$$(\alpha_1 + \alpha_2)\psi = \alpha_1\psi + \alpha_2\psi \quad (14)$$

for all ψ . The commutative and associative laws for the addition of observables follow at once from this definition and from the corresponding laws for the addition of ψ -symbols. We further define the product $\alpha_1\alpha_2$ of two observables α_1 and α_2 by the condition

$$(\alpha_1\alpha_2)\psi = \alpha_1(\alpha_2\psi) \quad (15)$$

for all ψ . The associative and distributive laws for the multiplication of observables follow at once from the definition, e.g. for the associative law we have

$$\begin{aligned} [(\alpha_1\alpha_2)\alpha_3]\psi &= (\alpha_1\alpha_2)(\alpha_3\psi) = \alpha_1[\alpha_2(\alpha_3\psi)] \\ &= \alpha_1[(\alpha_2\alpha_3)\psi] = [\alpha_1(\alpha_2\alpha_3)]\psi \end{aligned}$$

and since this holds for all ψ we must have

$$(\alpha_1\alpha_2)\alpha_3 = \alpha_1(\alpha_2\alpha_3).$$

However, the commutative law for the multiplication of observables in general does not hold, i.e. in general $\alpha_1\alpha_2$ is not equal to $\alpha_2\alpha_1$. In the

special case when $\alpha_1 \alpha_2$ is equal to $\alpha_2 \alpha_1$, we say that α_1 commutes with α_2 or that α_1 and α_2 commute. We say that three or more observables commute when each commutes with all the others.

Since the theory is to be symmetrical between the ψ 's and the ϕ 's it must be possible to multiply any observable α into any ϕ -symbol. The product, which we always write as $\phi\alpha$ with the ϕ on the left-hand side, must be of the nature of a ϕ and thus be capable of denoting a state and of being added to other ϕ 's. Corresponding to (12) and (13) we must have

$$(\phi_1 + \phi_2)\alpha = \phi_1\alpha + \phi_2\alpha$$

and

$$(c\phi)\alpha = c(\phi\alpha).$$

We require one more axiom in our symbolic algebra, namely, an associative axiom of multiplication which says that

$$(\phi\alpha)\psi = \phi(\alpha\psi),$$

so that either of these numbers may be written as $\phi\alpha\psi$ without brackets.

This final axiom enables us to prove that the sum or product of two observables, defined by (14) or (15), is the same as the sum or product defined in the analogous way with ϕ 's instead of ψ 's, i.e. by

$$\phi(\alpha_1 + \alpha_2) = \phi\alpha_1 + \phi\alpha_2 \quad (16)$$

or

$$\phi(\alpha_1 \alpha_2) = (\phi\alpha_1)\alpha_2$$

for all ϕ . In the case of the sum, for instance, if we take the definition (14) we can infer from it, with the help of (6), that

$$\phi(\alpha_1 + \alpha_2)\psi = \phi\alpha_1\psi + \phi\alpha_2\psi$$

or

$$[\phi(\alpha_1 + \alpha_2) - \phi\alpha_1 - \phi\alpha_2]\psi = 0$$

for all ϕ and ψ . Hence from (10) we must have

$$\phi(\alpha_1 + \alpha_2) - \phi\alpha_1 - \phi\alpha_2 = 0,$$

which is the required result (16). The case of the product is quite similar. A further similar argument enables one to deduce, from the assumption that if $\alpha\psi = 0$ for all ψ then $\alpha = 0$, the result that if $\phi\alpha = 0$ for all ϕ then $\alpha = 0$.

§ 10. Conjugate Complex Observables

It is convenient to count sums and products of any observables as other observables. This involves, as we shall see shortly, an extension of the meaning of an observable to include the analogues of complex functions of classical dynamical variables, or rather the values of

such complex functions at specified times. An observable is thus not necessarily a quantity capable of direct measurement by a single observation, but is a theoretical generalization of such a quantity.

More generally it is convenient to count any operator that can be multiplied into the ψ 's and ϕ 's in accordance with the foregoing axioms as an observable. Thus one can define an observable α by specifying the values of $\alpha\psi$ for all ψ , and these values may be chosen arbitrarily except for the condition (12). If one takes a complete set of independent ψ 's, ψ_r , say, a complete set being one such that any ψ can be expressed linearly in terms of its members, then the values of $\alpha\psi_r$, for the members of this set ψ_r , may be chosen quite arbitrarily, and the value of $\alpha\psi$ when ψ is not a member of the set is then determined by (12), so that α is determined. Again, instead of specifying the $\alpha\psi_r$'s, one could define α by specifying the numbers $\phi_s \alpha\psi_r$, which are quite arbitrary when the ϕ_s 's as well as the ψ_r 's form a complete independent set. The fact that α is uniquely determined in this way follows from (10).

Now let α be any observable and consider the equation

$$\overline{\phi_s \alpha \psi_r} = \phi_r \beta \psi_s \quad (17)$$

where ψ_r and ψ_s are any two ψ 's and ϕ_r and ϕ_s are their conjugate imaginaries. We can consider this equation as defining a new observable β , since we can assume (17) holds for a complete set of independent ψ_r 's and for a complete set of independent ψ_s 's, and since, as is easily verified, if (17) holds for two values of ψ_r , it must hold also for any linear combination of them, and similarly for ψ_s . In fact if (17) holds for $\psi_r = \psi_1$ and for $\psi_r = \psi_2$ we have the equations

$$\overline{\phi_s \alpha \psi_1} = \phi_1 \beta \psi_s, \quad \overline{\phi_s \alpha \psi_2} = \phi_2 \beta \psi_s,$$

from which we can deduce

$$\begin{aligned} \overline{\phi_s \alpha (c_1 \psi_1 + c_2 \psi_2)} &= \bar{c}_1 \overline{\phi_s \alpha \psi_1} + \bar{c}_2 \overline{\phi_s \alpha \psi_2} \\ &= \bar{c}_1 \phi_1 \beta \psi_s + \bar{c}_2 \phi_2 \beta \psi_s \\ &= (\bar{c}_1 \phi_1 + \bar{c}_2 \phi_2) \beta \psi_s, \end{aligned}$$

which shows that (17) holds also for $\psi_r = c_1 \psi_1 + c_2 \psi_2$.

The observable β defined by (17) is called the *conjugate complex* of the observable α and is written $\bar{\alpha}$. Thus

$$\overline{\phi_s \alpha \psi_r} = \phi_r \bar{\alpha} \psi_s. \quad (18)$$

The conjugate complex of $\bar{\alpha}$ is α . We use the words 'conjugate complex' and not 'conjugate imaginary' since it is permissible to add

together an observable and its conjugate complex, both being quantities of the same nature, so that one can split up any observable α into its real part, $\frac{1}{2}(\alpha + \bar{\alpha})$, and pure imaginary part, $\frac{1}{2}(\alpha - \bar{\alpha})$. The condition for an observable α to be real is

$$\overline{\phi_s \alpha \psi_r} = \phi_r \alpha \psi_s$$

In the special case when the observable α is a number, its conjugate complex defined by (18) is the ordinary conjugate complex number.

It will now be proved that if ψ_1 and ϕ_1 are conjugate imaginary symbols, then so also are $\alpha \psi_1$ and $\phi_1 \bar{\alpha}$ for any observable α . If we denote by ϕ the conjugate imaginary of $\alpha \psi_1$, then from (8)

$$\phi \psi_s = \overline{\phi_s \alpha \psi_1}$$

for arbitrary ψ_s . But from the definition (18)

$$\overline{\phi_s \alpha \psi_1} = \phi_1 \bar{\alpha} \psi_s.$$

Hence

$$\phi \psi_s = \phi_1 \bar{\alpha} \psi_s$$

for arbitrary ψ_s , so that from (10) (with ϕ 's and ψ 's interchanged)

$$\phi = \phi_1 \bar{\alpha}, \quad (20)$$

which was to be proved.

We shall now find the conjugate complex of the product $\alpha_1 \alpha_2$ of two observables α_1 and α_2 . The equation that defines this conjugate complex, $\overline{\alpha_1 \alpha_2}$ is

$$\phi_p \overline{\alpha_1 \alpha_2} \psi_q = \overline{\phi_q \alpha_1 \alpha_2 \psi_p} \quad (21)$$

for arbitrary ψ_p and ϕ_q . If in formula (8) we put

$$\phi_s = \phi_q \alpha_1, \quad \psi_r = \alpha_2 \psi_p,$$

which require, from the theorem of equation (20),

$$\psi_s = \bar{\alpha}_1 \psi_q, \quad \phi_r = \phi_p \bar{\alpha}_2,$$

we get

$$\phi_p \bar{\alpha}_2 \bar{\alpha}_1 \psi_q = \overline{\phi_q \alpha_1 \alpha_2 \psi_p}.$$

Comparing this with (21) we obtain, since these equations hold for arbitrary ϕ_p and ψ_q , the result

$$\alpha_1 \alpha_2 = \bar{\alpha}_2 \bar{\alpha}_1. \quad (22)$$

Thus to find the conjugate complex of a product we must take the conjugate complex of each factor and reverse their order. This rule holds also when there are more than two factors in the product, as may be proved by successive applications of the rule for two factors, e.g.

$$\alpha_1 \bar{\alpha}_2 \alpha_3 = \bar{\alpha}_3 \alpha_1 \bar{\alpha}_2 = \bar{\alpha}_3 \bar{\alpha}_2 \bar{\alpha}_1.$$

As a corollary of this theorem we have that if α_1 and α_2 are two real observables, then $\alpha_1 \alpha_2 + \alpha_2 \alpha_1$ is also real and $\alpha_1 \alpha_2 - \alpha_2 \alpha_1$ is pure

imaginary. Only when α_1 and α_2 commute is $\alpha_1\alpha_2$ also real. Equation (18) and the theorem of equation (20) show that it is a general rule that when one forms the conjugate imaginary or the conjugate complex of any permissible combination of the symbols denoting observables and states, one must reverse the order of the factors in a product and take the conjugate imaginary or conjugate complex of each factor.

§ 11. Physical Interpretation of Algebra of Observables

The axioms and assumptions that we have made about observables are so far purely mathematical and have no physical implications. The physical connexions, which cause these axioms and assumptions to become physical laws, will now be given. The observables that appear in the discussion in this section must be understood to be all real observables.

If a state ψ_r and an observable α are such that, when an observation is made of the observable with the system in this state the result is certain to be the number a , we assume this information can be expressed by the equation

$$\alpha\psi_r = a\psi_r. \quad (23)$$

Conversely, when an equation of this type is given we assume it has the physical meaning that a measurement of the observable α with the system in state ψ_r will certainly give for result the number a or that the observable α *has the value a* for the state ψ_r , to use a classical way of speaking which is permissible in this case. Equation (23) is equivalent to

$$\phi_r\alpha = a\phi_r, \quad (24)$$

provided α is real, since, from the theorem of equation (20), equation (24) is just the conjugate imaginary of equation (23). Thus the symmetry between the ϕ 's and ψ 's is maintained.

In the special case when the observable α is a number, then equation (23) holds for every state ψ_r with this same number for a . This means that the observable is of a trivial kind such that any measurement of it always gives one particular result, independent of the state of the system.

We can now deduce some physical results from the theory. For example, if for a given state ψ the observable α_1 has the value a_1 and the observable α_2 has the value a_2 , we have the equations

$$\alpha_1\psi = a_1\psi, \quad \alpha_2\psi = a_2\psi, \quad (25)$$

from which we can deduce that

$$(\alpha_1 + \alpha_2)\psi = (\alpha_1 + \alpha_2)\psi,$$

$$\alpha_1\alpha_2\psi = \alpha_1\alpha_2\psi,$$

and thus infer that for the state ψ the observable $\alpha_1 + \alpha_2$ has the value $\alpha_1 + \alpha_2$ and the observable $\alpha_1\alpha_2$ has the value $\alpha_1\alpha_2$. These results are necessary for the theory to be consistent, since the observations of α_1 and α_2 for the system in state ψ are compatible, as neither observation need cause a change in the state, so that one would expect the ordinary classical ideas of measurement to be valid. For the same reason we require the result, which may easily be deduced from the first of equations (25) by induction, that $f(\alpha_1)$ has the value $f(\alpha_1)$ for the state ψ , where f denotes any function expressible as a power series. We shall later define more general functions of an observable than are expressible as power series, and for these more general functions this result will still hold. In fact it will form the basis of the definition of these more general functions.

Again, if we are given that an observable α has the value a for each of two states ψ_1 and ψ_2 , we can write down the equations

$$\alpha\psi_1 = a\psi_1, \quad \alpha\psi_2 = a\psi_2,$$

from which we can deduce that

$$\alpha(c_1\psi_1 + c_2\psi_2) = a(c_1\psi_1 + c_2\psi_2).$$

Thus α has the value a also for any state obtainable by superposition of ψ_1 and ψ_2 . This result was deduced in § 6 from the definition of superposition and the fact that it is also deducible from the present analysis illustrates the self-consistency of the theory.

In classical mechanics an observable always has a particular value for any state. This is not so in quantum mechanics, where a special condition of the type (23) is necessary for an observable to have a particular value for a certain state. In general the measurement of an observable for a given state will lead to one or other of a number of possible results, according to a certain probability law. The question now to be considered is what can be said in the general case about an observable with respect to a state. If one has an observable α and one takes any two states ϕ_r, ψ_s , one can form the number $\phi_r\alpha\psi_s$. This is the only general way of forming numbers referring to an observable and particular states. Thus an observable has a numerical value associated with each pair of states, in sharp distinction to the classical theory, where an observable always has a

numerical value associated with a single state, namely, the value of the observable for that state.

We could, however, as a special case, take conjugate imaginary symbols ϕ_r and ψ_r , which both denote the same state, and form the number $\phi_r \alpha \psi_r$. We should then have a number completely determined by the observable α and the state ψ_r , provided the ϕ_r and ψ_r are normalized, since, as is easily verified, $\phi_r \alpha \psi_r$ remains invariant under any transformation of the type (4) with $b_r = \bar{a}_r$ that preserves the normalization. Thus it is possible to associate with the observable α a definite numerical value for a single state ψ_r , but it would not be convenient to define this number as the value of the observable α for the state ψ_r , for the following reason. If for a particular state a_1 is the value of an observable α_1 and a_2 is that of α_2 , then we should require $a_1 + a_2$ to be the value of $\alpha_1 + \alpha_2$ and $a_1 a_2$ to be that of $\alpha_1 \alpha_2$. The definition just proposed for the value of an observable for a state would give

$$a_1 = \phi_r \alpha_1 \psi_r, \quad a_2 = \phi_r \alpha_2 \psi_r,$$

from which we could deduce

$$a_1 + a_2 = \phi_r (\alpha_1 + \alpha_2) \psi_r,$$

and hence infer that $a_1 + a_2$ is the value of $\alpha_1 + \alpha_2$. We could not, however, deduce that

$$a_1 a_2 = \phi_r \alpha_1 \alpha_2 \psi_r,$$

which would, in fact, in general be untrue, so that we could not infer that $a_1 a_2$ is the value of $\alpha_1 \alpha_2$. Thus we cannot take $\phi_r \alpha \psi_r$ as a general definition of the value of an observable α for a state ψ_r . We must fall back on the equation (23) to give the definition of this value in the special cases when it exists.

The fact, however, that the proof fails only in the case of the product $\alpha_1 \alpha_2$ and not in the case of the sum $\alpha_1 + \alpha_2$ allows us to say that $\phi_r \alpha \psi_r$ is the *average* value of the observable α for the state ψ_r . This is so because the average of the sum of two quantities must equal the sum of their averages, but the average of their product need not equal the product of their averages. Thus our symbolic algebra allows us to define a certain number as being the average value of an observable for a particular state, without leading us to inconsistencies. The assumption that this so-defined average is really what one would obtain if one measured the observable a large number of times (the system having to be re-prepared each time, of course,

in order that it may be in the proper state) and worked out the average result, constitutes the main link connecting the symbolic algebra with physical facts. The other links previously given, *i.e.* the assumption that $|\phi_r \psi_s|^2$ is the probability of agreement of ϕ_r with ψ_s and the assumption that the equation $\alpha\psi = a\psi$ holds when an observation of α on the system in state ψ will certainly lead to the result a , will be shown later (§ 18) to be deducible from this main link as special cases.

If an observable α has the value a for a state ψ_r , so that equation (23) holds, we can deduce that

$$\phi_r \alpha \psi_r = \phi_r a \psi_r = a \phi_r \psi_r = a$$

if ϕ_r and ψ_r are normalized. Hence the average value of α for the state ψ_r is found to be a , as is necessary for the physical interpretation of the theory to be consistent. We cannot, of course, deduce the converse, *i.e.* deduce (23) from the equation $\phi_r \alpha \psi_r = a$.

The numbers $\phi_r \alpha \psi_s$ which the theory also gives us, where ϕ_r and ψ_s denote two different states, do not have any such direct physical interpretation as the numbers $\phi_r \alpha \psi_r$. We shall find later that $|\phi_r \alpha \psi_s|^2$ is, apart from a certain factor, the probability of a transition from state ψ_s to state ϕ_r being caused by a perturbing energy whose time integral is α . (See § 52.)

§ 12. Example of Algebra of Observables

As an example of the symbolic algebra of observables, which is the same as ordinary algebra except for the non-validity of the commutative law of multiplication, we shall consider some properties of two observables, p and q , that satisfy

$$qp - pq = i, \quad (26)$$

i being a root of minus one. From § 10 we see that it is possible for two real observables p and q to satisfy this relation. If we multiply (26) by q on the left, and then by q on the right, we obtain

$$q^2 p - qpq - iq$$

and

$$qpq - pq^2 = iq,$$

from which, by addition, we find

$$q^2 p - pq^2 = 2iq.$$

This result can be generalized. If we multiply (26) firstly by q^{n-1} on the left, secondly by q^{n-2} on the left and q on the right, thirdly by

q^{n-3} on the left and q^2 on the right, and so on until n -thly we multiply simply by q^{n-1} on the right, we get the equations

$$\begin{aligned} q^n p - q^{n-1} pq &= iq^{n-1} \\ q^{n-1} pq - q^{n-2} pq^2 &= iq^{n-1} \\ q^{n-2} pq^2 - q^{n-3} pq^3 &= iq^{n-1} \\ \cdot &\quad \cdot &\quad \cdot &\quad \cdot &\quad \cdot &\quad \cdot &\quad \cdot \\ \cdot &\quad \cdot &\quad \cdot &\quad \cdot &\quad \cdot &\quad \cdot &\quad \cdot \\ qpq^{n-1} - pq^n &= iq^{n-1}, \end{aligned}$$

which give, on addition, the result

$$q^n p - pq^n = niq^{n-1}.$$

This result may be written

$$q^n p - pq^n = idq^n/dq.$$

It follows that, if $f(q)$ is any function of q expressible as a power series,

$$fp - pf = idf/dq, \quad (27)$$

since this result must hold separately for each term in the expansion.

As a special case, we may take for f the power series

$$f(q) = \sum_{n=0}^{\infty} \frac{(ic)^n q^n}{n!},$$

where c is a number. We can define this to be e^{icq} and the ordinary exponential theorem will then hold, since no symbol that does not commute with q could occur in the proof of it to make a difference between the present and ordinary algebra. With this expression for f , (27) becomes

$$e^{icq} p - pe^{icq} = -ce^{icq}$$

or

$$e^{icq} p = (p - c)e^{icq}. \quad (28)$$

III

EIGENVALUES AND EIGENSTATES

§ 13. Definitions and Elementary Properties

In the present chapter we shall consider some of the properties of real observables. If we have any real observable α we can write down the equation

$$\alpha\psi = a\psi \quad (1)$$

where a is a number, and consider it as an equation for the two unknowns a and ψ . If a and ψ are any solution, we call them respectively an *eigenvalue* and an *eigen-ψ* of the observable α . It may easily be seen that the eigenvalues are all real numbers, since if we multiply (1) by the ϕ -symbol that is conjugate imaginary to ψ , we obtain

$$\phi\alpha\psi = a\phi\psi.$$

Now $\phi\alpha\psi$ and $\phi\psi$ are both real, as follows from equations (19) and (8) of the preceding chapter when one takes $r = s$, and hence a must be real. Analogous to (1) is the equation

$$\phi\alpha = a\phi. \quad (2)$$

If a and ψ are any solution of (1), then the same value of a and the ϕ that is conjugate imaginary to this ψ form a solution of (2), since equation (2) is then the conjugate imaginary of equation (1). We call the ϕ 's that solve (2) *eigen-ϕ*'s, and the states denoted by the eigen- ψ 's or eigen- ϕ 's we call *eigenstates* of the observable α . Each eigen- ψ , eigen- ϕ or eigenstate is associated with one definite eigenvalue, or, as we shall say, belongs to that eigenvalue.

The physical meaning of an eigenvalue is that there exists a state, namely, the eigenstate belonging to it, such that a measurement of the observable when the system is in this state will certainly give for result just this eigenvalue. *The eigenvalues of an observable are the possible results of a measurement of this observable.* Every possible result of the measurement of α must be an eigenvalue as it must satisfy (1) when one takes for the ψ in this equation the state of the system immediately after the observation has been made. The whole set of eigenvalues of an observable may consist of a discrete set of numbers, or a continuous range of numbers, or perhaps both. The calculation of eigenvalues is one of the main problems of quantum mechanics.

In the special case when the observable is a number, it has only one eigenvalue, namely, itself, and any state is an eigenstate. If α is any observable and c is a number, then, as follows at once from the definitions, each eigenvalue of $\alpha+c$ is greater by c than an eigenvalue of α and each eigenstate of $\alpha+c$ is an eigenstate of α . Similarly each eigenvalue of $c\alpha$ is c times an eigenvalue of α and each eigenstate of $c\alpha$ is an eigenstate of α .

The theorem will now be proved that *two eigenstates belonging to two different eigenvalues of an observable are orthogonal*. Suppose the eigenstate ψ_1 belongs to the eigenvalue a_1 and the eigenstate ψ_2 belongs to the eigenvalue a_2 . We then have the equations

$$\alpha\psi_1 = a_1\psi_1 \quad (3)$$

$$\phi_2\alpha = a_2\phi_2. \quad (4)$$

Multiplying (3) by ϕ_2 on the left-hand side and (4) by ψ_1 on the right-hand side, we obtain

$$\phi_2\alpha\psi_1 = a_1\phi_2\psi_1$$

and

$$\phi_2\alpha\psi_1 = a_2\phi_2\psi_1.$$

Hence

$$(a_1 - a_2)\phi_2\psi_1 = 0,$$

so that, if a_1 is not equal to a_2 , then $\phi_2\psi_1 = 0$ and the two states ψ_1 and ψ_2 are orthogonal. This theorem is required by the physical meaning of eigenstates, since for two eigenstates belonging to two different eigenvalues there exists an observation, namely, the measurement of the observable α , for which the result must certainly be different in the two cases, so that the two states are, by definition, orthogonal.

If ψ_1 and ψ_2 are two eigen- ψ 's belonging to the same eigenvalue, then it is evident that any linear combination of them ($c_1\psi_1 + c_2\psi_2$) must also be an eigen- ψ belonging to this eigenvalue. It will now be proved that no linear combination of eigen- ψ 's belonging to different eigenvalues can be an eigen- ψ , i.e. that *eigen- ψ 's belonging to different eigenvalues are all necessarily independent*. If this were not so we should have a relation of the type

$$\sum_r c_r \psi_r = 0, \quad (5)$$

with numerical coefficients c_r , between a number of eigen- ψ 's belonging to different eigenvalues. We can without loss of generality assume that there is no other independent relation of this type between these eigen- ψ 's, since if there were others we could eliminate

some of the ψ_r 's, which would leave a single relation of this type between the remainder. Multiplying (5) by α , we find

$$0 = \alpha \sum_r c_r \psi_r = \sum_r c_r \alpha \psi_r = \sum_r c_r a_r \psi_r, \quad (6)$$

if a_r is the eigenvalue belonging to ψ_r . Now (6) is a linear relation between the ψ_r 's with numerical coefficients and therefore, by hypothesis, must not be independent of (5). This requires that the a_r 's shall all be equal, so that the ψ_r 's must all belong to the same eigenvalue.

This theorem could have been inferred from the definition of superposition in § 6 together with the physical meaning of eigenstates. A relation of the type (5) implies that one of the eigenstates, ψ_1 say, is obtainable by superposition of the others $\psi_2, \psi_3 \dots$, so that any result that can be obtained from an observation of the system in state ψ_1 must have a finite probability of being the result when the observation is made on the system in at least one of the states $\psi_2, \psi_3 \dots$. This would not be the case if the observation consisted in the measurement of the observable α when the ψ_r 's all belong to different eigenvalues of α . Thus a relation of the type (5) is impossible.

§ 14. The Expansion Theorem

The expansion theorem of the theory of eigenvalues asserts that *an arbitrary ψ -symbol can be expanded in terms of eigen- ψ 's of any real observable*, thus

$$\psi = \sum_p \psi_p, \quad (7)$$

where the ψ_p 's are eigen- ψ 's of a real observable α . Such an expansion must be unique, since otherwise there would be a relation of the type (5) between eigen- ψ 's belonging to different eigenvalues. If the eigenvalues of α do not form a discrete set of numbers but a continuous range, or if they form both a continuous range and a discrete set, then the number of eigen- ψ 's occurring in (7) may be more than an enumerable number and equal to the number of points on a line. In such a case we may require an integral of the type

$$\psi = \int \psi_p dp \quad (8)$$

in order to express the general ψ , or we may require both a sum and an integral. The theory of ψ -symbols developed in the previous chapter does not give any rigorous definition for an integral of the type (8). In order to get such a definition one would have to introduce a number of new assumptions concerning limits and continuity for the ψ -symbols, which would be beyond the scope of the present work. For all physical purposes it is sufficient for one not to aim at

a rigorous theory when dealing with such things, but to content oneself with making use of rough intuitive notions about limits and continuity, such as could be obtained, for instance, from the vector picture of the ψ 's. These intuitive notions show that if one has a ψ -symbol ψ_p that involves a parameter p in some reasonably continuous way, one can differentiate or integrate ψ_p with respect to p and the result will be another ψ -symbol.

Under these circumstances one cannot, of course, attempt to give a rigorous deduction of the expansion theorem from the symbolic algebra. The following argument, however, makes the theorem appear plausible. Consider the ψ -symbol ψ_τ that is a function of the parameter τ and that satisfies the differential equation

$$\frac{\partial}{\partial \tau} \psi_\tau = i\alpha \psi_\tau. \quad (9)$$

If ψ_τ is given for one value of τ , then this equation fixes ψ_τ for a slightly greater value of τ . Thus we should expect this equation to have one solution, and only one, for any given initial value for ψ_τ , i.e. for ψ_τ equal to an arbitrary ψ_0 when $\tau = 0$. Suppose now that this solution can be expressed as a Fourier series or integral in τ , thus, if we take for definiteness the case of the integral,

$$\psi_\tau = \int e^{ip\tau} \psi_p \, dp, \quad (10)$$

where ψ_p is independent of τ , but involves the new parameter p . Substituting this expression for ψ_τ in (9), we obtain

$$\int ip e^{ip\tau} \psi_p \, dp = i\alpha \int e^{ip\tau} \psi_p \, dp$$

or

$$\int p e^{ip\tau} \psi_p \, dp = \int e^{ip\tau} \alpha \psi_p \, dp.$$

Since this equation holds for all values of τ we can equate coefficients of $e^{ip\tau}$, which gives

$$p \psi_p = \alpha \psi_p.$$

Thus ψ_p is an eigen- ψ of α belonging to the eigenvalue p . If we now put $\tau = 0$ in (10), we obtain

$$\psi_0 = \int \psi_p \, dp,$$

which expresses the arbitrary ψ_0 in terms of the eigen- ψ 's ψ_p in the form (8). The discrete terms such as occur in (7) would arise when the Fourier expansion (10) requires terms of a Fourier series.

The weak point in the above argument is the assumption of the possibility of a Fourier expansion (10) for ψ_τ . If one takes the vector picture and considers ψ_τ to be a vector varying continuously with τ , one would expect some kind of Fourier expansion to be possible,

except when the magnitude of the vector tends to infinity as $\tau \rightarrow \infty$, a possibility that may very well occur with an equation of motion of the type (9). One can, however, exclude this possibility by making use of the fact that α is a real observable. (For an observable that is not real the expansion theorem is not necessarily true.) If ϕ_τ is the ϕ -symbol that is conjugate imaginary to ψ_τ , it will satisfy the conjugate imaginary differential equation to (9), which is

$$\frac{\partial}{\partial \tau} \phi_\tau = -i\phi_\tau \alpha.$$

Hence

$$\begin{aligned} \frac{\partial}{\partial \tau} (\phi_\tau \psi_\tau) &= \phi_\tau \frac{\partial \psi_\tau}{\partial \tau} + \frac{\partial \phi_\tau}{\partial \tau} \psi_\tau \\ &= \phi_\tau i\alpha \psi_\tau - i\phi_\tau \alpha \psi_\tau = 0. \end{aligned} \quad (11)$$

Thus the square of the modulus of the vector ψ_τ , which is $\phi_\tau \psi_\tau$, remains constant.

From the above non-rigorous discussion one would expect the expansion theorem to follow rigorously from the symbolic algebra with the addition of suitable axioms about limits and continuity. The corresponding theorem for ϕ 's must then, of course, also hold. Throughout the rest of this chapter we shall, for definiteness, assume the expansions we have to deal with involve sums and not integrals. The theorems to be proved would still be true for integrals, only formal alterations in the proofs being required. These formal alterations would, however, require a new notation, and this will be given in the next chapter (see § 22).

§ 15. Functions of an Observable

The expansion theorem enables one to give a definition of a function of a real observable of the same degree of generality as that of an ordinary function of a real variable. Let α be a real observable and let ψ_p be one of its eigen- ϕ 's, belonging to the eigenvalue a_p , so that

$$\alpha \psi_p = a_p \psi_p.$$

It is evident, as was mentioned in § 11, that if $f(x)$ denotes any function of x expressible as a power series, then

$$f(\alpha) \psi_p = f(a_p) \psi_p. \quad (12)$$

We can assume that this relation holds for more general functions. If $f(x)$ denotes any function of the real variable x whose domain includes the point $x = a_p$, then the right-hand side of (12) has a meaning and we can define $f(\alpha) \psi_p$ by this right-hand side. If there

are several eigen- ψ 's belonging to the same eigenvalue a_p , say ψ'_p , $\psi''_p \dots$, so that there can exist linear relations between them of the type

$$\sum c' \psi'_p = 0,$$

where the coefficients c' are numbers, then the definition (12) is self-consistent, since it gives

$$f(\alpha) \sum c' \psi'_p = \sum c' f(\alpha) \psi'_p = \sum c' f(a_p) \psi'_p = 0.$$

Thus if the domain of the function $f(x)$ includes all the eigenvalues of α , we can give a meaning to $f(\alpha)$ multiplied into any eigen- ψ of α . Further, we can give a meaning to $f(\alpha)$ multiplied into an arbitrary ψ , since we can expand this arbitrary ψ in terms of eigen- ψ 's and multiply $f(\alpha)$ into each term of the expansion separately.

Thus *one can give a meaning to $f(\alpha)$ when $f(x)$ is any function of the real variable x , even an irregular or discontinuous one, whose domain includes all the eigenvalues of α .* If this domain contains other points besides the eigenvalues of α , then the values of $f(x)$ for these other points will not affect $f(\alpha)$. These results are a necessary consequence of the physical meaning of eigenvalues. If α is an observable quantity, then $f(\alpha)$ must also be observable when $f(x)$ is any function of the real variable x that has a meaning for all values of x that are possible results of the observation of α , *i.e.* all eigenvalues of α , since the same apparatus and experiment that measure α really also measure $f(\alpha)$.

It follows from (12) that every eigen- ψ of α is an eigen- ψ of $f(\alpha)$. The converse, that every eigen- ψ of $f(\alpha)$ is an eigen- ψ of α , is not true, except when α is a function (a single-valued function is of course understood) of $f(\alpha)$. Also it follows that the eigenvalues of $f(\alpha)$ are just this function f of the eigenvalues of α , *e.g.* the eigenvalues of α^2 are the squares of those of α . These results are obviously necessary for the physical meanings of eigenvalues and eigenstates to be tenable. Again, it may easily be deduced from the definition (12) that the sum or product of two functions of an observable is a function of that observable and that a function of a function of an observable is a function of that observable, which results are also physically necessary.

We can use the eigen- ϕ 's instead of the eigen- ψ 's in order to define $f(\alpha)$. We then have

$$\phi_p f(\alpha) = f(a_p) \phi_p,$$

where ϕ_p is any eigen- ϕ of α . This equation is, according to § 10, just the conjugate imaginary equation to (12) and is thus deducible from (12). The two definitions of $f(\alpha)$ are therefore equivalent.

The theorem will now be proved that *any observable that commutes with α commutes also with $f(\alpha)$* . This theorem is of course obvious when f is expressible as a power series. Let β be any observable that commutes with α , i.e. that satisfies $\beta\alpha = \alpha\beta$. Let ψ_p be an eigen- ψ of α belonging to the eigenvalue a_p and let ϕ_q be an eigen- ϕ belonging to the eigenvalue a_q , which may or may not equal a_p , so that

$$\alpha\psi_p = a_p\psi_p \quad \phi_q\alpha = a_q\phi_q.$$

We now have

$$\phi_q\beta\alpha\psi_p = \phi_q\beta a_p\psi_p = a_p\phi_q\beta\psi_p.$$

Again

$$\phi_q\beta\alpha\psi_p = \phi_q\alpha\beta\psi_p = a_q\phi_q\beta\psi_p.$$

Hence

$$(a_p - a_q)\phi_q\beta\psi_p = 0,$$

so that either $\phi_q\beta\psi_p = 0$, or $a_p = a_q$, which would give $f(a_p) = f(a_q)$. Thus in either case

$$[f(a_p) - f(a_q)]\phi_q\beta\psi_p = 0.$$

Now

$$\phi_q\beta f(\alpha)\psi_p = \phi_q\beta f(a_p)\psi_p = f(a_p)\phi_q\beta\psi_p$$

and again

$$\phi_q f(\alpha)\beta\psi_p = f(a_q)\phi_q\beta\psi_p.$$

Hence

$$\phi_q[\beta f(\alpha) - f(\alpha)\beta]\psi_p = [f(a_p) - f(a_q)]\phi_q\beta\psi_p = 0.$$

This result is true for any eigen- ψ , ψ_p , and is hence also true for an arbitrary ψ , which can be expanded in terms of eigen- ψ 's. Similarly it is true for any eigen- ϕ , ϕ_q , and is hence also true for an arbitrary ϕ , which can be expanded in terms of eigen- ϕ 's. Hence

$$\beta f(\alpha) - f(\alpha)\beta = 0,$$

which is the result required. In this proof it is not assumed that β is a real observable, although, of course, it is understood that α is real in order that a general function of α may have a meaning.

The converse theorem will now be proved, namely, *if every observable that commutes with a real observable α also commutes with another observable f , then f is a function of α* . It will first be shown that if ψ_p is any eigen- ψ of α , then it is also an eigen- ψ of f . We introduce an observable β satisfying the following conditions:

$$\beta\psi_q = 0,$$

whenever ψ_q is an eigen- ψ of α belonging to an eigenvalue a_q that differs from that of ψ_p , which is a_p ;

$$\beta\psi_p = \psi_p$$

and

$$\beta\psi'_p = 0,$$

whenever ψ'_p is one of a set of eigen- ψ 's of α belonging to the eigenvalue a_p , such that this set, together with ψ_p , form a complete

independent set of all eigen- ψ 's belonging to the eigenvalue a_p . We shall then have that ψ_p , the ψ'_p 's, and ψ_q 's form a complete set of independent ψ 's, so that β is completely defined by these equations. It is now easily verified that

$$\begin{aligned}\alpha\beta\psi_q &= 0 = \beta\alpha\psi_q \\ \alpha\beta\psi_p &= a_p\psi_p = \beta\alpha\psi_p \\ \alpha\beta\psi'_p &= 0 = \beta\alpha\psi'_p.\end{aligned}$$

Thus

$$\alpha\beta\psi = \beta\alpha\psi$$

for arbitrary ψ and β commutes with α . Hence, by hypothesis, β also commutes with f , so that

$$\beta f\psi_p = f\beta\psi_p = f\psi_p.$$

Now for an arbitrary ψ -symbol ψ one must have

$$\beta\psi = c\psi_p,$$

where c is a number, as one can easily see by expanding ψ in terms of ψ_p , the ψ'_p 's and ψ_q 's, and multiplying β into each term separately.

Hence

$$\beta f\psi_p = c\psi_p,$$

so that

$$f\psi_p = c\psi_p$$

and ψ_p is an eigen- ψ of f . To complete the proof that f is a function of α according to the above definition, it remains to be shown only that if two or more eigen- ψ 's belong to the same eigenvalue of α , then they also belong to the same eigenvalue of f . The functional relation between the eigenvalues of f and those of α will then specify the function that f is of α . Now if two or more eigen- ψ 's of α belong to the same eigenvalue of α , then any linear combination of them will be an eigen- ψ of α . From what has already been proved it follows that this linear combination must also be an eigen- ψ of f , which can be the case only if the eigen- ψ 's, that it is a linear combination of, all belong to the same eigenvalue of f .

§ 16. Examples of Functions of Observables

Some examples of elementary functions of a real observable α will now be considered. The reciprocal α^{-1} always exists when α has not the eigenvalue zero. By definition it satisfies

$$\alpha^{-1}\psi_p = a_p^{-1}\psi_p,$$

where ψ_p is an eigen- ψ of α belonging to the eigenvalue a_p . Hence

$$\alpha\alpha^{-1}\psi_p = \alpha a_p^{-1}\psi_p = \psi_p$$

and since this is true for all ψ_p we must have $\alpha\alpha^{-1} = 1$. Similarly $\alpha^{-1}\alpha = 1$. Either of these equations is sufficient to determine α^{-1} completely when this reciprocal exists according to the above definition. To prove this result, suppose there are two solutions, $(\alpha^{-1})_1$ and $(\alpha^{-1})_2$, of $\alpha\alpha^{-1} = 1$; so that

$$\alpha(\alpha^{-1})_1 = 1 \quad \alpha(\alpha^{-1})_2 = 1.$$

This gives

$$\alpha\xi = 0, \tag{13}$$

where

$$\xi = (\alpha^{-1})_1 - (\alpha^{-1})_2.$$

If α is such that there exists a ξ , not identically zero, satisfying (13), then α can have no reciprocal, according to the above definition, since if such a reciprocal α^{-1} exists we obtain, by multiplying (13) on the left-hand side by α^{-1} ,

$$0 = \alpha^{-1}\alpha\xi = \xi.$$

Hence $\xi = 0$ and our two solutions of $\alpha\alpha^{-1} = 1$ are identical.

As a second example we shall take the square root of α . This is defined by

$$\sqrt{\alpha}\psi_p = \pm\sqrt{a_p}\psi_p. \tag{14}$$

The square root of α always exists, but is a real observable only provided α has no negative eigenvalues. From (14) one obtains

$$\sqrt{\alpha}\sqrt{\alpha}\psi_p = \sqrt{a_p}\sqrt{a_p}\psi_p = \alpha\psi_p,$$

so that

$$\sqrt{\alpha}\sqrt{\alpha} = \alpha. \tag{15}$$

On account of the ambiguity of sign in (14), the square root of an observable is to a certain extent indeterminate. In order to determine a square root completely one must choose a particular sign for each eigenvalue a_p to insert in (14), which is the same as fixing the sign of the square root of a real variable whose domain consists of the eigenvalues a_p . One can choose the sign to vary as irregularly as one likes in passing from one eigenvalue to the next, and equation (14) will always define an observable $\sqrt{\alpha}$ satisfying (15) that can legitimately be called a square root of α . If the observable α has two eigen- ψ 's belonging to one and the same eigenvalue a_q , then we could define an observable $\sqrt{\alpha}$ by equation (14) with the + sign for one of these eigen- ψ 's and the - sign for the other, and with arbitrary signs for the eigen- ψ 's belonging to eigenvalues other than a_q . This observable would still satisfy (15), but it would not be a function of the observable α in accordance with our definition, which requires a unique coefficient on the right-hand side of (14) for each eigenvalue a_p , so that this coefficient will form a single-valued function of the

real variable a_p . The $\sqrt{\alpha}$ defined without this unique coefficient would not, for instance, satisfy the condition of commuting with any observable that commutes with α . Thus, unlike what we had in the case of the reciprocal, equation (15) is not sufficient for the definition of square-root functions, but must be supplemented by the condition that the observable that is being defined is actually a function of α . The number of different square-root functions is 2^n where n is the number of different eigenvalues of α . The most useful one is usually that, which exists only when all the eigenvalues of α are positive, for which the positive sign is taken in every case.

As an example of a non-analytical function we may take the modulus $|\alpha|$ of the observable α . This is defined by

$$|\alpha|\psi_p = |a_p|\psi_p$$

and is quite a proper observable, in spite of the fact that the corresponding function of a real variable is discontinuous, and may be used freely in the analysis when desired.

§ 17. Simultaneous Eigenstates

A state ψ may be simultaneously an eigenstate of two observables α and β , i.e. it may satisfy both

$$\alpha\psi = a\psi$$

and

$$\beta\psi = b\psi,$$

where a and b are numbers. We should then have

$$\alpha\beta\psi = ab\psi = \beta\alpha\psi$$

or

$$(\alpha\beta - \beta\alpha)\psi = 0.$$

This suggests that the chances for the existence of a simultaneous eigenstate of two observables α and β are most favourable when $\alpha\beta - \beta\alpha = 0$, i.e. when α and β commute. When α and β do not commute the possibility for the existence of a simultaneous eigenstate is not absolutely ruled out, but the occurrence of such a state is exceptional. On the other hand, when α and β commute there exist so many simultaneous eigenstates, that, as will now be proved, *an arbitrary state can be expanded in terms of them*. We thus get a generalization of the expansion theorem of § 14.

Let α and β be two observables that commute and let ψ_a be an eigen- ψ of α belonging to the eigenvalue a . By the expansion theorem of § 14 we can expand ψ_a in terms of eigen- ψ 's of β , thus

$$\psi_a = \sum_b \psi_b, \quad (16)$$

where ψ_b is an eigen- ψ of β belonging to the eigenvalue b . It will now be proved that each ψ_b in this expansion is an eigen- ψ also of α and is thus a simultaneous eigen- ψ of α and β . If $f(\beta)$ is any function of the observable β , we have

$$\begin{aligned}\alpha f(\beta)\psi_a &= \alpha \sum_b f(\beta)\psi_b \\ &= \alpha \sum_b f(b)\psi_b\end{aligned}$$

from the definition of a function given in § 15. Now from a theorem of § 15, since α commutes with β it must also commute with $f(\beta)$, so that

$$\begin{aligned}\alpha f(\beta)\psi_a &= f(\beta)\alpha\psi_a = f(\beta)a\psi_a \\ &= af(\beta) \sum_b \psi_b = a \sum_b f(b)\psi_b.\end{aligned}$$

Hence

$$\alpha \sum_b f(b)\psi_b = a \sum_b f(b)\psi_b. \quad (17)$$

Now $f(b)$ is an arbitrary function of the real variable b , so that for each value of b in the domain of b , $f(b)$ is an arbitrary number. Hence we can equate coefficients of $f(b)$ in (17), which gives

$$\alpha\psi_b = a\psi_b.$$

Thus each of the ψ_b 's in the expansion (16) is an eigen- ψ of α belonging to the same eigenvalue a as that of our original ψ_a and is thus a simultaneous eigen- ψ of α and β . Any eigen- ψ ψ_a of α can therefore be expanded in terms of these simultaneous eigen- ψ 's. But an arbitrary ψ can be expanded in terms of ψ_a 's, and hence an arbitrary ψ can be expanded in terms of simultaneous eigen- ψ 's.

The converse theorem is also easily proved, namely, *if two observables α and β are such that an arbitrary ψ can be expanded in terms of the simultaneous eigen- ψ 's of α and β , then α and β commute*. We have, in fact, if ψ_{ab} is a simultaneous eigen- ψ of α and β belonging to the eigenvalues a and b respectively, the equation

$$(\alpha\beta - \beta\alpha)\psi_{ab} = (ab - ba)\psi_{ab} = 0.$$

Hence

$$(\alpha\beta - \beta\alpha)\psi = 0,$$

where ψ is any ψ -symbol that can be expanded in terms of the ψ_{ab} 's. If this is true for an arbitrary ψ , we can infer that

$$\alpha\beta - \beta\alpha = 0,$$

as required.

The idea of simultaneous eigen- ψ 's may obviously be extended to more than two observables and the theorem just proved still holds, i.e. an arbitrary ψ can be expanded in terms of the simultaneous eigen- ψ 's of any set of observables that commute, and also its converse. The same arguments used for the proof in the case of two

observables are adequate for the general case, e.g. if we have three observables α, β, γ that commute, each with the other two, we can expand any simultaneous eigen- ψ of α and β in terms of eigen- ψ 's of γ and then show that each of these eigen- ψ 's of γ is also an eigen- ψ of α and β .

The fact that there is an expansion theorem for two or more observables that commute, the same as that for a single observable, means that a set of two or more observables that commute has many of the properties of a single observable and can for many purposes be counted as a single observable, the result of a measurement of which is expressible by two or more numbers. Thus the theory of functions of a single observable developed in § 15 can be applied without change to functions of two or more observables that commute. If $\alpha, \beta, \gamma, \dots$ are a set of observables that commute, we define a general function of them, $f(\alpha \beta \gamma \dots)$, by

$$f(\alpha \beta \gamma \dots) \psi_{abc\dots} = f(a b c \dots) \psi_{abc\dots},$$

where $\psi_{abc\dots}$ is a simultaneous eigen- ψ of $\alpha, \beta, \gamma, \dots$ belonging to the eigenvalues a, b, c, \dots respectively, and $f(a, b, c, \dots)$ is a function of the real variables a, b, c, \dots whose domains consist of the eigenvalues of $\alpha, \beta, \gamma, \dots$ respectively. The theorems given in § 15 about functions of single observables will apply also to functions of sets of observables that commute, the proofs being formally equivalent in the two cases. For example, we shall have the theorem that any observable that commutes with each of a set of commuting observables $\alpha, \beta, \gamma, \dots$ will commute also with any function of them, $f(\alpha \beta \gamma \dots)$.

If we take the maximum possible number of independent observables that commute, the condition of independence being that no one of them can be expressed as a function of the others, then there cannot be more than one simultaneous eigenstate for them all belonging to a specified set of eigenvalues. To prove this result, let α_r be the set of commuting observables and suppose there are two independent simultaneous eigen- ψ 's, ψ_1 and ψ_2 , of all the α_r 's belonging to the same set of eigenvalues. Introduce the new observable β defined by

$$\beta\psi_1 = \psi_1, \quad \beta\psi_2 = 0, \quad \beta\psi_3 = 0,$$

whenever ψ_3 is a simultaneous eigen- ψ belonging to a different set of eigenvalues. Then this β commutes with all the α 's and also it is not a function of them, as may be seen from the fact that any linear

combination of ψ_1 and ψ_2 is a simultaneous eigen- ψ of all the α 's but is not an eigen- ψ of β , so that the set of α 's does not contain the maximum possible number of independent commuting observables. Hence, when the set of α 's does satisfy the given conditions, each eigenstate must be uniquely determined by the eigenvalues to which it belongs. Such a set we call a *complete* set of commuting observables.

§ 18. Some Probability Theorems

We shall now determine *the probability of a given result being obtained when an observation is made on the system in a given state*. For this purpose the only physical assumption we shall make use of is that given in § 11 for the average value of an observable. To determine the probability that an observable shall be found to have the value a when a measurement of it is made for the system in a state ψ , we use the fact that if a measurement is made of $f(\alpha)$, any function of α , the average result obtained will be

$$\phi f(\alpha)\psi,$$

where ϕ is the conjugate imaginary of ψ , provided ϕ and ψ are normalized. Suppose ϕ and ψ to be expanded in terms of eigen- ϕ 's and eigen- ψ 's, thus

$$\phi = \sum_a \phi_a, \quad \psi = \sum_{a'} \psi_{a'}, \quad (18)$$

where ϕ_a belongs to the eigenvalue a and $\psi_{a'}$ to a' . The expression for the average of $f(\alpha)$ now becomes

$$\begin{aligned} \sum_a \phi_a f(\alpha) \sum_{a'} \psi_{a'} &= \sum_{aa'} f(a) \phi_a \psi_{a'} \\ &= \sum_a f(a) \phi_a \psi_a, \end{aligned} \quad (19)$$

when we use the theorem of § 13 that eigenstates belonging to different eigenvalues are orthogonal. Now if $P(a)$ is the probability of the observable α being found to have the value a , the average value of $f(\alpha)$ must be $\sum_a f(a)P(a)$, since the ordinary probability rules will apply in this case. Equating this expression to (19), we find

$$\sum_a f(a)P(a) = \sum_a f(a) \phi_a \psi_a.$$

This holds when $f(a)$ is an arbitrary function of the real variable a , so that we must be able to equate coefficients of $f(a)$, which gives

$$P(a) = \phi_a \psi_a. \quad (20)$$

We can easily verify that this expression for $P(a)$ gives unity for the total probability of α having any value, since from the normalizing condition for ϕ and ψ we find

$$\sum_a \phi_a \sum_{a'} \psi_{a'} = 1,$$

which reduces to

$$\sum_a \phi_a \psi_a = 1.$$

We can put the expression (20) in a different form by inserting numerical coefficients in the expansions (18) so that they read

$$\phi = \sum_a \bar{c}_a \phi_a, \quad \psi = \sum_{a'} c_{a'} \psi_{a'},$$

and taking the ϕ_a 's and ψ_a 's to be normalized. We then get for $P(a)$

$$P(a) = \bar{c}_a \phi_a c_a \psi_a = |c_a|^2,$$

so that the probability of α having any given value is equal to the square of the modulus of the corresponding coefficient in the expansion.

From this it follows at once that if the state ψ is an eigenstate belonging to the eigenvalue a , the probability of α having the value a is unity. Thus the result that if $\alpha\psi = a\psi$, α certainly has the value a for the state ψ , is deducible from the general assumption for the average value of an observable. A second immediate consequence is that any result, a say, for an observation of α on the system in the state $c_1\psi_1 + c_2\psi_2$ has a finite probability of being the result when this observation is made for either state ψ_1 or state ψ_2 , since if the term belonging to the eigenvalue a in the expansion of $c_1\psi_1 + c_2\psi_2$ in eigen- ψ 's of α does not vanish, that in the expansion either of ψ_1 or of ψ_2 must also not vanish. This shows that the definition of superposition given in § 6 is equivalent to that contained in the symbolic algebra, together with the interpretation of this algebra that $\phi\alpha\psi$ is the average of α .

The results we have just obtained all remain true when we replace the observable α by a set of two or more observables that commute, the proofs being formally unaltered. Thus, we shall have that if ψ is expanded in terms of simultaneous eigen- ψ 's of two observables, α and β , that commute, i.e. $\psi = \sum_{ab} \psi_{ab}$,

where ψ_{ab} is a simultaneous eigen- ψ belonging to the eigenvalues a and b for α and β respectively, then the probability that the results a and b shall be obtained from measurements of α and β for the state ψ will be $\phi_{ab} \psi_{ab}$ when ψ is normalized. The existence of a definite probability for these results, independent of the order in which the observations are made, requires that the observations shall not interfere with each other and suggests that *the condition that two observables commute is equivalent to the condition that the two observations are compatible*. A formal proof of this will now be given. Before we can do

this we must obtain a mathematical form for the condition that an observation is made with the minimum of disturbance, which we have hitherto discussed only qualitatively.

Consider an observation, consisting of the measurement of an observable α , to be made on a system in the state ψ . The state of the system after the observation must be an eigenstate of α , since the result of a measurement of α for this state must be a certainty. Now suppose the observation to be made in such a way that the state of the system afterwards is always one of those that occur in the expansion of the initial ψ in terms of eigen- ψ 's of α , i.e. one of the ψ_a 's in

$$\psi = \sum_a \psi_a.$$

This is permissible since there is one eigen- ψ ψ_a in the expansion for every eigenvalue a that has a finite probability of being the result of the observation. This observation of α may then conveniently be defined to be the one that causes the minimum of disturbance to the system. Observations that cause the minimum disturbance are thus those with the property that, by a superposition of all the possible states after the observation, the state before the observation may be formed, or those with the property that any result that can be obtained from any observation on the system in the initial state is a possible result when the same observation is made on the system in one of the final states. It is observations with this property that should be understood in the discussion on compatibility in § 4. Granting the existence of observations with this property, there is a physical necessity for the expansion theorem of § 14.

Now let α and β be two observables that commute and let any state ψ be expanded in terms of simultaneous eigen- ψ 's of α and β , thus

$$\psi = \sum_{ab} \psi_{ab}.$$

The expansion of ψ in terms of eigen- ψ 's of α must then be

$$\psi = \sum_a \psi_a, \quad (21)$$

where

$$\psi_a = \sum_b \psi_{ab}, \quad (22)$$

and similarly the expansion of ψ in terms of eigen- ψ 's of β must be

$$\psi = \sum_b \psi_b, \quad (23)$$

where

$$\psi_b = \sum_a \psi_{ab}. \quad (24)$$

The suffixes in each case denote the corresponding eigenvalues. If ψ is normalized, then the probability for this state of the result b

being obtained from a measurement of β will be $\phi_b \psi_b$. When this result is obtained, the state of the system after the observation will be ψ_b , if the observation is made with the minimum of disturbance according to the above definition. If an observation is now made of α for this final state ψ_b , the probability of the result a being obtained will be, from (24),

$$\phi_{ab} \psi_{ab} / \phi_b \psi_b,$$

the denominator arising from the fact that the symbol ψ_b is not normalized. Thus the probability of first the result b being obtained for β and then the result a for α will be, by multiplication, $\phi_{ab} \psi_{ab}$. The total probability of the result a being obtained for the second observation with any result for the first must therefore be

$$\sum_b \phi_{ab} \psi_{ab}.$$

If, now, an observation of α were made on the system in the initial state ψ , with no observation at all of β , the probability of the result a being obtained would be, from (21), $\phi_a \psi_a$. On account of (22), this must equal

$$\sum_b \phi_{ab} \sum_{b'} \psi_{ab'} = \sum_b \phi_{ab} \psi_{ab},$$

from the orthogonality theorem of § 12, which is the same as the probability that the result a shall be obtained for α after an observation of β . This is just the condition that α and β shall be compatible according to § 4.

The converse will now be proved, that if the measurements of two observables α and β are two compatible observations, then α and β commute. It was shown in § 4 that if the compatible observations α and β are both made on the system in any state ψ , the final state will be such that the result for either observation with this state will be a certainty, *i.e.* the final state will be a simultaneous eigenstate for α and β . If the observations are made with the minimum of disturbance according to the above definition, then the initial state ψ must be capable of being expanded in terms of all the possible final states. Thus an arbitrary ψ can be expanded in terms of simultaneous eigen- ψ 's of α and β , so that α and β must commute.

The identification of the condition of commutability of observables with that of the compatibility of the observations allows us to see a physical necessity for the theorem of § 15 that any observable that commutes with an observable α commutes also with $f(\alpha)$, any function of α . This theorem may now be stated in the form that any observation that is compatible with the observation of α is com-

patible also with the observation of $f(\alpha)$ and is thus physically obvious, since any observation of α is in the fact itself also an observation of $f(\alpha)$.

It will now be shown that the fact that the probability of agreement of two states ψ_1 and ϕ_2 is $|\phi_2 \psi_1|^2$, when ψ_1 and ϕ_2 are normalized, is deducible from the general assumption for the average of an observable. It has been shown that from this general assumption one can deduce that the probability of an observable α having the value a for the state ψ_1 is $|c_a|^2$, where c_a is the coefficient of the eigen- ψ belonging to the eigenvalue a in the expansion of ψ_1 in terms of eigen- ψ 's of α ,

$$\psi_1 = \sum c_a \psi_a, \quad (25)$$

when ψ_1 , and all the ψ_a 's are normalized. This result is still true when α denotes a set of commuting observables α_r , and ψ_a is a simultaneous eigen- ψ belonging to the set of eigenvalues a_r . There is one maximum observation, the result of which for the state ψ_2 is a certainty. This maximum observation will consist in the measurement of a set of commuting observables α_r , which set must be a complete set, in the sense defined at the end of the preceding section, if the observation is really a maximum one. The state ψ_2 is then a simultaneous eigen- ψ of all these observables α_r , and there is no other simultaneous eigen- ψ belonging to the same set of eigenvalues as ψ_2 does. That term in the expansion (25) which belongs to the same set of eigenvalues as ψ_2 must therefore be just ψ_2 itself or differ from it by a trivial numerical factor. The probability of agreement of ψ_1 with ψ_2 , which is the probability that the result of the observation of the α 's for state ψ_1 is the same as for state ψ_2 , is therefore $|c_{a2}|^2$, where c_{a2} is the coefficient of that ψ_a in (25) that is just ψ_2 . But from the orthogonality theorem, one finds that $\phi_2 \psi_1$ is equal to just this coefficient c_{a2} , so that the probability of agreement is $|\phi_2 \psi_1|^2$.

§ 19. Contact Transformations

The following important theorem in the theory of eigenvalues will now be proved. *If S is any observable having a reciprocal S^{-1} and α is any observable, then the eigenvalues of $S\alpha S^{-1}$ are the same as those of α .* Let a be any eigenvalue of α and let ψ_a be an eigen- ψ of α belonging to it, so that

$$\alpha\psi_a = a\psi_a.$$

This gives $S\alpha S^{-1}S\psi_a = S\alpha\psi_a = Sa\psi_a = aS\psi_a$.

Hence $S\psi_a$ is an eigen- ψ of $S\alpha S^{-1}$ belonging to the eigenvalue a .

Conversely, as may be shown in a similar way, if α is any eigenvalue of $S\alpha S^{-1}$ and ψ is an eigen- ψ of $S\alpha S^{-1}$ belonging to it, then α is also an eigenvalue of α and $S^{-1}\psi$ is an eigen- ψ of α belonging to it.

It is not necessary for this theorem that S should be a real observable. If S is not real we cannot use the general definition of a function of an observable in order to define S^{-1} , but must use instead the conditions $SS^{-1} = S^{-1}S = 1$, which are sufficient for the proof of the theorem. S can be any observable such that there exists an S^{-1} satisfying these conditions. It is also not necessary for the theorem to be true that α should be a real observable, but since the only eigenvalues of interest in quantum mechanics are those of real observables, the theorem is useful only when both α and $S\alpha S^{-1}$ are real. This imposes a condition on S . If $S\alpha S^{-1}$ is to be real whenever α is real, we must have, from the rule (22) of § 10,

$$S\alpha S^{-1} = \overline{S\alpha S^{-1}} = \overline{S^{-1}\alpha S} = \overline{S^{-1}\alpha S},$$

which requires, ignoring possible trivial numerical factors,

$$S^{-1} = \overline{S}, S = \overline{S^{-1}}.$$

Either of these conditions is a consequence of the other.

When S satisfies these conditions, the transformation from a set of observables α_r to the set $\beta_r = S\alpha_r S^{-1}$ is called a contact transformation of observables, since, as we shall see later, it is analogous to a contact transformation of classical mechanics. Each of the new observables β_r has the same eigenvalues as the corresponding original one α_r . Further, the transformation has other remarkable properties, namely, *if any algebraic relation holds between some of the α 's, the same relation holds between the corresponding β 's, and if one of the α 's is a function of another one according to the general definition, the same functional relation holds between the corresponding β 's*.

To prove the first of these two properties, we observe that any algebraic relation between the α 's may be written in a rational integral form of the type

$$\sum c\alpha_p \alpha_q \dots \alpha_z = 0,$$

the summation consisting of an arbitrary number of terms, each consisting of an arbitrary number of factors, and the c 's being arbitrary numerical coefficients. From this we deduce, by multiplying by S on the left and S^{-1} on the right, the result

$$\sum cS\alpha_p \alpha_q \dots \alpha_z S^{-1} = 0$$

or

$$\sum cS\alpha_p S^{-1}S\alpha_q S^{-1}\dots S\alpha_z S^{-1} = 0$$

or

$$\sum c\beta_p \beta_q \dots \beta_z = 0$$

which is the result required.

To prove the second of these properties, suppose $\alpha_2 = f(\alpha_1)$, where $f(a)$ is a function of the real variable a defined for each of the eigenvalues of α_1 . Since these are also the eigenvalues of β_1 , we can give a meaning to $f(\beta_1)$. Let ψ_a be an eigen- ψ of α_1 , belonging to the eigenvalue a . We then have

$$f(\alpha_1)\psi_a = f(a)\psi_a. \quad (26)$$

But $S\psi_a$ must be an eigen- ψ of $S\alpha_1 S^{-1}$ or β_1 , belonging to the eigenvalue a of β_1 , so that we must also have

$$f(\beta_1)S\psi_a = f(a)S\psi_a. \quad (27)$$

Multiplying (26) by S on the left, we obtain

$$\begin{aligned} Sf(\alpha_1)S^{-1}S\psi_a &= Sf(a)\psi_a \\ &= f(\beta_1)S\psi_a \end{aligned} \quad (28)$$

from (27). Now $S\psi_a$ is an arbitrary eigen- ψ of β_1 , so that any ψ can be expanded in terms of $S\psi_a$'s. Hence we can equate coefficients of $S\psi_a$ in (28), which gives

$$f(\beta_1) = Sf(\alpha_1)S^{-1} = S\alpha_2 S^{-1} = \beta_2$$

as required.

If two contact transformations are applied successively, the result is another contact transformation. To see how this comes about, consider the transformation $\beta_r = S\alpha_r S^{-1}$ from the α 's to the β 's and the transformation $\gamma_r = T\beta_r T^{-1}$ from the β 's to the γ 's. We have then

$$\gamma_r = TS\alpha_r S^{-1}T^{-1}.$$

Now

$$(TS)(S^{-1}T^{-1}) = 1$$

and

$$(S^{-1}T^{-1})(TS) = 1,$$

so that we can put

$$S^{-1}T^{-1} = (TS)^{-1}.$$

The connexion between the α 's and γ 's now becomes

$$\gamma_r = TS\alpha_r(TS)^{-1},$$

which is a contact transformation.

If the observable S in the transformation $\beta = S\alpha S^{-1}$ differs from unity only by an infinitesimal, we get an infinitesimal contact transformation. Suppose

$$S = 1 + iA,$$

where A is very small, so that its square may be neglected. (A small

observable is one whose eigenvalues are all small, or whose average for any state is small.) We then have

$$S^{-1} = 1 - iA,$$

since this gives $SS^{-1} = S^{-1}S = 1$ with neglect of A^2 . The transformation equation now becomes

$$\beta = (1 + iA)\alpha(1 - iA),$$

which gives $\beta - \alpha = i(A\alpha - \alpha A)$, (29)

with neglect of A^2 . This is the standard form for an infinitesimal contact transformation. In order that $\beta - \alpha$ may be a real observable when α is real, A must be a real observable.

As an example of contact transformation theory, we shall obtain some more information about the observables p and q of § 12, satisfying equation (26) of that section. We apply the theorem that p has the same eigenvalues as $S_p S^{-1}$, taking for S the expression e^{icq} , where c is a real number, which makes $S^{-1} = \bar{S}$. We now find

$$S_p S^{-1} = e^{icq} p e^{-icq} = (p - c) e^{icq} e^{-icq} = p - c,$$

with the help of equation (28) of § 12. Thus p has the same eigenvalues as $p - c$, which are just c less than those of p , so that if a is any eigenvalue of p , $a - c$ must be another. This is true for arbitrary c , so that p must have as eigenvalues all numbers from $-\infty$ to $+\infty$. Similarly it may be proved that q has as eigenvalues all numbers from $-\infty$ to $+\infty$. These results are necessary consequences of the single algebraic condition $qp - pq = i$.

IV

REPRESENTATIONS OF STATES AND OBSERVABLES

§ 20. General Properties

In the two preceding chapters we dealt with certain abstract symbols, denoting states and observables, whose exact nature was not specified, but which were assumed to obey certain definite laws. In the present chapter we shall consider *representations* of these abstract symbols, *i.e.* sets of numbers having properties that correspond completely to those of the symbols they represent. When once one has found such a representation and has understood the nature of the correspondence, one can obtain all the properties of the abstract symbols that one wants by dealing entirely with their representatives, to which, since they are just sets of numbers, ordinary mathematical methods apply. One cannot, of course, obtain in this way any relation between the abstract symbols that one could not obtain directly from the algebra of the abstract symbols without the help of a representation. One can, however, often obtain results much more easily and conveniently with the help of a representation than without it, and further the numbers occurring in a representation have often a very direct physical interpretation, so that representations are of great use in applications of the theory.

Suppose we have a complete set of independent ψ 's, the general member of the set being denoted by ψ_p . The fact that the set is complete means that every ψ can be expressed as a sum of members of the set in the form

$$\psi = \sum_p a_p \psi_p, \quad (1)$$

where the coefficients a_p are numbers. The fact that the members of the set are independent requires that an expansion of the form (1) is unique, since if an alternative expansion

$$\psi = \sum_p a'_p \psi_p$$

were possible, we could obtain by subtraction

$$0 = \sum_p (a_p - a'_p) \psi_p,$$

which can be true with independent ψ_p 's only if $a_p = a'_p$ for all p . Thus according to (1) each ψ determines uniquely a set of numbers a_p and, conversely, each set of arbitrary numbers a_p determines a ψ . There is a one-one correspondence between the ψ 's and the sets of numbers a_p .

If ψ_a corresponds to the set of numbers a_p and ψ_b to the set b_p , we have

$$\psi_a = \sum_p a_p \psi_p, \quad \psi_b = \sum_p b_p \psi_p,$$

and hence

$$\psi_a + \psi_b = \sum_p (a_p + b_p) \psi_p,$$

so that $\psi_a + \psi_b$ corresponds to the set $(a_p + b_p)$. Also, if c is any number, $c\psi_a$ corresponds to the set ca_p . Thus all the properties of the ψ 's of addition and multiplication by numbers are possessed also by the sets of numbers a_p corresponding to them. The sets of numbers thus form a representation of the ψ 's, each ψ being represented by one set a_p defined by (1). The ψ_p 's will be referred to as *the fundamental ψ 's of the representation*. If we take a different set of fundamental ψ 's, we shall get a different set of numbers to represent each ψ , so that we shall get a new representation. There is one representation for each complete set of independent ψ 's, since they may always be taken as fundamental ψ 's. In the vector picture of the ψ 's the numbers representing any ψ are its co-ordinates relative to certain axes (which may be oblique), which are determined by the fundamental ψ 's. The different representations are then the co-ordinates referred to different axes. A state is defined by the ratios of a set of numbers a_p to each other, since a ψ can be multiplied by an arbitrary number and will still represent the same state.

We shall now consider how an observable α is to be represented. If ψ_q is any fundamental ψ of a representation of ψ 's, we can form the product $\alpha\psi_q$ and expand it in terms of the fundamental ψ 's in the form (1), thus

$$\alpha\psi_q = \sum_p \psi_p \alpha_{pq}, \quad (2)$$

where the coefficients α_{pq} are numbers, which depend of course, as the notation implies, on the suffix q of the ψ on the left-hand side. We have put the coefficients α_{pq} in (2) on the right-hand side of their respective ψ_p 's, instead of following the usual practice of putting coefficients on the left, so that the order of the two suffixes may be more easily remembered. That suffix of α_{pq} which is nearer to the ψ is the same as the suffix of the ψ . This is an example of a rule which will be used very extensively in the future.

Each observable α determines uniquely through equation (2) a set of numbers α_{pq} . Conversely, each set of numbers α_{pq} determines an observable α . There is thus a one-one correspondence between observables α and sets of numbers α_{pq} . These sets of numbers are the representatives of the observables. The correspondence between the

properties of the sets of numbers and those of the observables will now be investigated.

Each set of numbers representing an observable is twofold, on account of the two suffixes, and may most conveniently be written as a matrix array, each number α_{pq} of the set being the element of the matrix in the p -th row and q -th column. Thus each observable is represented by a matrix. The number of rows and columns of the matrices is equal to the number of fundamental ψ 's of the representation and one row and one column correspond to each fundamental ψ . A row and column that correspond to the same fundamental ψ correspond to one another. An element of the matrix that lies in a row and column corresponding to one another, i.e. an element of the type α_{pp} , is called a diagonal element, since all such elements lie on a diagonal of the matrix when the rows and columns are arranged both in the same order.

If an observable α is represented by the matrix α_{pq} and an observable β by the matrix β_{pq} , then it is easily verified that the observable $\alpha + \beta$ is represented by the matrix $\alpha_{pq} + \beta_{pq}$, and the observable $c\alpha$, where c is a number, by $c\alpha_{pq}$. These results may be expressed in symbols by the equations

$$(\alpha + \beta)_{pq} = \alpha_{pq} + \beta_{pq} \quad (3)$$

$$(c\alpha)_{pq} = c\alpha_{pq}, \quad (4)$$

which are the ordinary rules for the addition of matrices and for the multiplication of matrices by numbers. Again, if the product $\alpha\beta$ is represented by the matrix $(\alpha\beta)_{pq}$, we have by definition

$$(\alpha\beta)\psi_q = \sum_p \psi_p (\alpha\beta)_{pq}. \quad (5)$$

But we have also

$$\begin{aligned} (\alpha\beta)\psi_q &= \alpha(\beta\psi_q) = \alpha \sum_r \psi_r \beta_{rq} \\ &= \sum_r (\alpha\psi_r) \beta_{rq} = \sum_{pr} \psi_p \alpha_{pr} \beta_{rq}. \end{aligned} \quad (6)$$

By equating the coefficients of ψ_p in the right-hand sides of (5) and (6), which is permissible since the ψ 's are all independent, we obtain

$$(\alpha\beta)_{pq} = \sum_r \alpha_{pr} \beta_{rq}. \quad (7)$$

Thus the matrix representing $\alpha\beta$ equals the matrix representing α multiplied by the matrix representing β , according to the rule for matrix multiplication. The particular arrangement of the suffixes of α_{pq} chosen in the defining equation (2) is necessary in order that this rule of matrix multiplication may hold. If instead of (2) we had put

$$\alpha\psi_q = \sum_p \alpha_{qp} \psi_p,$$

we should have found for the law of multiplication

$$(\alpha\beta)_{pq} = \sum_r \alpha_{rq} \beta_{pr}, \quad (8)$$

which is not so convenient as (7).

Equations (3), (4), and (7) show that the properties of observables of addition and multiplication are all faithfully reproduced by the properties of the matrices representing them, and justify our saying that the matrices do represent them. Matrices, like observables, satisfy all the laws of ordinary algebra except the commutative law of multiplication.

It has been mentioned that a number may be regarded as a special case of an observable. The matrix representing a number c has its elements c_{pq} defined by $c\psi_q = \sum_p \psi_p c_{pq}$, which gives

$$c_{pp} = c, \quad c_{pq} = 0 \quad (p \neq q).$$

Thus the matrix representing c is a diagonal matrix, *i.e.* all its elements vanish except the diagonal ones, and further all the diagonal elements are equal to c . We can put

$$c_{pq} = c\delta_{pq},$$

where the symbol δ_{pq} is defined by

$$\delta_{pp} = 1, \quad \delta_{pq} = 0 \quad (p \neq q). \quad (9)$$

The numbers δ_{pq} are the elements of the matrix representing unity, which matrix has the property that it leaves unchanged any matrix when multiplied into it on either the left- or right-hand side.

We shall now obtain the *law of multiplication of the representatives of an observable and a ψ -symbol*. Let ψ be represented by the set of numbers a_p , as according to (1), and let the ψ -symbol $\alpha\psi$, where α is any observable, be represented by the set of numbers b_q , so that

$$\alpha\psi = \sum_q \psi_q b_q.$$

We have from (1)

$$\alpha\psi = \alpha \sum_p \psi_p a_p$$

$$= \sum_{pq} \psi_q \alpha_{qp} a_p.$$

Hence, equating coefficients of ψ_q , we get

$$b_q = \sum_p \alpha_{qp} a_p, \quad (10)$$

which is the required multiplication law. It suggests that we should regard the set of numbers a_p as a matrix, having rows corresponding to the various fundamental ψ 's of the representation, but having only one column. Equation (10) would then be the law of multiplication of such a matrix with a square matrix α_{qp} .

The correspondence that we have found between the properties of observables and ψ -symbols and those of their representatives, which is exemplified in equations (3), (4), (7), (10), allows us to take over any equation between the abstract symbols into an equation between the representatives. Suppose, for instance, that we are given the equation

$$\alpha\beta\psi = \gamma\psi' + \psi'', \quad (11)$$

where α , β , and γ are three observables and ψ , ψ' , and ψ'' are three states. By equating the representatives of each side of this equation, making use of the law (10), we obtain

$$\sum_p (\alpha\beta)_{qp} a_p = \sum_p \gamma_{qp} a'_p + a''_q,$$

where a_p , a'_p , and a''_q represent ψ , ψ' , and ψ'' respectively. From (7) we now get

$$\sum_{pr} \alpha_{qr} \beta_{rp} a_p = \sum_p \gamma_{qp} a'_p + a''_q.$$

Each symbol in the original equation (11) is here replaced by its representative, occurring in the corresponding position. The suffixes are arranged according to very simple and easily remembered rules, each consecutive pair of factors in any term having a common suffix, the two positions of this suffix being consecutive in the scheme of suffixes, while the suffix that occurs first in any term is the same for every term. A summation is taken over each suffix that occurs twice in a term.

As examples of equations that can be taken over in this way may be mentioned any of the equations between the abstract symbols occurring in the theory of eigenvalues of the preceding chapter. Equation (1) of that chapter, for instance, gives

$$\sum_q \alpha_{pq} a_q = aa_p.$$

If the matrix α_{pq} is known, then we have here an ordinary set of simultaneous algebraic equations for the unknowns a_p and also the unknown a . Any value of a for which these equations have a solution (not identically zero) may be called an eigenvalue of the matrix α_{pq} . If we eliminate the unknowns a_p , in which the equations are linear and homogeneous, we get the determinantal equation

$$\begin{vmatrix} \alpha_{11} - a & \alpha_{12} & \alpha_{13} & \dots & \dots & \dots \\ \alpha_{21} & \alpha_{22} - a & \alpha_{23} & \dots & \dots & \dots \\ \alpha_{31} & \alpha_{32} & \alpha_{33} - a & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{vmatrix} = 0 \quad (12)$$

to determine the eigenvalues a . The eigenvalues of a matrix repre-

senting an observable must, of course, be the same as the eigenvalues of the observable itself.

§ 21. Orthogonal Representations

We have not yet considered how ϕ -symbols are to be represented. We can always treat ϕ 's analogously to ψ 's, so that we can take any complete set of independent ϕ 's, ϕ_p say, and call them the fundamental ϕ 's of a representation. If we then expand an arbitrary ϕ in terms of them, thus

$$\phi = \sum_p a_p * \phi_p, \quad (13)$$

the set of numbers $a_p *$ will form the representative of this ϕ . Again if α is any observable, we can multiply it into a fundamental ϕ , ϕ_p , obtaining a product $\phi_p \alpha$, which we can expand in terms of the fundamental ϕ 's, thus

$$\phi_p \alpha = \sum_q \alpha_{pq} \phi_q. \quad (14)$$

The coefficients α_{pq} will then form the matrix that represents α . It may easily be verified that the matrix laws of addition and multiplication, equations (3), (4), and (7), hold also for the representatives of observables in the present representation in terms of fundamental ϕ 's. It should be noticed that the arrangement of the suffixes in α_{pq} requires the coefficients on the right-hand side of (14) to occur on the left of their respective ϕ -symbols, the opposite to what it was for equation (2). The particular arrangement of suffixes chosen in (14), like that chosen in (2), is necessary in order that we may have the multiplication law (7), which obeys the suffix rule, instead of the multiplication law (8).

We can in this way get a representation of observables on the basis either of a set of fundamental ϕ 's or of a set of fundamental ψ 's. The question now arises whether a set of fundamental ϕ 's and a set of fundamental ψ 's can be such that they both give the same representative for each observable. If this is so, we could count them both as belonging to the same representation, so that we should have one representation comprising representatives of both ϕ 's and ψ 's as well as observables. A necessary condition for the fundamental ϕ 's and fundamental ψ 's to give the same representatives for observables is that they shall be labelled by the same set of suffixes $p, q, r \dots$, which suffixes will then label the rows and columns of the matrices. Thus to each fundamental ψ there will be a corresponding fundamental ϕ having the same suffix. According to the notation that we have used hitherto, when a ψ and a ϕ have the same suffix they are

conjugate imaginary symbols denoting the same state, but this will now no longer hold.

We have already used the same suffixes for the fundamental ϕ 's in (14) as for the fundamental ψ 's in (2), so that we can investigate the consequences of these equations on the assumption that the coefficients α_{pq} are the same in each, for every observable α . If in (14) we change the summed suffix q to r and then multiply by ψ_q on the right, we obtain

$$\phi_p \alpha \psi_q = \sum_r \alpha_{pr} \phi_r \psi_q. \quad (15)$$

Similarly, if in (2) we change the summed suffix p to r and then multiply by ϕ_p on the left, we obtain

$$\phi_p \alpha \psi_q = \sum_r \phi_p \psi_r \alpha_{rq}. \quad (16)$$

The right-hand sides of equations (15) and (16) can be equal for an arbitrary observable α , i.e. for arbitrary α_{pq} 's, only provided

$$\phi_p \psi_q = 0 \quad (p \neq q) \quad (17)$$

and

$$\phi_p \psi_p = c,$$

where c is a number independent of p . We may without essential loss of generality take $c = 1$, so that we have

$$\phi_p \psi_p = 1. \quad (18)$$

Equations (17) and (18) can be combined in the single equation

$$\phi_p \psi_q = \delta_{pq}. \quad (19)$$

This is the condition that a set of fundamental ϕ 's and a set of fundamental ψ 's may both be considered as belonging to one representation.

With the help of these conditions we can easily obtain explicit expressions for the coefficients in the expansions. Thus to determine the coefficients a_p occurring in (1) for the expansion of an arbitrary ψ we have

$$\begin{aligned} \phi_q \psi &= \phi_q \sum_p a_p \psi_p = \sum_p a_p \delta_{pq} \\ &= a_q. \end{aligned} \quad (20)$$

Similarly the general coefficient a_q^* in (13) for the expansion of an arbitrary ϕ is

$$a_q^* = \phi \psi_q. \quad (21)$$

Again, from (14) we obtain

$$\begin{aligned} \phi_p \alpha \psi_r &= \sum_q \alpha_{pq} \phi_q \psi_r \\ &= \alpha_{pr} \end{aligned} \quad (22)$$

which gives explicitly the elements of the matrix representing any observable. This result could also have been obtained from (2).

In obtaining a general representation for both ϕ 's and ψ 's as well as observables, we have had to abandon the notation of a ϕ and ψ which have the same suffix being conjugate imaginary symbols denoting the same state, and this results in the representation being inconvenient and not very useful. The fundamental ϕ 's and ψ 's may, however, be such that each fundamental ϕ and ψ with the same suffix are really conjugate imaginary symbols denoting the same state, in which special case there is no need to abandon this notation. Such a representation is a particularly useful one. It is called an orthogonal representation. The set of states denoted either by the fundamental ϕ 's or by the fundamental ψ 's may be called the fundamental states of the representation. The condition (17) shows that these fundamental states are all orthogonal to each other and condition (18) shows that the ϕ 's and ψ 's representing them are normalized.

The vector picture of ϕ 's and ψ 's provides us with a simple geometrical interpretation of an orthogonal representation. In this vector picture each ϕ -symbol and the conjugate *imaginary* ψ -symbol are to be pictured as conjugate *complex* vectors. We can without inconsistency suppose that each fundamental ϕ and the conjugate imaginary fundamental ψ of an orthogonal representation are to be both pictured by the same real vector. Condition (17) now shows that these real vectors are all mutually perpendicular and condition (18) that they are each of unit length, so that they form the basis for a rectangular Cartesian system of co-ordinates. The numbers representing an arbitrary ϕ or ψ are now its co-ordinates in this system. Since the system of co-ordinates is real, a ϕ and the conjugate imaginary ψ , pictured as conjugate complex vectors, should have conjugate complex co-ordinates, and thus they should be represented by conjugate complex sets of numbers. It is easily verified, by comparing equations (20) and (21), that this is the case. Thus a state is represented by the same set of numbers whether it is denoted by a ϕ or a ψ , apart from an uncertainty in the sign of i .

If α is a real observable, then from equation (22) we find that the elements of the matrix representing it satisfy

$$\alpha_{pr} = \overline{\alpha_{rp}}$$

in the case of an orthogonal representation. A matrix for which this condition holds is called Hermitian. If in addition all the matrix elements are real, we have $\alpha_{pr} = \alpha_{rp}$, i.e. the matrix is symmetrical. From (22) we also find that a diagonal element α_{pp} is equal to the

average value, according to § 11, of the observable for the corresponding fundamental state ψ_p . If α is not a real observable, then its conjugate complex observable $\bar{\alpha}$, defined in § 10, has matrix elements to represent it, given by

$$\bar{\alpha}_{pr} = \overline{\alpha_{rp}}. \quad (23)$$

The matrix $\bar{\alpha}_{pr}$ may be called the conjugate complex matrix to α_{pr} .

§ 22. The δ Function

We have assumed throughout the above investigation of representations that the number of fundamental ψ 's, if not finite, is at most infinite enumerable, so that each of them can be labelled by a suffix p taking only a discrete set of values. For most dynamical systems of interest this condition is not fulfilled, the total number of independent states being infinite and equal to the number of points on a line. In such cases we must label each of the fundamental ψ 's by a suffix p that can assume any value in a certain range. The condition (1), which expresses that any ψ is a linear function of the fundamental ψ 's, must now be rewritten with an integral instead of a sum, thus

$$\psi = \int a_p \psi_p dp. \quad (24)$$

The domain of integration is to be understood to be the whole range of p used for labelling fundamental ψ 's. The coefficients a_p form a function of the continuous variable p .

It is not strictly true that every ψ can be expressed in the form (24) when the coefficients a_p are restricted to be finite, which is, of course, implied when one says they form a function of the continuous variable p . An example of a ψ that cannot be expressed in this form is one of the fundamental ψ 's, ψ_q say, itself. Another example is $\partial\psi_q/\partial q$ when ψ_q involves the parameter q in a manner sufficiently continuous for this differential coefficient to exist. It would be inconvenient if throughout the subsequent theory we were continually being reminded of the fact that there are exceptional ψ 's which cannot be expressed in the form (24). We get over the difficulty by allowing infinities of certain types to occur in the coefficients a_p , which enables every ψ formally to be expressed in the required form. This is analogous to the device sometimes used in geometry, of avoiding the exception of parallel lines to the rule that two straight lines always meet in one point, by saying that parallel lines meet in a point at infinity.

We observe that those ψ 's that are not of the form of the right-

hand side of (24) with finite a_p can always be regarded as limits of ψ 's that are of this form. We can, for instance, express ψ_q by

$$\psi_q = \lim_{n \rightarrow \infty} \int a_{pn} \psi_p dp,$$

where the coefficients a_{pn} satisfy

$$\lim_{n \rightarrow \infty} \int a_{pn} dp = 1,$$

$$\lim_{n \rightarrow \infty} a_{pn} = 0 \text{ (for } p \neq q\text{).}$$

As one approaches the limit, a_{pn} becomes a function of p which vanishes for all values of p except those very close to q and which is so large for values of p in the immediate neighbourhood of q that its integral is unity. We can now say formally that

$$\psi_q = \int a_p \psi_p dp, \quad (24')$$

where

$$a_p = \lim_{n \rightarrow \infty} a_{pn}.$$

This a_p , we can say, is an improper function of the variable p , having the value zero for all values of p except q and the value infinity for $p = q$, the infinity being such that its integral is unity. It is thus a function of the two variables p and q which depends only on their difference, so that we can put

$$a_p = \delta(p - q), \quad (25)$$

where the improper function $\delta(x)$ is defined by

$$\int_{-\infty}^{\infty} \delta(x) dx = 1$$

$$\delta(x) = 0 \text{ (for } x \neq 0\text{).}$$

The introduction of the δ function into our analysis will not be in itself a source of lack of rigour in the theory, since any equation involving the δ function can be transcribed into an equivalent but usually more cumbersome form in which the δ function does not appear. The δ function is thus merely a convenient notation. The only lack of rigour in the theory arises from the fact that we perform operations on the abstract symbols, such as differentiation and integration with respect to parameters occurring in them, which are not rigorously defined. When these operations are permissible, the δ function may be used freely for dealing with the representatives of the abstract symbols, as though it were a continuous function, without leading to incorrect results. We can, in fact, even give a meaning to the δ function of an observable, provided it has a continuous range of eigenvalues, by means of the general definition of § 15.

Certain elementary properties of the δ function, which are deducible from, or at least consistent with, the definition, should be noted, namely,

$$\begin{aligned}\delta(-x) &= \delta(x) \\ x\delta(x) &= 0\end{aligned}\tag{26}$$

and

$$\int_{-\infty}^{\infty} f(x)\delta(x-a) dx = f(a),\tag{27}$$

where $f(x)$ is any continuous function of x and a is any number, and the range of integration is any range through the point a , the limits ∞ and $-\infty$ being put down merely for definiteness. Thus the operation of multiplying by $\delta(x-a)$ and integrating with respect to x is equivalent to the operation of substituting a for x . This is still true when the operation is applied, not to an ordinary function $f(x)$ of x , but to a ψ -symbol or an observable involving the parameter x , provided it is reasonably continuous in x . We are, in fact, making an application of this rule, with the ψ -symbol ψ_p for $f(x)$ and the number q for a , when we assert that (24') holds with a_p defined by (25). A further property of the δ function is

$$\int_{-\infty}^{\infty} \delta(a-x) dx \delta(x-b) = \delta(a-b).\tag{28}$$

To prove this relation, we regard the left-hand side as a function of the number b and put it equal to $F(b)$. We see at once that $F(b) = 0$ if b is not equal to a , and also we have

$$\begin{aligned}\int_{-\infty}^{\infty} F(b) db &= \int_{-\infty}^{\infty} \delta(a-x) dx \int_{-\infty}^{\infty} \delta(x-b) db \\ &= \int_{-\infty}^{\infty} \delta(a-x) dx = 1.\end{aligned}$$

Thus $F(b)$ satisfies all the conditions that define $\delta(b-a)$ and may hence be put equal to $\delta(b-a)$ or $\delta(a-b)$. Equation (28) would have been obtained from equation (27) if for $f(x)$ we had substituted the improper function $\delta(x-b)$. This is an example which illustrates how a δ function may be used as though it were a continuous function without leading to incorrect results.

In order to put $\partial\psi_q/\partial q$ in the form of the right-hand side of (24) it is necessary to use the derivative $\delta'(x)$ of the function $\delta(x)$. This derivative is, of course, an even more discontinuous and improper function than $\delta(x)$ itself, but in many cases it can also be used freely

as though it were a continuous function of x without leading to incorrect results. It has the elementary properties

$$\begin{aligned}\delta'(-x) &= -\delta'(x) \\ x\delta'(x) &= -\delta(x)\end{aligned}\quad (29)$$

and

$$\int_{-\infty}^{\infty} f(x)\delta'(x-a) dx = -f'(a), \quad (30)$$

for any differentiable function of x , which may be a ψ -symbol or observable involving x as a parameter. The second and third of these relations may be obtained by differentiating (26) with respect to x and (27) with respect to a respectively. The third one may also be verified by an integration by parts, thus

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\delta'(x-a) dx &= \left[f(x)\delta(x-a) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'(x)\delta(x-a) dx \\ &= -f'(a)\end{aligned}$$

from (27). A further property is

$$\int_{-\infty}^{\infty} \delta'(a-x) dx \delta(x-b) = \delta'(a-b), \quad (31)$$

which may be obtained by differentiation of (28) with respect to a . It may also be obtained from (27) if one puts b for a and then takes $\delta'(a-x)$ for $f(x)$, and is then an example of how the δ' function may be used as though it were a continuous function.

If for $f(x)$ in (30) we put ψ_p , p being the variable instead of x , and if we put q for a , we get

$$\int_{-\infty}^{\infty} \psi_p \delta'(p-q) dp = -\partial \psi_q / \partial q.$$

This shows that $\partial \psi_q / \partial q$ may be expressed in the form of the right-hand side of (24) with $-\delta'(p-q)$ for a_p . By making use of higher derivatives of the δ function, one can express $\partial^2 \psi_q / \partial q^2$, $\partial^3 \psi_q / \partial q^3$, &c., also in this form.

§ 23. Case of a Continuous Range of Fundamental States

We can now generalize the theory of the representation of states and observables to apply to systems for which the number of independent states is equal to the number of points on a line. The ψ on the left-hand side of (24) will be represented by the numbers a_p that occur as coefficients on the right-hand side, or by the function a_p of the

continuous variable p . Also if α is any observable, corresponding to (2) we can expand $\alpha\psi_q$ in the form

$$\alpha\psi_q = \int \psi_p dp \alpha_{pq}, \quad (32)$$

where the α_{pq} are numbers, and these numbers, which form a function of the two continuous variables p and q , will then represent the observable α . It is sometimes convenient to call this function of two variables a matrix, in order that one may use the same words in talking about the case (32) as about the case (2). The number of rows and columns of such a matrix is equal to the number of points on a line. Corresponding to the multiplication law (7), we now have

$$(\alpha\beta)_{pq} = \int \alpha_{pr} dr \beta_{rq}, \quad (33)$$

which may be proved in an analogous way. Similarly, corresponding to (10), we now have that the function b_p of p representing $\alpha\psi$ is given in terms of a_p , that representing ψ , by the relation

$$b_q = \int \alpha_{qp} dp a_p. \quad (34)$$

If we regard the number c as an observable, its representative c_{pq} will, by definition, be given by

$$c\psi_q = \int \psi_p dp c_{pq}, \quad (35)$$

so that

$$c_{pq} = c\delta(p-q). \quad (36)$$

The matrix representing unity is now that whose general element is $\delta(p-q)$ and it still, of course, has the property of leaving unchanged any matrix when multiplied into it on either the right- or left-hand side. If we compare these results with the corresponding ones for the case of discrete fundamental ψ 's, we see that the only difference is that the two-suffix δ -symbol, defined by (9), is replaced by the δ function of the difference of the two suffixes. It is a general rule that the two-suffix δ -symbol is always to be replaced by the δ function in this way when one passes from the case of sums to the case of integrals.

The connexion between the fundamental ψ 's and the fundamental ϕ 's of the same representation is now

$$\phi_p \psi_q = \delta(p-q), \quad (37)$$

which is obtained from (19) by replacing the two-suffix δ -symbol according to the rule. This condition (37) implies that $\phi_p \psi_p$ is infinite. Thus the law of § 8 that any ϕ -symbol can be multiplied into any ψ -symbol, giving a number as product, must be relaxed to allow the possibility of the product being infinite.

When each fundamental ϕ and fundamental ψ with the same suffix are conjugate imaginary symbols denoting the same state, we have, as before, an orthogonal representation. We shall now consider the meaning of equation (37) for an orthogonal representation. This condition (37) may be split up into the two conditions

$$\phi_p \psi_q = 0 \quad (p \neq q) \quad (38)$$

$$\int \phi_p \psi_q dp = 1. \quad (39)$$

The first of these, corresponding to (17), again expresses that any two fundamental states are orthogonal. The second, corresponding to (18), is sometimes taken as the definition of the normalization of ψ_q when the suffix q labelling the independent states ψ_q takes on a continuous range of values, instead of the condition $\phi_q \psi_q = 1$, which would now be mathematically useless, as it would require the ϕ 's and ψ 's in (37) to be all multiplied by infinitely small coefficients. If, however, one changes the definition of normalization in this way, one must remember that the laws for the physical interpretation of the theory hold only for the old definition. The general law given at the end of § 11, that $\phi_q \alpha \psi_q$ is the average value of the observable α for the state ψ_q provided $\phi_q \psi_q = 1$, is of universal validity, for the continuous as well as for the discrete case. It is true that for the continuous case $\phi_q \alpha \psi_q$ will in general be zero when $\phi_q \psi_q = 1$, but, as the applications of the theory will show, this is what the physics then requires. Only the ratios of the averages of different observables are then of interest, and for the calculation of these the normalizing condition (39) is useful.

With the help of (37) we obtain from (24)

$$\phi_q \psi = \phi_q \int a_p \psi_p dp = \int a_p \delta(p-q) dp = a_q, \quad (40)$$

by making an application of (27). This result, corresponding to (20), gives explicitly the coefficients on the right-hand side of (24) representing ψ . The conjugate imaginary ϕ is represented by the numbers $a_q^* = \phi \psi_q$, corresponding to (21), which are the conjugate complex numbers to a_q in the case of an orthogonal representation. Again, from (32) we obtain

$$\phi_r \alpha \psi_q = \int \phi_r \psi_p dp \alpha_{pq} = \int \delta(r-p) dp \alpha_{pq} = \alpha_{rq}, \quad (41)$$

which, corresponding to (22), gives explicitly the elements of the matrix representing α . We no longer, however, have the result that a diagonal element α_{qq} is for an orthogonal representation the average value of α for the state ψ_q , since the normalizing condition (37) which

is here used is not the correct one for physical interpretation. This result would give, if for example we took α equal to unity, the value $\delta(q-q) = \infty$, whereas the average value of unity must of course be unity.

§ 24. The Weight Function

It is sometimes convenient to modify equations (24), (32), which define the representatives of a state and observable, by the introduction of a *weight function*. We can take any function ρ_p of the variable p which is defined throughout the range of p used for labelling the fundamental states and which has no zero values, and put instead of (24), (32)

$$\psi = \int a_p \psi_p \rho_p dp \quad (42)$$

$$\alpha\psi_q = \int \psi_p \rho_p dp \alpha_{pq}. \quad (43)$$

We can now consider the new coefficients a_p and α_{pq} to be the representatives of the state and observable. This does not give any essential generalization of the theory of representation, since the new representatives are connected with the original ones by very simple relations. It is merely a device which is convenient in certain applications of the theory, usually for increasing the symmetry of the equations, or for making more direct the physical interpretation of the representatives which will be given in § 28. We could, of course, adopt the same device in the case of a discrete set of fundamental states, but there do not seem to be any examples for which it is then of any value.

When the weight function ρ is introduced it must appear, not only in the expansions (42), (43), but in all formulas which involve an integral over the parameter p that labels the fundamental ψ 's, e.g. in the multiplication law for two observables, equation (33), which becomes

$$(\alpha\beta)_{pq} = \int \alpha_{pr} \rho_r dr \beta_{rq},$$

and in that for an observable with a ψ , equation (34), which becomes

$$b_q = \int \alpha_{qp} \rho_p dp a_p.$$

Again, the number c , regarded as an observable, is no longer represented by the right-hand side of (36), since instead of (35) we now have

$$c\psi_q = \int \psi_p \rho_p dp c_{pq},$$

which gives

$$c_{pq} = c\rho_p^{-1}\delta(p-q) = c\rho_q^{-1}\delta(p-q).$$

The unit matrix is thus changed from $\delta(p-q)$ to $\rho_p^{-1}\delta(p-q)$. This suggests that equation (37) should be changed to

$$\phi_p \psi_q = \rho_p^{-1} \delta(p-q), \quad (44)$$

a conclusion which is confirmed when one notes that the normalizing equation (39) must be changed to

$$\int \phi_p \psi_q \rho_p dp = 1. \quad (45)$$

We can now see what changes must be made in the representatives of states and observables when the weight function is introduced. If we multiply the ϕ_p and ψ_p of equations (37) and (39) by $\rho_p^{-\frac{1}{2}}$, they will then satisfy equations (44) and (45). We must then multiply the a_p of equation (24) by $\rho_p^{-\frac{1}{2}}$ in order that it may satisfy (42) and the α_{pq} of equation (32) by $(\rho_p \rho_q)^{-\frac{1}{2}}$ in order that it may satisfy (43). These results are particular cases of the general rule that any symbol involving the suffixes p, q, \dots gets multiplied by $(\rho_p \rho_q \dots)^{-\frac{1}{2}}$ when the weight function is introduced. From this rule one can see the necessity for the insertion of the factor ρ_p in every integral with respect to the variable p , when one bears in mind that the integrand must contain the suffix p twice.

§ 25. General Case of Representation

In most of the applications of quantum mechanics the atomic system dealt with has a still larger number of independent states than we have hitherto considered. The fundamental states of a representation can then be labelled conveniently only by means of several suffixes p_1, p_2, \dots, p_n , which can take on any values within a given domain of the n -dimensional p -space. The generalizations which must now be made in the preceding theory are quite obvious. We must have, for instance, instead of (24) and (32), the expansions

$$\psi = \iint \dots a_{p_1 p_2} \dots \psi_{p_1 p_2} \dots dp_1 dp_2 \dots \quad (46)$$

$$\alpha \psi_{q_1 q_2} \dots = \iint \dots \psi_{p_1 p_2} \dots dp_1 dp_2 \dots \alpha_{p_1 p_2} \dots q_1 q_2 \dots \quad (47)$$

A state ψ is now represented by $a_{p_1 p_2} \dots$, a function of the n -variables p_1, p_2, \dots , and an observable α by $\alpha_{p_1 p_2} \dots q_1 q_2 \dots$, a 'matrix' whose rows and columns are both labelled by these same variables. The ψ -symbol $\psi_{q_1 q_2} \dots$, one of the fundamental states, is represented by

$$\delta(p_1 - q_1) \delta(p_2 - q_2) \dots \delta(p_n - q_n), \quad (48)$$

as may easily be verified by substituting this expression for $a_{p_1 p_2} \dots$ in (46) and carrying out the integrations one by one with the help

of (27). It is always this product (48) that replaces the $\delta(p-q)$ of the one-dimensional case. In the same way the ψ -symbol

$$\frac{\partial}{\partial q_m} \psi_{q_1 q_2 \dots} \quad (m = 1, 2, \dots n)$$

is represented by

$$-\delta(p_1 - q_1)\delta(p_2 - q_2) \dots \delta(p_{m-1} - q_{m-1})\delta'(p_m - q_m)\delta(p_{m+1} - q_{m+1}) \dots \\ \delta(p_n - q_n), \quad (49)$$

as may easily be verified with the help of (30). This expression differs from (48), apart from the minus sign, only in the m -th factor.

We must make a still further generalization in order to include all the cases of representation that occur in practice, namely, we must allow both sums and integrals to occur together. In the one-dimensional case, for instance, we can have

$$\psi = \sum_P a_P \psi_P + \int a_p \psi_p dp. \quad (50)$$

The discrete set of numbers a_P together with the continuous set a_p now represent the state ψ . They may be considered as a function of a variable whose domain consists of a continuous range together with some discrete points. In the many-dimensional case we can have sums for some of the variables and integrals for others. The general rule applying to every case of representation is that a state is represented by a function whose domain is such that every point of it corresponds to one of the fundamental states. There is no restriction on the number of points in this domain or on their arrangement in the p -space that labels them. Thus the domain may consist of discrete points together with a number of continuous regions each having any number of dimensions. An observable is represented by a matrix whose rows and whose columns are in one-one correspondence with the points of this domain.

The equations of our previous theory of representation can all be taken over without difficulty, but cannot very well be written down in a form that includes all cases without an elaborate notation. We shall therefore take simply the case when (50) holds as an illustration. Corresponding to (2) and (32), we now have, for the definition of the representative of an observable,

$$\begin{aligned} \alpha \psi_Q &= \sum_P \psi_P \alpha_{PQ} + \int \psi_p dp \alpha_{pQ} \\ \alpha \psi_q &= \sum_P \psi_P \alpha_{Pq} + \int \psi_p dp \alpha_{pq}. \end{aligned} \quad \} (51)$$

There are thus four kinds of coefficients in the representative of an

observable, typified by α_{PQ} , α_{pQ} , α_{Pq} , α_{pq} , corresponding to the different cases of discrete or continuous values for the suffixes. Again, corresponding to (7) and (33), we now have for the multiplication law for the representatives of observables,

$$\begin{aligned} (\alpha\beta)_{PQ} &= \sum_R \alpha_{PR} \beta_{RQ} + \int \alpha_{Pr} dr \beta_{rQ} \\ (\alpha\beta)_{pQ} &= \sum_R \alpha_{pR} \beta_{RQ} + \int \alpha_{pr} dr \beta_{rQ} \\ (\alpha\beta)_{Pq} &= \sum_R \alpha_{PR} \beta_{Rq} + \int \alpha_{Pr} dr \beta_{rq} \\ (\alpha\beta)_{pq} &= \sum_R \alpha_{pR} \beta_{Rq} + \int \alpha_{pr} dr \beta_{rq}. \end{aligned}$$

In each case there is a sum over R and an integral over r . The conditions (19) and (37) become

$$\begin{array}{ll} \phi_P \psi_Q = \delta_{PQ} & \phi_p \psi_Q = 0 \\ \phi_P \psi_q = 0 & \phi_p \psi_q = \delta(p-q). \end{array}$$

These examples are sufficient to show how each equation is to be interpreted in any of the various kinds of representation that may arise.

We can make a final generalization by introducing a weight function in the general case. This weight function ρ may be an arbitrary function of the variables p_m that label the fundamental ψ 's, provided it never vanishes. It will always appear along with the differentials dp_m in any integration and will also appear, to the power of -1 , in the unit matrix.

TRANSFORMATION THEORY

§ 26. Eigenstates as Fundamental States of a Representation
 In the preceding chapter the idea of a representation of the abstract symbols was introduced and was treated entirely from a general mathematical point of view, the representatives being like co-ordinates of the symbols referred to a general co-ordinate system. We must now consider particular representations, *i.e.* co-ordinates referred to particular co-ordinate systems, which must be singled out and specified in a certain way. We shall find, incidentally, that our representatives now often have direct physical interpretations. We shall be concerned here and throughout the future work only with orthogonal representations.

An orthogonal representation is built up on the basis of a complete set of orthogonal states, forming the fundamental states. Such a set of states is obtained most easily with the help of the theory of eigenvalues of Chapter III. If we take a set of real observables that all commute, their simultaneous eigenstates form a complete set and any two belonging to two different sets of eigenvalues are orthogonal. If the set of commuting observables is a complete one, then, as shown in § 17, there is only one eigenstate for each set of eigenvalues, so that the eigenstates must now all be orthogonal. These eigenstates can therefore be taken to be the fundamental states of a representation. Each of them is associated with one set of eigenvalues, which may conveniently be used for labelling it, instead of the arbitrary suffixes p_m of the preceding chapter, which have no physical meaning. Thus if the commuting observables are $\xi_1, \xi_2, \dots, \xi_n$ and if we denote the eigenvalues of ξ_m by ξ'_m, ξ''_m, \dots a fundamental ψ may be written $\psi(\xi'_1 \xi'_2 \dots \xi'_n)$, or simply $\psi(\xi')$ for brevity. In the same way a fundamental ϕ may be written $\phi(\xi'')$. The fundamental ϕ that is conjugate imaginary to $\psi(\xi')$ will be $\phi(\xi')$.

The notation of primes and multiple primes to denote the eigenvalues of an observable is very convenient and will be used generally in the future. A new notation for the representatives of states and observables will now be introduced, which will greatly increase the symmetry in our equations. A general ψ -symbol ψ is represented by a set of numbers, each of which is associated with one of the funda-

mental ψ 's and thus with one set of eigenvalues. That particular one of the set of numbers which is associated with the eigenvalues $\xi'_1 \xi'_2 \dots \xi'_n$ will be written $(\xi'_1 \xi'_2 \dots \xi'_n|)$, or $(\xi'|)$ for brevity. When it is necessary to particularize the ψ -symbol by a suffix, k say, we can insert this suffix in the representative of ψ_k to the right of the vertical line, thus $(\xi'_1 \xi'_2 \dots \xi'_n|k)$ or $(\xi'|k)$. The reason for this notation is that, as we shall see later, there is a remarkable symmetry in the way $(\xi'|k)$ involves the set of numbers ξ' , referring to one of the fundamental ψ 's, on the one hand, and the parameter k which specifies the ψ that is being represented, on the other. This symmetry is exactly expressed when one puts the ξ' 's and the k to the left and right respectively. In a corresponding way we shall write the representative of a general ϕ -symbol as $(|\xi')$ and of a particular one, ϕ_k , as $(k|\xi')$. For the representative of an observable α , we shall write the matrix element α_{pq} , associated with the fundamental states ψ_p and ψ_q , as $(\xi'_1 \xi'_2 \dots \xi'_n|\alpha|\xi''_1 \xi''_2 \dots \xi''_n)$, or as $(\xi'|\alpha|\xi'')$ for brevity, where the ξ' 's and ξ'' 's are the eigenvalues belonging to the fundamental states ψ_p and ψ_q respectively, or $\psi(\xi')$ and $\psi(\xi'')$, as they would be written in the new notation.

Some of the equations of the preceding chapter will now be written in the new notation to illustrate how it runs. Equations (3) and (4) become

$$\begin{aligned} (\xi'|\alpha + \beta|\xi'') &= (\xi'|\alpha|\xi'') + (\xi'|\beta|\xi'') \\ (\xi'|c\alpha|\xi'') &= c(\xi'|\alpha|\xi''). \end{aligned}$$

Equation (1) or (24), defining the representative of a ψ -symbol, becomes, if we take for definiteness the case when each of the ξ'_m 's has a continuous range of values,

$$\psi = \int \psi(\xi') d\xi' (\xi'|), \quad (1)$$

where $d\xi'$ is short for the product $d\xi'_1 d\xi'_2 \dots d\xi'_n$ and only one integral sign is written to denote integration over all these variables. It should be noted how, when one puts the $d\xi'$ in the proper place, all the ξ' 's in (1) occur together. This is the new form of the suffix rule given near the end of § 20. Equation (2) or (32) of the preceding chapter, defining the representatives of an observable α , becomes in the same way

$$\alpha\psi(\xi'') = \int \psi(\xi') d\xi' (\xi'|\alpha|\xi''). \quad (2)$$

Again, the multiplication law for the representatives of two observables, equation (7) or (33), becomes

$$(\xi'|\alpha\beta|\xi'') = \int (\xi'|\alpha|\xi''') d\xi''' (\xi'''|\beta|\xi'')$$

and that for the representatives of an observable and a ψ , equation (10) or (34), becomes

$$(\xi' | l) = \int (\xi' | \alpha | \xi'') d\xi'' (\xi'' | k), \quad (3)$$

where k specifies the ψ -symbol ψ_k and l specifies $\psi_l = \alpha\psi_k$. The conjugate complex $\bar{\alpha}$ of an observable α is now represented by

$$(\xi' | \bar{\alpha} | \xi'') = (\bar{\xi''} | \alpha | \xi'), \quad (4)$$

corresponding to (23), and the representatives $(\xi' |)$ and $(| \xi')$ of a ψ and its conjugate imaginary ϕ are conjugate complex quantities.

The representation we are now considering is built up from a number of commuting observables $\xi_1, \xi_2, \dots, \xi_n$, whose simultaneous eigen- ψ 's are taken as fundamental ψ 's. Let us determine how one of these observables, ξ_m say, is itself represented. Putting ξ_m for α in (2), we get

$$\xi_m \psi(\xi'') = \int \psi(\xi') d\xi' (\xi' | \xi_m | \xi''). \quad (5)$$

But since $\psi(\xi'')$ is an eigen- ψ of ξ_m , belonging to the eigenvalue ξ_m'' we have

$$\xi_m \psi(\xi'') = \xi_m'' \psi(\xi'') = \int \psi(\xi') d\xi' \xi_m' \delta(\xi' - \xi''), \quad (6)$$

where $\delta(\xi' - \xi'')$ is short for the product $\delta(\xi'_1 - \xi''_1)\delta(\xi'_2 - \xi''_2) \dots \delta(\xi'_n - \xi''_n)$. Equating coefficients on the right-hand sides of (5) and (6), we obtain

$$(\xi' | \xi_m | \xi'') = \xi_m' \delta(\xi' - \xi''). \quad (7)$$

This, of course, is equal to $\xi_m'' \delta(\xi' - \xi'')$ and is therefore symmetrical between the singly and doubly primed symbols.

If the ξ' 's take on discrete sets of values instead of continuous ranges, we should obtain instead of (7)

$$(\xi' | \xi_m | \xi'') = \xi_m' \delta_{\xi' \xi''},$$

where $\delta_{\xi' \xi''}$ is short for the product $\delta_{\xi'_1 \xi''_1} \delta_{\xi'_2 \xi''_2} \dots \delta_{\xi'_n \xi''_n}$. *Thus the observable ξ_m is represented by a diagonal matrix, whose diagonal elements are its eigenvalues ξ_m' .* A diagonal matrix, in the case of continuous ranges of rows and columns, may conveniently be defined as one whose general element (ξ', ξ'') involves the δ function $\delta(\xi' - \xi'')$ as a factor, like the right-hand side of (7), and the coefficient of the δ function may be defined as the general diagonal element. With these definitions the above law in italics for the representative of ξ_m holds in all cases. The appropriateness of this definition for a diagonal matrix in the continuous case rests on the fact that, as is easily verified, it makes two diagonal matrices always commute, which is one of the most important properties of diagonal matrices in the discrete case. For this reason it would not be sufficient to

define a diagonal matrix in the continuous case merely as one whose general element (ξ' , ξ'') vanishes except when the ξ'' 's differ infinitely little from the ξ' 's.

If $f(\xi)$ is any function of the ξ 's, then its representative is found to be, by a similar argument to that leading to (7),

$$(\xi' | f(\xi) | \xi'') = f(\xi') \delta(\xi' - \xi''). \quad (8)$$

The coefficient $f(\xi')$ must, of course, have a meaning since the function f must be defined for each of the eigenvalues of the ξ 's. Thus the representation based on the simultaneous eigen- ψ 's of a set of observables as fundamental ψ 's is such that the representative of each of the ξ 's and of any function of them is a diagonal matrix. Conversely, every diagonal matrix in this representation represents a function of the ξ 's, this function being specified by the general diagonal element (ξ' , ξ') regarded as a function of the variables ξ' .

Thus if we take any set of observables that commute, there will exist a representation in which each of these observables simultaneously is represented by a diagonal matrix. If the set of observables is a complete one, then the representation will be completely determined by these observables, except for arbitrary phases which arise from the fact that a simultaneous eigen- ψ of these observables may be multiplied by any numerical factor of modulus unity without any of the conditions defining it being invalidated. For example, we can multiply each $\psi(\xi')$ by $\exp[-if(\xi')]$, where $f(\xi')$ is an arbitrary real function of the ξ' 's. This will require every representative of a state, ($\xi' |$), to be multiplied by $\exp if(\xi')$ and every representative of an observable, ($\xi' | \alpha | \xi''$), to be multiplied by $\exp i[f(\xi') - f(\xi'')]$. A diagonal element ($\xi' | \alpha | \xi'$) remains unaltered by this transformation, as is necessary on account of its having the physical meaning of an average. The arbitrary phases which thus arise in the representatives are usually unimportant and trivial, so that we may count a representation as being completely determined by the observables that are diagonal in it. This fact is already implied in our notation, since the only indication in a representative of the representation to which it belongs are the letters denoting the observables that are diagonal.

The representations considered in this section, in which each fundamental ψ is a simultaneous eigen- ψ of a set of real commuting observables, are not of a special kind, since every orthogonal representation has this property. In fact, if we take any representation, having ψ_p, ψ_q, \dots as fundamental ψ 's, we can then form any diagonal

matrix whose general element ξ_{pq} is of the form $a_p \delta(p-q)$, where a_p is a real function of p , and consider this diagonal matrix as representing an observable ξ . This observable will be real if the representation is orthogonal. We shall then have

$$\xi\psi_q = \int \psi_p dp \xi_{pq} = \int \psi_p dp a_p \delta(p-q) = a_q \psi_q,$$

so that each fundamental ψ , ψ_q , is an eigen- ψ of ξ . In the many-dimensional case, when several suffixes p or q are required to label a fundamental ψ , we can take several diagonal matrices and each will represent an observable ξ for which the fundamental ψ 's are all eigen- ψ 's. We can obtain in this way a sufficient number of observables ξ having the fundamental ψ 's as eigen- ψ 's to form a complete set. The notation and methods of the present section can then be applied.

§ 27. Canonical Transformations

If we take two representations, based respectively on the fundamental ψ 's $\psi(\xi')$, which are the simultaneous eigen- ψ 's of a set of commuting observables ξ_m , and the fundamental ψ 's $\psi(\eta')$, which are the simultaneous eigen- ψ 's of a set of commuting observables η_m , then an arbitrary ψ will have the two representatives $(\xi'|)$ and $(\eta'|)$, which are functions of the sets of variables ξ'_m and η'_m respectively. Since a ψ is completely determined by its representative in any one representation, there must be a connexion between the two representatives $(\xi'|)$ and $(\eta'|)$ such that either is determined by the other. We shall now investigate the form of this connexion.

From the definition of the representative $(\eta'|)$ we have, if we take for definiteness the case of integrals,

$$\psi = \int \psi(\eta') d\eta' (\eta'|). \quad (9)$$

Now each fundamental ψ of the η -representation, $\psi(\eta')$, will itself have a representative in the ξ -representation. We may write this representative $(\xi'| \eta')$, with η' on the right to show which ψ it represents. We shall then have

$$\psi(\eta') = \int \psi(\xi') d\xi' (\xi'| \eta') \quad (10)$$

for the definition of $(\xi'| \eta')$. Substituting this value for $\psi(\eta')$ in the right-hand side of (9), we get

$$\psi = \iint \psi(\xi') d\xi' (\xi'| \eta') d\eta' (\eta'|),$$

which gives, on comparison with equation (1) which defines $(\xi'|)$,

$$(\xi'|) = \int (\xi'| \eta') d\eta' (\eta'|). \quad (11)$$

This is the transformation equation which gives the ξ -representative of a ψ -symbol in terms of its η -representative. The corresponding equation which gives $(\eta'|)$ in terms of $(\xi'|)$ may be shown in the same way to be

$$(\eta'|) = \int (\eta'|\xi') d\xi' (\xi'|), \quad (12)$$

where $(\eta'|\xi')$ is the representative of the fundamental ψ , $\psi(\xi')$, in the η -representation.

The two representatives $(\xi'|)$ and $(\eta'|)$ are thus linear functions of one another. The expressions $(\xi'|\eta')$ and $(\eta'|\xi')$ which enable us to pass from one to the other will be called transformation functions. They are each functions of the two sets of variables ξ' and η' . We can obtain an explicit expression for $(\xi'|\eta')$ by multiplying equation (10) by $\phi(\xi'')$ to the left, a process corresponding to that used for getting equation (40) of the preceding chapter. The result is

$$(\xi'|\eta') = \phi(\xi')\psi(\eta'). \quad (13)$$

Similarly it may be shown that

$$(\eta'|\xi') = \phi(\eta')\psi(\xi'). \quad (14)$$

Hence $(\xi'|\eta')$ and $(\eta'|\xi')$ are conjugate complex quantities.

The transformation functions must satisfy certain conditions in order that (11) and (12) may be consistent. If we substitute for $(\eta'|)$ in (11) its value given by (12), we get

$$(\xi'|) = \iint (\xi'|\eta') d\eta' (\eta'|\xi'') d\xi'' (\xi''|).$$

But we have also

$$(\xi'|) = \int \delta(\xi' - \xi'') d\xi'' (\xi''|).$$

Since these equations must hold for an arbitrary function $(\xi''|)$ of the variables ξ'' , we can equate the coefficients of $(\xi''|)$ on their right-hand sides. This gives

$$\int (\xi'|\eta') d\eta' (\eta'|\xi'') = \delta(\xi' - \xi''). \quad (15)$$

An alternative way of obtaining this result is to apply equation (11) to the ψ -symbol $\psi(\xi'')$. Since the η -representative of this ψ -symbol is $(\eta'|\xi'')$, the right-hand side of (11) becomes $\int (\xi'|\eta') d\eta' (\eta'|\xi'')$, while the left-hand side becomes the ξ -representative of $\psi(\xi'')$, which is, of course, $\delta(\xi' - \xi'')$. The equation corresponding to (15) in which ξ and η have changed places, namely,

$$\int (\eta'|\xi') d\xi' (\xi'|\eta'') = \delta(\eta' - \eta''), \quad (16)$$

may be similarly obtained. Equations (15) and (16) are the only conditions which the transformation functions must satisfy identically. They are of the nature of orthogonality and normalization conditions.

The transformation of the representatives of ϕ -symbols may be treated in the same way. We should then find, for instance, the equation

$$(|\eta'|) = \int (|\xi'|) d\xi' (\xi'|\eta')$$

as the transformation equation which gives the representative $(|\eta'|)$ of an arbitrary ϕ -symbol in terms of its representative $(|\xi'|)$, where the quantity $(\xi'|\eta')$ is now defined as the η -representative of the fundamental ϕ , $\phi(\xi')$, i.e. by the equation

$$\phi(\xi') = \int (\xi'|\eta') d\eta' \phi(\eta').$$

If we multiply this equation by $\psi(\eta'')$ on the right, we obtain, as an explicit expression for this $(\xi'|\eta')$,

$$\phi(\xi')\psi(\eta'') = (\xi'|\eta''),$$

which is the same as (13). Thus this quantity $(\xi'|\eta')$, defined as the η -representative of $\phi(\xi')$, is the same as our previous one defined as the ξ -representative of $\psi(\eta')$, so that our notation of using the same symbol for them both is justified. The symmetry which thus exists in the way the quantity $(\xi'|\eta')$ involves the ξ' 's and η' 's is the same as that which was referred to in the preceding section when the new notation for the representative of a state was introduced, since any representative $(\xi'|k)$ of a specified ψ -symbol ψ_k , when suitably normalized, may be regarded as the transformation function connecting the ξ -representation with a representation in which ψ_k is one of the fundamental states.

Owing to the arbitrary phases occurring in representations, there will be a corresponding amount of arbitrariness in the transformation functions. If the fundamental states $\psi(\xi')$, $\psi(\eta')$ are multiplied by $\exp[-if(\xi')]$, $\exp[-ig(\eta')]$ respectively, f and g being arbitrary real functions, the transformation function $(\xi'|\eta')$ will get multiplied by $\exp\{-i[f(\xi')-g(\eta')]\}$. Thus the modulus of the transformation function is quite definite, the indeterminacy being only in its phase.

The connexion between the representatives of an observable α in the two representations may be easily obtained in a variety of different ways. We can, for instance, use the explicit expression for the representative of α given by equation (41) of the preceding chapter. Applying this to the ξ -representation, we get

$$(\xi'|\alpha|\xi'') = \phi(\xi')\alpha\psi(\xi'').$$

If we now substitute for the right-hand side, which consists of the

product of three abstract symbols, their representatives in the η -representation, we get

$$(\xi'|\alpha|\xi'') = \iint (\xi'|\eta') d\eta' (\eta'|\alpha|\eta'') d\eta'' (\eta''|\xi''), \quad (17)$$

which gives the ξ -representative in terms of the η -representative. Similarly we may obtain the result

$$(\eta'|\alpha|\eta'') = \iint (\eta'|\xi') d\xi' (\xi'|\alpha|\xi'') d\xi'' (\xi''|\eta''), \quad (18)$$

giving the η -representative in terms of the ξ -representative. These are the transformation equations for the representatives of an observable. Either representative is a linear function of the other, and the same transformation functions are required for passing from one to the other as for the representatives of states.

If we now take a third representation, ζ say, we shall have transformation functions $(\zeta'|\xi')$, $(\xi'|\zeta')$, connecting it with the ξ -representation, and transformation functions $(\zeta'|\eta')$, $(\eta'|\zeta')$, connecting it with the η -representation. There are simple relations between the transformation functions. Equation (13), with ζ instead of η , gives us

$$(\xi'|\zeta') = \phi(\xi')\psi(\zeta').$$

If we substitute for the right-hand side, which consists of the product of two abstract symbols, their representatives in the η -representation, we get

$$(\xi'|\zeta') = \int (\xi'|\eta') d\eta' (\eta'|\zeta'). \quad (19)$$

The conjugate complex equation, which could be deduced independently in the same way, is

$$(\zeta'|\xi') = \int (\zeta'|\eta') d\eta' (\eta'|\xi'). \quad (20)$$

Equations (19) and (20) give the ξ , ζ transformation functions in terms of the ξ , η and η , ζ ones.

If we multiply equation (17) by $d\xi''(\xi''|\eta'')$, putting the new factor on the right-hand side of each term in order to maintain the 'fluency' of the notation, and integrate with respect to ξ'' , we obtain

$$\begin{aligned} \int (\xi'|\alpha|\xi'') d\xi'' (\xi''|\eta'') &= \iiint (\xi'|\eta') d\eta' (\eta'|\alpha|\eta'') d\eta'' (\eta''|\xi'') d\xi'' (\xi''|\eta'') \\ &= \iint (\xi'|\eta') d\eta' (\eta'|\alpha|\eta'') d\eta'' \delta(\eta'' - \eta'') \end{aligned}$$

with the help of (16). Hence

$$\int (\xi'|\alpha|\xi'') d\xi'' (\xi''|\eta'') = \int (\xi'|\eta') d\eta' (\eta'|\alpha|\eta''). \quad (21)$$

We shall call either side of this equation $(\xi'|\alpha|\eta'')$ and consider as it the representative of the observable α in a *mixed representation* (ξ, η) . It is, in fact, a matrix sufficient to determine the observable α and differs from the representative matrices we have previously con-

sidered only in that its rows and its columns refer to two different sets of fundamental states and are therefore no longer in one-one correspondence with each other. The representative matrices of two observables in mixed representations can be added provided they are both in the same mixed representation, *i.e.* we have

$$(\xi'|\alpha + \beta|\eta') = (\xi'|\alpha|\eta') + (\xi'|\beta|\eta').$$

Also they can be multiplied if they are in two different mixed representations such that the columns (specified by the letter on the right-hand side) of the first factor refer to the same set of fundamental states as, and are thus in one-one correspondence with, the rows of the second, *i.e.* we can multiply $(\xi'|\alpha|\eta')$ into $(\eta'|\beta|\zeta')$ to give a product

$$(\xi'|\alpha\beta|\zeta') = \int (\xi'|\alpha|\eta') d\eta' (\eta'|\beta|\zeta').$$

It should be noticed that the representative of unity in the mixed (ξ, η) representation, *i.e.* $(\xi'|1|\eta')$, is just the transformation function $(\xi'|\eta')$ itself, as follows at once from the definition (21). The terms 'diagonal matrix' and 'diagonal element' of course have no meaning when applied to representative matrices in mixed representations. Again, the representatives of the ξ 's and η 's themselves in the mixed (ξ, η) representation are given by the following expressions, as is easily verified by using the left- and right-hand sides of (21) respectively:

$$\begin{aligned} (\xi'|\xi_m|\eta') &= \xi'_m (\xi'|\eta') \\ (\xi'|\eta_m|\eta') &= (\xi'|\eta') \eta'_m. \end{aligned} \quad \} (22)$$

These representatives are thus expressible directly in terms of the transformation function.

The equations of this section have all been written down for the case when the parameters ξ' , η' , ..., labelling fundamental states take on continuous ranges of values. The necessary modifications to be made when some or all of them take on discrete sets, or both discrete sets and continuous ranges, of values are obvious. If in one representation ξ the ξ 's take on, say, continuous ranges of values, then it is not necessary that in another representation η , applying to the same dynamical system, the η 's should also take on continuous ranges of values, although if in one representation the number of fundamental states is finite, then it must be the same in any other representation.

The transformations here discussed from one representation to another may be called canonical transformations. One must take care not to confuse them with contact transformations, defined in § 19, as was frequently done in the earlier literature on quantum

mechanics. The two kinds of transformation are mathematically of the same form, as one sees if one writes the canonical transformation equations (17) and (18) symbolically with S and S^{-1} for the transformation functions $(\xi'|\eta')$ and $(\eta'|\xi')$, but they have quite different meanings. The canonical transformation is a transformation from one representation of observables to another representation of the *same* observables, while the contact transformation is a transformation from one set of observables to another *different* set of observables. For the contact transformation the new observables are connected with one another by the same algebraic and functional relationships as the original ones, while the corresponding results for the canonical transformation merely express the condition that the new representatives are entitled to be called representatives of the same observables. The contact transformation has its analogue in classical mechanics, as has been already mentioned, but the canonical transformation, which is the more important one in quantum mechanics, has, of course, no such analogue, since in the classical theory we do not deal with representations.

§ 28. Probability Amplitudes

Suppose observations to be made of each of a set of commuting observables ξ_m when the system is in a given state ψ . The probability of any given set of results being obtained is equal to, according to § 18, the square of the modulus of the corresponding coefficient in the expansion of ψ (which is assumed to be normalized) in terms of normalized simultaneous eigen- ψ 's of the observables ξ_m . If the observables ξ_m form a complete set, there will be only one simultaneous eigen- ψ for each set of eigenvalues ξ'_m and the coefficients in the expansion of ψ will form a representative of ψ , denoted by $(\xi'|)$. The probability of the set of results ξ'_m being obtained now becomes $|(\xi'|)|^2$. There is thus a physical meaning for the ξ -representative of any normalized ψ , or at least for the modulus of this representative, in terms of the probability for a given result being obtained for a maximum observation consisting in measuring the complete set of observables ξ_m . The same physical meaning can, of course, be given to the representative of any normalized ϕ , which is just the conjugate complex of that of the conjugate imaginary ψ .

Take now the case when ψ is one of the fundamental ψ 's, $\psi(\eta')$, of another representation η . The probability of the results ξ' being

obtained is now given by $|(\xi'|\eta')|^2$, i.e. by the square of the modulus of the transformation function. But the state $\psi(\eta')$ is the one for which the observables η certainly have the values η' . Thus $|(\xi'|\eta')|^2$ gives the probability of the observables ξ having the values ξ' when the η 's are known to have the values η' . For this reason the expression $(\xi'|\eta')$ is called by P. Jordan a *probability amplitude*. There is, as we saw in the preceding section, an uncertainty in its phase, but its modulus is quite definite. The square of its modulus is an ordinary probability. Since

$$|(\xi'|\eta')|^2 = (\xi'|\eta')(\eta'|\xi') = |(\eta'|\xi')|^2,$$

we have the reciprocal theorem, that *the probability of the ξ 's having the values ξ' when the η 's are given to have the values η' is equal to the probability of the η 's having the values η' when the ξ 's are given to have the values ξ' .*

When the ξ 's take on continuous ranges of values, then, as mentioned in § 23, the fundamental ψ 's of a representation must be multiplied by an infinitely small numerical coefficient in order that they may be properly normalized for the purpose of physical interpretations. Further, the theorem of § 18 that we have just used, giving probabilities in terms of the coefficients of an expansion, is no longer true when the expansion consists of an integral. For these reasons the expression we have obtained for the probability of the ξ 's having particular values for a given state does not hold in the continuous case. But in the continuous case in practice we need to know *only the probability of the ξ 's having values lying within specified ranges*. The probability of their having particular values is zero, as could be deduced formally from the theory. The connexion between the probability for the state ψ of the ξ 's having values lying within small specified ranges and the representative of ψ , when the fundamental ψ 's are normalized in accordance with equation (37) or (39) of § 23, will now be obtained. The method used will be to obtain the case of continuous ξ 's as a limiting form of the case of discrete ξ 's when there are very many of them lying very close together.

Take for definiteness the case when there is only one ξ and suppose that it has a very large number of discrete eigenvalues ξ' lying very close together. Let the number of eigenvalues per unit range of ξ' be s , which can vary with ξ' in an arbitrary way. Suppose now that an arbitrary normalized ψ is expanded in terms the eigen- ψ 's, $\psi_{\xi'}$,

which are correctly normalized for the purpose of physical interpretations, i.e.

$$\phi_{\xi} \psi_{\xi} = 1, \quad (23)$$

so that we have

$$\psi = \sum_{\xi'} c_{\xi'} \psi_{\xi'}. \quad (24)$$

Then $|c_{\xi'}|^2$ is the probability of ξ having the value ξ' for this state ψ . We may assume that $c_{\xi'}$ varies only slowly from one value of ξ' to the next, so that the total probability of ξ having a value lying within the range ξ' to $\xi' + d\xi'$, which is small but still large compared with the interval between consecutive eigenvalues ξ' , will be approximately

$$P = |c_{\xi'}|^2 s' d\xi',$$

where s' is the value of s when ξ' is the value of its variable. With the same kind of approximation we can replace the sum in (24) by an integral, which gives us

$$\psi = \int c_{\xi'} \psi_{\xi'} s' d\xi'. \quad (25)$$

We must now introduce eigen- ψ 's, $\psi(\xi')$, that are normalized according to the rule for the continuous case, i.e.

$$\int \phi(\xi') \psi(\xi') d\xi' = 1. \quad (26)$$

The change in the representatives caused by this change in the normalization of the fundamental ψ 's will be of the same nature as that studied in § 24 caused by a change in the weight function, except that in the present case in the limit the change is infinite.

To compare (26) with (23), we deduce from (23) the equation

$$\sum_{\xi''} \phi_{\xi''} \psi_{\xi''} = 1,$$

which, written with an integral instead of a sum, gives

$$\int \phi_{\xi''} \psi_{\xi''} s'' d\xi'' = 1.$$

Since the integrand here vanishes except when $\xi'' = \xi'$, we can replace s'' by $(s's'')^{\frac{1}{2}}$. Thus we can take

$$\phi(\xi') = s'^{\frac{1}{2}} \phi_{\xi'}, \quad \psi(\xi'') = s''^{\frac{1}{2}} \psi_{\xi''},$$

and equation (26) will be satisfied. We now get from (25)

$$\psi = \int c_{\xi'} \psi(\xi') s'^{\frac{1}{2}} d\xi' = \int \psi(\xi') d\xi' (\xi'|),$$

where $(\xi'|)$, the representative of ψ according to the rule for the continuous case, has the value

$$(\xi'|) = c_{\xi'} s'^{\frac{1}{2}}.$$

The probability P now becomes $|(\xi'|)|^2 d\xi'$. Thus *the square of the modulus of the representative gives the probability, per unit range of ξ , of ξ having a given value*. In the case when there are several observ-

ables ξ , it may be shown in the same way that the probability of each ξ_m having a value between ξ'_m and $\xi'_m + d\xi'_n$ is

$$P = |(\xi'|)|^2 d\xi'_1 d\xi'_2 \dots d\xi'_n = |(\xi')|^2 a\xi .$$

Suppose now, in this case of continuous ξ' 's, that we have one of the fundamental ψ 's, $\psi(\eta')$, of the new representation. Suppose the η' 's to take on discrete sets of values. The norm conditions (15) and (16) now become

$$\sum_{\eta'} (\xi'|\eta')(\eta'|\xi'') = \delta(\xi' - \xi'') \quad (28)$$

$$\int (\eta'|\xi') d\xi' (\xi'|\eta'') = \delta_{\eta'\eta''}. \quad (29)$$

These are just the correct normalizing conditions for us to be able to apply the result (27). This is because the first of them gives

$$\phi(\xi')\psi(\xi'') = \delta(\xi' - \xi''), \quad (30)$$

[since equation (28) is just equation (30) written in terms of η -representatives instead of abstract symbols,] showing that the fundamental ψ 's of the ξ -representation are normalized in accordance with (26); while the second of them gives

$$\phi(\eta')\psi(\eta'') = \delta_{\eta'\eta''}, \quad (31)$$

[since equation (29) is just equation (31) written in terms of ξ -representatives instead of abstract symbols,] showing that $\phi(\eta')\psi(\eta') = 1$ or that $\psi(\eta')$ is correctly normalized for the purpose of physical interpretations. Hence we have the result that

$$|(\xi'|\eta')|^2 d\xi' \quad (32)$$

is the probability of the ξ 's having values between ξ' and $\xi' + d\xi'$ when the η 's are given to have the values η' . The transformation function is still a sort of probability amplitude. From (29) we obtain

$$\int |(\xi'|\eta')|^2 d\xi' = 1,$$

which shows that the total probability of ξ' having any value is unity, giving a check on the normalizing conditions.

When both the η 's and the ξ 's take on continuous ranges of values, the transformation function can no longer be used to give actual probabilities in any convenient way. It will still, however, give relative probabilities. Even when $(\xi'|\eta')$ is not normalized with respect to η' correctly for physical interpretations, the expression (32) will still give the probability of the ξ 's having values between ξ' and $\xi' + d\xi'$, apart from a factor independent of ξ' . It will be found in the applications that such relative probabilities are all that is then required.

The two main types of problem in quantum mechanics are to determine the possible results of an experiment and to determine the probability of occurrence of one of these possible results under given initial conditions. The first type consists in calculating the eigenvalues of an observable, while the second always reduces to calculating a probability amplitude or transformation function and taking the square of its modulus. A general method for calculating the transformation function connecting a set of ξ 's with a set of η 's, when algebraic relations between the ξ 's and η 's are given, is as follows. First obtain the matrices $(\xi'|\eta_m|\xi'')$ representing the η 's in the ξ -representation, the only conditions that these matrices need satisfy being the given algebraic relations. One can now use the equations

$$\int (\xi'|\eta_m|\xi'') d\xi'' (\xi''|\eta') = (\xi'|\eta')\eta'_m,$$

which follow at once from (21) and (22). These are linear integral equations in the variables ξ' for the unknowns $(\xi'|\eta')$. They are, in fact, the standard equations of the theory of eigenvalues and the solutions, when normalized, are just the transformation functions. These solutions are often called eigenfunctions of the matrix $(\xi'|\eta_m|\xi'')$, which determines them. An application of this method will be made in § 35 to a case in which the integral equations reduce to differential equations on account of $(\xi'|\eta_m|\xi'')$ involving the δ function and its derivatives.

§ 29. Example

We have seen in § 26 that if we have any set of observables ξ_m that commute with one another, then there exists a representation, called the ξ -representation, in which each of them is represented by a diagonal matrix, whose diagonal elements are then its eigenvalues. This fact is of very great value in applications of the theory and usually forms the starting-point in any calculation of representatives. To illustrate how it is used, two simple examples will now be given, which will later be found to be of physical importance.

The first will concern the observables p and q satisfying

$$qp - pq = i,$$

which were introduced in § 12. Our problem will be to find the eigenvalues of $p^2 + q^2$. We shall assume that p and q are both real observables. We can then infer by an elementary argument that $p^2 + q^2$ cannot have any negative eigenvalues. We see that the eigen-

values of p^2 cannot be negative, since they are the squares of the eigenvalues of p , which are all real. It follows that the average value of p^2 for any state ψ cannot be negative. Similarly the average of q^2 for this state ψ cannot be negative. Hence the average of p^2+q^2 for the state ψ also cannot be negative. Thus p^2+q^2 cannot have a negative eigenvalue, since if it did it would have a negative average value, equal to this eigenvalue, for the corresponding eigenstate.

Let

$$\begin{aligned} A &= (p+iq)(p-iq) \\ &= p^2+q^2+i(qp-pq) \\ &= p^2+q^2-1. \end{aligned}$$

We then have

$$(p-iq)(p+iq) = p^2+q^2+1 = A+2,$$

and hence

$$A(p+iq) = (p+iq)(p-iq)(p+iq) = (p+iq)(A+2).$$

We now rewrite this equation in terms of the representatives of the symbols it involves, in a representation in which A is diagonal. This gives

$$\Sigma_{A'''}(A'|A|A''')(A'''|p+iq|A'') = \Sigma_{A'''}(A'|p+iq|A''')(A'''|A+2|A''),$$

which, since $(A'|A|A''') = A'\delta_{A'A'''}$,

reduces to $A'(A'|p+iq|A'') = (A'|p+iq|A'')(A''+2)$.

Hence either $(A'|p+iq|A'') = 0$ or $A' = A''+2$.

We have by a direct application of the matrix law of multiplication, where A' is any eigenvalue of A ,

$$(A'|p+iq)(p-iq)|A') = \Sigma_{A''}(A'|p+iq|A'')(A''|p-iq|A'), \quad (33)$$

the summation being extended over all eigenvalues A'' . But we have seen that $(A'|p+iq|A'')$ vanishes unless $A'' = A'-2$. Thus all the terms in the summation vanish except the one for which $A'' = A'-2$. If, now, $A'-2$ is not an eigenvalue of A , then all the terms in the summation will vanish without exception, and we shall have

$$0 = (A'|p+iq)(p-iq)|A') = (A'|A|A') = A'.$$

We have therefore obtained the result that if A' is any eigenvalue of A , either $A'-2$ is another eigenvalue or $A' = 0$. Thus if A' is any eigenvalue, we shall have the series of eigenvalues A' , $A'-2$, $A'-4$, $A'-6\dots$, which cannot extend to $-\infty$ since, as we have already seen, there can be no negative eigenvalues for p^2+q^2 , which is equal to $A+1$. This series of eigenvalues must therefore terminate,

and can terminate only with the value zero. Thus the eigenvalues of A are $0, 2, 4, 6 \dots$, and those of p^2+q^2 are $1, 3, 5, 7 \dots$

The representatives of p and q can now easily be obtained. Equation (33) reduces to

$$A' = (A'|p+iq|A'-2)(A'-2|p-iq|A').$$

The two factors on the right here are conjugate complex quantities, on account of equation (4). Hence

$$(A'|p+iq|A'-2) = A'^{\frac{1}{2}}e^{i\gamma},$$

where γ' is a real function of A' . All the elements not of this type of the matrix representing $p+iq$ vanish. The conjugate complex observable $p-iq$ is represented by

$$(A'-2|p-iq|A') = A'^{\frac{1}{2}}e^{-i\gamma},$$

with all the matrix elements not of this type vanishing. Hence

$$\left. \begin{aligned} (A'|p|A'-2) &= \frac{1}{2}A'^{\frac{1}{2}}e^{i\gamma} & (A'|q|A'-2) &= -\frac{1}{2}iA'^{\frac{1}{2}}e^{i\gamma'} \\ (A'-2|p|A') &= \frac{1}{2}A'^{\frac{1}{2}}e^{-i\gamma} & (A'-2|q|A') &= \frac{1}{2}iA'^{\frac{1}{2}}e^{-i\gamma'} \end{aligned} \right\} \quad (34)$$

and all the matrix elements representing p and q that are not of these types vanish. The occurrence of the arbitrary phase γ' in these representatives for p and q is in accordance with the remark of § 26, that a representation is not completely determined by the observables that are represented by diagonal matrices.

The eigenvalues of A form, as we have seen, a discrete set and hence in the representation with A diagonal the number of fundamental states is enumerable. This is rather remarkable in view of the fact that we can obtain another representation in which the number of fundamental states is equal to the number of points on a line, for example, the representation in which p is diagonal, since, as shown in § 19, the eigenvalues of p include all numbers from $-\infty$ to ∞ . Thus by counting the number of independent states of a system in different ways, one may obtain different cardinal numbers as result.

§ 30. Second Example

Our second example will concern three observables α, β, γ that satisfy

$$\left. \begin{aligned} \alpha\beta - \beta\alpha &= i\gamma \\ \beta\gamma - \gamma\beta &= i\alpha \\ \gamma\alpha - \alpha\gamma &= i\beta. \end{aligned} \right\} \quad (35)$$

Let

$$\alpha^2 + \beta^2 + \gamma^2 = \theta.$$

Our problem will be to determine the eigenvalues of α , β , γ , and θ . We shall assume α , β , and γ are real. We can then infer that θ cannot have any negative eigenvalues, by a similar argument to that at the beginning of our previous example.

We have

$$\begin{aligned}\gamma\alpha^2 - \alpha^2\gamma &= (\gamma\alpha - \alpha\gamma)\alpha + \alpha(\gamma\alpha - \alpha\gamma) \\ &= i\beta\alpha + i\alpha\beta\end{aligned}$$

from the third of equations (35). Similarly

$$\begin{aligned}\gamma\beta^2 - \beta^2\gamma &= (\gamma\beta - \beta\gamma)\beta + \beta(\gamma\beta - \beta\gamma) \\ &= -i\alpha\beta - i\beta\alpha.\end{aligned}$$

Hence

$$\gamma(\alpha^2 + \beta^2) - (\alpha^2 + \beta^2)\gamma = 0,$$

so that

$$\gamma\theta - \theta\gamma = 0.$$

Thus θ commutes with γ , and therefore from symmetry it commutes also with α and β . Hence it commutes with any function of α , β , and γ .

We thus have an observable θ commuting with all the observables that occur in the problem. *Whenever we find an observable having this property, we should expect to be able to treat it simply as a number in all subsequent investigations*, as by so doing we do not invalidate any of the algebraic equations that it satisfies. A formal proof of the legitimacy of this proceeding is as follows. We use a representation in which θ is diagonal, together with certain other observables, κ say, so that any observable P is represented by $(\theta'\kappa'|P|\theta''\kappa'')$. From the condition

$$\theta P - P\theta = 0$$

we obtain

$$\theta'(\theta'\kappa'|P|\theta''\kappa'') - (\theta'\kappa'|P|\theta''\kappa'')\theta'' = 0.$$

Hence

$$(\theta'\kappa'|P|\theta''\kappa'') = 0$$

unless $\theta' = \theta''$. Thus all the matrix elements representing any observable in the problem vanish unless they are of the type $(\theta'\kappa'|P|\theta'\kappa'')$. It follows that when any equation between the observables is expressed in terms of their representatives, all the matrix elements throughout the equation will refer to one and the same value of θ' . This value for θ' need not be explicitly referred to in the notation for the matrix elements, so that we may write $(\theta'\kappa'|P|\theta'\kappa'')$ simply as $(\kappa'|P|\kappa'')$. The equations will now be of exactly the same form as if θ were a number, equal to this θ' , and we used a representation defined by the κ 's without the help of θ .

We shall apply this method to the present example. Thus we shall

consider θ to be a definite number and on this basis work out the eigenvalues of γ . Those of α and β will be the same, from symmetry. Any numerical value that we give to θ which is consistent with equations (35) will be an eigenvalue of θ . Since α and β are real, we can infer that the average value of γ^2 for any state cannot exceed θ and hence the eigenvalues of γ^2 cannot exceed θ . Thus the eigenvalues of γ cannot be greater than $\theta^{\frac{1}{2}}$ or less than $-\theta^{\frac{1}{2}}$. The fact that any numerical value that we take for θ must be positive or zero, since, as we have already seen, any eigenvalue of θ must be positive or zero, makes this restriction on the eigenvalues of γ reasonable. We have from (35)

$$\begin{aligned} (\alpha + i\beta)\gamma - \gamma(\alpha + i\beta) &= -i\beta - \alpha \\ &= -(\alpha + i\beta) \end{aligned}$$

or

$$(\alpha + i\beta)\gamma = (\gamma - 1)(\alpha + i\beta).$$

If we express this result in the γ -representation, we get

$$(\gamma'|\alpha + i\beta|\gamma'')\gamma'' = (\gamma' - 1)(\gamma'|\alpha + i\beta|\gamma'').$$

Hence either $(\gamma'|\alpha + i\beta|\gamma'') = 0$ or $\gamma'' = \gamma' - 1$. Now if γ' is any eigenvalue of γ ,

$$(\gamma' | (\alpha + i\beta)(\alpha - i\beta) | \gamma') = \sum_{\gamma''} (\gamma' | \alpha + i\beta | \gamma'') (\gamma'' | \alpha - i\beta | \gamma'), \quad (36)$$

the summation being over all eigenvalues γ'' . The terms on the right-hand side all vanish except the one for which $\gamma'' = \gamma' - 1$. If $\gamma' - 1$ is not an eigenvalue of γ , then they all vanish and we have

$$(\gamma' | (\alpha + i\beta)(\alpha - i\beta) | \gamma') = 0.$$

But

$$\begin{aligned} (\alpha + i\beta)(\alpha - i\beta) &= \alpha^2 + \beta^2 - i(\alpha\beta - \beta\alpha) \\ &= \alpha^2 + \beta^2 - \gamma \\ &= \theta - \gamma^2 + \gamma \\ &= \theta + \frac{1}{4} - (\gamma - \frac{1}{2})^2. \end{aligned}$$

Hence if $\gamma' - 1$ is not an eigenvalue of γ , we have

$$\begin{aligned} 0 &= (\gamma' | \theta + \frac{1}{4} - (\gamma - \frac{1}{2})^2 | \gamma') \\ &= \theta + \frac{1}{4} - (\gamma' - \frac{1}{2})^2 \end{aligned}$$

or

$$\gamma' = \frac{1}{2} \pm k,$$

where k is defined as the positive square root

$$k = (\theta + \frac{1}{4})^{\frac{1}{2}}. \quad (37)$$

Thus if γ' is any eigenvalue, we shall have the series of eigenvalues γ' , $\gamma' - 1$, $\gamma' - 2 \dots$, which must terminate since there can be no eigenvalue less than $-\theta^{\frac{1}{2}}$. The last member of the series must be

either $\frac{1}{2}+k$ or $\frac{1}{2}-k$, and since there is no eigenvalue greater than $\theta^{\frac{1}{2}}$, and thus none greater than k , it must be $\frac{1}{2}-k$. Thus the eigenvalues of γ are $\frac{1}{2}-k, \frac{3}{2}-k, \frac{5}{2}-k \dots$

If we reverse the order of the factors in the product whose representative occurs on the left-hand side of (36), we can deduce by a similar argument that if γ' is any eigenvalue of γ , either $\gamma'+1$ is another eigenvalue or $\gamma' = -\frac{1}{2} \pm k$, and we can infer from this that the eigenvalues of γ are $k-\frac{1}{2}, k-\frac{3}{2}, k-\frac{5}{2} \dots$. By combining these two results, we see that $\frac{1}{2}-k$ and $k-\frac{1}{2}$ must differ by an integer, so that k must be an integer or half an odd integer. The eigenvalues of γ are then

$$k-\frac{1}{2}, k-\frac{3}{2}, k-\frac{5}{2} \dots -k+\frac{3}{2}, -k+\frac{1}{2}, \quad (38)$$

which shows incidentally that k must not be zero, as follows also from its defining equation (37). The corresponding value for θ is $k^2-\frac{1}{4}$, so that the eigenvalues of θ are all of this form.

A new point that is brought out by this example is that if we have two observables that commute and choose arbitrarily one of the eigenvalues of each, then there will not necessarily exist a state for which each observable has its chosen eigenvalue, *i.e.* a state that is a simultaneous eigenstate belonging to these two eigenvalues. Thus the eigenvalues of γ include all integers and half odd integers, and those of θ include all numbers of the form $k^2-\frac{1}{4}$ where k is an integer not zero or half odd integer, but there exists a state for which γ and θ have the values γ' and $k^2-\frac{1}{4}$ respectively only provided γ' is one of the numbers (38). Such restrictions on the possible simultaneous eigenstates of two or more commuting observables do not in any way invalidate our general theory.

EQUATIONS OF MOTION AND QUANTUM CONDITIONS

§ 31. General Remarks

The theory that has been developed so far contains a complete account of the new concepts and mathematical machinery required in quantum mechanics and also all the general physical laws. Only the general properties of states and observables have, however, been discussed, no reference being made to the particular conditions that they satisfy in the case of a specified dynamical system. We must now consider the form of these particular conditions and so make the theory applicable to given physical problems. It should be understood that the assumptions that will now be made are on quite a different footing from the foregoing ones. We are now concerned not with general physical laws applying to the whole of nature, but with special assumptions referring to a given physical problem, such as the interaction of a certain number of electrons and atomic nuclei. These assumptions will show how the information that we are dealing with a certain number of particles of given masses interacting according to given laws of force is to be made use of, and will give us equations which may be considered as forming the mathematical specification of which dynamical system is under consideration. Future developments of the theory may show that these assumptions are only approximate and require modifications ; in fact, as they will now be formulated, they are not in agreement with the principle of relativity and will at any rate require modifications on this account when applied to rapidly moving particles. On the other hand, the assumptions of the four preceding chapters are so closely interconnected that one could hardly modify them in any way without getting an entirely different scheme of mechanics, and the successes of the theory are so great as to make it fairly certain that no such modifications will be required, at least for the purpose of explaining the ordinary physical and chemical properties of matter. The theory of these four chapters is in agreement with the principle of relativity ; in fact it is so general that it is independent of any special relations between space and time. We must, of course, for this to be true, adopt a more general definition of an observable than the value of a variable at some instant of time, which we can do by considering

an observable to be the quantity measured in *any* observation and to be defined by the way the observation is made, together with the positions of the various component parts of the observing apparatus and the times when they are set working, if necessary. An observable now need not refer to an instant of time in some frame of reference, so that there is no conflict with relativity on this account. For the non-relativistic theory of the present chapter the previous definition of an observable is adequate.

If we are dealing with a given dynamical system, we shall have given dynamical variables, whose values at any time are what we call observables, and we shall require conditions that will determine the values of these variables at all times when their values at some particular time are known. These conditions will be the equations of motion of the system. In classical mechanics they would be sufficient to form the mathematical specification of the dynamical system under consideration. This is not so, however, in quantum mechanics, where additional relations are necessary for this purpose, which take the form of equations connecting the values of the variables at a particular time, of such a nature that they can replace the commutative law of multiplication of the classical theory. These additional relations are called quantum conditions. *It is only when the quantum conditions are given as well as the equations of motion that we know as much about the variables as in the classical theory and can consider the dynamical system as mathematically completely specified.* The equations of motion and quantum conditions are very closely connected with each other, and one cannot make any progress in solving a problem until they are *both* known.

Our problem is now to determine the quantum conditions and equations of motion for any given dynamical system, such as that formed by given electrons and atomic nuclei interacting. It is known that classical mechanics gives an accurate description of dynamical systems under certain limiting conditions, *e.g.* when the masses are large. One would therefore expect to be able to obtain a theory of these systems when the limiting conditions do not hold by making some natural generalizations in the classical equations of motion and choosing quantum conditions that form natural generalizations of the classical conditions that all the variables commute. It will be found that one can in this way obtain a quantum theory of individual dynamical systems analogous to the classical theory. This quantum

theory will not, however, include all the systems with which one has to deal, but only a large and important class of them, there being systems in the quantum theory which have no classical analogues (*e.g.* that consisting of a photon interacting with an atom, which will be treated in Chapter XII), for the treatment of which we must in each case choose special quantum conditions and equations of motion, either by means of special theoretical considerations or to fit experimental facts.

§ 32. Poisson Brackets

The classical equations of motion which we have to generalize may be written in the form

$$\frac{dq_r}{dt} = \dot{q}_r = \frac{\partial H}{\partial p_r} \quad \dot{p}_r = -\frac{\partial H}{\partial q_r}, \quad (1)$$

where the q 's and p 's are a set of generalized co-ordinates and their canonically conjugate momenta and H is the Hamiltonian, which is a given function of the q 's and p 's for a given dynamical system and is equal to the energy when it does not involve the time explicitly. These equations of motion involve partial differential coefficients, which in general have no meaning for dynamical variables in the quantum theory. We get over this difficulty by observing that *the equations of motion (1), and also all other important equations of general classical dynamics, can be written in a form in which they involve partial differential coefficients only through Poisson Bracket expressions*, and that, as we shall now find, these bracket expressions have their analogues in the quantum theory. Any two variables ξ and η have a Poisson Bracket (abridged to P. B.), denoted by $[\xi, \eta]$ and defined in the classical theory by

$$[\xi, \eta] = \sum_r \left\{ \frac{\partial \xi}{\partial q_r} \frac{\partial \eta}{\partial p_r} - \frac{\partial \xi}{\partial p_r} \frac{\partial \eta}{\partial q_r} \right\}. \quad (2)$$

These P. B.'s owe their importance to the fact that they remain invariant under a contact transformation (*i.e.* a transformation to a new set of canonical variables p_r^* , q_r^* such that the form of the equations of motion (1) remains unaltered), which results in the equations of motion being expressible in terms of P. B.'s. We have in fact

$$\dot{q}_r = [q_r, H] \quad \dot{p}_r = [p_r, H], \quad (3)$$

and more generally, for any variable ξ ,

$$\begin{aligned} \dot{\xi} &= \sum_r \left\{ \frac{\partial \xi}{\partial q_r} \dot{q}_r + \frac{\partial \xi}{\partial p_r} \dot{p}_r \right\} = \sum_r \left\{ \frac{\partial \xi}{\partial q_r} \frac{\partial H}{\partial p_r} - \frac{\partial \xi}{\partial p_r} \frac{\partial H}{\partial q_r} \right\} \\ &= [\xi, H]. \end{aligned} \quad (4)$$

To find the quantum analogues of these P.B.'s we shall note some of their general properties and try to choose the quantum P.B.'s so that they shall have the same properties. The following relations follow at once from the definition (2).

$$[\xi, \eta] = -[\eta, \xi] \quad (5)$$

$$\{\xi, c\} = 0, \quad (6)$$

where c is a number,

$$\left. \begin{aligned} [\xi_1 + \xi_2, \eta] &= [\xi_1, \eta] + [\xi_2, \eta] \\ [\xi, \eta_1 + \eta_2] &= [\xi, \eta_1] + [\xi, \eta_2] \end{aligned} \right\} (7)$$

$$\left. \begin{aligned} [\xi_1 \xi_2, \eta] &= \sum_r \left\{ \left(\frac{\partial \xi_1}{\partial q_r} \xi_2 + \xi_1 \frac{\partial \xi_2}{\partial q_r} \right) \frac{\partial \eta}{\partial p_r} - \left(\frac{\partial \xi_1}{\partial p_r} \xi_2 + \xi_1 \frac{\partial \xi_2}{\partial p_r} \right) \frac{\partial \eta}{\partial q_r} \right\} \\ &= [\xi_1, \eta] \xi_2 + \xi_1 [\xi_2, \eta] \\ [\xi, \eta_1 \eta_2] &= [\xi, \eta_1] \eta_2 + \eta_1 [\xi, \eta_2]. \end{aligned} \right\} (8)$$

Also the identity

$$[\xi, [\eta, \zeta]] + [\eta, [\zeta, \xi]] + [\zeta, [\xi, \eta]] = 0 \quad (9)$$

is easily verified. Equations (7) express that the P.B. $[\xi, \eta]$ involves ξ and η linearly, while equations (8) correspond to the ordinary rules for differentiating a product.

We can define the quantum P.B. so that it also has all these properties, provided the order of the factors ξ_1 and ξ_2 in the first of equations (8) is preserved throughout the equation, as in the way we have here written it, and similarly for the η_1 and η_2 in the second of equations (8). These conditions are sufficient to determine the form of the quantum P.B. uniquely, as may be seen from the following argument. We can evaluate the P.B. $[\xi_1 \xi_2, \eta_1 \eta_2]$ in two different ways, since we can use either of the two formulas (8) first, thus,

$$\begin{aligned} [\xi_1 \xi_2, \eta_1 \eta_2] &= [\xi_1, \eta_1 \eta_2] \xi_2 + \xi_1 [\xi_2, \eta_1 \eta_2] \\ &= \{[\xi_1, \eta_1] \eta_2 + \eta_1 [\xi_1, \eta_2]\} \xi_2 + \xi_1 \{[\xi_2, \eta_1] \eta_2 + \eta_1 [\xi_2, \eta_2]\} \\ &= [\xi_1, \eta_1] \eta_2 \xi_2 + \eta_1 [\xi_1, \eta_2] \xi_2 + \xi_1 [\xi_2, \eta_1] \eta_2 + \xi_1 \eta_1 [\xi_2, \eta_2] \end{aligned}$$

and $[\xi_1 \xi_2, \eta_1 \eta_2] = [\xi_1 \xi_2, \eta_1] \eta_2 + \eta_1 [\xi_1 \xi_2, \eta_2]$

$$= [\xi_1, \eta_1] \xi_2 \eta_2 + \xi_1 [\xi_2, \eta_1] \eta_2 + \eta_1 [\xi_1, \eta_2] \xi_2 + \eta_1 \xi_1 [\xi_2, \eta_2].$$

Equating these two results, we obtain

$$[\xi_1, \eta_1] (\xi_2 \eta_2 - \eta_2 \xi_2) = (\xi_1 \eta_1 - \eta_1 \xi_1) [\xi_2, \eta_2].$$

Since this condition holds with ξ_1 and η_1 quite independent of ξ_2 and η_2 , we must have

$$\xi_1 \eta_1 - \eta_1 \xi_1 = i\hbar [\xi_1, \eta_1]$$

$$\xi_2 \eta_2 - \eta_2 \xi_2 = i\hbar [\xi_2, \eta_2],$$

where \hbar must not depend on ξ_1 and η_1 or ξ_2 and η_2 and also must commute with $(\xi_1 \eta_1 - \eta_1 \xi_1)$, so that it must be a number. We want the P.B. of two real variables to be real, as in the classical theory, which requires that \hbar shall be a real number when introduced, as here, with the coefficient i . We are thus led to the following *general formula for the quantum P.B.* $[\xi, \eta]$ of any two variables ξ and η ,

$$\xi\eta - \eta\xi = i\hbar [\xi, \eta], \quad (10)$$

in which \hbar is a new universal constant having the dimensions of action. In order that the theory may agree with experiment, we must take \hbar equal to $h/2\pi$, where h is the universal constant that was introduced by Planck. It is easily verified that the quantum P.B. defined by (10) satisfies all the conditions (5), (6), (7), (8), and (9). These conditions often provide a more convenient way of actually evaluating a complicated P.B., by enabling one to express it in terms of simpler P.B.'s whose values may be known, than that afforded by a direct application of (10).

§ 33. Equations of Motion and Quantum Conditions obtained from Analogy with the Classical Theory

The assumption that the P.B. defined by (10) is the analogue of the classical one enables us to take over the classical equations of motion (3) and (4) into the quantum theory and also any other classical equations expressible in terms of P.B.'s. Further, the assumption that the P.B.'s of the p 's and q 's, which P.B.'s in the classical theory have the values

$$\begin{aligned} [q_r, q_s] &= 0 & [p_r, p_s] &= 0 \\ [q_r, p_s] &= \delta_{rs}, \end{aligned} \quad \left. \right\} (11)$$

have these same values in the quantum theory, provides us with quantum conditions, since we can now, with the help of (5), (6), (7), (8), evaluate the P.B. $[\xi, \eta]$ of any two analytic functions ξ and η of the p 's and q 's and thus obtain, by using (10), an equation for $\xi\eta - \eta\xi$ capable of replacing the classical condition that $\xi\eta - \eta\xi = 0$. We have thus solved the problem of obtaining equations of motion and quantum conditions forming a natural generalization of the classical theory. *The classical theory is, in fact, given by the limiting case $\hbar = 0$ of the quantum theory.*

The quantum conditions and equations of motion may be written without the use of P.B.'s, if we eliminate the P.B.'s with the help

of their defining equation in the quantum theory, equation (10). We obtain in this way for the quantum conditions (11)

$$\left. \begin{aligned} q_r q_s - q_s q_r &= 0 & p_r p_s - p_s p_r &= 0 \\ q_r p_s - p_s q_r &= i\hbar \delta_{rs}, \end{aligned} \right\} \quad (12)$$

and for the equation of motion (4)

$$i\hbar \dot{\xi} = \xi H - H \xi. \quad (13)$$

The condition for a variable ξ to be constant is that it shall commute with the Hamiltonian H .

The notion of P. B.'s is more fundamental in the quantum theory than in the classical theory, as is shown by the fact that one can define a P. B. in the quantum theory without reference to a set of canonical variables, which is not possible in the classical theory. For this same reason the notion of a set of canonical variables is less important in the quantum theory than in the classical theory. The notion of canonical variables is in the classical theory a dynamical notion, but in the quantum theory it is merely an algebraic notion, as the conditions defining when variables are canonical are then expressible by algebraic equations (11) or (12). Equations (11) may be considered as defining canonical variables also in the classical theory, but they then have no meaning unless the q_r, p_r are functions of another set of variables q_r^*, p_r^* which are given to be canonical, as otherwise the P. B.'s are undefined. A transformation from one set of canonical variables to another is called in the classical theory a contact transformation, and this name may conveniently be taken over into the quantum theory. The transformations discussed in § 19 evidently do transform one set of canonical variables into another, since, as shown in § 19, they leave algebraic relations between the variables unaltered and the conditions for variables to be canonical in the quantum theory are algebraic.

It should be understood that the symbols $q, p, \&c.$, in the equations we are now dealing with really denote the values $q(t), p(t), \&c.$, of the variables at some particular time t that is not specifically mentioned, so that our equations are equations between observables depending on a parameter t . The $\dot{\xi}$ in (4) and (13), defined as the rate of change of the observable $\xi(t)$ with respect to the parameter t , is also an observable. For observables $\xi(t), \eta(t)$ depending on a parameter t , we have the laws

$$\frac{d}{dt}(\xi + \eta) = \frac{d}{dt}\xi + \frac{d}{dt}\eta \quad \frac{d}{dt}(\xi\eta) = \xi \frac{d\eta}{dt} + \frac{d\xi}{dt}\eta,$$

which are consistent with the general quantum equation of motion (4) or (13), on account of their analogy with (7) and (8) respectively.

It is legitimate for us to assume the quantum conditions (11) or (12) only for one particular time, and we must then deduce that they hold at all times from the equations of motion. We can do this by observing that, if equations (11) or (12) hold at one particular time t , then the time-rate of change of their left-hand sides must vanish at time t , so that they will hold also at time $t+dt$, or alternatively by observing that, from the general equation of motion (13), the values of the p 's and q 's at time $t+dt$ are connected with their values at time t by an infinitesimal contact transformation of the type (29) of § 19. In order that we may be able to consider the commutative law of multiplication of the classical theory as completely replaced by our quantum equations, it is necessary that we should be able to evaluate expressions of the form $\xi(t_1)\eta(t_2) - \eta(t_2)\xi(t_1)$. This we can do by using the equations of motion to express $\xi(t_1)$ and $\eta(t_2)$ in terms of the p 's and q 's at some one time t and then applying the quantum conditions (12).

The equation of motion (4) or (13) must be generalized when ξ involves the time explicitly as well as through the p 's and q 's. The classical generalization of (4) for this case is

$$\dot{\xi} = \frac{\partial \xi}{\partial t} + [\xi, H], \quad (14)$$

which may be taken over directly into the quantum theory. The generalization of (13) is thus

$$i\hbar \dot{\xi} = i\hbar \frac{\partial \xi}{\partial t} + \xi H - H\xi. \quad (15)$$

The Hamiltonian H is a constant when and only when it does not involve the time explicitly. The equations of motion are not affected by the addition to the Hamiltonian of an arbitrary numerical constant, even one that varies with the time.

We are now in a position to be able to work out all that we require for any dynamical system when this system is specified by a Hamiltonian function H , given in terms of the q 's and p 's and perhaps also containing t explicitly. It should be observed that the order of the factors of products in the expression for H may be important, since our variables do not now all commute, so that there is a greater variety of Hamiltonians in the quantum theory than in the classical theory. Thus for a given Hamiltonian of the classical theory there

is not in general a unique corresponding Hamiltonian of the quantum theory, so that when one is given a dynamical system in the classical theory it is in general meaningless to talk about the *same* system in the quantum theory. There are, however, exceptions to this, it being possible in many cases to use the same language for describing dynamical systems in the quantum theory as in the classical theory without practical ambiguity. For example, one can describe a dynamical system as that of a particle of mass m moving in a field of force derivable from a potential function V . The Hamiltonian for this system in the classical theory would be, when expressed in Cartesian co-ordinates,

$$H = \frac{1}{2m} (p_x^2 + p_y^2 + p_z^2) + V(xyz).$$

One can without ambiguity say that the same system in the quantum theory is that having this same Hamiltonian, since this Hamiltonian does not contain any product of the type xp_x for which the order of the factors is important. It should be remarked that this freedom from ambiguity in the passage from a classical Hamiltonian to a quantum one can be maintained only provided one uses always Cartesian co-ordinates, as in general different quantum Hamiltonians would be obtained, differing from one another by terms containing \hbar as a factor, if one were to take over the classical Hamiltonian expressed in different kinds of curvilinear co-ordinates.

§ 34. Schrödinger's Form for the Quantum Conditions

In this section and the following one some of the more important consequences of the quantum conditions (12) will be obtained. We shall here be concerned exclusively with the values of the variables q, p at one particular time, which will not be specifically mentioned.

Equation (26) of Chapter II is, apart from the numerical factor \hbar , the same as the quantum condition connecting any co-ordinate q_r with its conjugate momentum p_r . Thus we can take over the consequences of that equation and apply them to our present q_r and p_r , with insertion of the factor \hbar where necessary. Equation (27) of Chapter II gives us in this way

$$fp_r - p_rf = i\hbar df/dq_r, \quad (16)$$

where f is any function of q_r expressible as a power series. This equation evidently holds also when f is a function of the other q 's as

well as q_r , provided the total differential coefficient is replaced by a partial one. Again, from the argument at the end of § 19, we can infer that each q_r and p_r must have as eigenvalues all numbers from $-\infty$ to ∞ . This would actually be the case, for instance, if they were Cartesian co-ordinates and momenta of particles.

It will now be shown that, ignoring a certain indefiniteness, one can give a meaning to the operator $\partial/\partial q_r$, applied to a ψ -symbol, or *one can differentiate a ψ with respect to an observable q_r* . The simplest way of treating this problem is to suppose the ψ to be represented in a representation in which, amongst others, the observable q_r is diagonal. The representative $(q'_r|)$ of ψ will be a function of the variable q'_r , whose domain extends from $-\infty$ to ∞ , and can therefore be differentiated partially with respect to q'_r , giving another function $\partial(q'_r|)/\partial q'_r$ of q'_r defined for this same domain $-\infty$ to ∞ . This new function will represent a ψ -symbol, which we define to be $\partial\psi/\partial q_r$. It would, of course, be strictly correct to say that one can give a meaning to the operator $\partial/\partial q_r$ applied to a ψ -symbol only provided for each ψ there is one unique $\partial\psi/\partial q_r$, i.e. provided the above procedure for obtaining $\partial\psi/\partial q_r$ gives a result independent of which of all the possible representations in which q_r is diagonal we use, and this is not the case. There is thus an indefiniteness in the meaning of the operator $\partial/\partial q_r$ applied to a ψ -symbol, the extent of which we shall now investigate.

Let us take first the case of a system of one degree of freedom, so that there is only one co-ordinate q and only one variable q' in the representative $(q'|)$ of a ψ . By differentiating this representative we obtain $\partial(q'|)/\partial q'$, the representative of a possible $\partial\psi/\partial q$, say $(\partial\psi/\partial q)_a$. Now in the present one-dimensional case the most general canonical transformation we can make such that q remains diagonal is that which involves the multiplying of the representative of any ψ -symbol by an arbitrary phase factor. Thus the new representative of ψ will be of the form

$$(q'|)^* = e^{iF'}(q'|), \quad (17)$$

where F' is short for $F(q')$, a real function of q' . If we use this new representation to define $\partial\psi/\partial q$, we obtain a new $\partial\psi/\partial q$, say $(\partial\psi/\partial q)_b$, whose representative in the new representation is

$$\frac{\partial}{\partial q'}(q'|)^* = e^{iF'} \frac{\partial}{\partial q'}(q'|) + ie^{iF'} \frac{dF'}{dq'}(q'|).$$

The representative of $(\partial\psi/\partial q)_b$ in the original representation is therefore

$$e^{-iF'} \frac{\partial}{\partial q'} (q'|)^* = \frac{\partial}{\partial q'} (q'|) + i \frac{dF'}{dq'} (q'|),$$

and hence

$$\left(\frac{\partial\psi}{\partial q} \right)_b = \left(\frac{\partial\psi}{\partial q} \right)_a + i \frac{dF}{dq} \psi. \quad (18)$$

This is an equation giving the general connexion between two $(\partial\psi/\partial q)$'s. It shows that the indefiniteness in the operator $\partial/\partial q$ consists in the possible addition of an arbitrary pure imaginary function of q .

In the n -dimensional case the general canonical transformation which leaves a single q , q_r , say, diagonal is much more general than a mere change of phase and thus the indefiniteness in the operator $\partial/\partial q_r$ is much greater than in the one-dimensional case. Whenever we use this operator, however, we shall deal not with a single $\partial/\partial q_r$ alone, but with the whole set $\partial/\partial q_1, \partial/\partial q_2 \dots \partial/\partial q_n$ together, which will make only those meanings for the operators useful that arise from a representation in which all the q 's are simultaneously diagonal. The arbitrariness in this representation is then again merely that of the phase, like (17), and leads again to the form (18) for the connexion between two $(\partial\psi/\partial q_r)$'s, namely

$$\left(\frac{\partial\psi}{\partial q_r} \right)_b = \left(\frac{\partial\psi}{\partial q_r} \right)_a + i \frac{\partial F}{\partial q_r} \psi, \quad (19)$$

where F is now an arbitrary real function of all the q 's. Thus the indefiniteness in the operators $\partial/\partial q_r$ now consists in the possible addition to each simultaneously of a function of the q 's, of the form $i\partial F/\partial q_r$, for the r -th. This small amount of indefiniteness has, however, been attained only by our considering each $\partial/\partial q_r$ as not specified by the observable q_r alone, but by q_r as one of a given complete set of commuting observables $q_1, q_2 \dots q_n$.

The operators $\partial/\partial q_r$ applied to ψ -symbols are linear operators that can be applied to an arbitrary ψ and are thus just ordinary observables. We shall call $\partial/\partial q_r$, considered as an observable, π_r . The representative of π_r in the q -representation used for defining $\partial\psi/\partial q_r$ is

$$(q'|\pi_r|q'') = -\delta(q'_1 - q''_1) \delta(q'_2 - q''_2) \dots \delta(q'_{r-1} - q''_{r-1}) \delta'(q'_r - q''_r) \delta(q'_{r+1} - q''_{r+1}) \dots \delta(q'_n - q''_n), \quad (20)$$

which is similar to expression (49) of Chapter IV. The matrix representing π_r is thus antisymmetrical, showing, according to equa-

tion (23) of § 21, that π_r is a pure imaginary observable. The form of (20) shows also that when π_r is multiplied into a ϕ -symbol, the result is

$$\phi\pi_r = -\partial\phi/\partial q_r, \quad (21)$$

in which $\partial\phi/\partial q_r$ is defined through its q -representative in the same way as $\partial\psi/\partial q_r$ was.

The commutability relations connecting the π 's with each other and with the q 's will now be obtained. For this purpose we use the fact, which is easily verified, that the operators $\partial/\partial q_r$ applied to ψ -symbols obey the same laws as when applied to ordinary functions. Thus

$$\frac{\partial^2\psi}{\partial q_r \partial q_s} = \frac{\partial^2\psi}{\partial q_s \partial q_r}$$

or

$$\pi_r \pi_s \psi = \pi_s \pi_r \psi,$$

and hence

$$\pi_r \pi_s - \pi_s \pi_r = 0. \quad (22)$$

Again

$$\frac{\partial}{\partial q_s} (q_r \psi) = q_r \frac{\partial \psi}{\partial q_s} + \delta_{rs} \psi$$

or

$$\pi_s q_r \psi = q_r \pi_s \psi + \delta_{rs} \psi,$$

and hence

$$q_r \pi_s - \pi_s q_r = -\delta_{rs}. \quad (23)$$

More generally, if f is any differentiable function of the q 's,

$$\frac{\partial}{\partial q_s} (f\psi) = f \frac{\partial \psi}{\partial q_s} + \frac{\partial f}{\partial q_s} \psi$$

or

$$\pi_s f\psi = f\pi_s \psi + \partial f/\partial q_s \cdot \psi,$$

and hence

$$f\pi_s - \pi_s f = -\partial f/\partial q_s. \quad (24)$$

These relations (22), (23), (24) could have been obtained alternatively directly from the representatives (20), with the help of properties of the δ function given in § 22.

The relations (22), (23) for the π 's are, apart from a numerical factor $-i\hbar$, just the same as the quantum conditions (12) for the p 's. Thus the observables $-i\hbar\pi_r$ satisfy the same commutability relations with each other and with the q 's as do the p_r . Equation (24) now corresponds to (16), with the difference that (24) has been shown to be valid for any differentiable function f , not merely for one expressible as a power series. There exist many sets of observables π_r , owing to the indefiniteness in $\partial/\partial q_r$ discussed above, but each such set must satisfy (22) and (23) and thus give rise to a set of observables $-i\hbar\pi_r$ with the same commutability properties as the p_r 's. Any one of these

sets, π_r , is connected with any other, π_{ra} , according to equation (19), by the relation

$$\pi_r = \pi_{ra} + i\partial F/\partial q_r. \quad (25)$$

It will now be shown that *there exists one set of π_r 's such that $-i\hbar\pi_r$ is just equal to p_r .*

If we take any set of π_r 's, π_{ra} say, then from (23) and (12), $p_s + i\hbar\pi_{sa}$ must commute with each q_r and must therefore be a function of the q 's only, *i.e.*

$$p_s + i\hbar\pi_{sa} = f_s(q). \quad (26)$$

Each f_s must be a real function of the q 's, since both p_s and $-i\hbar\pi_{sa}$ are real observables. Again, from (12) and (22), we obtain

$$\begin{aligned} 0 &= p_r p_s - p_s p_r \\ &= (-i\hbar\pi_{ra} + f_r) (-i\hbar\pi_{sa} + f_s) - (-i\hbar\pi_{sa} + f_s) (-i\hbar\pi_{ra} + f_r) \\ &= -i\hbar[\pi_{ra} f_s + f_r \pi_{sa} - \pi_{sa} f_r - f_s \pi_{ra}], \end{aligned}$$

or

$$\pi_{sa} f_r - f_r \pi_{sa} = \pi_{ra} f_s - f_s \pi_{ra}.$$

With the help of (24) we now find

$$\partial f_r / \partial q_s = \partial f_s / \partial q_r,$$

which shows that the functions f_r are all of the form

$$f_r = \partial G / \partial q_r,$$

where G is a function of the q 's independent of r . Thus (26) becomes

$$p_r = -i\hbar\pi_{ra} + \frac{\partial G}{\partial q_r} = -i\hbar\left(\pi_{ra} + \frac{i}{\hbar} \frac{\partial G}{\partial q_r}\right).$$

We can introduce a new set of π_r 's according to equation (25) taking F equal to G/\hbar , since F is an arbitrary real function of the q 's and G is real. For these new π_r 's we shall then have

$$p_r = -i\hbar\pi_r. \quad (27)$$

Equation (27), which was discovered by Schrödinger, is a very important one in applications of quantum mechanics. It is a consequence only of the quantum conditions (12) and may be regarded as a new form in which these quantum conditions may be expressed. It shows that we can take a representation in which the q 's are diagonal and in which each observable p_r , when multiplied into a ψ -symbol, is represented by the operator $-i\hbar\partial/\partial q'_r$ operating on the representative (q') of this ψ -symbol. When p_r is multiplied towards the left into a ϕ -symbol, it is then represented by the operator $i\hbar\partial/\partial q'_r$ operating on the representative ($|q'$) of this ϕ -symbol. If $f(q_s, p_r)$ is

any function of the q 's and p 's, expressible as a power series in the p 's, then it is equivalent to

$$f(q_s, -i\hbar\pi_r), \quad (28)$$

obtained from $f(q_s, p_r)$ by substituting $-i\hbar\pi_r$ for each p_r . This is to be understood as meaning that when f is multiplied into a ψ -symbol, its representative is the operator $f(q'_s, -i\hbar\partial/\partial q'_r)$ operating on the representative $(q'|)$ of this ψ -symbol, and when multiplied into a ϕ -symbol, its representative is the operator $\tilde{f}(q'_s, i\hbar\partial/\partial q'_r)$ operating on the representative $(|q')$ of this ϕ -symbol, where \tilde{f} is the function obtained from f by reversing the order of all the factors in each term. The equation for determining the eigenvalues f' of f is thus

$$f\left(q'_s, -i\hbar \frac{\partial}{\partial q'_r}\right)(q'|) = f' \cdot (q'|), \quad (29)$$

which is an ordinary partial differential equation for the unknown function $(q'|)$ and unknown number f' . When f is the Hamiltonian or energy of the system (assumed not to involve the time explicitly), this becomes *Schrödinger's equation for the determination of the possible numerical values for the energy*.

Equation (27) shows up the meaning of the indeterminacy in a representation when only the observables that are to be diagonal in it are specified. Corresponding to each representation in which the q 's are diagonal there exists one set of observables *conjugate* to the q 's [i.e. satisfying the same conditions as the p 's in (12)], whose representatives are of the specially simple form $-i\hbar\partial/\partial q'_r$ [when multiplied into a representative $(q'|)$ of a ψ -symbol]. If we now take one particular set of observables conjugate to the q 's and require that the representatives of these shall be of the specially simple form $-i\hbar\partial/\partial q'_r$, the representation is then *completely* determined, except for a trivial phase factor $e^{i\gamma}$, where γ is independent of the q 's, since the function F in (25) is completely determined by the condition that $-i\hbar\pi_r$ must equal p_r , except for an arbitrary constant. The indeterminacy in a representation when only the diagonal observables are specified is of the same nature, although it cannot be discussed in the same way, when any of these diagonal observables has no canonical conjugate, as is the case, for instance, when its eigenvalues do not extend from $-\infty$ to ∞ .

From equations (27) and (24) we see that (16) holds also for functions f of the q 's that are not expressible as power series.

§ 35. The Transformation Function ($q'|p'$)

The result (27) which we deduced with reference to the q -representatives of ψ - or ϕ -symbols must be applicable also to the transformation functions connecting two representations of which one is the q -representation, since these transformation functions are nothing but the representatives in either of the representations of the fundamental ψ 's and ϕ 's of the other. For instance the transformation function ($q'|p'_r|\alpha'$) is the q -representative of $\psi(\alpha')$. Hence from (27) the representative of $p_r\psi(\alpha')$ is $-i\hbar\partial(q'|\alpha')/\partial q'_r$. This representative, equal to $\int (q'|p_r|q'') dq''(q''|\alpha')$, may be written ($q'|p_r|\alpha'$) in the notation of mixed representations of § 27, so that we have

$$(q'|p_r|\alpha') = -i\hbar\partial(q'|\alpha')/\partial q'_r. \quad (30)$$

Similarly, if $f(q_s, p_r)$ is any function of the q 's and p 's expressible as a power series in the p 's, we see from the result (28) that

$$(q'|f|\alpha') = f\left(q'_s, -i\hbar\frac{\partial}{\partial q'_r}\right)(q'|\alpha'). \quad (31)$$

Again, the transformation function ($\alpha'|q'$) is the q -representative of $\phi(\alpha')$, so that, remembering (21), we obtain from the result (27)

$$(\alpha'|p_r|q') = i\hbar\partial(\alpha'|q')/\partial q'_r \quad (32)$$

and from the result (28)

$$(\alpha'|f|q') = \tilde{f}\left(q'_s, i\hbar\frac{\partial}{\partial q'_r}\right)(\alpha'|q'). \quad (33)$$

We shall now apply (30) to calculate the transformation function ($q'|p'$) connecting a co-ordinate q with its conjugate momentum p . We have

$$(q'|p|p') = -i\hbar\partial(q'|p')/\partial q'.$$

But from equations (22) of Chapter V

$$(q'|p|p') = (q'|p')p'.$$

Hence

$$-i\hbar\partial(q'|p')/\partial q' = p'(q'|p').$$

This is a differential equation for the unknown function ($q'|p'$) of q' . Its general solution is

$$(q'|p') = a'e^{iq'p'/\hbar},$$

where a' is an arbitrary function of p' .

We can determine the modulus of a' by using the normalizing condition

$$\int_{-\infty}^{\infty} (p'|q') dq' (q'|p'') = \delta(p' - p'').$$

This gives, when we put

$$(p'|q') = \overline{(q'|p')} = \bar{a}' e^{-iq'p'/\hbar},$$

the equation

$$\bar{a}' a'' \int_{-\infty}^{\infty} e^{-iq'(p'-p'')/\hbar} dq' = \delta(p' - p''),$$

where a'' is the value of a' when p'' is substituted for the p' in it. By carrying out the integration with respect to q' we obtain

$$\begin{aligned} \frac{1}{\bar{a}' a''} \delta(p' - p'') &= \frac{i\hbar}{p' - p''} \left[e^{-iq(p'-p'')/\hbar} \right]_{q=-\infty}^{q=\infty} \\ &= \frac{2\hbar}{p' - p''} \left[\sin q(p' - p'')/\hbar \right]_{q=-\infty}^{q=\infty}. \end{aligned}$$

Integrating each side with respect to p'' , we now get

$$\begin{aligned} \frac{1}{\bar{a}' a'} &= 2\hbar \left[\int_{-\infty}^{\infty} \frac{\sin q(p' - p'')/\hbar}{p' - p''} dp'' \right]_{q=-\infty}^{q=\infty} \\ &= 2\hbar[\pi]_{q=\infty}^{q=-\infty} = 2\pi\hbar = \hbar. \end{aligned}$$

Thus

$$a' = \hbar^{-\frac{1}{2}} e^{i\gamma'},$$

where γ' is some real function of p' , and hence

$$(q'|p') = \hbar^{-\frac{1}{2}} e^{i\gamma'} e^{iq'p'/\hbar}.$$

By suitably choosing the arbitrary phase in the p -representation we can remove the phase factor $e^{i\gamma'}$, which will leave us with

$$(q'|p') = \hbar^{-\frac{1}{2}} e^{iq'p'/\hbar}. \quad (34)$$

There is no arbitrary phase in the q -representation, since this phase is fixed when we use equation (27) or (30).

Our result (34) shows that the p -and q -representatives of a ψ -symbol are given in terms of one another by the relations

$$\left. \begin{aligned} (p'|) &= \hbar^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-iq'p'/\hbar} dq' (q'|) \\ (q'|) &= \hbar^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{iq'p'/\hbar} dp' (p'|). \end{aligned} \right\} \quad (35)$$

Thus either of them is given by the components in the Fourier resolution of the other. The transformation function connecting the n q 's, q_1, q_2, \dots, q_n , with their n conjugate p 's, p_1, p_2, \dots, p_n , is given by simple multiplication,

$$\begin{aligned} (q'_1 q'_2 \dots q'_n | p'_1 p'_2 \dots p'_n) &= (q'_1 | p'_1) (q'_2 | p'_2) \dots (q'_n | p'_n) \\ &= \hbar^{-n/2} e^{i(p'_1 q'_1 + p'_2 q'_2 + \dots + p'_n q'_n)/\hbar} \end{aligned} \quad (36)$$

§ 36. The Space-displacement Operator

In § 34 we saw how to give a meaning to the operator $\partial/\partial q_r$ applied to a ψ -symbol. For this purpose we had to make use of a representation in which q_r is diagonal. There are, however, certain cases in which one can give a meaning to this operator independently of any representation, so that this meaning becomes of more fundamental importance. These are the cases in which q_r is the value (x say) at a particular time of one of the Cartesian co-ordinates of the particle when the system consists of a single particle, or of the centre of gravity of the whole system in the general case. The operator $\partial/\partial x$ applied to a state is then connected with the operator of displacement of the state in the direction of the x -axis, as will now be shown.

Let ψ_1 denote any state of the system, arising when the system is prepared in a certain way. We now introduce that state ψ_2 which is the same as ψ_1 except for being displaced through a distance δx (a number) in the direction of the x -axis at the time t . To define ψ_2 rigorously, we must suppose all the apparatus used in the preparation of ψ_1 and all the external forces acting on the system up to time t to be displaced through this distance δx , the external forces after time t being unchanged. The state of the system after time t , which state is completely defined in this way, will then be ψ_2 . We can now form the difference $\psi_2 - \psi_1$ and divide by δx and proceed to the limit $\delta x \rightarrow 0$. The result of this procedure will be a ψ -symbol which depends in some linear way on our initial ψ -symbol ψ_1 . Thus we shall have

$$\lim_{\delta x \rightarrow 0} (\psi_2 - \psi_1)/\delta x = d_x \psi_1,$$

where d_x is a linear operator, i.e. where

$$d_x(\psi_1 + \psi_3) = d_x \psi_1 + d_x \psi_3.$$

for arbitrary ψ_1 and ψ_3 . Our displacement procedure thus enables us to define a *displacement operator* d_x , which, being a linear operator that can be multiplied into any ψ -symbol, can be regarded as an observable.

The displacement operator d_x is not completely defined owing to the fact that the ψ -symbol ψ_2 is undefined to the extent of an arbitrary numerical factor. If we make the assumption that ψ_2 shall have the same 'length' as ψ_1 , i.e. that

$$\phi_2 \psi_2 = \phi_1 \psi_1,$$

then this arbitrary factor will be of the form $e^{i\gamma}$, where γ is a real

number. Thus if ψ_2^* is any alternative ψ_2 , we shall have $\psi_2^* = e^{i\gamma}\psi_2$. Our new displacement operator d_x^* will now be given by

$$\begin{aligned} d_x^*\psi_1 &= \lim_{\delta x \rightarrow 0} (e^{i\gamma}\psi_2 - \psi_1)/\delta x \\ &= \lim_{\delta x \rightarrow 0} \left\{ \frac{\psi_2 - \psi_1}{\delta x} + \frac{e^{i\gamma} - 1}{\delta x} \psi_2 \right\} \\ &= d_x\psi_1 + ia\psi_1, \end{aligned}$$

where a is a real number, equal to the limit (assumed to exist) of $\gamma/\delta x$. Thus

$$d_x^* = d_x + ia,$$

so that the indefiniteness in our displacement operator consists of merely an additive arbitrary pure imaginary number.

The series of operations by which, given any ψ -symbol ψ , we defined the ψ -symbol $d_x\psi$ may be applied also to any ϕ -symbol ϕ and will then give us the ϕ -symbol $d_x\phi$. When d_x is regarded as an observable it can be multiplied into a ϕ -symbol to give a product ϕd_x . The connexion between $d_x\phi$ and ϕd_x will now be obtained. Any product of the form $\phi\psi$ is a number which must remain unchanged when both the ϕ and ψ are displaced through the distance δx , and hence

$$d_x(\phi\psi) = 0.$$

Since d_x is of the nature of a differentiation, we can use the ordinary law for the differential coefficient of a product, which gives us

$$(d_x\phi)\psi + \phi(d_x\psi) = 0.$$

When we consider d_x as an observable, we have

$$\phi(d_x\psi) = (\phi d_x)\psi.$$

Hence

$$(d_x\phi)\psi = -(\phi d_x)\psi.$$

Since this is true for arbitrary ψ , we obtain

$$\phi d_x = -d_x\phi,$$

which is the required connexion. This result is analogous to (21). It shows us that the conjugate imaginary symbol to $d_x\psi_r$, which is, of course, just $d_x\phi_r$, is equal to $-\phi_r d_x$, and hence allows us to infer that d_x is a pure imaginary observable, like the π_r of (21).

We shall now obtain the connexion between our new operator d_x and the operator $\partial/\partial x$ defined according to § 34. We take a representation in which x is diagonal. We suppose further that the phase of this representation is independent of x , so that when a ψ -symbol is displaced in the direction of the x -axis, its representative $(x' |)$ is merely displaced an equal distance through the domain of the vari-

able x' . (If the phase were arbitrary, then when the ψ -symbol is displaced its representative would be changed in some more complicated way.) The representatives $(x'|1)$ and $(x'|2)$ of ψ_1 and ψ_2 are now connected by the relation

$$(x'|2) = (x' - \delta x|1).$$

Thus the representative of $d_x \psi_1$ will be

$$\lim_{\delta x \rightarrow 0} \frac{(x' - \delta x|1) - (x'|1)}{\delta x} = -\frac{\partial}{\partial x}(x'|1)$$

and hence

$$d_x = -\frac{\partial}{\partial x}. \quad (37)$$

Equation (37) holds, of course, only for one of the possible operators $\partial/\partial x$. The others will differ from this one in accordance with equation (18). It will now be shown that the one for which (37) holds is the same as the one which, considered as an observable π_x , satisfies (27) or

$$p_x = -i\hbar\pi_x,$$

p_x being the momentum conjugate to x . This will mean that, with d_x considered as an observable,

$$p_x = i\hbar d_x. \quad (38)$$

We prove this by observing that p_x and $i\hbar d_x$ satisfy the same commutability relations. When the ψ -symbol $x\psi$ is displaced through the distance δx the result must be $(x - \delta x)\psi_2$, in which x has been changed into $x - \delta x$, since the displacement of apparatus required for the definition of the displaced ψ -symbol causes apparatus that measures the observable x to become apparatus that measures $x - \delta x$. Thus from the definition of d_x

$$\begin{aligned} d_x x\psi_1 &= \lim_{\delta x \rightarrow 0} \{(x - \delta x)\psi_2 - x\psi_1\}/\delta x \\ &= xd_x\psi_1 - \psi_1. \end{aligned}$$

Hence

$$d_x x - xd_x = -1. \quad (39)$$

In the same way it may be shown that d_x commutes with y , z , p_x , p_y , p_z , and in fact with every dynamical variable (at time t) independent of x . Thus $p_x - i\hbar d_x$ commutes with everything and must be a number. We may take this number to be zero, on account of the arbitrary additive number arising in the definition of d_x , and thus obtain (38).

Equation (38), which connects our displacement operator d_x with

the momentum p_x , is an alternative way of expressing the quantum conditions (12) or (27), in so far as they refer to the centre of gravity of the whole system, and is perhaps the most fundamental of all ways of expressing them, showing most clearly the underlying physical assumption. This equation (38) is quite a plausible assumption for one to make for one's quantum conditions, apart from the fact that it is derivable from equations (12), which were set up from analogy with the classical theory, on account of its simplicity and generality and the fact that it leads at once to the law of the conservation of momentum. When there are no external forces acting on the system, we see from the definition of d_x that it does not depend on the time t . Equation (38) then shows that the momentum does not depend on t and is therefore constant.

§ 37. The Time-displacement Operator.

Corresponding to the space-displacement operator d_x of the preceding section, we now introduce an analogous time-displacement operator d_t , defined as follows. If ψ_1 is any ψ -symbol, we form the time-displaced ψ -symbol ψ_2 by supposing all the apparatus used in preparing ψ_1 to be set in motion a time δt later and all varying external forces acting on the system up to time t to be retarded a time δt . The state of the system after this time t will then be our ψ_2 . We now take the limit of $(\psi_2 - \psi_1)/\delta t$ and define it to be $d_t \psi_1$. We can consider d_t to be an observable and, as in the case of d_x , can show that it is a pure imaginary observable and that it is completely defined except for an arbitrary, pure imaginary, additive numerical constant.

By means of this d_t we shall deduce the equations of motion of the system. In this way we shall establish the form of these equations without anywhere making use of classical analogues. We introduce the real observable H defined by

$$-i\hbar d_t = H. \quad (40)$$

Thus

$$H\psi = -i\hbar d_t \psi \quad (41)$$

for arbitrary ψ . If we now take any observable ξ that is the value at time t of some dynamical variable and apply (41) to the ψ -symbol $\xi\psi$, we obtain

$$H\xi\psi = -i\hbar d_t \xi\psi.$$

We can evaluate the right-hand side here by the method used for deriving (39), or more directly by making use of the fact that the

ordinary law for the differentiation of a product applies to the operator d_t , so that

$$H\xi\psi = -i\hbar(d_t\xi)\psi - i\hbar\xi(d_t\psi).$$

It is now easily seen that $d_t\xi$ is just the ordinary time differential coefficient $\dot{\xi}$. (This is to be contrasted with the corresponding result for the d_x operator, namely, $d_x\xi = -\partial\xi/\partial x$.) We thus obtain

$$H\xi\psi = -i\hbar\dot{\xi}\psi + \xi H\psi,$$

which gives

$$i\hbar\dot{\xi} = \xi H - H\xi.$$

This is of the same form as (13), with for Hamiltonian just the H defined in terms of the time-displacement operator d_t by (40).

The above argument is quite general and shows that *the equations of motion for any dynamical system are expressible in terms of a Hamiltonian* in the form (13), whether this system is one that has an analogue in the classical theory and is describable in terms of canonical co-ordinates and momenta or not. The general dynamical system in quantum mechanics is thus one in which the dynamical variables satisfy arbitrary commutability relations, and there is a Hamiltonian which is an arbitrary real function of them. More generally still, we may have a system in which the Hamiltonian cannot be expressed as an analytic function of dynamical variables and can be specified only through its representative in some representation, which representative may be an arbitrary Hermitian matrix. An example of a system of this more general kind is provided by the problem, considered in Chapter XII, of the interaction of a photon with an atom.

Corresponding to equation (37) we can prove the result

$$d_t = -\frac{\partial}{\partial t} \quad (42)$$

We must first give a meaning to the operator $\partial/\partial t$ applied to a ψ -symbol, which we can do with the help of a representation in which a complete set of commuting observables $q_{1t}, q_{2t}, \dots, q_{nt}$ are diagonal, which observables must be the values at time t of a set of dynamical variables q_1, q_2, \dots, q_n (which need not necessarily have conjugate momenta p_1, p_2, \dots, p_n). The representative of any ψ -symbol ψ will now be a function of the n variables $q'_{1t}, q'_{2t}, \dots, q'_{nt}$, the form of this function depending in general on t . Thus we can regard this representative as a function of the $n+1$ variables $q'_1, q'_2, \dots, q'_n, t$, and as such can differentiate it partially with respect to t and define the resulting function to be the representative of $\partial\psi/\partial t$. We get in this

way a general definition of the operator $\partial/\partial t$ in which there is, of course, a considerable amount of indefiniteness, owing not only to the arbitrary phases of the representation but also to the fact that we can take different sets of q 's to be diagonal and will then in general get different results. We are interested, however, in only one of the operators $\partial/\partial t$, this being the one that is given when the phases of the representation do not depend explicitly on t , so that when a time displacement δt is applied to a state, the $q_{t+\delta t}$ -representative of the displaced state is the same function of its variables $q'_{t+\delta t}$ that the q_t -representative of the undisplaced state is of its variables q'_t . Thus to obtain the q_t -representative of the displaced state we must substitute $t-\delta t$ for t in the q_t -representative of the undisplaced state, considered as a function of the $n+1$ variables $q'_1, q'_2, \dots, q'_n, t$. There is now complete analogy with the x -displacement case, so that (42) follows in the same way as (37). The validity of (42) shows that the operator $\partial/\partial t$ defined by a representation with phases not explicitly dependent on t is independent of which set of q 's are diagonal in the representation. If we have one representation giving a $\partial/\partial t$ operator that satisfies (42), we can obtain another by making any canonical transformation for which the transformation function does not involve t .

From (41) and (42) we obtain

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi. \quad (43)$$

This may be regarded as an alternative way of expressing the equations of motion of the system. Expressed in terms of representatives, it gives us *

$$i\hbar \frac{\partial}{\partial t} (q'_t) = \int (q'_t | H | q''_t) dq''_t (q''_t), \quad (44)$$

an equation which shows how the representative (q'_t) of a state, considered as a function of the $n+1$ variables $q'_1, q'_2, \dots, q'_n, t$, varies with t . When the q_r have conjugate momenta p_r , it reduces to the ordinary differential equation

$$i\hbar \frac{\partial}{\partial t} (q'_t) = H \left(q'_t, -i\hbar \frac{\partial}{\partial q'_t} \right) (q'_t). \quad (45)$$

This equation was discovered by Schrödinger and is known as *Schrödinger's wave equation*. It is very useful in applications of

* The case of continuous q' 's is taken for definiteness, the usual modifications in the notation being required for the discrete case.

quantum mechanics since its solutions have an immediate physical interpretation, the square of the modulus of any solution giving the probability of the q 's having specified values for one particular state throughout all time. It is called a wave equation because in many elementary examples, as will be seen in the next chapter, its solutions are of the form of waves moving through q -space. For this same reason the solutions are called *wave functions*, even also in those examples where they have no resemblance to waves.

When the Hamiltonian does not involve the time explicitly, the wave equation in the form (45) or in the more general form (44) will have solutions that vary periodically with the time, according to the law

$$(q'|) = (q'|)_0 e^{-iW't/\hbar}, \quad (46)$$

where W' is a number and $(q'|)_0$ is independent of t . The equation that $(q'|)_0$ must satisfy is

$$\begin{aligned} W'(q'|)_0 &= \int (q'| H | q'') dq'' (q''|)_0 \\ &= H\left(q', -i\hbar \frac{\partial}{\partial q'}\right) (q'|)_0. \end{aligned}$$

But this is just the equation for determining the eigenvalues of H , namely, equation (29) with H for f . Thus W' is an eigenvalue of H or energy-level of the system and $(q'|)_0$ is an eigenfunction of H .

§ 38. Heisenberg's Matrices

In the preceding section we dealt with a q_t -representation, defined by observables q_t that are the values at time t of a set of dynamical variables q . We saw that if the phases of the representation are suitably chosen, then Schrödinger's equation holds, in the form (44) or (45), in which case the representation may conveniently be called a Schrödinger representation. The condition for the phases is such that, when a state is given a time-displacement δt , the $q_{t+\delta t}$ -representative of the displaced state is the same function of its variables $q'_{t+\delta t}$ as the q_t -representative of the undisplaced state is of its variables q'_t . This condition will hold in an analogous form for observables. If we take an observable ξ_t which is the value at time t of a dynamical variable ξ , then the displaced observable will be $\xi_{t+\delta t}$. We shall then have that the $q_{t+\delta t}$ -representative of the displaced observable, namely $(q'_{t+\delta t}|\xi_{t+\delta t}|q''_{t+\delta t})$, is the same function of its variables $q'_{t+\delta t}$, $q''_{t+\delta t}$ as the q_t -representative of the undisplaced observable, namely $(q'_t|\xi_t|q'')$, is of its variables q'_t , q''_t . This means simply that the form of the

function $(q'_t|\xi_t|q''_t)$ of the variables q'_t, q''_t is independent of t . More concisely, one can say that *the Schrödinger representative of ξ_t is independent of t* .

In general, when one wants a representation of observables, the Schrödinger one would not be a convenient one to take, since it refers to a definite time t and gives simple representatives only for those observables ξ_t referring to the same time t . A convenient representation would now be one which makes no reference to any time t , so that observables $\xi_{t_1}, \eta_{t_2} \dots$ referring to different times $t_1, t_2 \dots$ could all be represented simultaneously and would all be on the same footing. For such a representation we should have

$$\left(\alpha' \left| \frac{d\xi_t}{dt} \right| \alpha'' \right) = \frac{d}{dt} \left(\alpha' |\xi_t| \alpha'' \right). \quad (47)$$

Such a representation can easily be obtained when the Hamiltonian does not involve the time explicitly. In the general case it is not so easy and is therefore then not very useful.

When H does not involve the time explicitly we can take for the observables α that are diagonal in our representation a complete set of commuting dynamical variables that are constants of the motion. Then H will commute with the α 's and will be a function of them, represented by a diagonal matrix

$$(\alpha' | H | \alpha'') = H' \delta_{\alpha' \alpha''},$$

H' being written for $H(\alpha')$, for brevity. Our representation will now be one that is independent of t (provided the phases are independent of t), so that equation (47) holds. There is now a simple law for the variation of the matrix elements of ξ_t with t . From the equation of motion (13) we obtain

$$i\hbar (\alpha' | \dot{\xi} | \alpha'') = (\alpha' | \xi | \alpha'') H'' - H' (\alpha' | \xi | \alpha''),$$

which, with the help of (47), becomes

$$i\hbar \frac{d}{dt} (\alpha' | \xi | \alpha'') = -(H' - H'') (\alpha' | \xi | \alpha'').$$

Hence $(\alpha' | \xi | \alpha'')$ varies with t according to the law

$$(\alpha' | \xi | \alpha'') = (\alpha' | \xi | \alpha'')_0 e^{i(H' - H'')t/\hbar}, \quad (48)$$

$(\alpha' | \xi | \alpha'')_0$ being independent of t . The variation is thus periodic with the frequency

$$|H' - H''|/2\pi\hbar = |H' - H''|/h. \quad (49)$$

This scheme of matrices, in which the Hamiltonian is diagonal and the matrix elements all vary with the time according to the law (48),

was discovered by Heisenberg in 1925 and was historically the first form of quantum mechanics.

A diagonal element $(\alpha'|\xi|\alpha')$ does not vary with the time. This diagonal element is the average value of ξ for a fundamental state $\psi(\alpha')$ of the representation. Thus for each fundamental state $\psi(\alpha')$ the average value of any dynamical variable ξ is a constant. The probability of ξ having any specified value is therefore also constant, since this probability is determined by the average value of functions of ξ . Thus each $\psi(\alpha')$ is a stationary state according to the definition of § 3. *The fundamental states of a Heisenberg representation are stationary states.* Any eigenstate of H may be taken as a fundamental state of a Heisenberg representation and is therefore a stationary state.

The matrices of Heisenberg's representation fit in very well with the 'anschaulich' forms of quantum theory in existence before quantum mechanics, in particular with Bohr's theory of the atom. The fundamental states of the representation are Bohr's stationary states (which are really stationary, of course, only so long as one neglects the interaction of the atom with radiation) and the eigenvalues of H are Bohr's energy-levels. It follows that the frequency (49) of matrix elements referring to two states α' and α'' is that of the quantum of radiation emitted or absorbed according to Bohr's theory when the atom makes a jump from one of these states to the other, as was assumed by Heisenberg in his first work on quantum mechanics. There now arises a strong correspondence between the matrix elements representing any dynamical variable and the Fourier components of that variable in the classical theory for a multiply-periodic system. This correspondence led Heisenberg to the assumption that the rate of spontaneous emission of radiation of a system in the quantum theory can be obtained from the classical formula if one substitutes in this formula for the Fourier components of the total electric displacement of the system the corresponding matrix elements. According to this assumption, a system having an electric moment \mathbf{D} (a vector) will, when in the state α' , emit radiation of frequency $\nu = (H' - H'')/h$, where $H'' = H(\alpha'')$ is an energy-level, less than H' , of some state α'' , at the rate

$$\frac{4}{3} \frac{(2\pi\nu)^4}{c^3} |(\alpha'|\mathbf{D}|\alpha'')|^2. \quad (50)$$

Also the distribution of this radiation over the different directions

of emission and its state of polarization for each direction will be the same as that for a classical electric dipole of moment

$$(\alpha'|\mathbf{D}|\alpha'') + (\alpha''|\mathbf{D}|\alpha').$$

To interpret this rate of emission of radiant energy according to Bohr's theory, we must divide it by the quantum of energy of this frequency, namely $\hbar\nu$, and call it the probability per unit time of this quantum being spontaneously emitted, with the atomic system simultaneously dropping to the state α'' of lower energy. A justification for these assumptions of Heisenberg will be obtained in Chapter XII, where a quantum treatment of the interaction of an atomic system with radiation will be given.

By altering the phases in a Heisenberg representation we can pass to the Schrödinger representation in which the same α 's are diagonal. Let us see what is the connexion between the phases in the two cases. In the Schrödinger representation the representative of any state will satisfy the wave equation

$$i\hbar \frac{\partial}{\partial t}(\alpha'|) = \sum_{\alpha''} (\alpha'|H|\alpha'') (\alpha''|) = H'(\alpha'|),$$

which can in this case be integrated directly and gives

$$(\alpha'|) = (\alpha'|)_0 e^{-iH't/\hbar},$$

where $(\alpha'|)_0$ is independent of t . On the other hand, the representative of a state in the Heisenberg representation will not depend on t , since the representation and also, of course, the state do not in any way depend on t . Hence the phases of the Schrödinger representation are $e^{-iH't/\hbar}$ relative to those of the Heisenberg representation, a result which could have been obtained alternatively from a comparison of (48) with the fact that the Schrödinger representative of ξ_t is independent of t . There is thus a difference between the phases of the Heisenberg representation, which are *totally* independent of t , and those of the Schrödinger representation, which are *explicitly* independent of t . The explicit independence of t for the Schrödinger representation means simply that any matrix in this representation represents a function of the dynamical variables that does not involve t explicitly.

VII

ELEMENTARY APPLICATIONS

§ 39. The Free Particle

In this chapter we shall consider some simple dynamical systems according to quantum mechanics. The simplest of all systems is that of a particle in free space. For this system we may take as dynamical variables the three Cartesian co-ordinates x, y, z and their conjugate momenta p_x, p_y, p_z . The Hamiltonian in classical mechanics, when one takes into account the variation of the mass of the particle with its velocity required by the principle of relativity, is

$$H = c(m^2c^2 + p_x^2 + p_y^2 + p_z^2)^{\frac{1}{2}}, \quad (1)$$

where m is the rest-mass of the particle and c is the velocity of light, and the positive square root is taken. This Hamiltonian may be taken over into the quantum theory when one gives the meaning of § 16 to the positive square root, which one can do since the eigenvalues of $m^2c^2 + p_x^2 + p_y^2 + p_z^2$ are all positive.

The momenta p_x, p_y, p_z commute with H and are thus constants of the motion, as in the classical theory. Again, the co-ordinates x, y, z vary according to the equations

$$\dot{x} = [x, H] = \frac{c^2 p_x}{H} \quad \dot{y} = \frac{c^2 p_y}{H} \quad \dot{z} = \frac{c^2 p_z}{H}, \quad (2)$$

the same as in the classical theory. These equations may be verified in the quantum theory by an application of equation (16) of § 34, which equation, as remarked at the end of that section, holds also for functions that are not expressible as power series. The general proof of this equation, however, required the use of a representation. It is of interest to notice that we can deduce (2) by working in abstract symbols and not making any use of representations, in the following way. We have by a direct application of the quantum conditions

$$xH^2 - H^2x = c^2(xp_x^2 - p_x^2x) = 2ihc^2p_x \quad (3)$$

$$\text{or} \quad (xH - Hx)H + H(xH - Hx) = 2ihc^2p_x. \quad (4)$$

Now H commutes with p_x and hence from (3)

$$(xH^2 - H^2x)H - H(xH^2 - H^2x) = 0,$$

$$\text{which gives} \quad (xH - Hx)H^2 - H^2(xH - Hx) = 0.$$

We must now use the condition that $(m^2c^2 + p_x^2 + p_y^2 + p_z^2)^{\frac{1}{2}}$, being

defined as a square-root function, commutes with everything that commutes with $m^2c^2 + p_x^2 + p_y^2 + p_z^2$, i.e. H commutes with everything that commutes with H^2 . We have just seen that H^2 commutes with $xH - Hx$ and hence H must commute with $xH - Hx$. We can now infer from (4) that

$$xH - Hx = i\hbar c^2 p_x / H,$$

which gives the first of the equations (2). We thus have an illustration of the fact that any result that may be obtained with the help of a representation can also be obtained from the abstract symbols alone without reference to representations, but that the method with a representation may be much quicker and more convenient.

The Schrödinger equation for the Hamiltonian (1) is *

$$i\hbar \frac{\partial}{\partial t} (x|) = c \left\{ m^2 c^2 - \hbar^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right\}^{\frac{1}{2}} (x|), \quad (5)$$

where the x in $(x|)$ stands for x , y , and z . We have here on the right-hand side the square root of an operator involving $\partial/\partial x \dots$ which square root cannot be expressed as a power series that is valid for the whole range of eigenvalues of p_x , p_y , p_z , namely $-\infty$ to ∞ . To give a meaning to such a function of an operator we should in general have to make a canonical transformation to a representation in which the observable corresponding to this operator is diagonal, when the meaning would be as given in § 15. Our present example is, however, sufficiently simple for this not to be necessary. We can write down solutions of (5) immediately, namely

$$(x|) = a \exp i(p'_x x + p'_y y + p'_z z - W't)/\hbar, \quad (6)$$

where p'_x , p'_y , p'_z , W' are numbers satisfying

$$W'^2 = c^2(m^2c^2 + p'^2_x + p'^2_y + p'^2_z) \quad W' > 0$$

and a is an arbitrary number. The general solution of (5) can be expressed as a sum or integral of solutions of the form (6).

The state represented by (6) is an eigenstate for the components of momentum, belonging to the eigenvalues p'_x , p'_y , p'_z . The corresponding value for the energy is W' . The representative (6) is, in fact, of the same form as the transformation function (36) of § 35. Thus the state of a particle moving in free space with a given momentum is represented by plane waves of the type (6), the direc-

* The primes are omitted from the variables in the wave function. This is permissible when it does not lead to confusion.

tion of motion of the waves being determined by (p'_x, p'_y, p'_z) , the momentum of the particle. The probability of the particle being found in any specified volume $dxdydz$ at time t is proportional to $|(\chi)|^2dxdydz$ and is thus independent of the position of this volume. The wave-length λ of the waves is given by

$$\lambda = \hbar / (p'^2_x + p'^2_y + p'^2_z)^{1/2} = \hbar / P', \quad (7)$$

where P' is the magnitude of the momentum of the particle, and their frequency ν is given by

$$\nu = W'/\hbar. \quad (8)$$

Thus their velocity u is

$$u = \lambda\nu = W'/P' = c^2/v, \quad (9)$$

where v is the velocity of the particle.

The fact that the velocity of the waves and the velocity of the particle both lie in the same direction and are connected by the relation (9) holds, of course, in all Lorentz frames of reference. It was this relativity invariance which first led de Broglie, before the discovery of quantum mechanics, to postulate the existence of waves of the type (6) associated with the motion of a particle, which waves would control the particle in the same way in which light-waves control photons. The case of the photon may be obtained from that of the free particle by taking the rest-mass m equal to zero. The waves (6) then become just the light-waves associated with the photon, apart from polarization considerations and the fact that they involve an imaginary exponential instead of a sine or cosine.

§ 40. Wave Packets

By superposing a number of solutions of the type (6) belonging to different values of the momentum \mathbf{p}' lying in the neighbourhood of a given value, one can obtain a solution that, at every instant of time, vanishes (approximately) everywhere outside a certain finite region. Within this region the waves are approximately of a single wavelength, corresponding to the given value of \mathbf{p}' . This solution thus forms a group of waves or *wave packet*. The velocity V of such a wave packet is not equal to the velocity of the waves, but lies in the same direction and is given by the hydrodynamical formula for group velocity

$$V = \frac{d\nu}{d(1/\lambda)}.$$

With the help of (7) and (8), this becomes

$$V = \frac{dW'}{dP'} = c \frac{d}{dP'} (m^2 c^2 + P'^2)^{\frac{1}{2}} = \frac{c^2 P'}{W'} = v.$$

Thus *the group velocity is the same as the velocity of the particle.*

This important result was first obtained by de Broglie. It is capable of wide generalizations. If we have any dynamical system describable by a Hamiltonian $H(q, p)$, which is an arbitrary function of canonical q 's and p 's, then, if it is permissible to treat Planck's constant \hbar as small so that terms involving it as a factor may be neglected, *the Schrödinger equation will admit of solutions consisting of wave packets whose motions are along the trajectories of the classical theory.* The proof is as follows. The Schrödinger equation is

$$i\hbar \frac{\partial}{\partial t} (q|) = H\left(q, -i\hbar \frac{\partial}{\partial q}\right)(q|). \quad (10)$$

We express the Schrödinger function $(q|)$ as though it were of the form of waves, thus

$$(q|) = e^{is/\hbar} A,$$

where A and S are real functions of the q 's, which give the amplitude and phase respectively. The effect of the operator $-i\hbar\partial/\partial q_r$ on $(q|)$ is now

$$-i\hbar \frac{\partial}{\partial q_r} (q|) = e^{is/\hbar} \left(\frac{\partial S}{\partial q_r} - i\hbar \frac{\partial}{\partial q_r} \right) A \quad (11)$$

and that of the operator $i\hbar\partial/\partial t$ is

$$i\hbar \frac{\partial}{\partial t} (q|) = e^{is/\hbar} \left(-\frac{\partial S}{\partial t} + i\hbar \frac{\partial}{\partial t} \right) A.$$

If f is any function of the operators $-i\hbar\partial/\partial q_r$ expressible as a power series, one finds readily by repeated applications of (11)

$$f\left(-i\hbar \frac{\partial}{\partial q_r}\right)(q|) = e^{is/\hbar} f\left(\frac{\partial S}{\partial q_r} - i\hbar \frac{\partial}{\partial q_r}\right) A.$$

Thus (10) becomes, after removal of the factor $e^{is/\hbar}$,

$$\left(-\frac{\partial S}{\partial t} + i\hbar \frac{\partial}{\partial t}\right) A = H\left(q, \frac{\partial S}{\partial q} - i\hbar \frac{\partial}{\partial q}\right) A. \quad (12)$$

The right-hand side, considered as a function of the $(\partial S/\partial q - i\hbar\partial/\partial q)$'s may be expanded by Taylor's theorem as a power series in \hbar , which we are supposing to be a small number. The terms in this expansion are alternately real and pure imaginary. If we neglect all except the

first two and equate these to the real and pure imaginary parts of the left-hand side of (12), we obtain

$$-\frac{\partial S}{\partial t} = H\left(q, \frac{\partial S}{\partial q}\right) \quad (13)$$

and

$$-\frac{\partial A}{\partial t} = \sum_r \frac{\partial H(q, \partial S / \partial q)}{\partial (\partial S / \partial q_r)} \frac{\partial A}{\partial q_r}. \quad (14)$$

Equation (13) is just the Hamilton-Jacobi equation of classical mechanics. Thus *the phase of the Schrödinger wave function is given by the principal function S of the Hamilton-Jacobi theory when one counts \hbar as small*. Equation (14) is the one that governs the amplitude A of the wave function. It shows that for any solution S of (13) the amplitude remains constant along the trajectories given by

$$\frac{dq_r}{dt} = \frac{\partial H(q, \partial S / \partial q)}{\partial (\partial S / \partial q_r)} \quad (15)$$

but is otherwise arbitrary. Thus we can take A to vanish everywhere except on a certain group of neighbouring trajectories, along each of which it must have a constant value. We obtain in this way a solution of the wave equation that at any time vanishes everywhere outside a certain small region. There is a limit to how small this region may be, imposed by the approximations we have made. Our neglect of later terms in the Taylor expansion of the right-hand side of (12) is justified only provided

$$\hbar \frac{\partial}{\partial q} A \ll \frac{\partial S}{\partial q} A.$$

This requires that A shall vary by an appreciable fraction of itself only through a range of q in which S varies by many times \hbar , i.e. a range of q consisting of many wave-lengths of the wave function. Thus our solution of the wave equation that vanishes everywhere outside a certain small region is of the nature of a wave packet. The motion of this wave packet is given by the trajectories (15), which are, when one remembers that $\partial S / \partial q_r$ is playing the part of p_r , just the trajectories of classical mechanics.

For the system consisting of a free particle, a wave packet represents a state for which both the position and the momentum have definite numerical values to a certain limited degree of accuracy. Such a state is of the kind that usually occurs in practice, particularly if the particle has a large mass, since one usually knows roughly both the position and the momentum of a particle with which one is

dealing. If Δx is the order of magnitude of the size of the wave packet, then, when one resolves the packet into its Fourier components, the wave-lengths of the different components will be distributed over a range of order

$$\Delta\lambda = \lambda^2/\Delta x.$$

From (7) this corresponds to a distribution of the momentum of the particle over a range of order

$$\Delta p = \hbar/\lambda^2. \Delta\lambda = \hbar/\Delta x.$$

Thus we have

$$\Delta p \Delta x = \hbar, \quad (16)$$

which shows there is a theoretical limit to the accuracy with which both the position and momentum may have definite numerical values together. The relation (16) is known as *Heisenberg's principle of indeterminacy*. It shows how, the more accurately the position of a particle is known, the greater the indeterminacy in its momentum and *vice versa*. One would expect a principle of this type to hold simply from the quantum condition

$$xp - px = i\hbar.$$

It should be understood that (16) holds only in the most favourable case and that the indeterminacies may be much greater than is implied by this equation. In fact if one takes a wave packet for which (16) holds at one instant of time, in course of time this packet will spread and $\Delta p \Delta x$ will increase. For a discussion of this spreading and for a treatment of the motion of wave packets representing particles in fields of force, the reader is referred to papers by Kennard and Darwin.*

Heisenberg's principle of indeterminacy applies also to general dynamical systems describable by means of canonical q 's and p 's. We have seen that such systems have states represented by wave packets moving in q -space. Any such state is one for which both the q 's and the p 's have numerical values to a certain degree of accuracy, the orders of magnitude of the minimum indeterminacy Δq_r in a co-ordinate q_r and Δp_r in the conjugate momentum p_r , being connected by

$$\Delta p_r \Delta q_r = \hbar. \quad (17)$$

This general relation may be deduced in the same way as (16) from

* Kennard, *Zeits. f. Physik*, vol. xliv, p. 344; Darwin, *Roy. Soc. Proc. A*, vol. exvii, p. 258.

the connexion between the size of a wave packet and the indeterminacy in the wave-length of its waves, or it may be inferred directly from the quantum condition

$$q_r p_r - p_r q_r = i\hbar.$$

The states dealt with in classical mechanics, of a system composed of massive particles or bodies, are represented by these wave packets and (17) gives the limit of accuracy of the classical treatment.

§ 41. The Harmonic Oscillator in One Dimension

We shall now consider the problem of the harmonic oscillator in one dimension. The Hamiltonian for this system in classical mechanics is

$$H = 1/2m \cdot (p^2 + m^2\omega^2q^2) \quad (18)$$

where m is the mass of the oscillating particle and ω is another numerical constant, equal to 2π times the frequency. This Hamiltonian can be taken over into the quantum theory and must then be supplemented by the quantum condition

$$qp - pq = i\hbar \quad (19)$$

to give a completely determinate problem.

The equations of motion are easily verified to be the same as in the classical theory. We must now determine the eigenvalues of the Hamiltonian H . This question is the same as that dealt with in § 29, there being a difference only in the numerical constants, on account of the \hbar in (19) and the $2m$ and $m^2\omega^2$ in (18). The present q is $(\hbar/m\omega)^{\frac{1}{2}}$ times the q of § 29 and the present p is $(\hbar m\omega)^{\frac{1}{2}}$ times the p of § 29, which results in the present H being $\frac{1}{2}\hbar\omega$ times the $(p^2 + q^2)$ of § 29. Thus from the result that the $(p^2 + q^2)$ of § 29 has the eigenvalues $1, 3, 5 \dots$, we can infer that the present H has the eigenvalues

$$\frac{1}{2}\hbar\omega, \frac{3}{2}\hbar\omega, \frac{5}{2}\hbar\omega \dots$$

These are the possible values for the energy of a harmonic oscillator in the quantum theory.

We shall now obtain the Heisenberg matrices representing p and q . These can be obtained readily from equations (34) of § 29. Allowing for the change in the numerical constants and remembering that the A of § 29 is equal to $(2H/\hbar\omega - 1)$, we find

$$\left. \begin{aligned} (H'|p|H' - \hbar\omega) &= (\frac{1}{2}m)^{\frac{1}{2}}(H' - \frac{1}{2}\hbar\omega)^{\frac{1}{2}}e^{i(\omega t + \gamma)} \\ (H' - \hbar\omega|p|H') &= (\frac{1}{2}m)^{\frac{1}{2}}(H' - \frac{1}{2}\hbar\omega)^{\frac{1}{2}}e^{-i(\omega t + \gamma)} \\ (H'|q|H' - \hbar\omega) &= -i/(2m)^{\frac{1}{2}}\omega \cdot (H' - \frac{1}{2}\hbar\omega)^{\frac{1}{2}}e^{i(\omega t + \gamma)} \\ (H' - \hbar\omega|q|H') &= i/(2m)^{\frac{1}{2}}\omega \cdot (H' - \frac{1}{2}\hbar\omega)^{\frac{1}{2}}e^{-i(\omega t + \gamma)} \end{aligned} \right\} \quad (20)$$

when the correct time-factors are included. In the classical theory we have, when we express p and q as Fourier series,

$$\begin{aligned} p &= (2mH)^{\frac{1}{2}} \cos(\omega t + \gamma) = (\frac{1}{2}mH)^{\frac{1}{2}} \{e^{i(\omega t + \gamma)} + e^{-i(\omega t + \gamma)}\} \\ q &= (2H/m)^{\frac{1}{2}} \omega^{-1} \sin(\omega t + \gamma) = (H/2m)^{\frac{1}{2}} \omega^{-1} \{-ie^{i(\omega t + \gamma)} + ie^{-i(\omega t + \gamma)}\}. \end{aligned}$$

This shows up the correspondence between the Fourier components of the classical theory and the Heisenberg matrix elements. The classical Fourier components are, of course, equal to these matrix elements when one neglects \hbar .

If the oscillator carries an electric charge ϵ , its electric moment will be ϵq . According to Heisenberg's assumption, given in § 38, for the spontaneous emission of radiation, the oscillator will then emit only radiation of frequency $\omega/2\pi$ since all the matrix elements of q vanish except those mentioned in (20). This result is the same as in the classical theory. When the oscillator is in a state of energy $H' = (n + \frac{1}{2})\hbar\omega$, or, as we may say, when it is in its n -th quantum state, its rate of emission of radiation, according to (50) of § 38, will be

$$\frac{4}{3} \frac{\omega^4}{c^3} \frac{\epsilon^2}{2m\omega^2} (H' - \frac{1}{2}\hbar\omega) = \frac{2\hbar\epsilon^2\omega^3}{3mc^3} n \quad (21)$$

giving a probability $2\epsilon^2\omega^2/3mc^3 \cdot n$ per unit time of the oscillator jumping from state n to state $n-1$. In the state of lowest energy, for which $n = 0$, there is no emission of radiation.

In the classical treatment of periodic and multiply-periodic dynamical systems it is often convenient to make use of action and angle variables. We can introduce corresponding variables in the quantum theory. In our present problem of the harmonic oscillator we can define the action variable J by

$$J = H/\omega - \frac{1}{2}\hbar. \quad (22)$$

It is a constant of the motion and its eigenvalues are integral multiples of \hbar greater than or equal to zero. Thus its matrix representative in the Heisenberg representation is

0	0	0	0	0	.	.
0	\hbar	0	0	0	.	.
0	0	$2\hbar$	0	0	.	.
0	0	0	$3\hbar$	0	.	.
0	0	0	0	$4\hbar$.	.
.

when the rows and columns are arranged in order of ascending energy-levels. To define the angle variable we introduce the two matrices

$$\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 0 & \dots & & & \\ 1 & 0 & 0 & 0 & 0 & \dots & & & \\ 0 & 1 & 0 & 0 & 0 & \dots & & & \\ 0 & 0 & 1 & 0 & 0 & \dots & & & \\ 0 & 0 & 0 & 1 & 0 & \dots & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & & & \end{array} \quad \begin{array}{ccccccccc} 0 & 1 & 0 & 0 & 0 & \dots & & & \\ 0 & 0 & 1 & 0 & 0 & \dots & & & \\ 0 & 0 & 0 & 1 & 0 & \dots & & & \\ 0 & 0 & 0 & 0 & 1 & \dots & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & & & \end{array}$$

in which the non-vanishing elements are just to the left and just to the right of the principal diagonal respectively, and call the variables that they represent at time $t = 0$, e^{iw} and e^{-iw} respectively. These two matrices are conjugate complex, according to the definition of § 21, and thus represent conjugate complex observables, in agreement with what is implied by the notation of e^{iw} and e^{-iw} . This notation implies further, however, that the two matrices are the reciprocals of one another and this is not altogether true. The matrix representing the product $e^{-iw}e^{iw}$ is, in fact, just the unit matrix, but that representing $e^{iw}e^{-iw}$ differs from the unit matrix through having zero for its first diagonal element. Thus

$$e^{-iw}e^{iw} = 1 \quad e^{iw}e^{-iw} \neq 1. \quad (23)$$

The variables e^{iw} , e^{-iw} , defined above through their matrix representatives, are the best quantum analogues that we can get to the exponentials of i and $-i$ times the angle variable of the classical theory. They have many properties analogous to those of their classical counterparts and their only serious defect is that $e^{iw}e^{-iw}$ is not quite equal to unity. Thus, for example, we obtain at once from the matrices the relations

$$\left. \begin{aligned} e^{iw}J &= (J - \hbar)e^{iw} \\ e^{-iw}J &= (J + \hbar)e^{-iw} \end{aligned} \right\} (24)$$

which are equivalent to the classical relations

$$[e^{iw}, J] = ie^{iw} \quad [e^{-iw}, J] = -ie^{-iw}.$$

Equations (24), when compared with equation (28) of Chapter II, are seen to be consistent with the view that J and w are conjugate dynamical variables satisfying the relation

$$wJ - Jw = i\hbar,$$

although actually this relation is meaningless since we cannot define w itself but only $e^{\pm iw}$. Again, the dynamical variable e^{iw} at an arbitrary

time t must be represented by a matrix whose elements vary with t according to the Heisenberg law $e^{i(H'-H)t/\hbar}$. Since all the matrix elements vanish except those referring to consecutive energy-levels for which $H' - H'' = \hbar\omega$, every matrix element will vary with the time according to the law $e^{i\omega t}$. This corresponds to the fact that in the classical theory w increases linearly with t at the rate ω .

The co-ordinate and momentum q and p can be expressed in terms of the action and angle variables. The momentum p , for instance, is, according to (20), represented by the matrix

$$(\frac{1}{2}m\hbar\omega)^{\frac{1}{2}} \cdot \begin{vmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ 1 & 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & 2 & \dots \\ 0 & 0 & 0 & 2 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{vmatrix}$$

with disregard of trivial phase factors, and hence

$$\left. \begin{aligned} p &= (\frac{1}{2}m\omega)^{\frac{1}{2}}(J^{\frac{1}{2}}e^{iw} + e^{-iw}J^{\frac{1}{2}}). \\ \text{Similarly } q &= (2m\omega)^{-\frac{1}{2}}(-iJ^{\frac{1}{2}}e^{iw} + ie^{-iw}J^{\frac{1}{2}}). \end{aligned} \right\} (25)$$

We see from these equations that p and q , when expressed in terms of the action and angle variables, involve them only through the two combinations $J^{\frac{1}{2}}e^{iw}$ and $e^{-iw}J^{\frac{1}{2}}$. Further, all dynamical variables that we may have to deal with to obtain any physical result must be functions of p and q and will therefore, when expressed in terms of the action and angle variables, involve them only through the two quantities $J^{\frac{1}{2}}e^{iw}$ and $e^{-iw}J^{\frac{1}{2}}$. Now it is easily verified from the matrix representatives that these two quantities are respectively equal to

$$\left. \begin{aligned} J^{\frac{1}{2}}e^{iw} &= e^{iw}(J + \hbar)^{\frac{1}{2}} \\ e^{-iw}J^{\frac{1}{2}} &= (J + \hbar)^{\frac{1}{2}}e^{-iw} \end{aligned} \right\} (26)$$

and that their products in either order are

$$\begin{aligned} J^{\frac{1}{2}}e^{iw} \cdot e^{-iw}J^{\frac{1}{2}} &= J \\ e^{-iw}J^{\frac{1}{2}} \cdot J^{\frac{1}{2}}e^{iw} &= (J + \hbar)^{\frac{1}{2}}e^{-iw} \cdot e^{iw}(J + \hbar)^{\frac{1}{2}} = J + \hbar. \end{aligned}$$

These results hold in spite of the inequality in (23). They show that when we are dealing with dynamical variables of physical importance, which can involve the action and angle variables only through the two quantities $J^{\frac{1}{2}}e^{iw}$ and $e^{-iw}J^{\frac{1}{2}}$, we may count e^{iw} and e^{-iw} as truly reciprocal quantities without getting into error. Thus we can freely

use the action and angle variables in complete analogy with the classical theory without getting incorrect physical results.

The wave equation for the harmonic oscillator with Hamiltonian (18) is

$$i\hbar \frac{\partial}{\partial t} (q|) = \frac{1}{2m} \left\{ -\hbar^2 \frac{\partial^2}{\partial q^2} + m^2 \omega^2 q^2 \right\} (q|).$$

The wave functions representing stationary states will be the periodic solutions of this equation, for which the operator $i\hbar\partial/\partial t$ is the same as multiplication by the energy-level H' . They will thus satisfy

$$H'(q|) = \frac{1}{2m} \left\{ -\hbar^2 \frac{\partial^2}{\partial q^2} + m^2 \omega^2 q^2 \right\} (q|). \quad (27)$$

The general solution of this equation has been given by Schrödinger.* We shall here obtain some of the solutions representing states of lowest energy for use in the next section.

Equation (27) reduces to

$$\left\{ \frac{d^2}{dq^2} - \frac{q^2}{a^4} + \frac{2n+1}{a^2} \right\} (q|) = 0, \quad (28)$$

where a^2 is the number $\hbar/m\omega$ and H' has been put equal to $(n+\frac{1}{2})\hbar\omega$. Put

$$(q|) = f(q) e^{-q^2/2a^2}.$$

Equation (28) now becomes

$$\frac{d^2f}{dq^2} - 2 \frac{df}{dq} \frac{q}{a^2} + f \left[\frac{q^2}{a^4} - \frac{1}{a^2} \right] + \left[-\frac{q^2}{a^4} + \frac{2n+1}{a^2} \right] f = 0$$

or

$$\frac{d^2f}{dq^2} - 2 \frac{q}{a^2} \frac{df}{dq} + \frac{2n}{a^2} f = 0.$$

The solution of this equation, when n is any non-negative integer, is a finite power series in q . For

$$n = 0, 1, 2, 3, \dots$$

the solutions are easily verified to be

$$f(q) = 1, q, q^2 - \frac{1}{2}a^2, q^3 - \frac{3}{2}qa^2, \dots$$

The successive eigenfunctions are thus

$$\begin{aligned} (q|0) &= e^{-q^2/2a^2} & (q|1) &= qe^{-q^2/2a^2} \\ (q|2) &= (q^2 - \frac{1}{2}a^2)e^{-q^2/2a^2} & (q|3) &= (q^3 - \frac{3}{2}qa^2)e^{-q^2/2a^2} \dots \end{aligned} \quad \} \quad (29)$$

* Schrödinger, *Ann. d. Phys.*, vol. Ixxix, p. 514 (1926).

§ 42. The Harmonic Oscillator in Two Dimensions

Let us now suppose the harmonic oscillator of the preceding section can vibrate also in a second direction, at right angles to the first, with the same frequency $\omega/2\pi$. We shall then have a harmonic oscillator in two dimensions, whose Hamiltonian is

$$H = 1/2m \cdot (p_x^2 + p_y^2) + \frac{1}{2}m\omega^2(x^2 + y^2), \quad (30)$$

where x and y are the co-ordinates and p_x and p_y the conjugate momenta. The study of this system is of interest as it provides beautiful examples of the superposition of states and also it can be applied to the problem of the polarization of a photon.

The Hamiltonian (30) can be regarded as the sum of the Hamiltonians of two separate dynamical systems, namely, the two one-dimensional harmonic oscillators with the Hamiltonians

$$H_x = 1/2m \cdot p_x^2 + \frac{1}{2}m\omega^2x^2 \quad H_y = 1/2m \cdot p_y^2 + \frac{1}{2}m\omega^2y^2. \quad (31)$$

On account of this fact there is a simple connexion between the eigenfunctions of the H of (30), representing stationary states of the whole system, and those of the H_x and H_y of (31), representing stationary states of the component systems. Let us first consider the general case of a system whose Hamiltonian H can be regarded as the sum of the Hamiltonians H_1 and H_2 of two separate dynamical systems, i.e.

$$H = H_1 + H_2,$$

where all the observables in H_1 are different from and commute with all those in H_2 . We can now choose a complete set of commuting observables defining a representation, consisting of some observables q_1 that occur only in H_1 and some q_2 that occur only in H_2 . This will result in the representative of H being of the form

$$(q'_1 q'_2 | H | q''_1 q''_2) = (q'_1 | H_1 | q''_1) \delta(q'_2 - q''_2) + \delta(q'_1 - q''_1) (q'_2 | H_2 | q''_2), \quad (32)$$

if we take the case of continuous q' for definiteness. Now let $(q'_1 | H'_1)$ and $(q'_2 | H'_2)$ be eigenfunctions of H_1 and H_2 respectively, belonging to the eigenvalues H'_1 and H'_2 , so that

$$\int (q'_1 | H_1 | q''_1) dq''_1 (q''_1 | H'_1) = H'_1 (q'_1 | H'_1)$$

$$\int (q'_2 | H_2 | q''_2) dq''_2 (q''_2 | H'_2) = H'_2 (q'_2 | H'_2).$$

We shall then have from (32)

$$\begin{aligned} & \int \int (q'_1 q'_2 | H | q''_1 q''_2) dq''_1 dq''_2 (q''_1 | H'_1) (q''_2 | H'_2) \\ &= \int (q'_1 | H_1 | q''_1) dq''_1 (q''_1 | H'_1) (q'_2 | H'_2) + (q'_1 | H'_1) \int (q'_2 | H_2 | q''_2) dq''_2 (q''_2 | H'_2) \\ &= H'_1 (q'_1 | H'_1) (q'_2 | H'_2) + H'_2 (q'_1 | H'_1) (q'_2 | H'_2). \end{aligned}$$

This shows that the product $(q'_1 | H'_1) (q'_2 | H'_2)$ is an eigenfunction of H belonging to the eigenvalue $H'_1 + H'_2$. *The product of eigenfunctions of the Hamiltonians of each of the component systems is an eigenfunction of the Hamiltonian of the whole system, the corresponding eigenvalue being the sum of those for the components.* The physical meaning of this result is, of course, that when the component systems are in stationary states, the whole system is also in a stationary state, whose energy is the sum of those of the components and whose representative eigenfunction is the product of those of the components.

Let us apply this general result to our problem of the two-dimensional oscillator. We have already in the preceding section considered the eigenfunctions of Hamiltonians of the form of H_x and H_y . Let $(x|n_x)$ and $(y|n_y)$ be eigenfunctions of H_x and H_y , labelled by the quantum numbers n_x and n_y , the corresponding energy-levels being $H'_x = (n_x + \frac{1}{2})\hbar\omega$ and $H'_y = (n_y + \frac{1}{2})\hbar\omega$ respectively. Their product

$$(x|n_x)(y|n_y)$$

will then be an eigenfunction of the Hamiltonian H of (30), belonging to the eigenvalue

$$H' = H'_x + H'_y = (n_x + n_y + 1)\hbar\omega.$$

Thus the eigenvalues of H are integral multiples of $\hbar\omega$ greater than zero. Each of these eigenvalues (except the lowest one $\hbar\omega$) belongs to several eigenfunctions, corresponding to the various possible ways of choosing n_x and n_y to have a given integer as sum. There are thus several stationary states with the same energy. A system for which this is the case is called *degenerate*.

Let us now examine the eigenfunctions of some of the states of low energy, using the results (29) for the eigenfunctions for the one-dimensional oscillator. The state of lowest energy $\hbar\omega$ has the quantum numbers $n_x = 0$, $n_y = 0$ and is represented by the eigenfunction

$$(x|0)(y|0) = e^{-(x^2+y^2)/2a^2}. \quad (33)$$

There is only one state belonging to this energy-level, which is there-

fore non-degenerate. The next lowest energy-level $2\hbar\omega$ has two independent states belonging to it, corresponding to the two sets of quantum numbers $n_x = 1, n_y = 0$ and $n_x = 0, n_y = 1$. The two eigenfunctions are

$$\left. \begin{aligned} (x|1)(y|0) &= xe^{-(x^2+y^2)/2a^2} \\ (x|0)(y|1) &= ye^{-(x^2+y^2)/2a^2}. \end{aligned} \right\} \quad (34)$$

We can take any linear combination of these two eigenfunctions and get another eigenfunction representing another stationary state belonging to the same energy-level $2\hbar\omega$.

Our two-dimensional harmonic oscillator has circular symmetry about the origin in the xy plane. Hence, if we take a new set of rectangular Cartesian co-ordinates $x^* = x \cos \theta + y \sin \theta, y^* = x \sin \theta - y \cos \theta$, the wave functions in x^*, y^* will be of the same form as those in x, y . The stationary state of energy $2\hbar\omega$ for which the x^* component of oscillation is in the one-quantum state and the y^* component in the zero-quantum state, *i.e.* for which $n_{x^*} = 1, n_{y^*} = 0$, will therefore be represented by the eigenfunction

$$x^*e^{-(x^{*2}+y^{*2})/2a^2}.$$

But this is equal to

$$(x \cos \theta + y \sin \theta)e^{-(x^2+y^2)/2a^2}, \quad (35)$$

which is a linear combination of the two eigenfunctions (34). Thus the one-quantum state of linear oscillation in any direction can be obtained by a superposition of the two one-quantum states of linear oscillation in the x and y directions respectively.

The essential differences in the nature of this quantum superposition from that of classical superposition for the same dynamical system should be noted. In the classical theory if we superpose a state of linear oscillation of given energy in the x -direction with one of linear oscillation of the same energy in the y -direction, the resulting state will be of double the energy, instead of the same energy as in the quantum theory. Again, if this resulting state is one of linear oscillation, it must be in a direction at 45° to the original oscillations and cannot be in an arbitrary direction as in the quantum theory.

The example of quantum superposition just discussed is directly applicable to the problem of the polarization of a photon. A photon of given frequency moving in a given direction may be regarded as a harmonic electromagnetic oscillation in a one-quantum state. This oscillation may be resolved into two perpendicular directions, corre-

sponding to two independent states of linear polarization of the photon, so it forms a dynamical system formally the same as the two-dimensional oscillator investigated above. The wave functions (34), (35) may thus represent states of linear polarization of the photon. We see that the state of a photon linearly polarized in an arbitrary direction θ can be obtained by superposition of the states of polarization 0 and $\frac{1}{2}\pi$. The relative weights of these two states in the superposition process are given by the squares of the moduli of the coefficients of the wave functions (34) in the expression (35) and are thus as $\cos^2 \theta : \sin^2 \theta$, in agreement with the discussion in Chapter I.

We can superpose the two states of linear oscillation represented by the two eigenfunctions (34) in such a way as to get a state of circular oscillation in either direction about the origin, corresponding to a circularly polarized photon. To do this we must take the following linear combinations of the eigenfunctions (34),

$$(x+iy)e^{-(x^2+y^2)/2a^2}, \quad (x-iy)e^{-(x^2+y^2)/2a^2}. \quad (36)$$

These two new eigenfunctions will represent states of circular symmetry, as is at once apparent from the fact that they remain invariant, except for multiplication by a numerical factor, when one makes a transformation to the co-ordinates x^*, y^* . We can determine the direction of rotation for either of these eigenfunctions from a consideration of the angular momentum. We define the angular momentum, as in the classical theory, by $xp_y - yp_x$. It is represented by the operator $-i\hbar(x\partial/\partial y - y\partial/\partial x)$, which operator, when multiplied into the first of the eigenfunctions (36), gives the result

$$\begin{aligned} & -i\hbar\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)(x+iy)e^{-(x^2+y^2)/2a^2} = \\ &= -i\hbar x\left\{i - \frac{(x+iy)y}{a^2}\right\}e^{-(x^2+y^2)/2a^2} + i\hbar y\left\{1 - \frac{(x+iy)x}{a^2}\right\}e^{-(x^2+y^2)/2a^2} \\ &= \hbar(x+iy)e^{-(x^2+y^2)/2a^2}. \end{aligned}$$

This operator is thus equivalent to multiplication by \hbar , showing that the first of the eigenfunctions (36) represents a state for which the angular momentum has the value \hbar . The second must now from symmetry represent a state for which the angular momentum has the value $-\hbar$. It should be noticed that the states of linear oscillation represented by the eigenfunctions (34) are not states for which the angular momentum has the value zero, as it would in the classical theory, but are states for which there is an even chance of its having

the value \hbar or $-\hbar$. The state of lowest energy represented by (33) is one for which the angular momentum has the value zero.

We can deal in the same way with the two-quantum states of energy $3\hbar\omega$, of which there are three independent ones, corresponding to the three sets of quantum numbers $n_x = 2, n_y = 0; n_x = 1, n_y = 1; n_x = 0, n_y = 2$. The three eigenfunctions are

$$\left. \begin{aligned} (x|2)(y|0) &= (x^2 - \frac{1}{2}a^2)e^{-(x^2+y^2)/2a^2} \\ (x|1)(y|1) &= xy e^{-(x^2+y^2)/2a^2} \\ (x|0)(y|2) &= (y^2 - \frac{1}{2}a^2)e^{-(x^2+y^2)/2a^2}. \end{aligned} \right\} (37)$$

The two-quantum state of linear oscillation in any direction x^* will be represented by the eigenfunction

$$\begin{aligned} &(x^{*2} - \frac{1}{2}a^2)e^{-(x^{*2}+y^{*2})/2a^2} \\ &= \{(x \cos \theta + y \sin \theta)^2 - \frac{1}{2}a^2\}e^{-(x^2+y^2)/2a^2}, \end{aligned}$$

which is a linear combination of the three eigenfunctions (37). There are three two-quantum states of circular oscillation, represented by the eigenfunctions

$$\begin{aligned} (x+iy)^2 e^{-(x^2+y^2)/2a^2} &\quad \{(x+iy)(x-iy)-a^2\}e^{-(x^2+y^2)/2a^2} \\ &(x-iy)^2 e^{-(x^2+y^2)/2a^2}. \end{aligned}$$

It is easily verified that the angular momentum has the values $2\hbar, 0, -2\hbar$ for these three states respectively.

§ 43. The Spin of the Electron

In dealing with problems about electrons according to quantum mechanics, one finds one does not get agreement with experiment if one assumes the electrons to be simply point charges repelling one another according to the Coulomb law of force. It is necessary to make the assumption that each electron is spinning and so has an internal angular momentum, and also that it has a magnetic moment. To make the theory agree with experiment we must assume that the eigenvalues of the Cartesian component of the spin angular momentum in any direction are $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$, and that the magnetic moment of the electron (with its sign reversed) always lies in the same direction as the spin angular momentum and has as eigenvalues for its component in any direction the values $* e\hbar/2mc$ and $-e\hbar/2mc$. Thus if an electron in a certain state of spin has a spin angular momentum of $\frac{1}{2}\hbar$ in a particular direction, it will have a magnetic moment

* The e here, denoting minus the charge on an electron, is, of course, to be distinguished from the e denoting the base of exponentials.

$-e\hbar/2mc$ in this same direction. A theoretical reason for these assumptions will be provided by the relativity theory of the electron given in Chapter XIII. For the present we shall merely take them as empirical results and investigate their principal consequences.

Let s_x, s_y, s_z be the three Cartesian components of the spin angular momentum. We require quantum conditions for these three observables, to replace the classical conditions that they all commute. In § 44 the quantum conditions will be obtained for the three components of the angular momentum about a point of a single particle and also of a set of particles. It will be found that these quantum conditions are of the same form for a single particle as for a set of particles, which suggests that this form, namely equations (8) of § 44, is the general one governing any angular momentum, even the angular momentum of a spinning body. This gives us the quantum conditions

$$[s_y, s_z] = s_x \quad [s_z, s_x] = s_y \quad [s_x, s_y] = s_z, \quad (38)$$

for s_x, s_y, s_z , which may be written alternatively

$$s_y s_z - s_z s_y = i\hbar s_x \quad s_z s_x - s_x s_z = i\hbar s_y \quad s_x s_y - s_y s_x = i\hbar s_z \quad (39)$$

and combined in the single vector equation

$$\mathbf{s} \times \mathbf{s} = i\hbar \mathbf{s}.$$

There will be further algebraic relations satisfied by s_x, s_y, s_z , owing to the fact that each of these observables has only two eigenvalues $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$. Thus its square will have only the one eigenvalue $\frac{1}{4}\hbar^2$ and may therefore be put equal to the number $\frac{1}{4}\hbar^2$, i.e.

$$s_x^2 = s_y^2 = s_z^2 = \frac{1}{4}\hbar^2. \quad (40)$$

It is convenient to write

$$s_x = \frac{1}{2}\hbar\sigma_x \quad s_y = \frac{1}{2}\hbar\sigma_y \quad s_z = \frac{1}{2}\hbar\sigma_z,$$

introducing the three new observables σ_x, σ_y , and σ_z . The magnetic moment of the electron then has the components

$$- \frac{e\hbar}{2mc} \sigma_x, \quad - \frac{e\hbar}{2mc} \sigma_y, \quad - \frac{e\hbar}{2mc} \sigma_z,$$

so that these three observables $\sigma_x, \sigma_y, \sigma_z$ are sufficient to describe completely the spin of the electron. They form the components of a vector σ .

From (39) we find

$$\sigma_y \sigma_z - \sigma_z \sigma_y = 2i\sigma_x \quad \sigma_z \sigma_x - \sigma_x \sigma_z = 2i\sigma_y \quad \sigma_x \sigma_y - \sigma_y \sigma_x = 2i\sigma_z, \quad (41)$$

and from (40)

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1,$$

corresponding to the fact that each σ has just the two eigenvalues 1 and -1. From the first of equations (41)

$$\begin{aligned} 2i(\sigma_x \sigma_y + \sigma_y \sigma_x) &= (2i\sigma_x)\sigma_y + \sigma_y(2i\sigma_x) \\ &= (\sigma_y \sigma_z - \sigma_z \sigma_y)\sigma_y + \sigma_y(\sigma_y \sigma_z - \sigma_z \sigma_y) \\ &= -\sigma_z \sigma_y^2 + \sigma_y^2 \sigma_z = 0, \end{aligned}$$

so that

$$\sigma_x \sigma_y = -\sigma_y \sigma_x.$$

Two observables like these which satisfy the commutative law of multiplication except for a minus sign are said to *anticommute*. Thus σ_x anticommutes with σ_y and from symmetry any of the three observables $\sigma_x, \sigma_y, \sigma_z$ anticommutes with any other. We now obtain from (41)

$$\left. \begin{aligned} \sigma_y \sigma_z &= i\sigma_x = -\sigma_z \sigma_y \\ \sigma_z \sigma_x &= i\sigma_y = -\sigma_x \sigma_z \\ \sigma_x \sigma_y &= i\sigma_z = -\sigma_y \sigma_x \\ \sigma_x \sigma_y \sigma_z &= i. \end{aligned} \right\} \quad (42)$$

We must verify that the relations (42) are invariant under a rotation of axes, in order to show that our assumptions about the spin are permissible. Let the components of σ referred to a new set of mutually perpendicular axes be

$$\begin{aligned} \sigma_1 &= l_1 \sigma_x + m_1 \sigma_y + n_1 \sigma_z \\ \sigma_2 &= l_2 \sigma_x + m_2 \sigma_y + n_2 \sigma_z \\ \sigma_3 &= l_3 \sigma_x + m_3 \sigma_y + n_3 \sigma_z. \end{aligned}$$

From (42) we now obtain

$$\begin{aligned} \sigma_1^2 &= (l_1 \sigma_x + m_1 \sigma_y + n_1 \sigma_z)^2 \\ &= l_1^2 \sigma_x^2 + m_1^2 \sigma_y^2 + n_1^2 \sigma_z^2 + l_1 m_1 (\sigma_x \sigma_y + \sigma_y \sigma_x) + \\ &\quad + m_1 n_1 (\sigma_y \sigma_z + \sigma_z \sigma_y) + n_1 l_1 (\sigma_z \sigma_x + \sigma_x \sigma_z) \\ &= l_1^2 + m_1^2 + n_1^2 = 1. \end{aligned}$$

Again,

$$\begin{aligned} \sigma_2 \sigma_3 &= (l_2 \sigma_x + m_2 \sigma_y + n_2 \sigma_z)(l_3 \sigma_x + m_3 \sigma_y + n_3 \sigma_z) \\ &= l_2 l_3 \sigma_x^2 + m_2 m_3 \sigma_y^2 + n_2 n_3 \sigma_z^2 + l_2 m_3 \sigma_x \sigma_y + m_2 l_3 \sigma_y \sigma_x + \\ &\quad + m_2 n_3 \sigma_y \sigma_z + n_2 m_3 \sigma_z \sigma_y + n_2 l_3 \sigma_z \sigma_x + l_2 n_3 \sigma_x \sigma_z \\ &= l_2 l_3 + m_2 m_3 + n_2 n_3 + i(l_2 m_3 - m_2 l_3) \sigma_z + i(m_2 n_3 - n_2 m_3) \sigma_x + \\ &\quad + i(n_2 l_3 - l_2 n_3) \sigma_y \\ &= i(l_1 \sigma_x + m_1 \sigma_y + n_1 \sigma_z) = i\sigma_1. \end{aligned}$$

Thus $\sigma_1, \sigma_2, \sigma_3$ satisfy relations of the same form as (42).

We shall now obtain matrices to represent the spin observables σ_x , σ_y , σ_z . These matrices need have only two rows and columns, since the observables they represent have each only two eigenvalues. If we take a representation in which σ_z is diagonal, then σ_z will be represented by

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Let σ_x be represented by

$$\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}.$$

Since σ_x is a real observable this matrix must be Hermitian, so that a_1 and a_4 must be real and a_2 and a_3 conjugate complex numbers. The equation $\sigma_z\sigma_x = -\sigma_x\sigma_z$ now gives us

$$\begin{pmatrix} a_1 & a_2 \\ -a_3 & -a_4 \end{pmatrix} = -\begin{pmatrix} a_1 & -a_2 \\ a_3 & -a_4 \end{pmatrix},$$

so that $a_1 = a_4 = 0$. Hence σ_x is represented by a matrix of the form

$$\begin{pmatrix} 0 & a_2 \\ a_3 & 0 \end{pmatrix}.$$

The equation $\sigma_x^2 = 1$ now shows that $a_2 a_3 = 1$. Thus a_2 and a_3 , being conjugate complex numbers, must be of the form $e^{i\alpha}$ and $e^{-i\alpha}$ respectively, where α is a real number, so that σ_x is represented by a matrix of the form

$$\begin{pmatrix} 0 & e^{i\alpha} \\ e^{-i\alpha} & 0 \end{pmatrix}.$$

Similarly it may be shown that σ_y is also represented by a matrix of this form. By suitably choosing the phases in the representation, which is not completely determined by the condition that σ_z shall be diagonal, we can arrange that σ_x shall be represented by the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The representative of σ_y is then determined by the equation $\sigma_y = i\sigma_x\sigma_z$. We thus obtain finally the three matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

to represent σ_x , σ_y , and σ_z respectively, which matrices satisfy all the algebraic relations (42). The component of the spin vector σ in an

arbitrary direction specified by the direction cosines l, m, n is represented by

$$\begin{pmatrix} n & l-im \\ l+im & -n \end{pmatrix}. \quad (43)$$

In our representation with σ_z diagonal, a state of spin will be represented by a function $(\sigma'_z|)$ of the variable σ'_z , whose domain consists of only the two points $+1, -1$. This function is thus a pair of numbers. The state for which σ_z has the value unity will be represented by the function, $f_\alpha(\sigma'_z)$ say, consisting of the pair of numbers $1, 0$ and that for which it has the value -1 by the function, $f_\beta(\sigma'_z)$ say, consisting of the pair $0, 1$. Any function of the variable σ'_z , i.e. any pair of numbers, can be expressed as a linear combination of these two. Thus *any state of spin can be obtained by superposition of the two states for which σ_z equals $+1$ and -1 respectively.* For example, the state for which the component of σ in the direction l, m, n , represented by (43), has the value 1 is represented by the pair of numbers a, b which satisfy

$$\begin{aligned} na + (l-im)b &= a \\ (l+im)a - nb &= b. \end{aligned}$$

This gives

$$\frac{a}{b} = \frac{l-im}{1-n} = \frac{1+n}{l+im}.$$

If this state is regarded as a superposition of the two states for which σ_z equals $+1$ and -1 , the relative weights in the superposition process are as

$$|a|^2 : |b|^2 = |l-im|^2 : (1-n)^2 = 1+n : 1-n.$$

For the complete description of an electron we require the spin observables σ together with the Cartesian co-ordinates x, y, z and momenta p_x, p_y, p_z . The spin observables are assumed to commute with these co-ordinates and momenta. Thus a complete set of commuting observables for a system consisting of a single electron will be x, y, z, σ_z . In a representation in which these are diagonal, the representative of any state will be a function of four variables x', y', z', σ'_z . Since σ'_z has a domain consisting of only two points, this function of four variables is the same as two functions of three variables, namely the two functions

$$(x'y'z'|)_+ = (x', y', z', +1|) \quad (x'y'z'|)_- = (x', y', z', -1|).$$

Thus *the presence of the spin may be considered either as introducing a new variable into the wave function representing a state or as giving this wave function two components.*

VIII

MOTION IN A CENTRAL FIELD OF FORCE

§ 44. Properties of the Angular Momentum

An atom consists of a massive positively charged nucleus together with a number of electrons moving round it, under the influence of the attractive force of the nucleus and their own mutual repulsions. An exact treatment of this dynamical system would be a very difficult mathematical problem. One can, however, gain some insight into the main features of the system by making the rough approximation of regarding each electron as moving independently in a certain *central* field of force, namely that of the nucleus, assumed fixed, together with some kind of average of the forces due to the other electrons. Thus our present problem of the motion of a particle in a central field of force forms a corner-stone in the theory of the atom.

Let the Cartesian co-ordinates of the particle, referred to a system of axes with the centre of force as origin, be x , y , z and the corresponding components of momentum p_x , p_y , p_z . They satisfy the quantum conditions

$$[x, y] = 0 \quad [x, p_x] = 1 \quad [x, p_y] = 0,$$

&c. The Hamiltonian, with neglect of relativity mechanics, will be of the form

$$H = 1/2m \cdot (p_x^2 + p_y^2 + p_z^2) + V, \quad (1)$$

where V , the potential energy, is a function only of $(x^2 + y^2 + z^2)$.

We now introduce the components of angular momentum defined, as in the classical theory, by

$$m_x = yp_z - zp_y \quad m_y = zp_x - xp_z \quad m_z = xp_y - yp_x, \quad (2)$$

or by the vector equation

$$\mathbf{m} = \mathbf{x} \times \mathbf{p}.$$

From these equations we obtain at once the identity

$$m_x x + m_y y + m_z z = 0. \quad (3)$$

We must now evaluate the P.B.'s of the angular momentum components with the observables x , p_x , &c., and with each other. This we can do most conveniently with the help of the laws (7) and (8) of § 32, thus

$$[m_z, x] = [xp_y - yp_x, x] = -y[p_x, x] = y \quad \left. \right\} (4)$$

$$[m_z, y] = [xp_y - yp_x, y] = x[p_y, y] = -x$$

$$[m_z, z] = [xp_y - yp_x, z] = 0, \quad (5)$$

and similarly

$$[m_z, p_x] = p_y \quad [m_z, p_y] = -p_x \quad (6)$$

$$[m_z, p_z] = 0, \quad (7)$$

with corresponding relations for m_x and m_y . Again

$$\begin{aligned} [m_y, m_z] &= [zp_x - xp_z, m_z] = z[p_x, m_z] - [x, m_z]p_z \\ &= -zp_y + yp_z = m_x \\ [m_z, m_x] &= m_y \quad [m_x, m_y] = m_z. \end{aligned} \quad \left. \right\} (8)$$

These results are all the same as in the classical theory. The sign in the results (4), (6), and (8) may easily be remembered from the rule that the + sign occurs when the three observables, consisting of the two in the P. B. on the left-hand side and the one forming the result on the right, are in the cyclic order (xyz) and the - sign occurs otherwise.

From (4) and (5) we obtain

$$\begin{aligned} [m_z, x^2 + y^2 + z^2] &= x[m_z, x] + [m_z, x]x + y[m_z, y] + [m_z, y]y \\ &= xy + yx - yx - xy = 0. \end{aligned} \quad (9)$$

Similarly from (6) and (7) we find

$$[m_z, p_x^2 + p_y^2 + p_z^2] = 0. \quad (10)$$

Thus m_z commutes with $(x^2 + y^2 + z^2)$ and with $(p_x^2 + p_y^2 + p_z^2)$. It therefore commutes with the Hamiltonian H which, according to (1), is a function of these two observables only. Similarly m_x and m_y commute with H . Thus *the angular momentum is a constant of the motion*, as in the classical theory.

Equations (8) may be put in the vector form

$$\mathbf{m} \times \mathbf{m} = i\hbar \mathbf{m}. \quad (11)$$

If we have several particles with angular momenta $\mathbf{m}_1, \mathbf{m}_2 \dots$, each of them will satisfy (11), thus

$$\mathbf{m}_r \times \mathbf{m}_r = i\hbar \mathbf{m}_r.$$

Further, any one of these angular momenta will commute with any other, so that

$$\mathbf{m}_r \times \mathbf{m}_s + \mathbf{m}_s \times \mathbf{m}_r = 0 \quad (r \neq s).$$

Hence if $\mathbf{M} = \sum_r \mathbf{m}_r$ is the total angular momentum,

$$\begin{aligned} \mathbf{M} \times \mathbf{M} &= \sum_{rs} \mathbf{m}_r \times \mathbf{m}_s = \sum_r \mathbf{m}_r \times \mathbf{m}_r + \sum_{r < s} (\mathbf{m}_r \times \mathbf{m}_s + \mathbf{m}_s \times \mathbf{m}_r) \\ &= i\hbar \sum_r \mathbf{m}_r = i\hbar \mathbf{M}. \end{aligned}$$

This result is of the same form as (11), so that the components of the total angular momentum \mathbf{M} of any number of particles satisfy

the same commutability relations as those of the angular momentum of a single particle. Thus (11) or (8) may be regarded as *the general commutability relations satisfied by any angular momentum*. They certainly hold when the angular momentum is that of a number of particles, and may be assumed to hold also for the angular momentum of a spinning body, as was done in § 43 for the spinning electron.

We introduce the observable k defined as the positive square root

$$k = (m_x^2 + m_y^2 + m_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}}. \quad (12)$$

Equations (8) show that our observables m_x , m_y , m_z , if measured in units which make $\hbar = 1$, satisfy just the same conditions as the α , β , γ of § 30, the present k corresponding to the k of § 30. Thus we can apply the results of § 30 directly to our present observables. We obtain in this way that k commutes with m_x , m_y , m_z and that its eigenvalues are integral or half odd integral multiples of \hbar greater than zero. Also for any eigenvalue k' of k , the possible eigenvalues of m_x , m_y or m_z are

$$k' - \frac{1}{2}\hbar, k' - \frac{3}{2}\hbar, k' - \frac{5}{2}\hbar \dots - k' + \frac{1}{2}\hbar,$$

and are thus half odd integral or integral according as k' is integral or half odd integral. However, by using the further condition that m_x , m_y , m_z are of the form (2) we can show that their eigenvalues must be integral and thus that those of k must be half odd integral. We have, in fact, that m_z is represented by the operator $-i\hbar(x\partial/\partial y - y\partial/\partial x)$, which, if one makes the transformation $x = \rho \cos \phi$, $y = \rho \sin \phi$ to the cylindrical variables ρ , ϕ , becomes the operator $-i\hbar\partial/\partial\phi$. The general eigenfunction of this operator is of the form $f(\rho) e^{im'_z\phi/\hbar}$, m'_z being the eigenvalue and $f(\rho)$ being an arbitrary function of ρ . Now it is implied throughout our theory that an eigenfunction is a single-valued function of its variables and hence m'_z must be an integral multiple of \hbar . Similarly it may be shown that m_x and m_y have only integral eigenvalues. Thus the eigenvalues of the components of angular momentum of a particle moving in an orbit must be integral multiples of \hbar , although those of the components of angular momentum in the general case, which satisfy (8) but need not be of the form (2), may be either integral or half odd integral. Those assumed in § 43 for the components of spin angular momentum of an electron were half odd integral.

The components of angular momentum in different directions do not commute with each other, so that one cannot in general assign

numerical values to them simultaneously. One can at most give a numerical value to the component in one particular direction. The state of the system will then be one which, in the language of Bohr's theory, is *spacially quantized* in that direction. There is, however, one special case in which one can assign numerical values to all the components simultaneously, namely, one can give them all the value zero, since this will not contradict the commutability relations (8). The resulting state of zero angular momentum, with $k = \frac{1}{2}\hbar$, is then one that is spacially quantized simultaneously in all directions.

§ 45. Transition to Polar Co-ordinates

For further discussion of the problem of motion in a central field of force it is convenient to introduce polar observables. We introduce first the radius r , defined as the positive square root

$$r = (x^2 + y^2 + z^2)^{\frac{1}{2}}.$$

If we evaluate its P.B.'s with p_x , p_y , and p_z , we obtain, with the help of formula (16) of Chapter VI,

$$[r, p_x] = \frac{\partial r}{\partial x} = \frac{x}{r} \quad [r, p_y] = \frac{y}{r} \quad [r, p_z] = \frac{z}{r},$$

the same as in the classical theory. We could alternatively have evaluated these P.B.'s by the method given in § 39 for $[x, H]$.

We now introduce the observable p_r defined by

$$p_r = r^{-1}(xp_x + yp_y + zp_z - i\hbar). \quad (13)$$

Its P.B. with r is given by

$$\begin{aligned} r[r, p_r] &= [r, rp_r] = [r, xp_x + yp_y + zp_z] \\ &= x[r, p_x] + y[r, p_y] + z[r, p_z] \\ &= x \cdot x/r + y \cdot y/r + z \cdot z/r = r. \end{aligned}$$

Hence

$$[r, p_r] = 1$$

or

$$rp_r - p_r r = i\hbar,$$

so that p_r is canonically conjugate r . Now the eigenvalues of r , from its definition as a positive square root, must be all positive or zero, so that we have obtained a contradiction to the result, proved at the end of § 19, that an observable can have a canonical conjugate only if its eigenvalues include all numbers from $-\infty$ to ∞ . This inconsistency arises from the fact that the observable p_r , defined by (13) does not strictly exist, since r has the eigenvalue zero so that r^{-1} does not strictly exist. In spite of this defect the observable p_r is

a useful one for the study of motion in a central field of force. Our equations, which will often involve p_r , and will sometimes involve r^{-1} in other ways than through p_r , will be inaccurate, but only in so far as they apply to the one point $r = 0$, and this is too small a region of space to invalidate physical conclusions obtained from them.

The observable p_r , defined by (13) is a real one, since its conjugate complex \bar{p}_r is given by

$$\begin{aligned}\bar{p}_r r &= p_x x + p_y y + p_z z + i\hbar \\ &= xp_x + yp_y + zp_z - 2i\hbar \\ &= rp_r - i\hbar = p_r r,\end{aligned}$$

so that

$$\bar{p}_r = p_r.$$

We can easily verify that our two new observables r and p_r commute with the angular momentum. Equation (9) shows us that m_z commutes with r^2 . It must therefore commute also with r , since r is defined as a square-root function so that everything that commutes with r^2 commutes also with r . Again, for p_r , we have

$$\begin{aligned}r[p_r, m_z] &= [rp_r, m_z] = [xp_x + yp_y, m_z] \\ &= -yp_x - xp_y + xp_y + yp_x = 0.\end{aligned}$$

Thus r and p_r commute with m_z , and hence also with m_x and m_y and with k .

We can now express the Hamiltonian in terms of our radial observables r and p_r , and also k . We have, if Σ_{xyz} denotes a sum over cyclic permutations of the suffixes x, y, z ,

$$\begin{aligned}k^2 - \frac{1}{4}\hbar^2 &= \Sigma_{xyz} m_z^2 = \Sigma_{xyz} (xp_y - yp_x)^2 \\ &= \Sigma_{xyz} (xp_y xp_y + yp_x yp_x - xp_y yp_x - yp_x xp_y) \\ &= \Sigma_{xyz} (x^2 p_y^2 + y^2 p_x^2 - xp_x p_y y - yp_y p_x x + x^2 p_x^2 - xp_x p_x x - \\ &\quad - 2i\hbar xp_x) \\ &= (x^2 + y^2 + z^2)(p_x^2 + p_y^2 + p_z^2) - (xp_x + yp_y + zp_z) \times \\ &\quad \times (p_x x + p_y y + p_z z + 2i\hbar) \\ &= r^2(p_x^2 + p_y^2 + p_z^2) - (rp_r + i\hbar)rp_r \\ &= r^2(p_x^2 + p_y^2 + p_z^2) - r^2 p_r^2.\end{aligned}$$

Hence

$$H = \frac{1}{2m} \left(p_r^2 + \frac{k^2 - \frac{1}{4}\hbar^2}{r^2} \right) + V. \quad (14)$$

This form for H is such that k commutes not only with H , as is necessary since k is a constant of the motion, but also with every observable occurring in H , namely both r and p_r . Thus in dealing

with the Hamiltonian in this form we can treat k as a number. The permissible numbers we can take for k are its eigenvalues and are thus positive half odd integral multiples of \hbar . If we write down the Schrödinger equation for the stationary states, it will now read

$$\left\{ \frac{1}{2m} \left(-\hbar^2 \frac{\partial^2}{\partial r^2} + \frac{k^2 - \frac{1}{4}\hbar^2}{r^2} \right) + V \right\} (r|) = H' (r|), \quad (15)$$

the single variable r in the wave function $(r|)$ being sufficient when k is counted as a number. Any value of the parameter H' for which this equation, with a permissible value for k , has a solution (satisfying the boundary conditions to be discussed later) is a possible energy-level of the system. The energy-levels (except those for which $k = \frac{1}{2}\hbar$) are all degenerate and belong each to several independent stationary states, corresponding to the various possible eigenvalues of a Cartesian component of the angular momentum. The number of these states, for any value of k , is the odd number $2k/\hbar$.

If we write down the Schrödinger equation in the original Cartesian co-ordinates x, y, z , we shall have

$$\left\{ -\frac{\hbar^2}{2m} \nabla^2 + V \right\} (xyz|) = H' (xyz|), \quad (16)$$

where ∇^2 is the Laplacian operator $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$. This becomes, on transforming to polar co-ordinates r, θ, ϕ ,

$$\begin{aligned} \left\{ -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right) + V \right\} (r\theta\phi|) &= \\ &= H' (r\theta\phi|). \end{aligned}$$

The solutions of this equation are of the form

$$(r\theta\phi|) = \chi(r) S_n(\theta\phi),$$

where S_n is a spherical harmonic of order n satisfying

$$\left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right) S_n(\theta\phi) = -n(n+1) S_n(\theta\phi),$$

n being an integer, and $\chi(r)$ is a function of r only, satisfying

$$\left\{ -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{n(n+1)}{r^2} \right) + V \right\} \chi(r) = H' \chi(r). \quad (17)$$

This equation, like (15), is such that the values of H' for which it has a solution are the energy-levels of the system.

The equivalence of equations (15) and (17) may be seen from the fact that if in (15) we put $(r|) = r\chi(r)$ we obtain just equation (17) with $n = k/\hbar - \frac{1}{2}$. The fact that the two eigenfunctions $(r|)$ and $\chi(r)$

are not identical but differ by this factor r is due to their different physical interpretations. A solution $(r|)$ of (15) represents a state for which the probability of the particle lying in the spherical shell between r and $r+dr$ is proportional to $|(r)|^2 dr$. On the other hand, a solution $(xyz|)$ of (16) represents a state for which the probability of the particle lying in a small volume $dxdydz$ is $|(xyz)|^2 dxdydz$ or $|\chi(r)S_n(\theta\phi)|^2 dxdydz$, so that the probability of its lying in the spherical shell between r and $r+dr$ is proportional to $|\chi(r)|^2 r^2 dr$. Thus the physical interpretations require $(r|)$ to be proportional to $r\chi(r)$.

It should be noticed that not every solution of (17), when multiplied by the appropriate spherical harmonic, will give a solution of (16), as it may fail to satisfy (16) at the origin. One can see most clearly how this comes about by considering the special case for which the potential V vanishes, giving us the problem of the free particle. If we further take $H' = 0$, equation (16) reduces to

$$\nabla^2(xyz|) = 0 \quad (18)$$

and equation (17) to

$$\left\{ \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{n(n+1)}{r^2} \right\} \chi(r) = 0. \quad (19)$$

Now a solution of (19) for $n = 0$ is $\chi(r) = 1/r$, but this solution multiplied by the appropriate spherical harmonic $S_0 = 1$ does not satisfy (18), since, although $\nabla^2(1/r)$ vanishes for any finite value of r , its integral through any volume about the origin is 4π , and hence

$$\nabla^2(1/r) = 4\pi\delta(x)\delta(y)\delta(z).$$

Thus the solution $\chi(r) = 1/r$ of (19) does not represent a stationary state of the system. Again the solution $\chi(r) = 1/r^2$ of (19) for $n = 1$, when multiplied by the spherical harmonic $S_1 = \cos\theta$, gives a wave function $(xyz|)$, the integral of the square of whose modulus over any volume, however small, that contains the origin is infinite. This wave function must represent a state for which the particle is certainly at the origin and this cannot be a stationary state of zero energy for the problem of the free particle. Similarly for arbitrary n in equation (19), of the two solutions $\chi(r) = r^n$ and $\chi(r) = r^{-n-1}$, the second will not give the representative of a stationary state of the system.

It thus appears that equation (17) is not adequate to replace equation (16) as the necessary and sufficient condition for the representative of a stationary state. Equation (17) must be supplemented by a suitable boundary condition at the point $r = 0$. Any solution

$\chi(r)$ of (17) for which the integral $\int_0^\infty r^2 |\chi(r)|^2 dr$ is not convergent must certainly be rejected, and also some for which this integral is convergent, namely those which, when operated on by ∇^2 , give an infinite result involving the δ function at the origin. These conditions show that only those solutions are to be allowed which, if they tend to infinity as $r \rightarrow 0$, do so more slowly than $1/r$. The corresponding boundary condition for the function $(r|)$ of equation (15) is that it shall tend to zero as $r \rightarrow 0$.

There are also boundary conditions for the eigenfunction at $r = \infty$. If we are interested only in 'closed' states, i.e. states for which the particle does not go off to infinity, we must restrict the integral $\int_0^\infty |(r)|^2 dr$ or $\int_0^\infty r^2 |\chi(r)|^2 dr$ to be convergent. These closed states, however, are not the only ones that are physically permissible, as we can also have states in which the particle arrives from infinity, is scattered by the central field of force, and goes off to infinity again. For these states the wave function $(xyz|)$ may remain finite as $r \rightarrow \infty$. Such states will be dealt with in Chapter X under the heading of collision problems. In any case the eigenfunction $(xyz|)$ must not tend to infinity as $r \rightarrow \infty$, or it will represent a state that has no physical meaning.

§ 46. Energy-levels of the Hydrogen Atom

The above analysis may be applied to the problem of the hydrogen atom with neglect of the relativity variation of mass with velocity and the spin of the electron. The potential energy V is now $-e^2/r$, so that equation (15) becomes

$$\left\{ \frac{d^2}{dr^2} - \frac{k^2 - \frac{1}{4}}{r^2} + \frac{2me^2}{\hbar^2} \frac{1}{r} \right\} (r|) = -\frac{2mH'}{\hbar^2} (r|), \quad (20)$$

when written in terms of a new observable k , equal to \hbar^{-1} times the previous k . A thorough investigation of this equation has been given by Schrödinger.* We shall here obtain its eigenvalues H' from a consideration of its eigenfunctions expressed in the form of power series.

It is convenient to put

$$(r|) = f(r) e^{-r/a}, \quad (21)$$

introducing the new function $f(r)$, where a is one or other of the square roots

$$a = \pm \sqrt{-\hbar^2/2mH'}. \quad (22)$$

* Schrödinger, *Ann. d. Physik*, vol. Ixxix, p. 361 (1926).

Equation (20) now becomes

$$\left\{ \frac{d^2}{dr^2} - \frac{2}{a} \frac{d}{dr} - \frac{k^2 - \frac{1}{4}}{r^2} + \frac{2me^2}{\hbar^2} \cdot \frac{1}{r} \right\} f(r) = 0. \quad (23)$$

We look for a solution of this equation in the form of a power series

$$f(r) = \sum_s c_s r^s, \quad (24)$$

in which consecutive values for s differ by unity although these values themselves need not be integers. On substituting (24) in (23) we obtain

$$\sum_s c_s \{s(s-1)r^{s-2} - 2s/a \cdot r^{s-1} - (k^2 - \frac{1}{4})r^{s-2} + 2me^2/\hbar^2 \cdot r^{s-1}\} = 0,$$

which gives, on equating to zero the coefficient of r^{s-2} , the following relation between successive coefficients c_s ,

$$c_s [s(s-1) - (k^2 - \frac{1}{4})] = c_{s-1} [2(s-1)/a - 2me^2/\hbar^2]. \quad (25)$$

We saw in the preceding section that only those eigenfunctions ($r|$) are allowed that tend to zero with r and hence from (21) $f(r)$ must tend to zero with r . The series (24) must therefore terminate on the side of small s and the minimum value of s must be greater than zero. Now the only possible minimum values of s are those that make the coefficient of c_s in (25) vanish, *i.e.* $k + \frac{1}{2}$ and $-k + \frac{1}{2}$, and the second of these is negative or zero. Thus the minimum value of s must be $k + \frac{1}{2}$. Since k is always half an odd integer, the values of s will all be integers. The series (24) will in general extend to infinity on the side of large s . For large values of s the ratio of successive terms is

$$\frac{c_s}{c_{s-1}} r = \frac{2r}{sa}$$

according to (25). Thus the series (24) will always converge, as the ratios of the higher terms to one another are the same as for the series

$$\sum_s \frac{1}{s!} \left(\frac{2r}{a} \right)^s, \quad (26)$$

which converges to $e^{2r/a}$.

We must now examine how our solution ($r|$) behaves for large values of r . We must distinguish between the two cases of H' positive and H' negative. For H' negative, a given by (22) will be real. Suppose we take the positive value for a . Then as $r \rightarrow \infty$ the sum of the series (24) will tend to infinity according to the same law as the sum of the series (26), *i.e.* the law $e^{2r/a}$. Thus from (21) ($r|$) will tend to infinity according to the law $e^{r/a}$ and will not represent a physically

possible state. There is therefore in general no permissible solution of (20) for negative values of H' . An exception arises, however, whenever the series (24) terminates on the side of large s , in which case the boundary conditions are all satisfied. The condition for this termination of the series is that the coefficient of c_{s-1} in (25) shall vanish for some value of the suffix $s-1$ not less than its minimum value $k+\frac{1}{2}$, which is the same as the condition that

$$\frac{s}{a} - \frac{me^2}{\hbar^2} = 0$$

for some integer s not less than $k+\frac{1}{2}$. With the help of (22) this condition becomes

$$H' = -\frac{me^4}{2s^2\hbar^2}, \quad (27)$$

and is thus a condition for the energy-level H' . Since s may be any positive integer, the formula (27) gives a discrete set of negative energy-levels for the hydrogen atom. These are in agreement with experiment. Each of them (except the lowest one $s = 1$) is degenerate, as it may occur with various possible values for k , namely, any positive half odd integer less than s . This degeneracy is in addition to that mentioned in the preceding section arising from the various possible values for a component of angular momentum, which degeneracy occurs with any central field of force. The k degeneracy occurs only with an inverse square law of force and even then is removed when one takes relativity mechanics into account, as will be found in Chapter XIII. The solution of (20) when H' satisfies (27) tends to zero exponentially as $r \rightarrow \infty$ and thus represents a closed state, corresponding to an elliptic orbit in Bohr's theory.

For any positive values of H' , a given by (22) will be pure imaginary. The series (24), which is roughly the same as the series (26), will now have a sum that remains finite as $r \rightarrow \infty$. Thus $(r|)$ given by (21) will now remain finite as $r \rightarrow \infty$ and will therefore be a permissible solution of (20), since it will correspond to an eigenfunction ($xyz|$) that tends to zero according to the law $1/r$ as $r \rightarrow \infty$. Hence in addition to the discrete set of negative energy-levels (27), all positive energy-levels are allowed. The states of positive energy are not closed, since their representatives $(r|)$ do not make the integral $\int^{\infty} |(r)|^2 dr$ converge. These states correspond to the hyperbolic orbits of Bohr's theory.

§ 47. Selection Rules

When \mathbf{D} , the total electric displacement of a system, is represented in a Heisenberg representation, it often happens that a great many of its matrix elements, $(\alpha'|\mathbf{D}|\alpha'')$ say, vanish. In fact they may all vanish except those for which the α' 's and α'' 's are connected in a certain way. When this is the case, according to Heisenberg's interpretation of the matrix elements, a transition of the system with emission of radiation can take place only between two stationary states whose labels α' and α'' are connected in this way. There is then, as we say, a *selection rule* for the α 's, only certain selected transitions being allowed. In general we must consider separately the different Cartesian components D_x , D_y , D_z of \mathbf{D} and obtain for each of them the condition that its matrix element $(\alpha'|D|\alpha'')$ shall not vanish. We shall then often find that for those transitions $\alpha' \rightarrow \alpha''$ which can take place, *i.e.* for which the vector $(\alpha'|\mathbf{D}|\alpha'')$ does not vanish, some of the Cartesian components $(\alpha'|D_x|\alpha'')$, $(\alpha'|D_y|\alpha'')$, $(\alpha'|D_z|\alpha'')$ do vanish. There will then be conditions on the direction of emission and state of polarization of the emitted radiation, which conditions, according to Heisenberg's assumption, will be the same as the classical ones for the radiation emitted by an electric dipole whose magnitude and direction are given by the vector

$$(\alpha'|\mathbf{D}|\alpha'') + (\alpha''|\mathbf{D}|\alpha').$$

There is a general method for obtaining all selection rules, which is as follows. Let D be one of the Cartesian components of \mathbf{D} . We must obtain an algebraic equation connecting D and the α 's which does not involve any observables other than D and the α 's and which is linear in D . Such an equation will be of the form

$$\sum_r f_r D g_r = 0, \quad (28)$$

where the f_r 's and g_r 's are functions of the α 's only. When this equation is expressed in terms of representatives, it gives us

$$\sum_r f_r(\alpha') (\alpha'|D|\alpha'') g_r(\alpha'') = 0,$$

or

$$(\alpha'|D|\alpha'') \sum_r f_r(\alpha') g_r(\alpha'') = 0,$$

which shows that $(\alpha'|D|\alpha'') = 0$ unless

$$\sum_r f_r(\alpha') g_r(\alpha'') = 0. \quad (29)$$

This last equation, giving the connexion which must exist between α' and α'' in order that $(\alpha'|D|\alpha'')$ may not vanish, constitutes the selection rule, so far as the component D of \mathbf{D} is concerned.

We shall now obtain the selection rules for m_z and k for an electron moving in a central field of force. The components of electric displacement are here proportional to the Cartesian co-ordinates x, y, z . Taking first m_z , we have that m_z commutes with z , or that

$$m_z z - zm_z = 0.$$

This is an equation of the required type (28), giving us the selection rule

$$m'_z - m''_z = 0$$

for the z -component of the displacement. Again, from equations (8) we have

$$[m_z, [m_z, x]] = [m_z, y] = -x$$

or $m_z^2 x - 2m_z xm_z + xm_z^2 - \hbar^2 x = 0,$

which is also of the type (28) and gives us the selection rule

$$m'^2_z - 2m'_z m''_z + m''^2_z - \hbar^2 = 0$$

or $(m'_z - m''_z - \hbar)(m'_z - m''_z + \hbar) = 0$

for the x -component of the displacement. The selection rule for the y -component is the same. Thus our selection rules for m_z are that *for the emission of radiation with a polarization corresponding to an electric dipole in the z -direction, m'_z cannot change, while for that corresponding to an electric dipole in the x -direction or y -direction, m'_z must change by $\pm \hbar$.*

We can determine more accurately the state of polarization of the radiation emitted with a transition in which m'_z changes by $\pm \hbar$, by considering the condition for the non-vanishing of matrix elements of $x+iy$ and $x-iy$. We have

$$[m_z, x+iy] = y - ix = -i(x+iy)$$

or

$$m_z(x+iy) - (x+iy)(m_z + \hbar) = 0,$$

which is again of the type (28). It gives

$$m'_z - m''_z - \hbar = 0$$

as the condition that $(m'_z | x+iy | m''_z)$ shall not vanish. Similarly

$$m'_z - m''_z + \hbar = 0$$

is the condition that $(m'_z | x-iy | m''_z)$ shall not vanish. Hence

$$(m'_z | x-iy | m'_z - \hbar) = 0$$

or

$$(m'_z | x | m'_z - \hbar) = i(m'_z | y | m'_z - \hbar) = (a+ib)e^{i\omega t}$$

say, a , b , and ω being real, and similarly

$$(m'_z - \hbar|x|m'_z) = -i(m'_z - \hbar|y|m'_z) = (a - ib)e^{-i\omega t}.$$

Thus the vector $(m'_z|\mathbf{D}|m'_z - \hbar) + (m'_z - \hbar|\mathbf{D}|m'_z)$, which determines the state of polarization of the radiation emitted with transitions for which $m''_z = m'_z - \hbar$, has the following three components

$$\begin{aligned} (m'_z|x|m'_z - \hbar) + (m'_z - \hbar|x|m'_z) &= (a+ib)e^{i\omega t} + (a-ib)e^{-i\omega t} = \\ &= 2a \cos \omega t - 2b \sin \omega t \\ (m'_z|y|m'_z - \hbar) + (m'_z - \hbar|y|m'_z) &= -i(a+ib)e^{i\omega t} + i(a-ib)e^{-i\omega t} = \\ &= 2a \sin \omega t + 2b \cos \omega t \\ (m'_z|z|m'_z - \hbar) + (m'_z - \hbar|z|m'_z) &= 0. \end{aligned} \quad \left. \right\} (30)$$

From the form of these components we see that radiation emitted in the z -direction will be circularly polarized, that emitted in any direction in the xy plane will be linearly polarized in this plane, and that emitted in intermediate directions will be elliptically polarized. The direction of circular polarization for radiation emitted in the z -direction will depend on whether ω is positive or negative, and this will depend on which of the two states m'_z or $m''_z = m'_z - \hbar$ has the greater energy.

We shall now determine the selection rule for k . We have

$$\begin{aligned} [k^2, z] &= [m_x^2, z] + [m_y^2, z] \\ &= -ym_x - m_x y + xm_y + m_y x \\ &= 2(m_y x - m_x y + i\hbar z) \\ &= 2(m_y x - ym_x) = 2(xm_y - m_x y). \end{aligned}$$

Similarly

$$[k^2, x] = 2(ym_z - m_y z)$$

and

$$[k^2, y] = 2(m_x z - xm_z).$$

Hence

$$\begin{aligned} [k^2, [k^2, z]] &= 2[k^2, m_y x - m_x y + i\hbar z] \\ &= 2m_y [k^2, x] - 2m_x [k^2, y] + 2i\hbar [k^2, z] \\ &= 4m_y (ym_z - m_y z) - 4m_x (m_x z - xm_z) + 2(k^2 z - zk^2) \\ &= 4(m_x x + m_y y - m_z z)m_z - 4(m_x^2 + m_y^2 + m_z^2)z + 2(k^2 z - zk^2). \end{aligned}$$

The first term here vanishes, from (3), leaving us with

$$\begin{aligned} [k^2, [k^2, z]] &= -4(k^2 - \frac{1}{4}\hbar^2)z + 2(k^2 z - zk^2) \\ &= -2(k^2 z + zk^2) + \hbar^2 z, \end{aligned}$$

which gives

$$k^4 z - 2k^2 zk^2 + zk^4 - 2\hbar^2(k^2 z + zk^2) + \hbar^4 z = 0. \quad (31)$$

Similar equations hold for x and y . These equations are of the required type (28), and give us the selection rule

$$k'^4 - 2k'^2k''^2 + k''^4 - 2\hbar^2(k'^2 + k''^2) + \hbar^4 = 0$$

or

$$(k' + k'' + \hbar)(k' + k'' - \hbar)(k' - k'' + \hbar)(k' - k'' - \hbar) = 0.$$

A transition can take place between two states k' and k'' only if one of these four factors vanishes.

Now the first of the factors, $(k' + k'' + \hbar)$, can never vanish since the eigenvalues of k are all positive. The second, $(k' + k'' - \hbar)$, can vanish only if $k' = \frac{1}{2}\hbar$ and $k'' = \frac{1}{2}\hbar$. But transitions between two states with these values for k cannot occur on account of the selection rule for m_z , as may be seen from the following argument. If two states (labelled respectively with a single prime and a double prime) are such that $k' = \frac{1}{2}\hbar$ and $k'' = \frac{1}{2}\hbar$, then, according to the discussion at the end of § 44, each Cartesian component of the angular momentum must vanish for each of them, i.e. $m'_x = m'_y = m'_z = 0$ and $m''_x = m''_y = m''_z = 0$. The selection rule for m_z now shows that the matrix elements of x and y referring to the two states must vanish, as the value of m_z does not change during the transition, and the similar selection rule for m_x or m_y shows that the matrix element of z also vanishes. Thus transitions between the two states cannot occur. Our selection rule for k now reduces to

$$(k' - k'' + \hbar)(k' - k'' - \hbar) = 0,$$

showing that k must change by $\pm \hbar$. This selection rule may be written

$$k'^2 - 2k'k'' + k''^2 - \hbar^2 = 0,$$

and since this is the condition that a matrix element $(k'|z|k'')$ shall not vanish, we get the equation

$$k^2 z - 2kzk + zk^2 - \hbar^2 z = 0$$

or

$$[k, [k, z]] = -z, \quad (32)$$

a result which could not easily be obtained in a more direct way.

§ 48. The Zeeman Effect for the Hydrogen Atom

We shall now consider the system of a hydrogen atom in a uniform magnetic field. The Hamiltonian (1) with $V = -e^2/r$, which describes the hydrogen atom in no external field, gets modified by the magnetic field, the modification, according to classical mechanics,

consisting in the replacement of the components of momentum, p_x , p_y , p_z , by $p_x + e/c \cdot A_x$, $p_y + e/c \cdot A_y$, $p_z + e/c \cdot A_z$, where A_x , A_y , A_z are the components of the vector potential describing the field. For a uniform field of magnitude \mathcal{H} in the direction of the z -axis we may take $A_x = -\frac{1}{2}\mathcal{H}y$, $A_y = \frac{1}{2}\mathcal{H}x$, $A_z = 0$. The classical Hamiltonian will then be

$$H = \frac{1}{2m} \left\{ \left(p_x - \frac{1}{2} \frac{e}{c} \mathcal{H}y \right)^2 + \left(p_y + \frac{1}{2} \frac{e}{c} \mathcal{H}x \right)^2 + p_z^2 \right\} - \frac{e^2}{r}.$$

This classical Hamiltonian may be taken over into the quantum theory if we add on to it a term giving the effect of the spin of the electron. The electron has a magnetic moment $-e\hbar/2mc \cdot \sigma$, whose energy in the magnetic field will be $e\hbar\mathcal{H}/2mc \cdot \sigma_z$. Thus the quantum Hamiltonian will be

$$H = \frac{1}{2m} \left\{ \left(p_x - \frac{1}{2} \frac{e}{c} \mathcal{H}y \right)^2 + \left(p_y + \frac{1}{2} \frac{e}{c} \mathcal{H}x \right)^2 + p_z^2 \right\} - \frac{e^2}{r} + \frac{e\hbar\mathcal{H}}{2mc} \sigma_z. \quad (33)$$

There ought strictly to be other terms in this Hamiltonian giving the interaction of the magnetic moment of the electron with the electric field of the nucleus of the atom, but this effect is small, of the same order of magnitude as that of the relativity variation of the mass of the electron with its velocity, and will be neglected here. It will be taken into account in the relativity theory of the electron given in Chapter XIII.

If the magnetic field is not too large, we can neglect terms involving \mathcal{H}^2 , so that the Hamiltonian (33) reduces to

$$\begin{aligned} H &= \frac{1}{2m} (p_x^2 + p_y^2 + p_z^2) - \frac{e^2}{r} + \frac{e\mathcal{H}}{2mc} (xp_y - yp_x) + \frac{e\hbar\mathcal{H}}{2mc} \sigma_z \\ &= \frac{1}{2m} (p_x^2 + p_y^2 + p_z^2) - \frac{e^2}{r} + \frac{e\mathcal{H}}{2mc} (m_z + \hbar\sigma_z). \end{aligned} \quad (34)$$

The extra terms due to the magnetic field are now $e\mathcal{H}/2mc \cdot (m_z + \hbar\sigma_z)$. But these extra terms commute with the total Hamiltonian and are thus constants of the motion. This makes the problem very easy. The stationary states of the system, *i.e.* the eigenstates of the Hamiltonian (34), will be those eigenstates of the Hamiltonian for no field that are simultaneously eigenstates of the observables m_z and σ_z , or at least of the one observable $m_z + \hbar\sigma_z$, and the energy-levels of the system will be those for the system with no field, given by (27) if one considers only closed states, increased by an eigenvalue of $e\mathcal{H}/2mc \cdot (m_z + \hbar\sigma_z)$. Thus any stationary state of the system with no

field which is spacially quantized in the z -direction, i.e. for which m_z has the numerical value m'_z , an integral multiple of \hbar , and for which also σ_z has the numerical value $\sigma'_z = \pm 1$, will still be a stationary state when the field is applied. Its energy will be increased by an amount consisting of the sum of two parts, a part $e\mathcal{H}/2mc \cdot m'_z$ arising from the orbital motion, which may be considered as due to an orbital magnetic moment $-em'_z/2mc$, and a part $e\mathcal{H}/2mc \cdot \hbar\sigma'_z$ arising from the spin. The ratio of the orbital magnetic moment to the orbital angular momentum m'_z is $-e/2mc$, which is half the ratio of the spin magnetic moment to the spin angular momentum. This fact is sometimes referred to as the magnetic anomaly of the spin.

Since the energy-levels now involve m_z , the selection rule for m_z obtained in the preceding section becomes capable of direct comparison with experiment. According to this selection rule, m_z can change by \hbar , 0 or $-\hbar$ during an emission process. This means that the amount of energy emitted will differ by $-e\hbar\mathcal{H}/2mc$, 0 or $e\hbar\mathcal{H}/2mc$ respectively from the amount emitted when there is no field, since σ_z will not change as it commutes with the electric displacement of the system. Thus the frequency of the emitted radiation will differ by $-e\mathcal{H}/4\pi mc$, 0 or $e\mathcal{H}/4\pi mc$ from that for no field, so that each spectral line for no field gets split up into three components. If one considers the radiation emitted in the z -direction, then from (30) the two outer components will be circularly polarized while the central undisplaced one will be of zero intensity. These results are in agreement with experiment and also with the classical theory of the Zeeman effect. The agreement with the classical theory ceases, however, when one takes into account relativity mechanics and the interaction of the spin with the electric field of the nucleus.

§ 49. Combination of Angular Momenta

Suppose we have two particles moving in the central field of force, whose angular momenta are the vectors \mathbf{m} and μ . The magnitudes of these vectors are the observables k and κ , defined by (12) and

$$\kappa = (\mu_x^2 + \mu_y^2 + \mu_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}}$$

respectively. The total angular momentum will then be the vector $\mathbf{M} = \mathbf{m} + \mu$, whose magnitude is

$$K = (M_x^2 + M_y^2 + M_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}}.$$

Each of the observables k and κ commutes with all the components of \mathbf{m} , μ , and \mathbf{M} . Thus k , κ , K will commute with each other and

can be given numerical values simultaneously. Our problem now is to determine the possible numerical values for K when k and κ have given numerical values.

The easiest way of solving this problem is to suppose k and κ are equal to two given numbers, as we can do since they commute with all the observables mentioned in the problem, and then to use a matrix representation in which m_z and μ_z are diagonal. We can ignore all the observables describing the dynamical system that are not functions of the components of \mathbf{m} and $\boldsymbol{\mu}$. Our matrix representation will then have only a finite number of rows and columns, each labelled by a number m'_z having one of the values $k - \frac{1}{2}\hbar, k - \frac{3}{2}\hbar, \dots, -k + \frac{1}{2}\hbar$ and a number μ'_z having one of the values $\kappa - \frac{1}{2}\hbar, \kappa - \frac{3}{2}\hbar, \dots, -\kappa + \frac{1}{2}\hbar$. The possible values of $M'_z = m'_z + \mu'_z$ will then be $k + \kappa - \hbar, k + \kappa - 2\hbar, k + \kappa - 3\hbar, \dots, -k - \kappa + \hbar, -k - \kappa - \hbar, \dots, -k - \kappa - 2\hbar, -k - \kappa - 3\hbar, \dots, k - \kappa, k - \kappa - \hbar, \dots, -k + \kappa, -k + \kappa - \hbar, \dots, -k + \kappa - 2\hbar, -k + \kappa - 3\hbar, \dots, 2\kappa, 2\kappa - 1, \dots, 1$. The number of times each of them occurs is given by the following scheme (if one assumes for definiteness that $k \geq \kappa$),

$$\left. \begin{array}{ccccccc} k + \kappa - \hbar, & k + \kappa - 2\hbar, & k + \kappa - 3\hbar, & \dots & k - \kappa, & k - \kappa - \hbar, & \dots \\ 1 & 2 & 3 & \dots & 2\kappa & 2\kappa & \dots \\ & & & & -k + \kappa, & -k + \kappa - \hbar, & \dots & -k - \kappa + \hbar \\ & & & & 2\kappa & 2\kappa - 1 & \dots & 1 \end{array} \right\} (35)$$

If we now make a canonical transformation to a representation in which K and M_z are diagonal, the number of rows and columns of the matrices for which M_z has a given value M'_z must remain unaltered. If K', K'', \dots are the possible values for K , there will be a set of rows and columns having the M_z -values $K' - \frac{1}{2}\hbar, K' - \frac{3}{2}\hbar, \dots, K' + \frac{1}{2}\hbar$, together with a set having the M_z -values $K'' - \frac{1}{2}\hbar, K'' - \frac{3}{2}\hbar, \dots, K'' + \frac{1}{2}\hbar$, &c. Comparing this distribution of M_z -values with (35), we see that the possible values for K must be

$$k + \kappa - \frac{1}{2}\hbar, k + \kappa - \frac{3}{2}\hbar, k + \kappa - \frac{5}{2}\hbar, \dots, k - \kappa + \frac{1}{2}\hbar. \quad (36)$$

This result is a quite general one applying to the combination of any two angular momenta, not necessarily the orbital angular momenta of two particles. For example, it could be applied to the orbital angular momentum and spin of an electron. In this case, since the spin angular momentum has the magnitude $\kappa = \hbar$, it shows that when the orbital angular momentum has the magnitude k , the combined angular momentum can have only one or other of the two magnitudes $k \pm \frac{1}{2}\hbar$.

We now have a general method for dealing with complicated

atomic systems. For an isolated system the total angular momentum \mathbf{M} is always a constant of the motion and its resultant K together with one of its components M_z will be two commuting constants of the motion. We try to express \mathbf{M} as the sum of two angular momenta \mathbf{m} and μ whose magnitudes k and κ are constants of the motion. If we can do this, then we try to express either of the parts, \mathbf{m} say, itself as a sum of two angular momenta, \mathbf{m}_1 and \mathbf{m}_2 say, whose magnitudes k_1 and k_2 are constants of the motion, and so on. We obtain in this way a series of constants of the motion $M_z, K, k, \kappa, k_1, k_2 \dots$ which all commute with each other and may, if there are enough of them, be taken as defining a Heisenberg representation. The possible numerical values for the $K, k, \kappa \dots$ specifying a row and column are restricted by the general rule (36). The energy will be some function of $K, k, \kappa, k_1, k_2, \dots$ but independent of M_z . In general one cannot secure that k, κ, k_1, k_2 are exactly constants of the motion, but one may be able to choose them so that they are approximately so and then apply a perturbation method, as discussed in the next chapter.

We shall now obtain the selection rule for the magnitude K of the total angular momentum \mathbf{M} of a general atomic system. Let \mathbf{m} be the orbital angular momentum of one of the electrons, whose coordinates are x, y, z , say, and let $\mathbf{M} - \mathbf{m} = \mu$. It is not necessary for the present discussion that the magnitudes k and κ of the two angular momenta \mathbf{m} and μ into which M has been split up should be constants of the motion. We must obtain the condition that the (K', K'') matrix element of x, y , or z shall not vanish. This is evidently the same as the condition that the (K', K'') matrix element of λ_1, λ_2 , or λ_3 shall not vanish, where λ_1, λ_2 , and λ_3 are any three independent linear functions of x, y , and z with numerical coefficients, or more generally with any coefficients that commute with K and are thus represented by matrices which are diagonal with respect to K . Let

$$\begin{aligned}\lambda_0 &= M_x x + M_y y + M_z z \\ \lambda_x &= M_y z - M_z y - i\hbar x \\ \lambda_y &= M_z x - M_x z - i\hbar y \\ \lambda_z &= M_x y - M_y x - i\hbar z.\end{aligned}$$

We have

$$\begin{aligned}M_x \lambda_x + M_y \lambda_y + M_z \lambda_z &= \Sigma_{xyz} (M_x M_y z - M_x M_z y - i\hbar M_x x) \\ &= \Sigma_{xyz} (M_x M_y - M_y M_x - i\hbar M_z) z = 0 \quad (37)\end{aligned}$$

from the general condition (11) for angular momentum. Thus λ_x , λ_y , and λ_z are not linearly independent functions of x , y , and z . Any two of them, however, together with λ_0 are three linearly independent functions of x , y , and z and may be taken as the above λ_1 , λ_2 , and λ_3 , since the coefficients M_x , M_y , M_z all commute with K . Our problem thus reduces to finding the condition that the (K', K'') matrix elements of λ_0 , λ_x , λ_y , and λ_z shall not vanish. The physical meanings of these λ 's are that λ_0 is proportional to the component of the vector (x, y, z) in the direction of the vector \mathbf{M} and λ_x , λ_y , λ_z are proportional to the Cartesian components of the component of (x, y, z) perpendicular to \mathbf{M} .

From (4) together with the condition that x , y , and z commute with μ we obtain

$$\left. \begin{aligned} [M_z, x] &= [m_z + \mu_z, x] = y \\ [M_z, y] &= -x \quad [M_z, z] = 0. \end{aligned} \right\} (38)$$

Hence

$$\begin{aligned} [M_z, \lambda_0] &= [M_z, M_x]x + M_x[M_z, x] + [M_z, M_y]y + M_y[M_z, y] \\ &= M_yx + M_xy - M_xy - M_yx = 0. \end{aligned}$$

Thus λ_0 commutes with M_z , and from symmetry it must commute also with M_x and M_y , so that it must commute with K . It follows that only the diagonal elements $(K'| \lambda_0 | K')$ of λ_0 can differ from zero, so the selection rule is that K cannot change so far as this component of the electric displacement is concerned.

With further applications of (38) we obtain

$$\begin{aligned} [M_z, \lambda_x] &= [M_z, M_y]z - M_z[M_z, y] - i\hbar[M_z, x] \\ &= -M_xz + M_zx - i\hbar y = \lambda_y \\ [M_z, \lambda_y] &= M_z[M_z, x] - [M_z, M_x]z - i\hbar[M_z, y] \\ &= M_zy - M_yz + i\hbar x = -\lambda_x \\ [M_z, \lambda_z] &= [M_z, M_x]y + M_x[M_z, y] - [M_z, M_y]x - M_y[M_z, x] \\ &= M_yy - M_xx + M_xx - M_yy = 0. \end{aligned}$$

These relations between M_z and λ_x , λ_y , λ_z are of exactly the same form as the relations (4), (5) between m_z and x , y , z and also (37) is of the same form as (3). The observables λ_x , λ_y , λ_z thus have the same properties relative to the angular momentum \mathbf{M} that x , y , z have relative to \mathbf{m} . The deduction in § 47 of the selection rule for k when the electric displacement is proportional to (x, y, z) can therefore be taken over and applied to the selection rule for K when the electric dis-

placement is proportional to $(\lambda_x, \lambda_y, \lambda_z)$. We find in this way that, so far as $\lambda_x, \lambda_y, \lambda_z$ are concerned, the selection rule for K is that it must change by $\pm \hbar$.

Collecting results, we have as the selection rule for K that it must change by 0 or $\pm \hbar$. We have considered the electric displacement produced by only one of the electrons, but the same selection rule must hold for each of them and thus also for the total electric displacement.

IX

PERTURBATION THEORY

§ 50. General Remarks

In the preceding two chapters exact treatments were given of some simple dynamical systems in the quantum theory. Most quantum problems, however, cannot be solved exactly with the present resources of mathematics, as they lead to equations whose solutions cannot be expressed in finite terms with the help of the ordinary functions of analysis. For such problems one must use a perturbation method. This consists in splitting up the Hamiltonian into two parts, one of which must be simple and the other small. The first part may then be considered as the Hamiltonian of a simplified or *unperturbed* system, which can be dealt with exactly, and the addition of the second will then require small corrections, of the nature of a perturbation, in the solution for the unperturbed system. If this second part contains a small numerical factor ϵ , we can obtain the solution of our equations for the perturbed system in the form of a power series in ϵ , which, provided it converges, will give the answer to our problem with any desired accuracy. Even when the series does not converge, the first approximation obtained by means of it is usually fairly accurate.

There are two distinct methods in perturbation theory. In one of these the perturbation is considered as causing a *modification of the states* of the unperturbed system. In the other we do not consider any modification to be made in the states of the unperturbed system, but we suppose that the perturbed system, instead of remaining in *one* of these states, is continually changing from one to another, or making transitions, under the influence of the perturbation. Which method is to be used in any particular case depends on the nature of the problem to be solved. The first method is useful usually only when both the Hamiltonian for the undisturbed system and the perturbing energy (the correction in this Hamiltonian) do not involve the time explicitly, and is then applied to the stationary states. It can then be used for calculating things that do not refer to any definite time, such as the energy-levels of the stationary states of the perturbed system, or, in the case of collision problems, the probability of scattering through a given angle. The second method must, on

the other hand, be used for solving all problems involving a consideration of time, such as those about the transient phenomena that occur when the perturbation is suddenly applied, or more generally problems in which the perturbation varies with the time in any way (*i.e.* in which the perturbing energy involves the time explicitly in an arbitrary way). Again, this second method must be used in collision problems, even though the perturbing energy does not here involve the time explicitly, if one wishes to calculate absorption and emission probabilities, since these probabilities, unlike a scattering probability, cannot be defined without reference to a state of affairs that varies with the time.

§ 51. The Change in the Energy-levels caused by a Perturbation

The first of the above-mentioned methods will now be applied to the calculation of the changes in the energy-levels of a system caused by a perturbation. The perturbing energy, like the Hamiltonian for the unperturbed system, must now not involve the time explicitly. Our problem has a meaning, of course, only provided the energy-levels of the unperturbed system are discrete and the differences between them are large compared with the changes in them caused by the perturbation. This fact results in the treatment of perturbation problems by the first method having some different features according to whether the energy-levels of the unperturbed system are discrete or continuous.

Let the Hamiltonian of the perturbed system be

$$H = H_0 + V, \quad (1)$$

H_0 being the Hamiltonian of the unperturbed system and V the small perturbing energy. By hypothesis each eigenvalue H' of H lies very close to one and only one eigenvalue H''_0 of H_0 . It is convenient to use the same number of primes to specify any eigenvalue of H and the eigenvalue of H_0 to which it lies very close. Thus we shall have H'' differing from H''_0 by a small quantity of order V and differing from H'_0 by a quantity that is not small unless $H'_0 = H''_0$. We must now take care always to use different numbers of primes to specify eigenvalues of H and H_0 which we do not want to lie very close together.

Let $\psi(H')$ be an eigen- ψ of H belonging to the eigenvalue H' , so that

$$H\psi(H') = H'\psi(H'). \quad (2)$$

This means that $\psi(H')$ denotes a stationary state of the perturbed system of energy H' . Again, let $\psi(H''_0)$ be an eigen- ψ of H_0 (at some particular time t) belonging to the eigenvalue H''_0 , so that

$$H_0 \psi(H''_0) = H''_0 \psi(H''_0). \quad (3)$$

This $\psi(H''_0)$ will denote a non-stationary state of the perturbed system, and indeed a different non-stationary state for each different value of the above t , but for the unperturbed system it will denote a stationary state of energy H''_0 .

Now suppose that for the unperturbed system there is only one stationary state for each energy-level H''_0 , i.e. the unperturbed system is non-degenerate. This requires that H_0 shall have only one independent eigen- ψ belonging to any eigenvalue H''_0 (which is a condition governing only the form of the observable H_0 and independent of whether we are considering the perturbed or the unperturbed system). From our assumption that the changes in the energy-levels caused by the perturbation are small compared with the differences of the energy-levels of the unperturbed system, there must be only one independent eigen- ψ of H belonging to any eigenvalue H' , so that the perturbed system is also non-degenerate. The fact that the perturbing energy V is small, or that H_0 (at time t) and H are two nearly equal observables, will require, not only that their eigenvalues are nearly equal, but also that corresponding eigen- ψ 's are nearly equal, apart from numerical factors. Thus we shall have

$$\psi(H') = c\psi(H'_0) + \psi_1, \quad (4)$$

where c is a number and ψ_1 is a small ψ -symbol. We may assume ψ_1 to be orthogonal to $\psi(H'_0)$, since if it were not so it could be expressed as the sum of two parts, one of which is orthogonal to $\psi(H'_0)$ while the other is a numerical multiple of $\psi(H'_0)$ which can be absorbed in the first term of the right-hand side of (4). We can now take $c = 1$, so that we have

$$\psi(H') = \psi(H'_0) + \psi_1, \quad (5)$$

where ψ_1 is small and orthogonal to $\psi(H'_0)$.

From (1), (2), and (5) we now obtain

$$\{H_0 + V\}\{\psi(H'_0) + \psi_1\} = H\psi(H') = H'\psi(H') = H'\{\psi(H'_0) + \psi_1\}.$$

With the help of (3), this gives

$$H'_0 \psi(H'_0) + H_0 \psi_1 + V\psi(H'_0) + V\psi_1 = H'\psi(H'_0) + H'\psi_1.$$

If we neglect the second-order term $V\psi_1$, this reduces to

$$\{H' - H'_0\}\psi(H'_0) + \{H' - H_0\}\psi_1 = V\psi(H'_0). \quad (6)$$

If we now multiply this equation throughout by $\phi(H'_0)$, the conjugate imaginary symbol to $\psi(H'_0)$, on the left, the second term will contribute nothing, since

$$\phi(H'_0)\{H' - H_0\}\psi_1 = \phi(H'_0)\{H' - H'_0\}\psi_1 = \{H' - H'_0\}\phi(H'_0)\psi_1 = 0,$$

on account of $\phi(H'_0)$ and ψ_1 being orthogonal. We shall thus be left with

$$H' - H'_0 = \phi(H'_0)V\psi(H'_0), \quad (7)$$

assuming $\phi(H'_0)$ and $\psi(H'_0)$ to be normalized.

This result gives us the first-order change in the energy-level of any state caused by the perturbation. It shows that *the first-order change in the energy-level is equal to the average value of the perturbing energy for the unperturbed stationary state*. When formulated in this way, this result in quantum perturbation theory is the same as in the classical theory and as in the old quantum mechanics of Bohr's theory. One can say alternatively that the first-order change in an energy-level is equal to the corresponding diagonal element of the matrix representing the perturbing energy in a representation in which the Hamiltonian for the unperturbed system is diagonal *i.e.* in a Heisenberg representation for the unperturbed system.

We must now consider the case when the unperturbed system is degenerate, so that there are several eigen- ψ 's of H_0 belonging to the same eigenvalue H'_0 . The perturbation may now, perhaps, be such that the perturbed system is non-degenerate, or that it is not so much degenerate as the unperturbed system. This means that each energy-level H'_0 of the unperturbed system gets split up by the perturbation into several energy-levels H' all lying close to H'_0 .* We shall now have that every eigen- ψ of H is approximately equal to an eigen- ψ of H_0 , but the converse, that every eigen- ψ of H_0 is approximately equal to an eigen- ψ of H , will not be true, as may be seen from the following argument. If ψ_a and ψ_b are two eigen- ψ 's of H_0 belonging to the same eigenvalue and are approximately equal respectively to two eigen- ψ 's of H belonging to two different eigen-

* To distinguish these energy-levels one from another we should require some more elaborate notation, since according to the present notation they must all be specified by the same number of primes, namely, by the number of primes specifying the energy-level of the unperturbed system from which they arise. For our present purposes, however, this more elaborate notation is not required.

values, then any linear combination of them, $a\psi_a + b\psi_b$, will also be an eigen- ψ of H_0 but will not be approximately equal to any eigen- ψ of H . The problem of finding which eigen- ψ 's of H_0 are approximately equal to eigen- ψ 's of H is the analogue of the problem of secular perturbations in classical mechanics.

Any eigen- ψ of H_0 belonging to the eigenvalue H'_0 is expressible as a linear combination of a complete set of such eigen- ψ 's. We shall choose such a set consisting of the simultaneous eigen- ψ 's, $\psi(H'_0\xi')$, of H_0 and a number of observables ξ that commute with H_0 and with each other and that together with H_0 form a complete commuting set of observables. Any eigen- ψ $\psi(H'_0)$ is now expressible in the form

$$\psi(H'_0) = \sum_{\xi'} \psi(H'_0\xi') (\xi' |),$$

where the coefficients $(\xi' |)$ are numbers forming a representative of $\psi(H'_0)$. Any eigen- ψ $\psi(H')$ of H , belonging to some eigenvalue H' that lies close to H'_0 , is approximately equal to some $\psi(H'_0)$ and is therefore of the form

$$\psi(H') = \sum_{\xi'} \psi(H'_0\xi') (\xi' |) + \psi_1, \quad (8)$$

where ψ_1 is small. As in the non-degenerate case, we may assume that ψ_1 is orthogonal to each $\psi(H'_0\xi')$, since if it is not it can be expressed as the sum of two parts, one of which is orthogonal to the $\psi(H'_0\xi')$'s while the other is a linear combination of them. We now obtain with the help of (1), (2), and (3)

$$\begin{aligned} \{H_0 + V\} \{\sum_{\xi'} \psi(H'_0\xi') (\xi' |) + \psi_1\} &= H\psi(H') = H'\psi(H') \\ &= H' \{\sum_{\xi'} \psi(H'_0\xi') (\xi' |) + \psi_1\} \end{aligned}$$

$$\text{or } \{H' - H'_0\} \sum_{\xi'} \psi(H'_0\xi') (\xi' |) + \{H' - H_0\}\psi_1 = V \sum_{\xi'} \psi(H'_0\xi') (\xi' |),$$

with neglect of the second-order term $V\psi_1$. If we multiply this equation throughout by $\phi(H'_0\xi'')$ on the left, we shall again have the term $\phi(H'_0\xi'')\{H' - H_0\}\psi_1$ vanishing and shall be left with

$$\{H' - H'_0\} (\xi'' |) = \sum_{\xi'} \phi(H'_0\xi'') V \psi(H'_0\xi') (\xi' |),$$

provided the $\psi(H'_0\xi')$ are normalized. This result is the same as

$$\{H' - H'_0\} (\xi' |) = \sum_{\xi'} (H'_0\xi' | V | H'_0\xi'') (\xi'' |), \quad (9)$$

where $(H'_0\xi' | V | H'_0\xi'')$ is an element of the matrix representing V in the (H_0, ξ) -representation.

Equation (9) is of the form of the standard equation of the theory of eigenvalues. It shows that $H' - H'_0$ is an eigenvalue of the matrix

$(H'_0 \xi' | V | H'_0 \xi'')$. This matrix is a part of the representative of the perturbing energy V in a Heisenberg representation for the unperturbed system, namely the part consisting of those elements that refer to the same unperturbed energy level H'_0 for their row and column. Each change of the energy-level H'_0 caused by the perturbation is an eigenvalue of this matrix and further the eigenfunctions, namely the quantities $(\xi'|)$, are just the coefficients required in (8) to give us those linear functions of the eigen- ψ 's of H_0 belonging to the eigenvalue H'_0 that are approximately eigen- ψ 's of H and approximately represent stationary states of the perturbed system. We have thus obtained to the first order the energy-levels and stationary states of the perturbed system. It should be noticed that these first-order results are independent of the values of all those matrix elements of the perturbing energy which refer to two different energy-levels H'_0 and H''_0 of the unperturbed system.

One can use this perturbation method for the calculation of the higher approximations if required. General recurrence formulas giving the n -th order corrections in terms of those of lower order have been obtained by Born, Heisenberg, and Jordan.*

§ 52. The Perturbation considered as causing Transitions

We shall now consider the second of the two perturbation methods mentioned in § 50. We suppose again that we have an unperturbed system governed by a Hamiltonian H_0 which does not involve the time explicitly, and a perturbing energy V which can now be an arbitrary function of the time. The Hamiltonian for the perturbed system is again $H = H_0 + V$. For the present method it does not make any essential difference whether the energy-levels of the unperturbed system, *i.e.* the eigenvalues of H_0 , form a discrete or continuous set. We shall, however, take the discrete case, for definiteness.

We introduce an α -representation in which a complete set of commuting observables α are diagonal, each of which is the value at time t of some dynamical variable that is a constant of the motion for the unperturbed system. This means that H_0 at time t commutes with each of the α 's and is thus represented by a diagonal matrix

$$(\alpha' | H_0 | \alpha'') = H'_0 \delta_{\alpha' \alpha''}. \quad (10)$$

* *Zeits. f. Physik*, vol. xxxv, p. 565 (1925).

If the phases of the representation are such that the Schrödinger equation holds, we have, using stars to distinguish the representatives in this case,

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (\alpha' |)^* &= \sum_{\alpha''} (\alpha' | H_0 + V | \alpha'')^* (\alpha'' |)^* \\ &= H'_0 (\alpha' |)^* + \sum_{\alpha''} (\alpha' | V | \alpha'')^* (\alpha'' |)^*. \end{aligned} \quad (11)$$

For our present purpose, however, it is more convenient to choose these phases to be those of the Heisenberg representation for the undisturbed system, so that our representative $(\alpha' |)$ of a state is connected with the Schrödinger one $(\alpha' |)^*$ by the relation

$$(\alpha' |)^* = e^{-iH'_0 t/\hbar} (\alpha' |), \quad (12)$$

which was obtained at the end of § 38. The two representatives of an observable will be connected in the same way by

$$(\alpha' | \xi | \alpha'')^* = e^{-i(H'_0 - H''_0)t/\hbar} (\alpha' | \xi | \alpha'').$$

The representative (10) of H_0 is, of course, the same in either case, since it is diagonal.

Our new representative $(\alpha' |)$ does not satisfy the Schrödinger equation, of course, but satisfies instead the following equation, obtained by substituting (12) in (11),

$$\begin{aligned} i\hbar \left[-iH'_0/\hbar \cdot e^{-iH'_0 t/\hbar} (\alpha' |) + e^{-iH'_0 t/\hbar} \frac{\partial}{\partial t} (\alpha' |) \right] &= \\ &= H'_0 e^{-iH'_0 t/\hbar} (\alpha' |) + \sum_{\alpha''} (\alpha' | V | \alpha'')^* e^{-iH''_0 t/\hbar} (\alpha'' |), \end{aligned}$$

which reduces to

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (\alpha' |) &= \sum_{\alpha''} e^{i(H'_0 - H''_0)t/\hbar} (\alpha' | V | \alpha'')^* (\alpha'' |) \\ &= \sum_{\alpha''} (\alpha' | V | \alpha'') (\alpha'' |). \end{aligned} \quad (13)$$

The Schrödinger representative $(\alpha' | V | \alpha'')^*$ of the perturbing energy V does not depend on t , except in so far as V itself involves t explicitly, while the representative $(\alpha' | V | \alpha'')$ appearing in our equation (13) varies rapidly with t , according to the Heisenberg law $e^{i(H'_0 - H''_0)t/\hbar}$ when one neglects the explicit dependence of V on t .

Equation (13) is the fundamental equation of the present method in perturbation theory. It is an exact equation, no use having yet been made of the fact that the perturbation is small. It shows how the representative of a state of a perturbed system varies with the time when the representation is chosen so that the whole of this variation is caused by the perturbation, and thus expresses most

clearly the way in which the perturbation may be considered as causing a continual change in the state of the system. At any instant the probability of the α 's having specified values α' is

$$P' = |(\alpha')|^2 \quad (14)$$

provided P' is normalized.

We shall now obtain an approximate solution to equation (13) for a given initial value of the representative $(\alpha'|)$ of the state. Since V is small, the rate of change of $(\alpha'|)$ is small and $(\alpha'|)$ remains approximately equal to its initial value, at any rate for times that do not differ too much from the initial time. We can thus obtain a first approximation by substituting for $(\alpha''|)$ in the right-hand side of (13) its initial value and then performing a simple integration. We may then obtain a second approximation by substituting the first approximation in the right-hand side of (13), and so on indefinitely.

Let the initial value of $(\alpha'|)$, i.e. the value at time $t = 0$, be $a'_0(\alpha')$, or a'_0 say, for brevity. We shall then have in the first approximation for the value of $(\alpha'|)$ at an arbitrary time τ ,

$$\begin{aligned} (\alpha'|)_\tau &= a'_0 - i/\hbar \cdot \sum_{\alpha''} \int_0^\tau (\alpha' | V | \alpha'')_t a''_0 dt \\ &= a'_0 + a'_{1\tau} \end{aligned}$$

say, $a'_{1\tau}$ being the first-order correction, whose value at time τ is

$$a'_{1\tau} = -i/\hbar \sum_{\alpha''} a''_0 \int_0^\tau (\alpha' | V | \alpha'')_t dt. \quad (15)$$

The second approximation at an arbitrary time T will now be

$$\begin{aligned} (\alpha'|)_T &= a'_0 - i/\hbar \cdot \sum_{\alpha''} \int_0^T (\alpha' | V | \alpha'')_\tau [a''_0 + a''_{1\tau}] d\tau \\ &= a'_0 + a'_{1T} + a'_{2T}, \end{aligned}$$

where a'_{2T} , the second-order correction, has the value at time T

$$\begin{aligned} a'_{2T} &= -i/\hbar \cdot \sum_{\alpha''} \int_0^T (\alpha' | V | \alpha'')_\tau a''_{1\tau} d\tau \\ &= -1/\hbar^2 \cdot \sum_{\alpha'' \alpha'''} a'''_0 \int_0^T (\alpha' | V | \alpha'')_\tau d\tau \int_0^\tau (\alpha'' | V | \alpha''')_t dt. \quad (16) \end{aligned}$$

The probability (14) of the α 's having the values α' at any time is now, to the second order of accuracy,

$$\begin{aligned} P' &= (a'_0 + a'_{1T} + a'_{2T})(\bar{a}'_0 + \bar{a}'_{1T} + \bar{a}'_{2T}) \\ &= a'_0 \bar{a}'_0 + (a'_1 \bar{a}'_0 + a'_0 \bar{a}'_1) + (a'_2 \bar{a}'_0 + a'_1 \bar{a}'_1 + a'_0 \bar{a}'_2) + \dots \\ &= P'_0 + P'_1 + P'_2 + \dots, \end{aligned} \quad (17)$$

P'_0 being the initial value of this probability and P'_1 and P'_2 being the first and second order corrections.

Suppose now that we are given, not the initial value a'_0 of $(\alpha'|)$, but only the initial probability P'_0 of the α 's having any specified values α' , and want to calculate the probability at any subsequent time of the α 's having specified values. We now know only the modulus of $(\alpha'|)$ and not its phase, so that we must average over all phases. This averaging results in a considerable simplification in the expression (17) for P' , since this expression is bilinear in a_0 and \bar{a}_0 [both a_1 and a_2 being linear functions of a_0 according to (15) and (16)], and thus consists of a sum of terms of the form $a''_0 \bar{a}'''_0$. The average of $a''_0 \bar{a}'''_0$ or $a_0(\alpha'') \bar{a}_0(\alpha''')$ will vanish except when $\alpha''' = \alpha''$, so that the only surviving terms will be those of the form $a''_0 \bar{a}''_0$. In this way P'_1 reduces to

$$\begin{aligned} P'_{1\tau} &= a'_{1\tau} \bar{a}'_0 + a'_0 \bar{a}'_{1\tau} \\ &= \left[-i/\hbar \cdot a'_0 \int_0^\tau (\alpha' | V | \alpha')_t dt \right] \bar{a}'_0 + a'_0 \left[i/\hbar \cdot \bar{a}'_0 \int_0^\tau (\alpha' | V | \alpha')_t dt \right] \\ &= 0. \end{aligned}$$

Similarly P'_2 reduces to

$$\begin{aligned} P'_{2T} &= a'_{2T} \bar{a}'_0 + a'_{1T} \bar{a}'_{1T} + a'_0 \bar{a}'_{2T} \\ &= -1/\hbar^2 \cdot a'_0 \bar{a}'_0 \sum_{\alpha''} \int_0^T (\alpha' | V | \alpha'')_\tau d\tau \int_0^\tau (\alpha'' | V | \alpha')_t dt + \\ &\quad + 1/\hbar^2 \cdot \sum_{\alpha''} a''_0 \bar{a}''_0 \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2 - \\ &\quad - 1/\hbar^2 \cdot a'_0 \bar{a}'_0 \sum_{\alpha''} \int_0^T (\alpha'' | V | \alpha')_\tau d\tau \int_0^\tau (\alpha' | V | \alpha'')_t dt, \end{aligned}$$

use being made, in dealing with the third term, of the fact that the matrix $(\alpha' | V | \alpha'')$ is Hermitian. If we interchange t and τ in this third term, we can combine it with the first term to give

$$\begin{aligned} &- |a'_0|^2 / \hbar^2 \cdot \sum_{\alpha''} \left[\int_0^T d\tau \int_0^\tau dt + \int_0^T dt \int_0^\tau d\tau \right] (\alpha' | V | \alpha'')_\tau (\alpha'' | V | \alpha')_t = \\ &= - |a'_0|^2 / \hbar^2 \cdot \sum_{\alpha''} \int_0^T d\tau \int_0^T dt (\alpha' | V | \alpha'')_\tau (\alpha'' | V | \alpha')_t \\ &\quad |a'_0|^2 / \hbar^2 \cdot \sum_{\alpha''} \left| \int_0^\tau (\alpha' | V | \alpha'')_t dt \right|^2. \end{aligned}$$

Thus our expression for P'_2 becomes

$$\begin{aligned} P'_{2T} &= 1/\hbar^2 \cdot \sum_{\alpha''} \{ |a''_0|^2 - |a'_0|^2 \} \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2 \\ &= 1/\hbar^2 \cdot \sum_{\alpha''} \{ P''_0 - P'_0 \} \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2, \end{aligned}$$

and the probability P' of the α 's having the values α' is, to the second order of accuracy,

$$P'_T = P'_0 + 1/\hbar^2 \cdot \sum_{\alpha''} \{ P''_0 - P'_0 \} \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2. \quad (18)$$

This result is capable of a simple interpretation. If we suppose that initially the α 's certainly have the values α'' , so that $P''_0 = 1$, $P'_0 = 0$ for $\alpha' \neq \alpha''$, (in which special case the averaging over the phases of the a_0 's is not necessary), then the right-hand side of (18) reduces to the single term

$$1/\hbar^2 \cdot \left| \int_0^T (\alpha' | V | \alpha'')_t dt \right|^2 = P(\alpha'' \alpha') \quad (19)$$

say. This may be interpreted as the probability of the system making a transition from the state α'' to the state α' under the influence of the perturbation V during the interval of time 0 to T . It is symmetrical between α' and α'' . Returning now to the general case, we see that (18) may be regarded as expressing that the change in the probability of the α 's having the values α' during the time interval 0 to T , namely $P'_T - P'_0$, is made up of the total probability $\sum_{\alpha''} P''_0 P(\alpha'' \alpha')$ of the system jumping into the state α' from some other state α'' , minus the total probability $P'_0 \sum_{\alpha''} P(\alpha' \alpha'')$ of its jumping out of the state α' , during this time interval. Thus the ordinary laws of probability apply, showing that there is no interference between the different transition processes. If we had not averaged over the initial phases, then there would have been such interference.

The integrand in (19) is the representative in a certain representation of the perturbing energy at time t . This representation is one that does not depend very much on t , since if we put $V = 0$ it would become the Heisenberg representation and would not depend on t at all. Hence we can, without spoiling the order of accuracy of our result, replace the integral in (19) by $(\alpha' | \int_0^T V_t dt | \alpha'')$ and obtain an alternative expression for the transition probability

$$P(\alpha'' \alpha') = 1/\hbar^2 \cdot \left| \left(\alpha' | \int_0^T V_t dt | \alpha'' \right) \right|^2. \quad (20)$$

This provides a simple physical meaning for the non-diagonal elements of the matrix representing an observable when this observable can be regarded as the time integral of a perturbing energy.

§ 53. Application to Radiation

In the preceding section a general theory of the perturbation of an atomic system was developed, in which the perturbing energy could vary with the time in an arbitrary way. A perturbation of this kind can be realized in practice by allowing incident electromagnetic radiation to fall on the system. Let us see what our result (19) or (20) reduces to in this case.

If we neglect the effects of the magnetic field of the incident radiation, and if we further assume that the wave-lengths of the harmonic components of this radiation are all large compared with the dimensions of the atomic system, then the perturbing energy is simply the scalar product

$$V = (\mathbf{D}, \mathcal{E}), \quad (21)$$

where \mathbf{D} is the total electric displacement of the system and \mathcal{E} is the electric force of the incident radiation. We suppose \mathcal{E} to be a given function of the time. If we take for simplicity the case when the incident radiation is plane polarized with its electric vector in a certain direction and let D denote the Cartesian component of \mathbf{D} in this direction, the expression (21) for V reduces to the ordinary product

$$V = D\mathcal{E},$$

where \mathcal{E} is the magnitude of the vector \mathcal{E} . The matrix elements of V are

$$(\alpha' | V | \alpha'') = (\alpha' | D | \alpha'') \mathcal{E},$$

since \mathcal{E} is a number. Now $(\alpha' | D | \alpha'')$ varies with the time t according to the Heisenberg law

$$(\alpha' | D | \alpha'') = (\alpha' | D | \alpha'')_0 e^{i(H'_0 - H''_0)t/\hbar},$$

$(\alpha' | D | \alpha'')_0$ being constant, and hence our expression (19) for the transition probability becomes

$$P(\alpha' \alpha'') = 1/\hbar^2 \cdot |(\alpha' | D | \alpha'')|^2 \left| \int_0^T \mathcal{E}_t e^{i(H'_0 - H''_0)t/\hbar} dt \right|^2. \quad (22)$$

If the incident radiation during the time interval 0 to T is resolved into its Fourier components, the energy crossing unit area per unit

frequency range about the frequency ν will be, according to classical electrodynamics,

$$E_\nu = c/2\pi \cdot \left| \int_0^T \mathcal{E}_t e^{2\pi i \nu t} dt \right|^2. \quad (23)$$

Comparing this with (22), we see that the transition probability between two states α' and α'' with energies H'_0 and H''_0 depends on that Fourier component of the incident radiation whose frequency is $\nu = |H'_0 - H''_0|/h$, in agreement with Bohr's theory. The magnitude of this transition probability is connected with the intensity of the Fourier component through the relation

$$P(\alpha' \alpha'') = 2\pi/c\hbar^2 \cdot |(\alpha' | D | \alpha'')|^2 E_\nu. \quad (24)$$

This relation gives the probability of the system, if initially in the state of lower energy, of absorbing radiation and being carried to the upper state, and if initially in the upper state, of being *stimulated* by the incident radiation to emit and fall to the lower state. The present theory does not account for the fact that the system, if in the upper state with no incident radiation, can emit spontaneously and fall to the lower state.

The existence of the phenomenon of stimulated emission was inferred by Einstein,* long before the discovery of quantum mechanics, from a consideration of thermodynamic equilibrium between atoms and a field of black-body radiation satisfying Planck's law. Einstein showed that the transition probability for stimulated emission must equal that for absorption between the same pair of states and deduced a relation connecting this transition probability with that for spontaneous emission. Heisenberg's assumption for the spontaneous emission probability, given in § 38, together with Einstein's theory, will therefore provide us with values for the transition probabilities for absorption and stimulated emission. These values are in agreement with (24). Thus the theory of the present section gives a partial justification for Heisenberg's assumption. The complete justification will be provided by the general theory of Chapter XII, in which the electromagnetic field will be treated as a dynamical system interacting with the atom according to the laws of quantum mechanics. This general theory will not only confirm the result (24) for absorption and stimulated emission, but will also give the required value for the spontaneous emission probability.

* Einstein, *Phys. Zeits.*, vol. xviii, p. 121 (1917).

§ 54. Transitions caused by a Perturbation Independent of the Time

The perturbation method of § 52 is still valid when the perturbing energy V does not involve the time t explicitly. Since the total Hamiltonian H in this case does not involve t explicitly, we could now, if desired, deal with the system by the perturbation method of § 51 and find its stationary states. Whether this method would be convenient or not would depend on what we want to find out about the system. If what we have to calculate makes an explicit reference to the time, *e.g.* if we have to calculate the wave function at one time when we are given its value at another time, the method of § 52 would be the more convenient one.

Let us see what the result (19) for the transition probability becomes when V does not involve t explicitly. The matrix element $(\alpha' | V | \alpha'')$ now varies with t according to the Heisenberg law and thus its time integral is

$$\begin{aligned} \int_0^T (\alpha' | V | \alpha'')_t dt &= (\alpha' | V | \alpha'')_0 \int_0^T e^{i(H'_0 - H''_0)t/\hbar} dt \\ &= (\alpha' | V | \alpha'')_0 \frac{e^{i(H'_0 - H''_0)T/\hbar} - 1}{i(H'_0 - H''_0)/\hbar}, \end{aligned}$$

provided $H'_0 \neq H''_0$. Thus the transition probability (19) becomes

$$\begin{aligned} P(\alpha' \alpha'') &= |(\alpha' | V | \alpha'')|^2 [e^{i(H'_0 - H''_0)T/\hbar} - 1] [e^{-i(H'_0 - H''_0)T/\hbar} - 1] / (H'_0 - H''_0)^2 \\ &= 2|(\alpha' | V | \alpha'')|^2 [1 - \cos\{(H'_0 - H''_0)T/\hbar\}] / (H'_0 - H''_0)^2. \quad (25) \end{aligned}$$

If H''_0 differs appreciably from H' this transition probability is small and remains so for all values of T . This result is required by the law of the conservation of energy. The total energy H is constant and hence the proper-energy H_0 (*i.e.* the energy with neglect of the part V due to the perturbation), being approximately equal to H , must be approximately constant. This means that if H_0 initially has the numerical value H'_0 , at any later time there must be only a small probability of its having a numerical value differing considerably from H'_0 .

On the other hand, when the initial state α' is such that there exists another state α'' having the same or very nearly the same proper-energy H_0 , the probability of a transition to the final state α'' may be quite large. The case of physical interest now is that in which there is a continuous range of final states α'' having a con-

tinuous range of proper-energy levels H''_0 passing through the value H'_0 of the proper-energy of the initial state. The initial state must not be one of the continuous range of final states, but may be either a separate discrete state or one of another continuous range of states. We shall now have, remembering the rules of § 28 for the interpretation of probability amplitudes with continuous ranges of states, that, with $P(\alpha'\alpha'')$ having the value (25), the probability of a transition to a final state within the small range α'' to $\alpha''+d\alpha''$ will be $P(\alpha'\alpha'') d\alpha''$ when the initial state α' is discrete and will be proportional to this quantity when α' is one of a continuous range.

We may suppose that the α 's describing the final state, which are any complete set of commuting dynamical variables that all commute with H_0 , consist of H_0 itself together with a number of other dynamical variables β . (The β 's need have no meaning for the initial state α' .) We shall suppose for definiteness that the β 's have only discrete eigenvalues. The total probability of a transition to a final state α'' for which the β 's have the values β'' and H_0 has any value, (there will be a strong probability of its having a value near the initial value H'_0 ,) will now be (or be proportional to)

$$\begin{aligned} \int P(\alpha'\alpha'') dH''_0 &= \\ &= 2 \int_{-\infty}^{\infty} |(\alpha'| V | H'_0 \beta'')|^2 [1 - \cos\{(H'_0 - H''_0)T/\hbar\}] / (H'_0 - H''_0)^2 \cdot dH''_0 \quad (26) \\ &= 2T/\hbar \cdot \int_{-\infty}^{\infty} |(\alpha'| V | H'_0 + \hbar x/T, \beta'')|^2 [1 - \cos x] / x^2 \cdot dx \end{aligned}$$

if one makes the substitution $(H''_0 - H'_0)T/\hbar = x$. For large values of T this reduces to

$$\begin{aligned} 2T/\hbar \cdot |(\alpha'| V | H'_0 \beta'')|^2 \int_{-\infty}^{\infty} [1 - \cos x] / x^2 \cdot dx \\ = 2\pi T/\hbar \cdot |(\alpha'| V | H'_0 \beta'')|^2. \quad (27) \end{aligned}$$

Thus the total probability up to time T of a transition to a final state for which the β 's have the values β'' is proportional to T . There is therefore a definite probability coefficient, or probability per unit time, for the transition process under consideration, having the value

$$2\pi/\hbar \cdot |(\alpha'| V | H'_0 \beta'')|^2. \quad (28)$$

It is proportional to the square of the modulus of the matrix element, associated with this transition, of the perturbing energy.

In order that the approximations used in deriving (27) may be

valid, the time T must be not too small and not too large. It must be large compared with the periods of the atomic system in order that the evaluation of the integral (26) leading to the result (27) may be correct, while it must not be excessively large or else the general formula (19) will break down. In fact one could make the probability (27) greater than unity by taking T large enough. The upper limit to T is fixed by the condition that the probability (19) or (27) must be small compared with unity. There is no difficulty in T satisfying both these conditions simultaneously provided the perturbing energy V is sufficiently small.

§ 55. The Anomalous Zeeman Effect

One of the simplest examples of the perturbation method of § 51 is the calculation of the change in the energy-levels of a general atom caused by a uniform magnetic field. The problem of a hydrogen atom in a uniform magnetic field has already been dealt with in § 48 and was so simple that perturbation theory was unnecessary. The case of a general atom is not much more complicated when we make a few approximations such that we can set up a simple model for the atom.

We first of all consider the atom in the absence of the magnetic field along the lines indicated in § 49 and look for angular momenta that are constants of the motion. The total angular momentum of the atom, the vector \mathbf{j} say, is certainly a constant of the motion. This angular momentum may be regarded as the sum of two parts, the total orbital angular momentum of all the electrons, \mathbf{l} say, and the total spin angular momentum, \mathbf{s} say. Thus we have $\mathbf{j} = \mathbf{l} + \mathbf{s}$. Now the effect of the spin magnetic moments on the motion of the electrons is small compared with the effect of the Coulomb forces and may be neglected as a first approximation. With this approximation the spin angular momentum of each electron is a constant of the motion, there being no forces tending to change its orientation. Thus \mathbf{s} , and hence also \mathbf{l} , will be constants of the motion. We now have the three constant angular momenta \mathbf{l} , \mathbf{s} , and \mathbf{j} , related in the same way as the \mathbf{m} , μ , and \mathbf{M} of § 49. The magnitudes, l , s , and j say, of these angular momenta will be given by

$$\begin{aligned} l &= (l_x^2 + l_y^2 + l_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}} \\ s &= (s_x^2 + s_y^2 + s_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}} \\ j &= (j_x^2 + j_y^2 + j_z^2 + \frac{1}{4}\hbar^2)^{\frac{1}{2}}, \end{aligned}$$

corresponding to equation (12) of Chapter VIII, and from (36) of that chapter we see that with given numerical values for l and s the possible numerical values for j are

$$l+s-\frac{1}{2}\hbar, \quad l+s-\frac{3}{2}\hbar, \quad \dots |l-s|+\frac{1}{2}\hbar.$$

Let us consider a stationary state for which l , s , and j have definite numerical values in agreement with the above scheme. The energy of this state will depend on l , but one might think that with neglect of the spin magnetic moments it would be independent of s , and also of the direction of the vector \mathbf{s} relative to \mathbf{l} , and thus of j . It will be found in Chapter XI, however, that the energy depends very much on the magnitude s of the vector \mathbf{s} , although independent of its direction when one neglects the spin magnetic moments, on account of certain phenomena arising from the fact that the electrons are indistinguishable one from another. There are thus different energy-levels of the system for each different value of l and s . This means that l and s are functions of the energy, according to the general definition of a function given in § 15, since the l and s of a stationary state are fixed when the energy of that state is fixed.

We can now take into account the effect of the spin magnetic moments, treating it as a small perturbation according to the method of § 51. The energy of the unperturbed system will still be approximately a constant of the motion and hence l and s , being functions of this energy, will still be approximately constants of the motion. The directions of the vectors \mathbf{l} and \mathbf{s} , however, not being functions of the unperturbed energy, need not now be approximately constants of the motion and may undergo large secular variations. Since the vector \mathbf{j} is constant, the only possible variation of \mathbf{l} and \mathbf{s} is a precession about the vector \mathbf{j} . We thus have an approximate model of the atom consisting of the two vectors \mathbf{l} and \mathbf{s} of constant lengths precessing about their sum \mathbf{j} , which is a fixed vector. The energy is determined mainly by the magnitudes of \mathbf{l} and \mathbf{s} and depends only slightly on their relative directions, specified by j . Thus states with the same l and s and different j will have only slightly different energy-levels, forming what is called a *multiplet* term.

Let us now suppose our atom to be subjected to a uniform magnetic field of magnitude \mathcal{H} in the direction of the z -axis. The extra energy due to this magnetic field will consist of a term

$$e\mathcal{H}/2mc \cdot (m_z + \hbar\sigma_z), \quad (29)$$

like the last term in equation (34) of Chapter VIII, contributed by each electron, and will thus be altogether

$$e\mathcal{H}/2mc \cdot \Sigma (m_z + \hbar\sigma_z) = e\mathcal{H}/2mc \cdot (l_z + 2s_z) = e\mathcal{H}/2mc \cdot (j_z + s_z). \quad (30)$$

This is our perturbing energy V . We shall now use the method of § 51 to determine the changes in the energy-levels caused by this V . The method will be legitimate only provided the field is so weak that V is small compared with the energy differences within a multiplet.

Our unperturbed system is degenerate, on account of the direction of the vector \mathbf{j} being undetermined. We must therefore take, from the representative of V in a Heisenberg representation for the unperturbed system, those matrix elements that refer to one particular energy-level for their row and column, and obtain the eigenvalues of the matrix thus formed. We can do this best by first splitting up V into two parts, one of which is a constant of the unperturbed motion, so that its representative contains only matrix elements referring to the same unperturbed energy-level for their row and column, while the representative of the other contains only matrix elements referring to two different unperturbed energy-levels for their row and column, so that this second part does not affect the first-order perturbation. The term involving j_z in (30) is a constant of the unperturbed motion and thus belongs entirely to the first part. For the term involving s_z we have

$$s_z(j_x^2 + j_y^2 + j_z^2) = j_z(s_x j_x + s_y j_y + s_z j_z) + (s_z j_x - j_z s_x) j_x + (s_z j_y - j_z s_y) j_y$$

or

$$s_z = \frac{j_z}{j^2 - \frac{1}{4}\hbar^2} \left[(j^2 - \frac{1}{4}\hbar^2) - (l^2 - \frac{1}{4}\hbar^2) + (s^2 - \frac{1}{4}\hbar^2) \right] + \frac{1}{j^2 - \frac{1}{4}\hbar^2} [\gamma_y j_x - \gamma_x j_y], \quad \dots (31)$$

where

$$\left. \begin{aligned} \gamma_x &= s_z j_y - j_z s_y = s_z l_y - l_z s_y = l_y s_z - l_z s_y \\ \gamma_y &= j_z s_x - s_z j_x = l_z s_x - s_z l_x = l_z s_x - l_x s_z. \end{aligned} \right\} \quad (32)$$

The first term in this expression for s_z is a constant of the unperturbed motion and thus belongs entirely to the first part, while the second term, as we shall now see, belongs entirely to the second part.

Corresponding to (32) we can introduce

$$\gamma_z = l_x s_y - l_y s_x.$$

It can now easily be verified that

$$j_x \gamma_x + j_y \gamma_y + j_z \gamma_z = 0$$

and that

$$[j_z, \gamma_x] = \gamma_y, \quad [j_z, \gamma_y] = -\gamma_x, \quad [j_z, \gamma_z] = 0.$$

These relations are of the same form as the relations (3), (4), and (5) of Chapter VIII, so that our $\gamma_x, \gamma_y, \gamma_z$ are connected with the angular momentum \mathbf{j} in the same way in which the x, y, z of Chapter VIII were connected with the angular momentum \mathbf{m} . We can thus take over the analysis of § 47, in which the condition was obtained for the non-vanishing of a matrix element of x, y , and z in a representation in which k is diagonal. We find in this way that the only non-vanishing matrix elements of γ_x, γ_y , and γ_z in a representation in which j is diagonal are those referring to transitions in which j changes by $\pm \hbar$. The coefficients of γ_x and γ_y in the second term on the right-hand side of (31) commute with j , so that the representative of the whole of this term will contain only matrix elements referring to transitions in which j changes by $\pm \hbar$, and thus referring to two different energy-levels of the unperturbed system.

Hence the perturbing energy V becomes, when we neglect that part of it whose representative consists of matrix elements referring to two different unperturbed energy-levels,

$$\frac{e\mathcal{H}}{2mc} j_z \left\{ 1 + \frac{1}{2} \frac{(j^2 - \frac{1}{4}\hbar^2) - (l^2 - \frac{1}{4}\hbar^2) + (s^2 - \frac{1}{4}\hbar^2)}{j^2 - \frac{1}{4}\hbar^2} \right\}. \quad (33)$$

The eigenvalues of this give the first-order changes in the energy-levels. We can make the representative of this expression diagonal by choosing our representation such that j_z is diagonal, *i.e.* by taking the fundamental states to be spacially quantized in the z -direction. The expression (33) then gives us directly the first-order changes in the energy-levels caused by the magnetic field. This expression is known as Landé's formula.

The result (33) holds only provided the perturbing energy V is small compared with the energy differences within a multiplet. For larger values of V a more complicated theory is required. For very strong fields, however, for which V is large compared with the energy differences within a multiplet, the theory is again very simple. We may now neglect altogether the energy of the spin magnetic moments for the atom with no external field, so that for our unperturbed system the vectors \mathbf{I} and \mathbf{s} themselves are constants of the motion,

and not merely their magnitudes l and s . Our perturbing energy V , which is still $e\mathcal{H}/2mc \cdot (j_z + s_z)$, is now a constant of the motion for the unperturbed system, so that its eigenvalues give directly the changes in the energy-levels. These eigenvalues are integral or half-odd integral multiples of $e\mathcal{H}\hbar/2mc$ according to whether the number of electrons is even or odd.

X

COLLISION PROBLEMS

§ 56. General Remarks

In this chapter we shall investigate problems connected with a particle which, coming from infinity, encounters or ‘collides with’ some atomic system and, after being scattered through a certain angle, goes off to infinity again. The atomic system which does the scattering we shall call, for brevity, the *scatterer*. We thus have a dynamical system composed of an incident particle and a scatterer interacting with each other, which we must deal with according to the laws of quantum mechanics, and for which we must, in particular, calculate the probability of scattering through any given angle. This problem was first solved by Born by a method substantially equivalent to that of the next section. We must take into account the possibility that the scatterer, considered as a system by itself, may have a number of different stationary states and that if it is initially in one of these states when the particle arrives from infinity, it may be left in a different one when the particle goes off to infinity again. The colliding particle may thus induce transitions in the scatterer.

The Hamiltonian for the whole system of scatterer plus particle will not involve the time explicitly, so that this whole system will have stationary states represented by periodic solutions of Schrödinger’s wave equation. The meaning of these stationary states requires a little care to be properly understood. It is evident that for *any* state of the system the particle will spend nearly all its time at infinity, so that the time average of the probability of the particle being in any finite volume will be zero. Now for a *stationary* state the probability of the particle being in a given finite volume, like any other result of observation, must be independent of the time, and hence this probability will equal its time average, which we have seen is zero. We shall thus be interested only in the relative probabilities of the particle being in different finite volumes, their absolute values being all zero. Mathematically we have that if the ψ denoting a stationary state is normalized correctly for physical interpretation, *i.e.* such that $\phi\psi = 1$, and if we let Q denote that observable, which is a certain function of the position of the particle (at a given time), that is equal to unity if the particle is in a given finite volume and

zero otherwise, then $\phi Q\psi = 0$, meaning that the average value of Q , i.e. the probability of the particle being in the given volume, is zero. It would therefore be more convenient for us to denote the stationary state by a ψ normalized to infinity, i.e. for which $\phi\psi = \infty$, the infinity being such as to make $\phi Q\psi$ finite. This finite $\phi Q\psi$ would then give the relative probability of the particle being in the given volume.

In picturing a state of a system denoted by a ψ which is not normalized correctly for physical interpretation, but for which $\phi\psi = n$ say, it may be convenient to suppose that we have n similar systems all occupying the same space but with no interaction between them, so that each one follows out its own motion independently of the others. We can then interpret $\phi\alpha\psi$, where α is any observable, directly as the total α for all the n systems. In applying these ideas to the above-mentioned ψ normalized to infinity, denoting a stationary state of the system of scatterer plus colliding particle, we should picture an infinite number of such systems with the scatterers all located at the same point and the particles distributed continuously throughout space. The number of particles in a given finite volume would be pictured as $\phi Q\psi$, Q being the observable defined above, which has the value unity when the particle is in the given volume and zero otherwise. If the ψ is represented by a Schrödinger wave function involving the Cartesian co-ordinates of the particle, then the square of the modulus of the wave function could be interpreted directly as the density of particles in the picture. One must remember, however, that *each of these particles has its own individual scatterer*. Different particles may belong to scatterers in different states. There will thus be one particle density for each state of the scatterer, namely, the density of those particles belonging to scatterers in that state. This is taken account of by the wave function involving variables describing the state of the scatterer in addition to those describing the position of the particle.

For determining scattering coefficients we have to investigate *stationary states* of the whole system of scatterer plus particle. For instance, if we want to determine the probability of scattering in various directions when the scatterer is initially in a given stationary state and the incident particle has initially a given velocity in a given direction, we must investigate that stationary state of the whole system whose picture, according to the above method, contains at great distances from the point of location of the scatterers only

particles moving with the given initial velocity and direction and belonging each to a scatterer in the given initial stationary state, together with particles moving *outward* from the point of location of the scatterers and belonging possibly to scatterers in various stationary states. This picture corresponds closely to the actual state of affairs in an experimental determination of scattering coefficients, with the difference that the picture really describes only one *actual* system of scatterer plus particle. The distribution of outward moving particles at infinity in the picture gives us immediately all the information about scattering coefficients that could be obtained by experiment. For practical calculations about the stationary state described by this picture one may use the perturbation method of § 51, taking as unperturbed system, for example, that for which there is no interaction between the scatterer and particle.

In dealing with collision problems, a further possibility to be taken into consideration is that the scatterer may perhaps be capable of absorbing and re-emitting the particle. This possibility arises when there exists one or more *states of absorption* of the whole system, a state of absorption being an *approximately* stationary state which, at a certain time, is closed in the sense of § 45 (*i.e.* the probability of the particle being at a greater distance than r from the scatterer tends to zero as $r \rightarrow \infty$). Since a state of absorption is only approximately stationary, its property of being closed will be only a transient one and after a sufficient lapse of time there will be a finite probability of the particle being on its way to infinity. Physically this means there is a finite probability of spontaneous emission of the particle. The fact that we had to use the word ‘approximately’ in stating the conditions required for the phenomena of emission and absorption to be able to occur shows that these conditions are not expressible in exact mathematical language. One can give a meaning to these phenomena only when one is using a perturbation method. They occur when the unperturbed system (of scatterer plus particle) has stationary states that are closed. The perturbation now spoils the stationary property of these states and gives rise to spontaneous emission and its converse absorption.

For calculating absorption and emission probabilities it is necessary to deal with *non-stationary states* of the system, in contradistinction to the case for scattering coefficients, so that the perturbation method of § 52 must be used. Thus for calculating an emission coefficient

we must consider the non-stationary states of absorption described above. Again, since an absorption is always followed by a re-emission, it cannot be distinguished from a scattering in any experiment involving a steady state of affairs, corresponding to a stationary state of the system. The distinction can be made only by reference to a non-steady state of affairs, *e.g.* by use of a stream of incident particles that has a sharp beginning, so that the scattered particles will appear immediately after the incident particles meet the scatterers, while those that have been absorbed and re-emitted will begin to appear only some time later. This stream of particles would then be the picture of a certain non-stationary ψ , normalized to infinity, which could be used for obtaining the absorption coefficient.

§ 57. The Scattering Coefficient

We shall now consider the calculation of scattering coefficients, taking first the case when there is no absorption and emission, which means that our unperturbed system has no closed stationary states. We may conveniently take this unperturbed system to be that for which there is no interaction between the scatterer and particle. Its Hamiltonian will thus be of the form

$$H_0 = H_s + W, \quad (1)$$

where H_s is that for the scatterer alone and W that for the particle alone, namely

$$W = 1/2m \cdot (p_x^2 + p_y^2 + p_z^2). \quad (2)$$

The perturbing energy V , assumed small, will now be a function of the Cartesian co-ordinates of the particle x, y, z and also, perhaps, of its momenta p_x, p_y, p_z , together with dynamical variables describing the scatterer.

Since we are now interested only in *stationary* states of the whole system, we can use the perturbation method of § 51. Our unperturbed system now necessarily has a continuous range of energy-levels, since it contains a free particle, and this gives rise to certain modifications in the perturbation method. The question of the change in the energy-levels caused by the perturbation, which was the main question of § 51, no longer has a meaning, and the convention in § 51 of using the same number of primes to denote nearly equal eigenvalues of H_0 and H now drops out. Again the problem of secular perturbations cannot now arise, since if the unperturbed system is degenerate the perturbed one, which must also have a continuous

range of energy-levels, will also be degenerate to exactly the same extent. Any eigen- ψ of the unperturbed Hamiltonian H_0 , belonging to the eigenvalue H'_0 say, will be approximately equal to some eigen- ψ of H , and indeed to each of an infinity of eigen- ψ 's of H belonging to a small range of eigenvalues H' approximately equal to H'_0 . (The meaning of two ψ -symbols being approximately equal cannot be accurately defined in the case of continuous eigenvalues without a more rigorous theory than that aimed at in the present work. It should be noticed, though, that this meaning is such that two eigen- ψ 's of an observable belonging to two nearly equal eigenvalues may be approximately equal, in spite of the fact that they are orthogonal.)

We again express the stationary state $\psi(H')$ of the perturbed system as the sum of an eigen- ψ $\psi(H'_0)$ of the unperturbed Hamiltonian and a small correction ψ_1 . We can no longer, however, take ψ_1 to be orthogonal to $\psi(H'_0)$, as in equation (5) of § 51. The reason for this is that when we introduce our ψ_1 as in equation (4) of § 51 and then express this ψ_1 as the sum of two parts, one a numerical multiple of $\psi(H'_0)$, and the other orthogonal to $\psi(H'_0)$, these parts may both be large, in the case of continuous eigenvalues H'_0 , in spite of their sum being small. For example, these parts could be respectively of the form $\psi(H'_0)$ and $-\psi(H'_0 + \delta H'_0)$. Thus we cannot have our ψ_1 both small and orthogonal to $\psi(H'_0)$ and we prefer to have it small. To make up for this lack of simplicity in ψ_1 we can now take H'_0 exactly equal to H' . Let us call this number H'_0 or H' , equal to the energy of the stationary state we are seeking, E . We now have the equation

$$(E - H_0)\psi(H') = V\psi(H') \quad (3)$$

which gives

$$(E - H_0)\psi_1 = V\psi(H')$$

or

$$(E - H_s - W)\psi_1 = V\psi(H'_0) \quad (4)$$

from (1), with neglect of the second-order term $V\psi_1$. We shall use this equation (4) for determining the stationary states of the perturbed system to the first order.

Let α denote a complete set of commuting variables describing the scatterer, which are constants of the motion when the scatterer is alone, and may thus be used for labelling the stationary states of the scatterer. This requires that H_s shall commute with the α 's and be a function of them. We can now take a representation of the whole

system in which the α 's and x, y, z , the co-ordinates of the particle, are diagonal. This will make H_s diagonal. Let $\psi(H'_0)$ be represented by $(x\alpha|0)$ and ψ_1 by $(x\alpha|1)$, the single variable x being written in the wave function to denote x, y , and z . In the same way the single differential dx will be written to denote the product $dxdydz$. Equation (4), written in terms of representatives, becomes, with the help of (2),

$$\{E - H_s(\alpha') + \hbar^2/2m \cdot \nabla^2\}(x\alpha'|1) = \sum_{\alpha''} \int (x\alpha'|V|x''\alpha'') dx'' (x''\alpha''|0). \quad (5)$$

Suppose that the incident particle has the momentum \mathbf{p}^0 and that the initial stationary state of the scatterer is α^0 . The stationary state $\psi(H'_0)$ of our unperturbed system is now the one for which $\mathbf{p} = \mathbf{p}^0$ and $\alpha = \alpha^0$, and hence its representative is of the form

$$(x\alpha|0) = \delta_{\alpha\alpha^0} e^{i(\mathbf{p}^0, \mathbf{x})/\hbar}. \quad (6)$$

This makes equation (5) reduce to

$$\{E - H_s(\alpha') + \hbar^2/2m \cdot \nabla^2\}(x\alpha'|1) = \int (x\alpha'|V|x^0\alpha^0) dx^0 e^{i(\mathbf{p}^0, \mathbf{x}^0)/\hbar}$$

$$\text{or } \{k^2 + \nabla^2\}(x\alpha'|1) = F, \quad (7)$$

where

$$k^2 = 2m/\hbar^2 \cdot \{E - H_s(\alpha')\} \quad (8)$$

and

$$F = 2m/\hbar^2 \cdot \int (x\alpha'|V|x^0\alpha^0) dx^0 e^{i(\mathbf{p}^0, \mathbf{x}^0)/\hbar}, \quad (9)$$

a definite function of x, y, z , and α' . We must also have

$$E = H'_0 = H_s(\alpha^0) + \mathbf{p}^{02}/2m. \quad (10)$$

Our problem now is to obtain a solution $(x\alpha'|1)$ of (7) which, for values of x, y, z denoting points far from the scatterer, represents only outward moving particles. The square of its modulus, $|(x\alpha'|1)|^2$, will then give the density of scattered particles belonging to scatterers in the state α' when the density of the incident particles is $|(x\alpha|0)|^2$, which is unity. If we transform to polar co-ordinates r, θ, ϕ , equation (7) becomes

$$\left\{ k^2 + \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right\} (r\theta\phi\alpha'|1) = F. \quad (11)$$

Now F must tend to zero as $r \rightarrow \infty$, on account of the physical fact that the interaction energy between the scatterer and particle must tend to zero as the distance between them tends to infinity. If we neglect F in (11) altogether, an approximate solution for large r is

$$(r\theta\phi\alpha'|1) = u(\theta\phi\alpha')/r \cdot e^{ikr}, \quad (12)$$

where u is an arbitrary function of θ, ϕ , and α' , since this expression substituted in the left-hand side of (11) gives a result of order r^{-3} . When we do not neglect F , the solution of (11) will still be of the

form (12) for large r , provided F tends to zero sufficiently rapidly as $r \rightarrow \infty$, but the function u will now be definite and determined by the solution for smaller values of r .

For values α' of the α 's such that k^2 , defined by (8), is positive, the k in (12) must be chosen to be the positive square root of k^2 , in order that (12) may represent only outward moving particles, i.e. particles for which the radial component of momentum p_r , represented by $-i\hbar\partial/\partial r$, has a positive value. We now have that the density of scattered particles belonging to scatterers in state α' , equal to the square of the modulus of (12), falls off with increasing r according to the inverse square law, as is physically necessary, and their angular distribution is given by $|u(\theta\phi\alpha')|^2$. Further, the magnitude, P' say, of the momentum of these scattered particles is equal to $k\hbar$, since the exponential in (12) must be of the form $e^{iTr/\hbar}$, so that their energy is equal to

$$\frac{P'^2}{2m} = \frac{k^2\hbar^2}{2m} = E - H_s(\alpha') = H_s(\alpha^0) - H_s(\alpha') + \mathbf{p}^0{}^2/2m,$$

with the help of (8) and (10). This is just the energy of an incident particle, namely $\mathbf{p}^0{}^2/2m$, reduced by the increase in energy of the scatterer, namely $H_s(\alpha') - H_s(\alpha^0)$, in agreement with the law of conservation of energy. For values α' of the α 's such that k^2 is negative there are no scattered particles, the total initial energy being insufficient for the scatterer to be left in the state α' .

We must now evaluate $u(\theta\phi\alpha')$ for a set of values α' for the α 's such that k^2 is positive, and obtain the angular distribution of the scattered particles belonging to scatterers in state α' . It is sufficient to evaluate u for the direction $\theta = 0$ of the pole of the polar co-ordinates, since this direction is arbitrary. We make use of Green's theorem, which states that for any two functions of position A and B the volume integral $\int (A \nabla^2 B - B \nabla^2 A) dx$ taken over any volume equals the surface integral $\int (A \partial B / \partial n - B \partial A / \partial n) dS$ taken over the boundary of the volume, $\partial/\partial n$ denoting differentiation along the normal to the surface. We take

$$A = e^{-ikr \cos \theta} \quad B = (r\theta\phi\alpha' | 1)$$

and apply the theorem to a large sphere with the origin as centre. The volume integrand is thus

$$\begin{aligned} & e^{-ikr \cos \theta} \nabla^2(r\theta\phi\alpha' | 1) - (r\theta\phi\alpha' | 1) \nabla^2 e^{-ikr \cos \theta} \\ &= e^{-ikr \cos \theta} (\nabla^2 + k^2)(r\theta\phi\alpha' | 1) = e^{-ikr \cos \theta} F \end{aligned}$$

from (7) or (11), while the surface integrand is, with the help of (12),

$$\begin{aligned} & e^{-ikr\cos\theta} \frac{\partial}{\partial r} (r\theta\phi\alpha' | 1) - (r\theta\phi\alpha' | 1) \frac{\partial}{\partial r} e^{-ikr\cos\theta} \\ &= e^{-ikr\cos\theta} u \left(-\frac{1}{r^2} + \frac{ik}{r} \right) e^{ikr} + i \frac{u}{r} e^{ikr} k \cos\theta e^{-ikr\cos\theta} \\ &= iku/r \cdot (1 + \cos\theta) e^{ikr(1-\cos\theta)} \end{aligned}$$

with neglect of r^{-2} . Hence we get

$$\begin{aligned} \int e^{-ikr\cos\theta} F dx &= \int_0^{2\pi} d\phi \int_0^\pi r^2 \sin\theta d\theta \cdot iku/r \cdot (1 + \cos\theta) e^{ikr(1-\cos\theta)} \\ &= ikr \int_0^{2\pi} d\phi \int_0^2 dy \cdot u(\theta\phi\alpha') (2 - \gamma) e^{ikry}, \end{aligned}$$

where $\gamma = 1 - \cos\theta$, the volume integral on the left being taken over the whole of space. The right-hand side becomes, on being integrated by parts with respect to γ ,

$$\int_0^{2\pi} d\phi \left\{ \left[u(\theta\phi\alpha') (2 - \gamma) e^{ikry} \right]_{\gamma=0}^{\gamma=2} - \int_0^2 dy \cdot e^{ikry} \frac{\partial}{\partial y} [u(\theta\phi\alpha') (2 - \gamma)] \right\}.$$

The second term in the {} brackets is of the order of magnitude of r^{-1} , as would be revealed by further partial integrations, and may therefore be neglected. We are thus left with

$$\int e^{-ikr\cos\theta} F dx = -2 \int_0^{2\pi} d\phi u(0\phi\alpha') = -4\pi u(0\phi\alpha'),$$

giving the value of $u(\theta\phi\alpha')$ for the direction $\theta = 0$.

This result may be written

$$u(0\phi\alpha') = -1/4\pi \cdot \int e^{-iP'r\cos\theta/\hbar} F dx, \quad (13)$$

since $P' = k\hbar$. If the vector \mathbf{p}' denotes the momentum of the scattered electrons coming off in a certain direction (and is thus of magnitude P'), the value of u for this direction will be

$$u(\theta'\phi'\alpha') = -1/4\pi \cdot \int e^{-i(\mathbf{p}', \mathbf{x})/\hbar} F dx,$$

as follows from (13) if one takes this direction to be the pole of the polar co-ordinates. This becomes, with the help of (9),

$$\begin{aligned} u(\theta'\phi'\alpha') &= -m/2\pi\hbar^2 \cdot \iint e^{-i(\mathbf{p}', \mathbf{x})/\hbar} dx (x\alpha' | V | x^0\alpha^0) dx^0 e^{i(\mathbf{p}', \mathbf{x}^0)/\hbar} \\ &= -2\pi m\hbar (p'\alpha' | V | p^0\alpha^0), \end{aligned} \quad (14)$$

when one makes a transformation from the co-ordinates x to the momenta p of the particle, using the transformation function (36) of Chapter VI. The single letter p is here used to denote the three components of momentum.

The density of scattered particles belonging to scatterers in state α' is now given by $|u(\theta'\phi'\alpha')|^2/r^2$. Since their velocity is P'/m , the rate at which these particles appear per unit solid angle about the direction of the vector \mathbf{p}' will be $P'/m \cdot |u(\theta'\phi'\alpha')|^2$. The density of the incident particles is, as we have seen, unity, so that the number of incident particles crossing unit area per unit time is equal to their velocity P^0/m where P^0 is the magnitude of \mathbf{p}^0 . Hence the effective area that must be hit by an incident particle in order to be scattered in a unit solid angle about the direction \mathbf{p}' and then belong to a scatterer in state α' will be

$$P'/P^0 \cdot |u(\theta'\phi'\alpha')|^2 = 4\pi^2 m^2 h^2 P'/P^0 \cdot |(p'\alpha' | V | p^0\alpha^0)|^2. \quad (15)$$

This is the scattering coefficient for transitions $\alpha^0 \rightarrow \alpha'$ of the scatterer. It depends on that matrix element $(p'\alpha' | V | p^0\alpha^0)$ of the perturbing energy V whose column $p^0\alpha^0$ and whose row $p'\alpha'$ refer respectively to the initial and final states of the unperturbed system, between which the scattering transition process takes place. The result (15) is thus in some ways analogous to the result (19) or (20) of Chapter IX, although the numerical coefficients are different in the two cases, corresponding to the different natures of the two transition processes.

§ 58. Solution with the p -Representation

The result (15) for the scattering coefficient makes a reference only to that representation in which the momentum p is diagonal. One would thus expect to be able to get a more direct proof of the result by working all the time in the p -representation, instead of working in the x -representation and transforming at the end to the p -representation, as was done in § 57. This would not at first sight appear to be a great improvement, as the lack of directness of the x -representation method is offset by its greater 'Anschaulichkeit', it being possible to picture the square of the modulus of the x -representative of a state as the density of a stream of particles in process of being scattered. The x -representation method has, however, other more serious disadvantages. One of the main applications of the theory of collisions is to the case of photons as incident particles. Now a photon is not a simple particle but has a polarization. It is evident from classical electromagnetic theory that a photon with a definite momentum, *i.e.* one moving in a definite direction with a definite frequency, may have a definite state of polarization (linear, circular,

&c.), while a photon with a definite position, which is to be pictured as an electromagnetic disturbance confined to a very small volume, cannot have any definite polarization. These facts mean that the polarization observable of a photon commutes with its momentum but not with its position. This results in the p -representation method being immediately applicable to the case of photons, it being only necessary to introduce the polarizing variable into the representatives and treat it along with the α 's describing the scatterer, while the x -representation method is not applicable. Further, in dealing with photons it is necessary to take the relativity variation of mass with velocity into account. This can easily be done in the p -representation method, but not so easily in the x -representation method.

Equation (4) still holds when the relativity variation of mass with velocity is taken into account for the particle, but W is now given by

$$W^2/c^2 = m^2c^2 + P^2 = m^2c^2 + p_x^2 + p_y^2 + p_z^2 \quad (16)$$

instead of by (2). Written in terms of p -representatives, equation (4) becomes

$$\{E - H_s(\alpha') - W\}(p\alpha'|1) = \sum_{\alpha''} \int (p\alpha' | V | p''\alpha'') dp'' (p''\alpha''|0),$$

W being here understood as a definite function of p_x, p_y, p_z given by (16). This may be written

$$\{W' - W\}(p\alpha'|1) = \sum_{\alpha''} \int (p\alpha' | V | p''\alpha'') dp'' (p''\alpha''|0), \quad (17)$$

where

$$W' = E - H_s(\alpha') \quad (18)$$

and is the energy required by the law of conservation of energy for a scattered particle belonging to a scatterer in state α' . The p -representative of $\psi(H'_0)$, obtained by transforming (6) with the transformation function (36) of Chapter VI, is

$$(p\alpha|0) = h^{\frac{1}{2}} \delta_{\alpha\alpha^0} \delta(p - p^0), \quad (19)$$

as may be verified most easily by transforming this back to the x -representation. The $\delta(p - p^0)$ means the product

$$\delta(p_x - p_x^0) \delta(p_y - p_y^0) \delta(p_z - p_z^0).$$

Equation (17) now becomes

$$\{W' - W\}(p\alpha'|1) = h^{\frac{1}{2}} (p\alpha' | V | p^0\alpha^0). \quad (20)$$

We now make a canonical transformation from the Cartesian co-ordinates p_x, p_y, p_z of \mathbf{p} to its polar co-ordinates P, ω, χ , given by

$$p_x = P \cos \omega \quad p_y = P \sin \omega \cos \chi \quad p_z = P \sin \omega \sin \chi.$$

If in the new representation we take the weight function $P^2 \sin \omega$, then the weight attached to any volume of p -space will be the same

as in the previous p -representation, so that the canonical transformation will mean simply a relabelling of the rows and columns of the matrices without any alteration of the matrix elements or of the set of numbers representing a state. Thus (20) will become in the new representation

$$\{W' - W\}(P\omega\chi\alpha'|1) = \hbar^{\frac{1}{2}}(P\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0), \quad (21)$$

W being now a function of the single variable P .

The coefficient of $(P\omega\chi\alpha'|)$, namely $\{W' - W\}$, is now simply a multiplying factor and not a differential operator as it was with the x -representation method. We can therefore divide out by this factor and obtain an explicit expression for $(P\omega\chi\alpha'|1)$. When, however, α' is such that W' , defined by (18), is greater than mc^2 , this factor will have the value zero for a certain point in the domain of the variable P , namely the point $P=P'$, given in terms of W' by (16). The function $(P\omega\chi\alpha'|1)$ will then have a singularity at this point. This singularity shows that $(P\omega\chi\alpha'|)$ represents an infinite number of particles moving about at great distances from the scatterers with energies indefinitely close to W' and it is therefore this singularity that we have to study to get the angular distribution of the particles at infinity.

The result of dividing out (21) by the factor $\{W' - W\}$ is

$$(P\omega\chi\alpha'|1) = \hbar^{\frac{1}{2}}(P\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0)/\{W' - W\} + \lambda(\omega\chi\alpha')\delta(W' - W), \quad \dots \quad (22)$$

where λ is an arbitrary function of ω , χ and α' , since when an arbitrary multiple of $\delta(W' - W)$ is multiplied by $W' - W$ the product will vanish. To give a meaning to the first term on the right-hand side of (22), we make the convention that its integral with respect to P over a range that includes the value P' is the limit when $\epsilon \rightarrow 0$ of the integral when the small domain $P'-\epsilon$ to $P'+\epsilon$ is excluded from the range of integration. This is sufficient to make the meaning of (22) precise, since we are interested effectively only in the integrals of the representatives of states when the representation has continuous ranges of rows and columns. We see that equation (21) is inadequate to determine the representative $(P\omega\chi\alpha'|1)$ completely, on account of the arbitrary function λ occurring in (22). We must choose this λ such that $(P\omega\chi\alpha'|1)$ represents only outward moving particles, since we want the only inward moving particles to be those represented by (19).

Let us take first the general case when the representative $(P\omega\chi|)$ of a state of the particle satisfies an equation of the type

$$\{W' - W\}(P\omega\chi|) = f(P\omega\chi) \quad (23)$$

where $f(P\omega\chi)$ is any function of P , ω and χ , and W' is a number greater than mc^2 , so that $(P\omega\chi|)$ is of the form

$$(P\omega\chi|) = f(P\omega\chi)/\{W' - W\} + \lambda(\omega\chi)\delta(W' - W), \quad (24)$$

and let us determine now what λ must be in order that $(P\omega\chi|)$ may represent only outward moving particles. We can do this by transforming $(P\omega\chi|)$ to the x -representation, or rather the $(r\theta\phi)$ -representation, and comparing it with (12) for large values of r . The transformation function is

$$(r\theta\phi|P\omega\chi) = h^{-\frac{1}{2}}e^{i(p, x)/\hbar} = h^{-\frac{1}{2}}e^{iPr[\cos\omega\cos\theta + \sin\omega\sin\theta\cos(\chi-\phi)]/\hbar}.$$

For the direction $\theta = 0$ we find

$$\begin{aligned} (r0\phi|) &= h^{-\frac{1}{2}} \int_0^\infty P^2 dP \int_0^{2\pi} d\chi \int_0^\pi \sin\omega d\omega e^{iPr\cos\omega/\hbar} (P\omega\chi|) \\ &= h^{-\frac{1}{2}} \int_0^\infty P^2 dP \int_0^{2\pi} d\chi \left\{ - \left[\frac{e^{iPr\cos\omega/\hbar}}{iPr/\hbar} (P\omega\chi|) \right]_{\omega=0}^{\omega=\pi} + \right. \\ &\quad \left. + \int_0^\pi d\omega \frac{e^{iPr\cos\omega/\hbar}}{iPr/\hbar} \frac{\partial}{\partial\omega} (P\omega\chi|) \right\}. \end{aligned}$$

The second term in the {} brackets is of order r^{-2} , as may be verified by further partial integrations with respect to ω , and can therefore be neglected. We are left with

$$\begin{aligned} (r0\phi|) &= ih^{-\frac{1}{2}}(2\pi r)^{-1} \int_0^\infty P dP \int_0^{2\pi} d\chi \{ e^{-iPr/\hbar} (P\pi\chi|) - e^{iPr/\hbar} (P0\chi|) \} \\ &= ih^{-\frac{1}{2}}r^{-1} \int_0^\infty P dP \{ e^{-iPr/\hbar} (P\pi\chi|) - e^{iPr/\hbar} (P0\chi|) \}. \end{aligned} \quad (25)$$

When we substitute for $(P\omega\chi|)$ its value given by (24), the first term in the integrand in (25) gives

$$ih^{-\frac{1}{2}}r^{-1} \int_0^\infty P dP e^{-iPr/\hbar} \{ f(P\pi\chi)/(W' - W) + \lambda(\pi\chi)\delta(W' - W) \}. \quad (26)$$

The term involving $\delta(W' - W)$ here may be integrated immediately and gives, when one uses the relation $P dP = W dW/c^2$, which follows from (16),

$$\begin{aligned} ih^{-\frac{1}{2}}c^{-2}r^{-1} \int_{mc^2}^\infty W dW e^{-iPr/\hbar} \lambda(\pi\chi)\delta(W' - W) &= \\ &= ih^{-\frac{1}{2}}c^{-2}r^{-1} W' \lambda(\pi\chi) e^{-iPr/\hbar}. \end{aligned} \quad (27)$$

To integrate the other term in (26) we use the formula that

$$\int_0^\infty g(P) \frac{e^{-iPr/\hbar}}{P' - P} dP = g(P') \int_0^\infty \frac{e^{-iPr/\hbar}}{P' - P} dP, \quad (28)$$

with neglect of terms involving r^{-1} , for any continuous function $g(P)$, which formula holds since $\int_0^\infty K(P)e^{-iPr/\hbar} dP$ is of order r^{-1} for any continuous function $K(P)$ and since the difference

$$g(P)/(P' - P) - g(P')/(P' - P)$$

is continuous. The right-hand side of (28), when evaluated with neglect of terms involving r^{-1} , and also with neglect of the small domain $P' - \epsilon$ to $P' + \epsilon$ in the domain of integration, gives

$$\begin{aligned} g(P') \int_{-\infty}^\infty \frac{e^{-iPr/\hbar}}{P' - P} dP &= g(P') e^{-iP'r/\hbar} \int_{-\infty}^\infty \frac{e^{i(P' - P)r/\hbar}}{P' - P} dP \\ &= ig(P') e^{-iP'r/\hbar} \int_{-\infty}^\infty \frac{\sin(P' - P)r/\hbar}{P' - P} dP = i\pi g(P') e^{-iP'r/\hbar}. \end{aligned} \quad (29)$$

In our present example $g(P)$ is

$$g(P) = ih^{-\frac{1}{2}}r^{-1}Pf(P\pi\chi)(P' - P)/(W' - W),$$

which has the limiting value when $P = P'$,

$$g(P') = ih^{-\frac{1}{2}}r^{-1}P'f(P'\pi\chi)W'/P'c^2 = ih^{-\frac{1}{2}}c^{-2}r^{-1}W'f(P'\pi\chi).$$

Substituting this in (29) and adding on the expression (27), we obtain the following value for the integral (26)

$$h^{-\frac{1}{2}}c^{-2}r^{-1}W' \{-\pi f(P'\pi\chi) + i\lambda(\pi\chi)\} e^{-iP'r/\hbar}. \quad (30)$$

Similarly the second term in the integrand in (25) gives

$$h^{-\frac{1}{2}}c^{-2}r^{-1}W' \{-\pi f(P'0\chi) - i\lambda(0\chi)\} e^{iP'r/\hbar}. \quad (31)$$

The sum of these two expressions is the value of $(r0\phi|)$ when r is large.

We require that $(r0\phi|)$ shall represent only outward moving particles, and hence it must be of the form of a multiple of $e^{iP'r/\hbar}$. Thus (30) must vanish, so that

$$\lambda(\pi\chi) = -i\pi f(P'\pi\chi). \quad (32)$$

We see in this way that the condition that $(r0\phi|)$ shall represent only outward moving particles in the direction $\theta = 0$ fixes the value of λ for the opposite direction $\theta = \pi$. Since the direction $\theta = 0$ or $\omega = 0$ of the pole of our polar co-ordinates is not in any way singular, we can generalize (32) to

$$\lambda(\omega\chi) = -i\pi f(P'\omega\chi), \quad (33)$$

which gives the value of λ for an arbitrary direction. This value substituted in (24) gives a result that may be written

$$(P\omega\chi|) = f(P\omega\chi)\{1/(W' - W) - i\pi\delta(W' - W)\}, \quad (34)$$

since one can substitute P' for P in the coefficient of a term involving $\delta(W' - W)$ as a factor without changing the value of the term. *The condition that $(P\omega\chi|)$ shall represent only outward moving particles is thus that it shall contain the factor*

$$\{1/(W' - W) - i\pi\delta(W' - W)\}. \quad (35)$$

With λ given by (33), expression (30) vanishes and the value of $(r\phi|)$ for large r is given by expression (31) alone, thus

$$(r\phi|) = -2\pi\hbar^{-\frac{1}{2}}c^{-2}r^{-1}W'f(P'\omega\chi)e^{iP'r/\hbar}.$$

This may be generalized to

$$(r\theta\phi|) = -2\pi\hbar^{-\frac{1}{2}}c^{-2}r^{-1}W'f(P'\omega\chi)e^{iP'r/\hbar},$$

giving the value of $(r\theta\phi|)$ for any direction θ, ϕ in terms of $f(P'\omega\chi)$ for the same direction labelled by ω, χ . This is of the form (12) with

$$u(\theta\phi) = -2\pi\hbar^{-\frac{1}{2}}c^{-2}W'f(P'\omega\chi)$$

and thus represents a distribution of outward moving particles of momentum P' whose number is

$$\frac{c^2P'}{W'} |u|^2 = \frac{4\pi^2W'P'}{\hbar c^2} |f(P'\omega\chi)|^2 \quad (36)$$

per unit solid angle per unit time. This distribution is thus that represented by the $(P\omega\chi|)$ of (34).

From this general result we can infer that, whenever we have a representative $(P\omega\chi|)$ representing only outward moving particles and satisfying an equation of the type (23), the number per unit solid angle per unit time of these particles is given by (36). If this $(P\omega\chi|)$ occurs in a problem in which the number of incident particles is one per unit volume, it will correspond to a scattering coefficient of amount

$$\frac{4\pi^2W^0W'P'}{\hbar c^4P_0} |f(P'\omega\chi)|^2. \quad (37)$$

It is only the value of the function $f(P\omega\chi)$ for the point $P = P'$ that is of importance.

If we now apply this general theory to our equations (21) and (22), we have

$$f(P\omega\chi) = h^{\frac{1}{2}}(P\omega\chi\alpha'|V|P^0\omega^0\chi^0\alpha^0).$$

Hence from (37) the scattering coefficient is

$$4\pi^2 h^2 W^0 W' P' / c^4 P^0 \cdot |(P' \omega \chi \alpha' | V | P^0 \omega^0 \chi^0 \alpha^0)|^2. \quad (38)$$

If one neglects relativity and puts $W^0 W' / c^4 = m^2$, this result reduces to the result (15) obtained in the preceding section by means of Green's theorem.

§ 59. Dispersive Scattering

We shall now determine the scattering when the incident particle is capable of being absorbed, that is, when our unperturbed system of scatterer plus particle has closed stationary states with the particle absorbed. The existence of these closed states for the unperturbed system will be found to have a considerable effect on the scattering for the perturbed system, and indeed an effect that depends very much on the energy of the incident particle, giving rise to the phenomenon of dispersion in optics when the incident particle is taken to be a photon.

We use a representation for which the fundamental states are the stationary states of the unperturbed system, as was the case for the p -representation of the preceding section. These stationary states are now the states $\psi(p'\alpha')$ for which the particle has a definite momentum p' and the scatterer is in a definite state α' , together with the closed states, ψ_k say, which form a separate discrete set. We shall assume that these states are all independent and orthogonal, so that our representation is of the usual orthogonal type. This assumption is probably not justifiable when the particle is an electron or atomic nucleus, since in this case for an absorbed state ψ_k the particle will still certainly be somewhere, so that one would expect to be able to expand ψ_k in terms of the eigen- ψ 's $\psi(x'\alpha')$ of x , y , z , and the α 's, and hence also in terms of the $\psi(p'\alpha')$. On the other hand, when the particle is a photon it will no longer exist for the absorbed states, which are then certainly independent of and orthogonal to the states $\psi(p'\alpha')$ for which the particle does exist. Thus the assumption is justified in this case, which is the important practical one.

The representative of a state will now consist of a discrete set of numbers ($k|$) referring to the fundamental states ψ_k together with the three-dimensional continuous ranges of numbers ($p'\alpha'|$) referring to the $\psi(p'\alpha')$, there being one such range for each set of values α' for the α 's. Similarly the matrices representing observables will now contain discrete rows and columns labelled by k together with continuous

ranges labelled by (p, α) . Thus, for example, the matrix representing V , the perturbing energy, will have elements $(k'|V|k'')$, $(k'|V|p''\alpha'')$, $(p'\alpha'|V|k'')$, and $(p'\alpha'|V|p''\alpha'')$.

Since we are concerned with scattering, we must still deal with *stationary* states of the whole system, which will still be given by an equation of the type (3). We shall now, however, have to work to the second order of accuracy, so that we cannot simply use the first-order equation (4). The exact equation (3) gives, when written in terms of representatives,

$$\left. \begin{aligned} \{W' - W\}(p\alpha') &= \\ &= \sum_{\alpha''} \int (p\alpha'|V|p''\alpha'') dp'' (p''\alpha'') + \sum_{k''} (p\alpha'|V|k'')(k''|) \end{aligned} \right\} (39)$$

$$\{E - E_k\}(k|) = \sum_{\alpha''} \int (k|V|p''\alpha'') dp'' (p''\alpha'') + \sum_{k''} (k|V|k'')(k''|),$$

where W' is given by (18) and E_k is the energy of the stationary state ψ_k of the unperturbed system. If we suppose the exact $\psi(H')$ to be expressed as the sum of $\psi(H'_0)$, a first-order correction ψ_1 , a second-order correction ψ_2 , and so on, thus

$$\psi(H') = \psi(H'_0) + \psi_1 + \psi_2 + \dots,$$

the r -th-order correction will be given in terms of the $(r-1)$ -th by

$$(E - H_0)\psi_r = V\psi_{r-1}.$$

Thus its representative $(p\alpha'|r)$, $(k|r)$ will be given by

$$\left. \begin{aligned} \{W' - W\}(p\alpha'|r) &= \\ &= \sum_{\alpha''} \int (p\alpha'|V|p''\alpha'') dp'' (p''\alpha''|r-1) + \sum_{k''} (p\alpha'|V|k'')(k''|r-1) \\ \{E - E_k\}(k|r) &= \\ &= \sum_{\alpha''} \int (k|V|p''\alpha'') dp'' (p''\alpha''|r-1) + \sum_{k''} (k|V|k'')(k''|r-1). \end{aligned} \right\} (40)$$

For $r = 1$ these equations are just the generalization of (17) when there exist absorbed states ψ_k . The unperturbed stationary state $\psi(p^0\alpha^0)$ will now be represented by

$$(p\alpha|0) = h^{\frac{1}{2}}\delta_{\alpha\alpha^0}\delta(p-p^0) \quad (k|0) = 0, \quad (41)$$

instead of merely by (19), so the first-order correction will be given by

$$\{W' - W\}(p\alpha'|1) = h^{\frac{1}{2}}(p\alpha'|V|p^0\alpha^0) \quad (42)$$

$$\{E - E_k\}(k|1) = h^{\frac{1}{2}}(k|V|p^0\alpha^0). \quad (43)$$

We may assume that the matrix elements $(k'|V|k'')$ of V vanish, since these matrix elements are not essential to the phenomena under investigation, and if they did not vanish it would mean simply that the absorbed states ψ_k had not been suitably chosen. We shall further assume that the matrix elements $(p'\alpha'|V|p''\alpha'')$ are of the second order

of smallness when the matrix elements $(k'|V|p''\alpha'')$, $(p'\alpha'|V|k'')$ are taken to be of the first order of smallness. This assumption will be justified for the case of photons in Chapter XII. We now have from (43) and (42) that $(k|1)$ is of the first order of smallness, provided E does not lie near one of the discrete set of energy-levels E_k , and $(p\alpha|1)$ is of the second order. The value of $(p\alpha|2)$ to the second order will thus be given, from the first of equations (40), by

$$\{W' - W\}(p\alpha'|2) = \hbar^2 \sum_{k'} (p\alpha'|V|k'')(k''|V|p^0\alpha^0)/(E - E_k).$$

The total correction of the second-order, arising partly from $(p\alpha|1)$ and partly from $(p\alpha|2)$, therefore satisfies

$$\begin{aligned} \{W' - W\}\{(p\alpha'|1) + (p\alpha'|2)\} &= \\ &= \hbar^2\{(p\alpha'|V|p^0\alpha^0) + \sum_k (p\alpha'|V|k)(k|V|p^0\alpha^0)/(E - E_k)\}. \end{aligned}$$

This equation is of the type (23), provided α' is such that $W' > mc^2$, which means that α' as a final state for the scatterer is not inconsistent with the law of conservation of energy. We can therefore infer from the general result (37) that the scattering coefficient is

$$\frac{4\pi^2\hbar^2 W^0 W' P'}{c^4 P^0} \left\{ (p'\alpha'|V|p^0\alpha^0) + \sum_k \frac{(p'\alpha'|V|k)(k|V|p^0\alpha^0)}{E - E_k} \right\}^2. \quad (44)$$

The scattering may now be considered as composed of two parts, a part that arises from the matrix element $(p'\alpha'|V|p^0\alpha^0)$ of the perturbing energy and a part that arises from the matrix elements $(p'\alpha'|V|k)$ and $(k|V|p^0\alpha^0)$. The first part, which is the same as our previously obtained result (38), may be called the true scattering. The second part may be considered as arising from an absorption of the incident particle into some state k , followed immediately by a re-emission in a different direction. The fact that we have to add the two terms before taking the square of the modulus denotes interference between the two kinds of scattering. There is no experimental way of separating the two kinds, the distinction between them being only mathematical.

§ 60. Resonance Scattering

Suppose the energy of the incident particle to be varied continuously while the initial state α^0 of the scatterer is kept fixed, so that the total energy E varies continuously. The formula (44) now shows that as E approaches one of the discrete set of energy-levels E_k , the scattering becomes very large. In fact, according to formula (44)

the scattering should be infinite when E is exactly equal to an E_k . An infinite scattering coefficient is, of course, physically impossible, so that we can infer that the approximations used in deriving (44) are no longer legitimate when E is close to an E_k . To investigate the scattering in this case we must therefore go back to the exact equations (39) and use a different method of approximating to their solution.

Let us take one particular E_k and consider the case when E is close to it. The large term in the scattering coefficient (44) now arises from those elements of the matrix representing V that lie in row k or in column k , i.e. those of the type $(k|V|p\alpha)$ or $(p\alpha|V|k)$. The scattering arising from the other matrix elements of V is of a smaller order of magnitude. This suggests that in our exact equations (39) we should make the approximation of neglecting all the matrix elements of V except the important ones, which are those of the type $(p\alpha'|V|k)$ or $(k|V|p\alpha')$, where α' is a state of the scatterer that has not too much energy to be disallowed as a final state by the law of conservation of energy. These equations then reduce to

$$\{W' - W\}(p\alpha'|) = (p\alpha'|V|k)(k|) \quad (45)$$

$$\{E - E_k\}(k|) = \sum_{\alpha'} \int (k|V|p\alpha') dp (p\alpha'|), \quad (46)$$

the α' summation being over those values of α' for which W' given by (18) is $> mc^2$. These equations are now sufficiently simple for us to be able to solve exactly without further approximation.

From (45) we obtain by division

$$(p\alpha'|) = (p\alpha'|V|k)(k|)/\{W' - W\} + \lambda\delta(W' - W). \quad (47)$$

We must choose λ , which may be any function of the momentum \mathbf{p} and α' , such that (47) represents the incident particles (19) together with only outward moving particles. [The right-hand side of (19), with α' substituted for α , is actually of the form $\lambda\delta(W' - W)$, since the conditions $\alpha' = \alpha^0$ and $\mathbf{p} = \mathbf{p}^0$ for this right-hand side not to vanish lead to $W' = E - H_s(\alpha') = E - H_s(\alpha^0) = W^0$ and $W = W^0$, which together give $W' = W$.] Thus (47) must be

$$(p\alpha'|) = h^{\frac{1}{2}}\delta_{\alpha'\alpha^0}\delta(p' - p^0) + (p\alpha'|V|k)(k|)\{1/(W' - W) - i\pi\delta(W' - W)\}, \quad \dots (48)$$

and from the general formula (37) the scattering coefficient will be

$$4\pi^2 W^0 W' P' / hc^4 P^0 \cdot |(p'\alpha'|V|k)|^2 |(k)|^2. \quad (49)$$

It remains for us to determine the value of $(k|)$. We can do this by substituting for $(p\alpha'|)$ in (46) its value given by (48). This gives

$$\{E - E_k\}(k|) =$$

$$= h^{\frac{1}{2}}(k|V|p^0\alpha^0) + (k|)\sum_{\alpha'} \int |(k|V|p\alpha')|^2 \{1/(W' - W) - i\pi\delta(W' - W)\} dp \\ = h^{\frac{1}{2}}(k|V|p^0\alpha^0) + (k|)\{a - ib\},$$

where $a = \sum_{\alpha'} \int |(k|V|p\alpha')|^2 dp / (W' - W)$ (50)

and $b = \pi \sum_{\alpha'} \int |(k|V|p\alpha')|^2 \delta(W' - W) dp \\ = \pi \sum_{\alpha'} \int \int \int |(k|V|P\omega\chi\alpha')|^2 \delta(W' - W) P^2 dP \sin \omega d\omega d\chi \\ = \pi \sum_{\alpha'} P' W' c^{-2} \int \int |(k|V|P'\omega\chi\alpha')|^2 \sin \omega d\omega d\chi.$ (51)

Thus $(k|) = h^{\frac{1}{2}}(k|V|p^0\alpha^0) / \{E - E_k - a + ib\}.$ (52)

Note that a and b are real and that b is positive.

This value for $(k|)$ substituted in (49) gives for the scattering coefficient

$$\frac{4\pi^2 h^2 W^0 W' P'}{c^4 P^0} \frac{|(p'\alpha'|V|k)|^2 |(k|V|p^0\alpha^0)|^2}{(E - E_k - a)^2 + b^2}. \quad (53)$$

One can obtain the total effective area that the incident particle must hit in order to be scattered anywhere by integrating (53) over all directions of scattering, *i.e.* by integrating over all directions of the vector \mathbf{p}' with its magnitude kept fixed at P' , and then summing over all α' that are to be taken into consideration, *i.e.* for which $W' > mc^2$. This gives, with the help of (51), the result

$$\frac{4\pi h^2 W^0}{c^2 P^0} \frac{b |(k|V|p^0\alpha^0)|^2}{(E - E_k - a)^2 + b^2}. \quad (54)$$

If we suppose E to vary continuously through the value E_k , the main variation of (53) or (54) will be due to the small denominator $(E - E_k - a)^2 + b^2$. If we neglect the dependence of the other factors in (53) and (54) on E , then the maximum scattering will occur when E has the value $E_k + a$ and the scattering will be half its maximum when E differs from this value by an amount b . The large amount of scattering that occurs for values of the energy of the incident particle that make E nearly equal to E_k give rise to the phenomenon of an absorption line. The centre of the line is displaced by an amount a from the resonance energy of the incident particle, *i.e.* the energy which would make the total energy just E_k , while the quantity b is what is sometimes called the half-width of the line.

§ 61. Emission and Absorption

For studying emission and absorption we must consider non-stationary states of the system and must use the perturbation method of § 52. To determine the coefficient of spontaneous emission we must take a state for which the particle is initially absorbed, so that the representative of the state is then

$$(k|) = 1 \quad (p\alpha|) = 0,$$

and determine the probability that at some later time the particle shall be on its way to infinity with a definite momentum. The method of § 54 can now be applied. From the result (28) of that section we see that the probability per unit time per unit range of ω and χ of the particle being emitted in any direction ω' , χ' with the scatterer being left in state α' is

$$2\pi/\hbar \cdot |(k|V|W'\omega'\chi'\alpha')|^2, \quad (55)$$

provided, of course, that α' is such that the energy W' , given by (18), of the particle is greater than mc^2 . For values of α' that do not satisfy this condition there is no emission possible. The matrix element $(k|V|W'\omega'\chi'\alpha')$ here must refer to a representation in which W , ω , χ , and α are diagonal with the weight function unity. The matrix elements of V appearing in the three preceding sections refer to a representation in which p_x , p_y , p_z are diagonal with the weight function unity, or P , ω , χ are diagonal with the weight function $P^2 \sin \omega$. They would thus refer to a representation in which W , ω , χ are diagonal with the weight function $dP/dW \cdot P^2 \sin \omega = WP/c^2 \cdot \sin \omega$. Thus the matrix element $(k|V|W'\omega'\chi'\alpha')$ in (55) is equal to $(W'P'/c^2 \cdot \sin \omega')^{\frac{1}{2}}$ times our previous matrix element $(k|V|P'\omega'\chi'\alpha')$ or $(k|V|p'\alpha')$, so that (55) is equal to

$$\frac{2\pi}{\hbar} \frac{W'P'}{c^2} \sin \omega' |(k|V|p'\alpha')|^2.$$

The probability of emission per unit solid angle per unit time, with the scatterer simultaneously dropping to state α' , is thus

$$\frac{2\pi}{\hbar} \frac{W'P'}{c^2} |(k|V|p'\alpha')|^2. \quad (56)$$

To obtain the total probability per unit time of the particle being emitted in any direction, with any final state for the scatterer, we must integrate (56) over all angles ω' , χ' and sum over all states α' whose energy $H_s(\alpha')$ is such that $H_s(\alpha') + mc^2 < E_k$. The result is just $2b/\hbar$, where b is defined by (51). There is thus this simple rela-

tion between the total emission coefficient and the half-breadth b of the absorption line.

Let us now consider absorption. This requires that we shall study a state for which initially the particle is certainly not absorbed but is incident with a definite momentum. Thus the initial representative of the state must be of the form (41). We must now determine the probability of the particle being absorbed after time T . Since our final state ψ_k is not one of a continuous range, we cannot use directly the result (28) of § 54. If, however, we take

$$(p\alpha|)_0 = \delta_{\alpha\alpha^0} \delta(p - p^0) \quad (k|)_0 = 0 \quad (57)$$

as the initial representative of the state, the analysis of §§ 52 and 54 is still applicable as far as equation (25) and shows us that the probability of the particle being absorbed into state ψ_k after time T is

$$2|(k|V|p^0\alpha^0)|^2[1 - \cos\{(E_k - E)T/\hbar\}]/(E_k - E)^2.$$

This corresponds to a distribution of incident particles of density \hbar^{-3} , owing to the omission of the factor \hbar^3 from (57), as compared with (41). The probability of there being an absorption after time T when there is one incident particle crossing unit area per unit time is therefore

$$2\hbar^3 W^0/c^2 P^0 \cdot |(k|V|p^0\alpha^0)|^2[1 - \cos\{(E_k - E)T/\hbar\}]/(E_k - E)^2. \quad (58)$$

To obtain the absorption coefficient we must consider the incident particles not all to have exactly the same energy $W^0 = E - H_s(\alpha^0)$, but to have a distribution of energy values about the correct value $E_k - H_s(\alpha^0)$ required for absorption. If we take a beam of incident particles consisting of one crossing unit area per unit time per unit energy range, the probability of there being an absorption after time T will be given by the integral of (58) with respect to E . This integral may be evaluated in the same way as (26) of § 56 and is equal to

$$4\pi^2 \hbar^2 W^0 T / c^2 P^0 \cdot |(k|V|p^0\alpha^0)|^2.$$

The probability per unit time of an absorption taking place with an incident beam of one particle per unit area per unit time per unit energy range is therefore

$$4\pi^2 \hbar^2 W^0 / c^2 P^0 \cdot |(k|V|p^0\alpha^0)|^2, \quad (59)$$

which is the absorption coefficient.

The connexion between the absorption and emission coefficients (59) and (56) and the resonance scattering coefficients calculated in the preceding section should be noted. When the incident beam does

not consist of particles all with the same energy, but consists of a unit distribution of particles per unit energy range crossing unit area per unit time, the total number of incident particles with energies near an absorption line that get scattered will be given by the integral of (54) with respect to E . If one neglects the dependence of the numerator of (54) on E , this integral will, since

$$\int_{-\infty}^{\infty} \frac{b}{(E - E_k - a)^2 + b^2} dE = \pi,$$

have just the value (59). Thus *the total number of scattered particles in the neighbourhood of an absorption line is equal to the total number absorbed*. We can therefore regard all these scattered particles as absorbed particles that are subsequently re-emitted in a different direction. Further, the number of particles in the neighbourhood of the absorption line that get scattered per unit solid angle about a given direction p' and then belong to scatterers in state α' will be given by the integral with respect to E of (53), which integral has in the same way the value

$$\frac{4\pi^2 h^2 W^0 W' P'}{c^4 P^0} \frac{\pi}{b} |(p' \alpha' | V | k)|^2 |(k | V | p^0 \alpha^0)|^2.$$

This is just equal to the absorption coefficient (59) multiplied by the emission coefficient (56) divided by $2b/\hbar$, the total emission coefficient. This is in agreement with the point of view of regarding the resonance scattered particles as those that are absorbed and then re-emitted, according to which point of view the fraction of the total number of absorbed particles that are re-emitted in a unit solid angle about a given direction would be just the emission coefficient for this direction divided by the total emission coefficient, provided the absorption and emission processes are governed independently each by its own probability law.

XI

SYSTEMS CONTAINING SEVERAL SIMILAR PARTICLES

§ 62. Symmetrical and Antisymmetrical States

If a system in atomic physics contains a number of particles of the same kind, e.g. a number of electrons, the particles are absolutely indistinguishable one from another. No observable change is made when two of them are interchanged. This circumstance gives rise to some curious phenomena in quantum mechanics having no analogue in the classical theory, which arise from the fact that in quantum mechanics a transition may occur resulting in merely the interchange of two similar particles, which transition then could not be detected by any observational means. A satisfactory theory ought, of course, to count two observationally indistinguishable states as the same state and to deny that any transition does occur when two similar particles exchange places. We shall find that such a theory can be developed in agreement with the principles of quantum mechanics.

Suppose we have a system containing n similar particles. We may take as our dynamical variables a set of variables ξ_1 describing the first particle, the corresponding set ξ_2 describing the second particle, and so on up to the set ξ_n describing the n -th particle. We shall then have the ξ_r 's commuting with the ξ_s 's for $r \neq s$. (We may require certain extra variables, describing what the system consists of in addition to the n similar particles, but it is not necessary to mention these explicitly in the present chapter.) The Hamiltonian describing the motion of the system will now be expressible as a function of the $\xi_1, \xi_2 \dots \xi_n$. The fact that the particles are similar requires that *the Hamiltonian shall be a symmetrical function of the $\xi_1, \xi_2 \dots \xi_n$, i.e. it shall remain unchanged when the sets of variables ξ_r are interchanged or permuted in any way.* This condition must hold no matter what perturbations are applied to the system.

We may take a representation with observables $q_1, q_2 \dots q_n$ diagonal, which are such that the q_1 's are the values at time t of certain commuting dynamical variables describing the first particle, the q_2 's are the values at time t of the corresponding variables describing the second particle, and so on. We may further choose the phases of the representation in the same way for each of the particles. (This means, for example, that if a certain momentum p_1 describing the

first particle is represented by $-i\hbar\partial/\partial q_1$, the corresponding momentum p_r describing the r -th particle must be represented by $-i\hbar\partial/\partial q_r$.) The representation will then treat all the particles on the same footing. The condition that the Hamiltonian H is symmetrical between all the particles may now be expressed by the condition that its representative $(q'_1 q'_2 \dots q'_n | H | q''_1 q''_2 \dots q''_n)$, or $(q' | H | q'')$ for brevity, is symmetrical between all the q 's, *i.e.* that it remains unchanged if any permutation is applied to the q' 's and the same permutation to the q'' 's. This condition may be expressed analytically thus,

$$(q' | H | q'') = (Pq' | H | Pg''), \quad (1)$$

where P denotes any permutation of the numbers $1, 2 \dots n$ and Pq' denotes the set of numbers obtained by applying the permutation P to the suffixes of $q'_1, q'_2 \dots q'_n$.

Let $(q'_1 q'_2 \dots q'_n |)$ or $(q' |)$ be the wave function representing any state. It will satisfy the wave equation

$$i\hbar \frac{\partial}{\partial t} (q' |) = \int (q' | H | q'') dq'' (q'' |). \quad (2)$$

If we apply any permutation P to the variables q' in $(q' |)$ we shall obtain a function $(Pg' |)$ satisfying

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (Pg' |) &= \int (Pg' | H | q'') dq'' (q'' |) \\ &= \int (Pg' | H | Pg'') dq'' (Pg'' |), \end{aligned}$$

since we can apply any permutation to the variables of integration q'' in the integrand without changing the value of the integral. With the help of (1) this becomes

$$i\hbar \frac{\partial}{\partial t} (Pg' |) = \int (q' | H | q'') dq'' (Pg'' |), \quad (3)$$

which shows that $(Pg' |)$ is a solution of the wave equation (2). Hence *if we apply any permutation to the variables in a solution of the wave equation we obtain another solution.*

Suppose we take a state whose representative $(q' |)$ at some particular time t is a symmetrical function of all the q' 's, so that

$$(q' |) = (Pg' |) \quad (4)$$

for any P . The right-hand sides of (2) and (3) are now equal, so that

$$\frac{\partial}{\partial t} (q' |) = \frac{\partial}{\partial t} (Pg' |).$$

This equation is the time derivative of (4) and shows that if (4) holds at one particular time it holds also at a slightly later time, and thus by induction it holds at all times. Thus *if a wave function is initially symmetrical it always remains symmetrical.*

Similarly we may take a state whose representative (q'') at some particular time is antisymmetrical, i.e. $(q'_1 q'_2 \dots q'_n)$ changes sign with interchange of any pair of q'' 's. We shall then have

$$(q'') = \pm (Pq''), \quad (5)$$

the + or - sign being taken according to whether the permutation P is even or odd (i.e. according to whether P can be built up from an even or an odd number of simple interchanges). The same argument as before now shows that *if a wave function is initially antisymmetrical it always remains antisymmetrical.*

Let us make a canonical transformation to a Q -representation which, like the original q -representation, treats all the particles on the same footing. This means that the Q 's consist of corresponding sets of observables $Q_1, Q_2 \dots Q_n$ describing the first, second . . . n th particle respectively and that the phases are chosen in the same way for each of the particles. The transformation function will now be of the form

$$(Q'_1 Q'_2 \dots Q'_n | q'_1 q'_2 \dots q'_n) = (Q'_1 | q'_1) (Q'_2 | q'_2) \dots (Q'_n | q'_n), \quad (6)$$

in which each factor $(Q'_r | q'_r)$ is the same function of its variables Q'_r, q'_r . This condition gives, if we denote $(Q'_1 Q'_2 \dots Q'_n | q'_1 q'_2 \dots q'_n)$ by $(Q' | q')$ for brevity,

$$(Q' | q') = (PQ' | Pq'), \quad (7)$$

for an arbitrary permutation P . The new representative of any state is given by

$$(Q' |) = \int (Q' | q') dq' (q' |). \quad (8)$$

From this equation we can deduce that

$$\begin{aligned} (PQ' |) &= \int (PQ' | q') dq' (q' |) \\ &= \int (PQ' | Pq') dq' (Pq' |) \\ &= \int (Q' | q') dq' (Pq' |) \end{aligned} \quad (9)$$

with the help of (7). Now if $(q' |)$ is symmetrical, so that equation (4) holds, the right-hand sides of (8) and (9) are equal. We then have $(Q' |) = (PQ' |)$, so that $(Q' |)$ is also symmetrical. Similarly if $(q' |)$ is antisymmetrical, $(Q' |)$ is also antisymmetrical. Thus *the property of the representative of a state of being symmetrical or antisymmetrical*

remains invariant under a canonical transformation. This invariance, together with the fact proved above that a wave function if initially symmetrical or antisymmetrical always remains so, shows that the property of being symmetrical or antisymmetrical is a property of the states themselves and not merely a property of their representatives. Thus we can talk about symmetrical and antisymmetrical states.

The invariance and permanence of the symmetry properties of the states means that for some particular kind of particle it is quite possible for only symmetrical or only antisymmetrical states to occur in nature. Whether this is the case cannot be decided by any general theoretical considerations, but can be settled only by reference to special experimentally determined facts about the particles in question. For photons one can settle the question by making use of Planck's radiation law. Only when one assumes *the symmetrical states for photons* does one get a statistical mechanics leading to Planck's law for radiation in statistical equilibrium. This statistical mechanics is known as the Einstein-Bose statistics, as it was first introduced by Bose and Einstein before the arrival of the modern quantum mechanics.

For electrons we use the fact that, if we make the approximation of regarding the electrons in an atom as each moving in its own 'orbit' (*i.e.* as being each describable by its own wave function involving only its own variables), then no two electrons will ever be in the same orbit. This fact, which is known as Pauli's exclusion principle, may be inferred from general experimental evidence on atomic structure. Let us see how to fit it in with the theory. If the wave functions representing the different orbits are

$$(q'|\alpha_1), (q'|\alpha_2), \dots (q'|\alpha_n),$$

a wave function representing the whole atom will be given by the product

$$(q'_1|\alpha_1)(q'_2|\alpha_2) \dots (q'_n|\alpha_n) \cdots (q'|\alpha) \quad (10)$$

say, for brevity. Other wave functions representing the same distribution of electrons over the various orbits may be obtained by applying any permutation to the α 's in (10). There will be altogether $n!$ such wave functions, the general one being $(q'|P\alpha)$. Any linear combination of these wave functions will also represent the same electron distribution. One such linear combination is the sum

$$\sum_P \frac{1}{n!} (q'|P\alpha), \quad (11)$$

which is symmetrical between all the q'' 's. Another is

$$\Sigma_P \pm (q'|P\alpha), \quad (12)$$

the + or - sign being taken according to whether P is an even or odd permutation, and this one is antisymmetrical. The antisymmetrical wave function (12) has the property that it vanishes identically if two of the α 's are equal. Hence if we assume that *for electrons only antisymmetrical states occur*, we shall get the result that there are no states with two electrons in the same orbit, which is just Pauli's exclusion principle. This assumption is the only one we can make which will lead to Pauli's exclusion principle.

In this way we can see that for photons we must take the symmetrical states and for electrons the antisymmetrical states. These are special cases of an empirical rule, which appears to hold without exception, according to which only the symmetrical or only the antisymmetrical states occur according to whether the particles in question carry a charge of an even or an odd multiple of the electronic charge. When only the symmetrical or only the antisymmetrical states are allowed for a particular kind of particle, the theory can no longer make a distinction between two states which differ only through a permutation of the particles, so that the difficulties mentioned at the beginning of this section disappear.

§ 63. Permutations as Observables

Let us now build up a general theory for a system containing n similar particles when states with any kind of symmetry properties are allowed, *i.e.* when there is no restriction to only symmetrical or only antisymmetrical states. The general state now will not be symmetrical or antisymmetrical, nor will it be expressible linearly in terms of symmetrical and antisymmetrical states when $n > 2$.

If P denotes any permutation and ψ any ψ -symbol, we can give a meaning to $P\psi$, the ψ -symbol obtained by operating on ψ with P . We define $P\psi$ to be the ψ -symbol whose representative is (Pg'') , obtained by applying the permutation P to the representative (q'') of ψ . This $P\psi$ is independent of the representation used for defining it, as follows from equation (9). Further, the operation by which $P\psi$ is obtained from ψ is a linear one. Hence we can regard $P\psi$ as the product of an observable P with ψ , *i.e.* we can regard the permutation P as an observable.

There are $n!$ permutations, each of which can be regarded as an

observable. One of them, P_1 say, is the identical permutation, which is equal to unity. If ψ denotes a symmetrical state, we have

$$P\psi = \psi \quad (13)$$

for any P , and hence a symmetrical ψ is an eigen- ψ of every permutation belonging to the eigenvalue unity. Similarly an anti-symmetrical ψ is an eigen- ψ of every permutation belonging to the eigenvalue ± 1 according to whether the permutation is even or odd. The product of any two permutations is a third permutation and hence any function of the permutations is reducible to a linear function of them. Any permutation P has a reciprocal P^{-1} satisfying $PP^{-1} = P^{-1}P = P_1 = 1$.

A permutation P , like any other observable, can be represented by a matrix. Its q -representative $(q'|P|q'')$ will satisfy

$$\int (q'|P|q'') dq'' (q'') = (Pq')$$

and hence

$$(q'|P|q'') = \delta(Pq' - q'') \quad (14)$$

$$= \delta(q' - P^{-1}q''). \quad (15)$$

The δ function in (14) or (15) denotes the product of n factors of the type $\delta(\{Pq'\}_r - q'_r)$ or $\delta(q'_r - \{P^{-1}q''\}_r)$ respectively. The conjugate complex of P is given by

$$(q'|\bar{P}|q'') = (q''|P|q') \cdots \delta(q'' - P^{-1}q') \\ = (q'|\bar{P}^{-1}|q'')$$

from (15) and (14), so that

$$\bar{P} = P^{-1} \quad (16)$$

Thus a permutation is not in general a real observable, its conjugate complex being equal to its reciprocal.

Any permutation of the numbers 1, 2, 3, ..., n may be expressed in the cyclic notation, e.g. with $n = 8$

$$P_a = (143)(27)(58)(6), \quad (17)$$

in which each number is to be replaced by the succeeding number in a bracket, unless it is the last in a bracket, when it is to be replaced by the first in that bracket. Thus P_a changes the numbers 12345678 into 47138625. The type of any permutation is specified by the partition of the number n which is provided by the number of numbers in each of the brackets. Thus the type of P_a is specified by the partition $8 = 3 + 2 + 2 + 1$. Permutations of the same type, i.e. corre-

sponding to the same partition, we shall call *similar*. Thus, for example, P_a in (17) is similar to

$$P_b = (871)(35)(46)(2). \quad (18)$$

The whole of the $n!$ possible permutations may be divided into sets of similar permutations, each such set being called a *class*. The permutation $P_1 = 1$ forms a class by itself. Any permutation is similar to its reciprocal.

When two permutations P_a and P_b are similar, either of them P_b may be obtained by making a certain permutation P in the other P_a . Thus, in our example (17), (18) we can take P to be the permutation that changes 14327586 into 87135462, *i.e.* the permutation

$$P = (18623)(475).$$

We then have the algebraic relation between P_a and P_b

$$P_b = PP_aP^{-1}. \quad (19)$$

To verify this, we observe that the product $P_a\psi$ of P_a with any ψ is changed into $P_b\psi$ if one applies the permutation P to the P_a in the product but not to the ψ . If we multiply the product by P on the left, we are applying this permutation to the whole ψ -symbol $P_a\psi$ and thus to both the P_a and the ψ , so that we must insert another factor P^{-1} between the P_a and the ψ , giving us $PP_aP^{-1}\psi$ to equate to $P_b\psi$. An alternative proof consists in noting that when the permutation P is applied to the representative $\delta(P_aq' - q'')$ of P_a , it gives $\delta(PP_aq' - Pq'')$ or $\delta(PP_aP^{-1}q' - q'')$, which is just the representative of PP_aP^{-1} .

Equation (19) is the general formula showing when two permutations P_a and P_b are similar. Of course P is not uniquely determined when P_a and P_b are given, but the existence of any P satisfying (19) is sufficient to show that P_a and P_b are similar.

§ 64. Permutations as Constants of the Motion

A permutation P may be considered as an observable at each instant of time and may therefore be considered as a dynamical variable. Let us see how P varies with the time. The fact that the Hamiltonian is symmetrical leads at once to the equation

$$PH = HP, \quad (20)$$

as may be verified by a similar argument to that used for equation

(19), or alternatively by a direct application of the matrix representatives. Thus from (14)

$$(q'|PH|q'') = \int \delta(Pq' - q''') dq''' (q'''|H|q'') = (Pq'|H|q'')$$

and from (15)

$$(q'|HP|q'') = \int (q'|H|q''') dq''' \delta(q''' - P^{-1}q'') = (q'|H|P^{-1}q''),$$

and the two right-hand sides are now equal from (1). Equation (20) shows that *each permutation is a constant of the motion*. The P 's are still constants when arbitrary perturbations are applied to the system, provided the perturbing energy to be added to the Hamiltonian is symmetrical. Thus the constancy of the P 's is absolute.

In dealing with any system in quantum mechanics, when we have found a constant of the motion α , we know that if for any state, α initially has the numerical value α' , then it always has this value, so that we can assign different numbers α' to the different states and so obtain a classification of the states. The procedure is not so straightforward, however, when we have several constants of the motion α which do not commute (as is the case with our permutations P), since we cannot assign numerical values for all the α 's simultaneously to any state. Let us first take the case of a system whose Hamiltonian does not involve the time explicitly. The existence of constants of the motion α which do not commute is then a sign that the system is degenerate. We must now look for a function β of the α 's which has one and the same numerical value β' for all those states belonging to one energy-level H' , so that we can use β for classifying the energy-levels of the system. We can express the condition for β by saying that it must be a function of H , according to the general definition of a function of an observable, so that β must commute with every observable that commutes with H , *i.e.* with every constant of the motion. If the α 's are the only constants of the motion, or if they are a set that commute with all other independent constants of the motion, our problem reduces to finding a function β of the α 's which commutes with all the α 's. We can then assign a numerical value β' for β to each energy-level of the system. If we can find several such functions β , they must all commute with each other, so that we can give them all numerical values simultaneously and obtain a complete classification of the energy-levels. When the Hamiltonian involves the time explicitly one cannot talk about energy-levels, but the β 's will still give a useful classification for the states.

We follow this method in dealing with our permutations P . We must find a function χ of the P 's such that $P\chi P^{-1} = \chi$ for every P . It is evident that a possible χ is ΣP_c , the sum of all the permutations in a certain class c , i.e. the sum of a set of similar permutations, since ΣPP_cP^{-1} must consist of the same permutations summed in a different order. There will be one such χ for each class. Further, there can be no other independent χ , since an arbitrary function of the P 's can be expressed as a linear function of them with numerical coefficients, and it will not then commute with every P unless the coefficients of similar P 's are always the same. We thus obtain all the χ 's that can be used for classifying the states. It is convenient to define each χ as an average instead of a sum, thus

$$\chi_c = n_c^{-1} \Sigma P_c,$$

where n_c is the number of P 's in the class c . An alternative expression for χ_c is

$$\chi_c = n!^{-1} \Sigma_P PP_c P^{-1}, \quad (21)$$

the summation being extended over all the $n!$ permutations P . For each permutation P there is one χ , $\chi(P)$ say, equal to the average of all permutations similar to P . One of the χ 's is $\chi(P_1) = 1$.

The constants of the motion $\chi_1, \chi_2, \dots, \chi_m$ obtained in this way will each have a definite numerical value for every stationary state of the system, in the case when the Hamiltonian does not involve the time explicitly, and also in the general case can be used for classifying the states, there being one set of states for every permissible set of numerical values $\chi'_1, \chi'_2, \dots, \chi'_m$ for the χ 's. Since the χ 's are absolute constants of the motion, these sets of states will be *exclusive*, i.e. transitions will never take place from a state in one set to a state in another.

The permissible sets of values χ' that one can give to the χ 's are limited by the fact that there exist algebraic relations between the χ 's. The product of any two χ 's, $\chi_p \chi_q$, is of course expressible as a linear function of the P 's, and since it commutes with every P it must be expressible as a linear function of the χ 's, thus

$$\chi_p \chi_q = a_1 \chi_1 + a_2 \chi_2 + \dots + a_m \chi_m, \quad (22)$$

where the a 's are numbers. Any numerical values χ' that one gives to the χ 's must be eigenvalues of the χ 's and must satisfy these same algebraic equations. For every solution χ' of these equations there is one exclusive set of states. One solution is evidently $\chi'_p = 1$ for

every χ_p , and this gives the set of symmetrical states satisfying (13). A second obvious solution is $\chi'_p = \pm 1$, the + or - sign being taken according to whether the permutations in the class p are even or odd, and this gives the set of antisymmetrical states. The other solutions may be worked out in any special case by ordinary algebraic methods, as the coefficients a in (22) may be obtained directly by a consideration of the types of permutation to which the χ 's concerned refer. Any solution is, apart from a certain factor, what is called in group theory a *character* of the group of permutations. The χ 's are all real observables, since each P and its conjugate complex P^{-1} are similar and will occur added together in the definition of any χ , so that the χ 's must be all real numbers.

The number of possible solutions of the equations (22) may easily be determined, since it must equal the number of different eigenvalues of an arbitrary function B of the χ 's. We can express B as a linear function of the χ 's with the help of equations (22); thus

$$B = b_1\chi_1 + b_2\chi_2 + \dots + b_m\chi_m. \quad (23)$$

Similarly we can express each of the quantities $B^2, B^3 \dots B^m$ as a linear function of the χ 's. From these m equations, together with the equation $\chi(P_1) = 1$, we can eliminate the m unknowns $\chi_1, \chi_2 \dots \chi_m$, obtaining as result an algebraic equation of degree m for B ,

$$B^m + c_1 B^{m-1} + c_2 B^{m-2} + \dots + c_m = 0.$$

The m solutions of this equation give the m possible eigenvalues for B , each of which will, according to (23), be a linear function of $b_1, b_2 \dots b_m$ whose coefficients are a permissible set of values $\chi'_1, \chi'_2 \dots \chi'_m$. These sets of values χ' thus obtained must be all different, since if there were fewer than m different permissible sets of values χ' for the χ 's there would exist a linear function of the χ 's every one of whose eigenvalues vanishes, which would mean that the linear function itself vanishes and the χ 's are not linearly independent. Thus the number of permissible sets of numerical values for the χ 's is just equal to m , which is the number of classes of permutations or the number of partitions of n . This number is therefore the number of exclusive sets of states.

The properties of the P 's which are not properties of the χ 's will only describe the degeneracy of the states, in the case of a system whose Hamiltonian does not involve the time explicitly. If ψ denotes any stationary state, $f(P)\psi$, where $f(P)$ is any function of

the permutations, will denote another stationary state belonging to the same energy-level, except when it vanishes identically. By expanding $f(P)\psi$ in terms of a complete set of independent stationary states belonging to this energy-level, we get a representation of $f(P)$ and thus of each P . In this way we see that, if we obtain a matrix representation of all the P 's consistent with each of the χ 's being a certain number χ' , then the number of rows and columns of the matrices will be the degree of degeneracy of the states in the exclusive set χ' , *i.e.* the number of independent states belonging to each energy-level. This degeneracy is an *essential* one and cannot be removed by any perturbation that is symmetrical between all the similar particles. The states ψ and $f(P)\psi$ are observationally indistinguishable, since any observation that can actually be made must consist in measuring an observable that is symmetrical between the similar particles and therefore commutes with $f(P)$. This remark applies also when the Hamiltonian involves the time explicitly.

§ 65. Determination of Energy-levels

Let us apply the perturbation method of § 51 and make a first-order calculation of the energy-levels in the case when the Hamiltonian does not involve the time explicitly. We suppose that for our unperturbed states each of the similar particles has its own 'orbit', represented by a wave function $(q'|\alpha)$ involving only the co-ordinates q' of this one particle. We shall have altogether n orbits, one for each particle, which we assume for the present to be all different, and label $\alpha_1, \alpha_2 \dots \alpha_n$. The wave function representing an unperturbed state of the whole system will then be the product (10). If we apply an arbitrary permutation P_a to the α 's, we shall obtain another wave function

$$(q'_1|\alpha_r)(q'_2|\alpha_s)\dots(q'_n|\alpha_t) = (q'|P_a\alpha), \quad (24)$$

representing another unperturbed state with the same energy. There are thus altogether $n!$ unperturbed states with this energy, if we assume there are no other causes of degeneracy. According to the method of § 51 when the unperturbed system is degenerate, we must consider those elements of the matrix representing the perturbing energy V that refer to two states with the same energy, *i.e.* those of the type $(P_a\alpha|V|P_b\alpha)$ where P_a and P_b are two permutations of the α 's. These will form a matrix with $n!$ rows and columns,

whose eigenvalues are the first-order corrections in the energy-levels.

It is necessary in the present discussion to distinguish between the two kinds of permutations, those of the q 's and those of the α 's. The essential difference between them can perhaps be seen most clearly in the following way. Let us consider a permutation in the general case, say that consisting of the interchange of 2 and 3. This may be interpreted either as the interchange of the objects 2 and 3 or as the interchange of the objects in the places 2 and 3, these two operations producing in general quite different results. The first of these interpretations is the one we have been using up to the present, the objects concerned being the q 's in the representative of a state. A permutation with this interpretation can be applied to an arbitrary function of the q 's. A permutation with the second interpretation has a meaning, however, when applied to a function of the q 's only if each of the q 's has a definite specifiable place in the function. This is not the case for a general function of the q 's, but it is the case for any of the $n!$ functions of the type (24), the place of each q being specified by the α with which it is bracketed. Any permutation applied to the q 's in given places now produces the same result as the reciprocal permutation applied to the α 's. A permutation of the q 's (*i.e.* one with the first interpretation), since it can be applied to any function of the q 's, *i.e.* to the representative of any ψ -symbol, may be regarded as an ordinary observable. On the other hand, a permutation of places or of the α 's can be considered as an observable only in a very restricted sense, since it has a meaning only when multiplied into a ψ -symbol whose representative is one of the $n!$ wave functions (24) or some linear combination of them. We denote such a permutation of the α 's, considered as an observable in this restricted sense, by the symbol P^α .

We can form algebraic functions of the observables P^α which will be other observables in the same restricted sense. In particular we can form $\chi(P_a^\alpha)$, the average of all P^α 's similar to P_a^α . This must equal $\chi(P_a)$, the average of the similar permutations of the q 's, since the total set of all permutations of a given type must evidently be the same whether the permutations are applied to the objects q or to the places α .

If we set up arbitrarily a one-one correspondence between the q 's and the α 's, as is done automatically when we label both the q 's and

the α 's by the numbers 1, 2, 3, ..., n , as in (10), then, if we have any permutation of the q 's, we can give a meaning to this same permutation of the α 's. This meaning is such that

$$(q|\alpha) = (Pq|P\alpha).$$

In this equation we can apply a permutation P_a to the α 's on both sides, which will give us

$$(q|P_a \alpha) = (Pq|P_a P\alpha), \quad (25)$$

an equation which shows us the connexion between permutations of the q 's and those of the α 's when applied to the wave function (24).

The matrix $(P_a \alpha|V|P_b \alpha)$, which we must now study, may be obtained from the matrix $(q'|V|q'')$ representing V by a canonical transformation, in which the transformation functions are just $(q'|P_a \alpha)$, the wave function (24), and its conjugate complex $(P_a \alpha|q')$, provided these functions are properly normalized. Thus

$$(P_a \alpha|V|P_b \alpha) = \iint (P_a \alpha|q') dq' (q'|V|q'') dq'' (q''|P_b \alpha). \quad (26)$$

Again, for arbitrary P ,

$$\begin{aligned} (P_a P\alpha|V|P_b P\alpha) &= \iint (P_a P\alpha|q') dq' (q'|V|q'') dq'' (q''|P_b P\alpha) \\ &= \iint (P_a P\alpha|Pq') dq' (Pq'|V|Pq'') dq'' (Pq''|P_b P\alpha), \end{aligned}$$

when we apply the permutation P to the variables of integration q' and q'' . With the help of (25), this reduces to

$$(P_a P\alpha|V|P_b P\alpha) = \iint (P_a \alpha|q') dq' (Pq'|V|Pq'') dq'' (q''|P_b \alpha). \quad (27)$$

Now since V is symmetrical between all the particles, we must have

$$(q'|V|q'') = (Pq'|V|Pq''),$$

like (1), and hence, comparing (26) and (27), we obtain

$$(P_a \alpha|V|P_b \alpha) = (P_a P\alpha|V|P_b P\alpha). \quad (28)$$

Let $(P\alpha|V|\alpha) = V_P$ for brevity. Then, taking $P = P_b^{-1}$ in (28), we obtain

$$(P_a \alpha|V|P_b \alpha) = (P_a P_b^{-1}|V|\alpha) = V_{P_a P_b^{-1}}.$$

Thus the general matrix element $(P_a \alpha|V|P_b \alpha)$ depends only on the ratio $P_a P_b^{-1}$, and of the total of $(n!)^2$ matrix elements there are only $n!$ different ones. The coefficient of any V_P in this matrix will be a matrix, each of whose elements is 0 or 1, the 1 occurring when

$$(P_a \alpha|V|P_b \alpha) = V_P,$$

i.e. when $P_a P_b^{-1} = P$. But this matrix, multiplied into any wave function $(q|P_b \alpha)$, gives the result $(q|P_a \alpha)$ with $P_a P_b^{-1} = P$, i.e. it gives the result $(q|PP_b \alpha)$, so that it is precisely the matrix repre-

senting the observable P^α or the permutation P applied to the α 's. Thus the whole matrix $(P_a^\alpha | V | P_b^\alpha)$ is equal to the matrix representing $\sum_P V_P P^\alpha$, where the summation is over all the $n!$ permutations P , and we can put

$$V = \sum_P V_P P^\alpha. \quad (29)$$

This formula shows that the perturbing energy V is equal to a linear function of the permutation observables P^α with numerical coefficients V_P . It is, of course, only an approximate formula, as it holds only with neglect of those matrix elements of V that refer to two different energy-levels of the unperturbed system. It can, however, be used for the calculation of the energy-levels in the first approximation, and is very convenient for this purpose as the expression $\sum_P V_P P^\alpha$ is easily handled. This expression, it should be remembered, is an observable only in the restricted sense mentioned above, but this sense is sufficiently general for equation (29) to be valid with neglect of those matrix elements of V referring to two different energy-levels of the unperturbed system.

As an example of an application of (29) we shall determine the average energy of all those states arising from a given state of the unperturbed system that belong to one exclusive set. This requires us to calculate the average eigenvalue of V when the χ 's have specified numerical values χ' . Now the average eigenvalue of P_a^α equals that of $P^\alpha P_a^\alpha P^{\alpha^{-1}}$ for arbitrary P^α and thus equals that of $n!^{-1} \sum_{P^\alpha} P^\alpha P_a^\alpha P^{\alpha^{-1}}$, which is $\chi'(P_a^\alpha)$ or $\chi'(P_a)$. Hence the average eigenvalue of V is $\sum_P V_P \chi'(P)$. A similar method could be used for calculating the average eigenvalue of any function of V , it being only necessary to replace each P^α by $\chi(P)$ to perform the averaging.

The number of energy-levels in an exclusive set $\chi = \chi'$ that arise from a given state of the unperturbed system is equal to the number of eigenvalues of (29) that are consistent with the equations $\chi = \chi'$. This number is the number of rows and columns in a representation of the P 's in which each $\chi = \chi'$, which number, from the result at the end of the preceding section, is just the degree of degeneracy of the states in this set.

The modifications required in the theory when the orbits $\alpha_1, \alpha_2, \dots, \alpha_n$ of the undisturbed system are not all different may easily be made. Suppose, for example, that α_1 and α_2 are the same. Then the permutation P_{12}^α that causes an interchange of α_1 and α_2 must equal unity. Only functions of the P^α 's that commute with P_{12}^α now have

a meaning. This, however, is sufficient for us to be able to follow out the same sort of argument as before, and obtain a result of the same form (29). The term in the summation in (29) that involves the permutation P_{12}^α now does not occur, since it could be added on to the term involving the identical permutation P_1^α . For the remaining terms, any two terms P_a^α and P_b^α must have the same coefficient if the permutations P_a^α and P_b^α can be obtained from one another by the interchange of α_1 and α_2 . This results in $\sum_P V_P P^\alpha$ commuting with P_{12}^α and thus having a meaning. The condition $P_{12}^\alpha = 1$ will impose restrictions on the possible numerical values χ' that the χ 's can have and will reduce the number of characters.

§ 66. Application to Electrons

Let us now consider the case when the similar particles are electrons. This requires, according to Pauli's exclusion principle discussed in § 62, that we take into account only the antisymmetrical states. It is now necessary to make explicit reference to the spin properties of the electrons. The effect of the spin on the motion of an electron in an electromagnetic field is not very great. There will be additional forces on the electron due to its magnetic moment, requiring additional terms in the Hamiltonian. The spin angular momentum will not have any direct action on the motion, but it will come into play when there are forces tending to rotate the magnetic moment, since the magnetic moment and angular momentum are constrained to be always in the same direction. These effects are all small, however, of the same order of magnitude as that of the relativity variation of mass with velocity, so there would be no point in taking them into account in a non-relativity theory. The importance of the spin lies not in these small effects on the motion of the electron, but in the fact that it gives two internal states to the electron, corresponding to the two possible values of the spin component in any assigned direction, which causes a doubling in the number of independent states of an electron moving in a given field. This fact has far-reaching consequences when combined with Pauli's exclusion principle.

Let us take a representation in which the diagonal observables q_r , describing the r -th electron are its three Cartesian co-ordinates x, y, z , and the z -component σ_z of its spin vector σ , which was introduced in § 43. The representative of a state will now be

$$(x_1, x_2 \dots x_n, \sigma_1, \sigma_2 \dots \sigma_n |) = (x\sigma |), \quad (30)$$

the single variable x being written instead of x, y, z and the suffix z being dropped from σ_z 's that occur in representatives. The exclusion principle requires that (30) shall be antisymmetrical in the x 's and σ 's together, *i.e.* if any permutation is applied to the x 's and also to the σ 's, (30) must remain unchanged or change sign according to whether the permutation is even or odd. In symbols

$$(x, \sigma|) = \pm (Px, P\sigma|) \quad (31)$$

for any permutation P . Thus *even if we neglect the spin forces in the Hamiltonian, we must take the spin variables into account in order to determine what states are allowed by the exclusion principle.*

If the theory of the three preceding sections is applied directly to the case of electrons, it will not give anything of interest, since all the allowed states are eigenstates of any permutation belonging to the eigenvalue ± 1 . We may, however, consider permutations P which operate on the x -variables alone in the representative of a state, and apply our theory to these. Such permutations may also be considered as observables. Further, they are also constants of the motion when we neglect the terms in the Hamiltonian that arise from the spin forces, which neglect results in the Hamiltonian not involving the spin observables σ . Hence with these permutations P we can again introduce the χ 's, equal to the average of all of the P 's in each class, and assert that for any permissible set of numerical values χ' for the χ 's there will be one exclusive set of states. Thus there exist these exclusive sets of states for systems containing many electrons even when we restrict ourselves to a consideration of only those states that satisfy Pauli's principle. The exclusiveness of the sets of states is now, of course, only approximate, since the χ 's are constants only so long as we neglect the spin forces. There will actually be a small probability for a transition from a state in one set to a state in another.

From (31) we obtain

$$PP^\sigma = \pm 1, \quad (32)$$

where P denotes any permutation which operates on the x -variables and P^σ the same permutation operating on the σ -variables in the representative of a state. There is thus a simple connexion between the P 's and P^σ 's, which means that instead of studying the observables P we can get all the results we want, *e.g.* the characters χ' , by studying the observables P^σ . The P^σ 's are much easier to study on

account of the fact that the σ variables in the wave function have domains consisting each of only the two points 1 and -1 , which are the two eigenvalues of each σ_z . This fact results in there being fewer characters χ' for the group of permutations of the σ -variables than for the group of general permutations, since it prevents a function of the variables $\sigma_1, \sigma_2, \dots, \sigma_n$ from being antisymmetrical in more than two of them.

The study of the observables P^σ is made specially easy by the fact that we can express them as algebraic functions of the observables σ . Consider the quantity

$$O_{12} = \frac{1}{2}\{1 + (\sigma_1, \sigma_2)\}.$$

With the help of equations (42) of § 43 we find readily that

$$(\sigma_1, \sigma_2)^2 = (\sigma_{1x} \sigma_{2x} + \sigma_{1y} \sigma_{2y} + \sigma_{1z} \sigma_{2z})^2 = 3 - 2(\sigma_1, \sigma_2), \quad (33)$$

and hence that

$$O_{12}^2 = \frac{1}{4}\{1 + 2(\sigma_1, \sigma_2) + (\sigma_1, \sigma_2)^2\} = 1. \quad (34)$$

Again, we find

$$\begin{aligned} O_{12} \sigma_{1x} &= \frac{1}{2}\{\sigma_{1x} + \sigma_{2x} - i\sigma_{1z} \sigma_{2y} - i\sigma_{1y} \sigma_{2z}\} \\ \sigma_{2x} O_{12} &= \frac{1}{2}\{\sigma_{2x} + \sigma_{1x} + i\sigma_{1y} \sigma_{2z} - i\sigma_{1z} \sigma_{2y}\} \end{aligned}$$

and hence

$$O_{12} \sigma_{1x} = \sigma_{2x} O_{12}.$$

Similar relations hold for σ_{1y} and σ_{1z} so that we have

$$O_{12} \sigma_1 = \sigma_2 O_{12}$$

or

$$O_{12} \sigma_1 O_{12}^{-1} = \sigma_2.$$

From this we can obtain with the help of (34)

$$O_{12} \sigma_2 O_{12}^{-1} = \sigma_1.$$

These commutability relations for O_{12} with σ_1 and σ_2 are precisely the same as those for P_{12}^σ , the permutation consisting of the interchange of the spin variables of electrons 1 and 2. Thus we can put

$$O_{12} = c P_{12}^\sigma,$$

where c is a number. Equation (34) shows that $c = \pm 1$. To determine which of these values for c is the correct one, we observe that the eigenvalues of P_{12}^σ are 1, 1, 1, -1 , corresponding to the fact that there exist three independent symmetrical and one antisymmetrical function of the two variables σ_{1z}, σ_{2z} , namely, with the notation of § 43, the three symmetrical functions $f_\alpha(\sigma_1) f_\alpha(\sigma_2)$, $f_\beta(\sigma_1) f_\beta(\sigma_2)$, $f_\alpha(\sigma_1) f_\beta(\sigma_2) + f_\beta(\sigma_1) f_\alpha(\sigma_2)$, and the one antisymmetrical function

$f_\alpha(\sigma_1)f_\beta(\sigma_2) - f_\beta(\sigma_1)f_\alpha(\sigma_2)$. Thus the mean of the eigenvalues of P_{12}^σ is $\frac{1}{2}$. Now the mean of the eigenvalues of (σ_1, σ_2) is evidently zero and hence the mean of the eigenvalues of O_{12} is $\frac{1}{2}$. Thus we must have $c = +1$, and so we can put

$$P_{12}^\sigma = \frac{1}{2}\{1 + (\sigma_1, \sigma_2)\}.$$

In this way any permutation P^σ consisting simply of an interchange can be expressed as an algebraic function of the σ 's. Any other permutation P^σ can be expressed as a product of interchanges and can therefore also be expressed as a function of the σ 's. With the help of (32) we can now express the P 's as algebraic functions of the σ 's and eliminate the P^σ 's from the discussion. We have, since the — sign must be taken in (32) when the permutations are interchanges and since the square of an interchange is unity,

$$P_{12} = -\frac{1}{2}\{1 + (\sigma_1, \sigma_2)\}. \quad (35)$$

The formula (35) may conveniently be used for the evaluation of the characters χ' which define the exclusive sets of states. We have for example, for the permutations consisting of interchanges

$$\chi_{12} = \chi(P_{12}) = -\frac{1}{2}\left\{1 + \frac{2}{n(n-1)} \sum_{r < t} (\sigma_r, \sigma_t)\right\}.$$

If we introduce the observable s to describe the magnitude of the total spin angular momentum, $\frac{1}{2}\sum_r \sigma_r$ in units of \hbar , through the formula

$$s^2 - \frac{1}{4} = (\frac{1}{2}\sum_r \sigma_r, \frac{1}{2}\sum_t \sigma_t),$$

analogous to equation (12) of Chapter VIII, we have

$$\begin{aligned} 2\sum_{r < t} (\sigma_r, \sigma_t) &= (\sum_r \sigma_r, \sum_t \sigma_t) - \sum_r (\sigma_r, \sigma_r) \\ &= 4s^2 - 1 - 3n. \end{aligned}$$

Hence

$$\chi_{12} = -\frac{1}{2}\left\{1 + \frac{4s^2 - 1 - 3n}{n(n-1)}\right\} = -\frac{n(n-4) + 4s^2 - 1}{2n(n-1)}. \quad (36)$$

Thus χ_{12} is expressible as a function of the observable s and of n the number of electrons. Any of the other χ 's could be evaluated on similar lines and would have to be a function of s and n only, since there are no other symmetrical functions of all the σ observables which could be involved. There is therefore one set of numerical values χ' for the χ 's, and thus one exclusive set of states, for each eigenvalue s' of s . The eigenvalues of s are

$$\frac{1}{2}n + \frac{1}{2}, \frac{1}{2}n - \frac{1}{2}, \frac{1}{2}n - \frac{3}{2} \dots$$

the series terminating with $\frac{1}{2}$ or 1.

We see in this way that each of the stationary states of a system with several electrons is an eigenstate of s , the magnitude in units of \hbar , of the total spin angular momentum $\frac{1}{2}\Sigma_r \sigma_r$, belonging to a definite eigenvalue s' . For any given s' there will be $2s'$ possible values for a component of the total spin vector in any direction and these will correspond to $2s'$ independent stationary states with the same energy. When we do not neglect the forces due to the spin magnetic moments these $2s'$ states will in general be split up into $2s'$ states with slightly different energies, and will thus form a multiplet of multiplicity $2s'$. Transitions in which s' changes, *i.e.* transitions from one multiplicity to another, cannot occur when the spin forces are neglected and will have only a small probability of occurrence when the spin forces are not neglected.

We can determine the energy-levels of a system with several electrons to the first approximation by using formula (29). If we consider only the Coulomb forces between the electrons, then the interaction energy V will consist of a sum of parts each referring to only two electrons, which will result in all the matrix elements V_P vanishing except those for which P is the identical permutation or is simply an interchange of two electrons. Thus (29) will reduce to

$$V = V_1 + \sum_{r < s} V_{rs} P_{rs}^\alpha, \quad (37)$$

V_{rs} being the matrix element referring to the interchange of orbits r and s . Since the P^α 's have the same properties as the P 's, any function of the P^α 's will have the same eigenvalues as the corresponding function of the P 's, so that the right-hand side of (37) will have the same eigenvalues as

$$V_1 + \sum_{r < s} V_{rs} P_{rs}$$

or

$$V_1 - \frac{1}{2} \sum_{r < s} V_{rs} \{1 + (\sigma_r, \sigma_s)\} \quad (38)$$

from (35). The eigenvalues of (38) will give the first-order corrections in the energy-levels. The form of (38) shows that a model which assumes a coupling energy between the spins of the various electrons, of magnitude $-\frac{1}{2}V_{rs}(\sigma_r, \sigma_s)$ for the electrons in the r and s orbits, would meet with a fair amount of success. This coupling energy is much greater than that of the spin magnetic moments. Such models of the atom were in use before the justification by quantum mechanics was obtained.

If two of the orbits of our unperturbed system are the same, say the orbits α_1 and α_2 are the same, we must take only those eigenvalues

of (37) that are consistent with $P_{12}^{\alpha} = 1$, or those eigenvalues of (38) consistent with $P_{12} = 1$ or $P_{12}^{\sigma} = -1$. This means we must take only those eigenvalues of (38) belonging to eigenfunctions that are simultaneously eigenfunctions of P_{12}^{σ} belonging to the eigenvalue -1 , *i.e.* eigenfunctions that are antisymmetrical in σ_1 and σ_2 . Thus we may say that the two electrons in the orbits α_1 and α_2 have their spins antiparallel. The case of more than two orbits the same cannot occur with electrons.

XII

THEORY OF RADIATION

§ 67. Theory of Einstein-Bose Assemblies

In Chapter X a theory was given of the scattering, absorption, and emission of a particle by an atomic system. The interaction of the particle and atomic system was assumed to be describable by an interaction energy V appearing in the Hamiltonian, which interaction energy had to be small but was otherwise arbitrary. If we could determine the energy of interaction between a photon and an atom or molecule, we could apply the methods of Chapter X immediately to the case when the incident particle is a photon. In this way we could obtain a theory of the interaction of light with an atomic system. We cannot determine this energy of interaction directly from analogy with the classical theory, in the way we obtained the Hamiltonians for most of the systems dealt with up to the present, since the phenomenon of the interaction of a photon with an atom has no analogue in the classical theory. We must proceed in a more indirect way. We know that the interaction of an atom with a field of radiation can be described approximately by classical electrodynamics when the field of radiation consists of a large number of photons. Our method is therefore to assume an arbitrary interaction energy V between a single photon and the atom and then in terms of V to investigate the interaction of a large number of photons with the atom. By comparing this interaction with that given by classical electrodynamics we can then obtain V .

Our problem now is thus to deal in general terms with the interaction of a large number of photons with an atom. This problem, it is important to observe, is a generalization of that of Chapter X, in spite of the fact that we then often considered a large number of incident particles. The incident particles of Chapter X were all independent and *each had its own scatterer*. In fact they were only introduced to help us to picture one actual incident particle interacting with one scatterer. We now have a large number of actual photons all interacting with the same atom. Also our photons are not independent of one another since, even if there are no forces between them describable by an interaction energy, they are, as we saw in the preceding chapter, such that only states that are sym-

metrical between them occur in nature, *i.e.* they satisfy the Einstein-Bose statistics.

Let us first consider the problem of an assembly of n similar systems of any kind that satisfy the Einstein-Bose statistics and are all perturbed by some external field of force. If we take a representation in which sets of observables q_1, q_2, \dots, q_n describing the first, second, ... last system respectively, are diagonal, the representative $(q'_1 q'_2 \dots q'_n)$ of any state must be symmetrical in the variables q'_1, q'_2, \dots, q'_n . Suppose the eigenvalues of any of the q 's, q_r , say, are $q^{(1)}, q^{(2)}, q^{(3)}, \dots$, which we assume for definiteness to be discrete. These eigenvalues must be the same for each of the n systems, *i.e.* they must be independent of r . (They will each be in general a set of numbers, consisting of an eigenvalue of each of the set of commuting observables q_r .) If we now have any *symmetrical* function of the variables q'_1, q'_2, \dots, q'_n , each point in the domain of this function can be specified by n'_1, n'_2, n'_3, \dots , the numbers of q' 's equal to $q^{(1)}, q^{(2)}, q^{(3)}, \dots$ respectively. The variables $n'_1, n'_2, n'_3 \dots$ will do just as well as the variables q'_1, q'_2, \dots, q'_n , so long as we are dealing only with symmetrical functions. Thus the representatives of states of our assembly satisfying the Einstein-Bose statistics may be expressed as functions of the variables $n'_1, n'_2, n'_3 \dots$ instead of the variables q'_1, q'_2, \dots, q'_n . This change is effectively a canonical transformation to a new representation in which the rows and columns of the matrices are labelled by the observables n_1, n_2, n_3, \dots which observables are the numbers of systems with q 's equal to $q^{(1)}, q^{(2)}, q^{(3)} \dots$ respectively, or, as we may say, the numbers of systems in the states $q^{(1)}, q^{(2)}, q^{(3)} \dots$. Since the new observables $n_1, n_2, n_3 \dots$ are functions of the q_1, q_2, \dots, q_n (non-analytic functions, it is true), the transformation is of the trivial kind consisting essentially of a relabelling of the rows and columns and the only change to be made in the representative of a state will be that arising from the change in the weights of the different points of its domain. To determine this change we use the condition

$$\sum_{n_1 n_2 \dots} |(n_1 n_2 \dots)|^2 = \sum_{q_1 q_2 \dots q_n} |(q_1 q_2 \dots q_n)|^2,$$

from which we can infer that

$$|(n_1 n_2 \dots)|^2 = \sum |(q_1 q_2 \dots q_n)|^2, \quad (1)$$

the summation in (1) being over all values of the q 's such that n_1 of them are equal to $q^{(1)}$, n_2 equal to $q^{(2)}$, and so on. The number of

terms in the summation in (1) is $n!/(n_1! n_2! n_3! \dots)$ and they are all equal, on account of $(q_1 q_2 \dots q_n|)$ being symmetrical. It is thus clear that we must take

$$(n_1 n_2 \dots |) = [n!/n_1! n_2! n_3! \dots]^{\frac{1}{2}} (q_1 q_2 \dots q_n|). \quad (2)$$

The question of interest now is to express the Hamiltonian of the system in terms of the new observables $n_1, n_2, n_3 \dots$. We can do this by writing down its representative in the q -representation and transforming to the n -representation. Since the transformation is of an unusual kind, the most convenient way of making it is to write down the whole Schrödinger equation and to transform that. This Schrödinger equation is

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (q_1 q_2 \dots q_n|) &= \\ &= \sum_{q'_1 q'_2 \dots q'_n} (q_1 q_2 \dots q_n | H | q'_1 q'_2 \dots q'_n) (q'_1 q'_2 \dots q'_n |). \end{aligned} \quad (3)$$

The Hamiltonian H is of the form

$$H = \sum_r U_r,$$

where U_r is the energy associated with the r -th system, consisting of its proper energy together with its interaction energy with the external field of force, and is a function of the dynamical variables of the r -th system only. The representative of U_r in the q_r -representation will be $(q'_r | U_r | q''_r)$, which will be a matrix independent of r , i.e. the same for each of the n systems. Its elements may also be written $(q^{(a)} | U | q^{(b)})$ or U_{ab} for brevity. The representative of U_r in the complete q -representation will be

$$\begin{aligned} (q'_1 q'_2 \dots q'_n | U_r | q''_1 q''_2 \dots q''_n) &= \\ &= (q'_r | U_r | q''_r) \delta_{q'_1 q''_1} \delta_{q'_2 q''_2} \dots \delta_{q'_{r-1} q''_{r-1}} \delta_{q'_{r+1} q''_{r+1}} \dots \delta_{q'_n q''_n}. \end{aligned}$$

This makes the Schrödinger equation (3) reduce to

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (q_1 q_2 \dots q_n|) &= \sum_r [(q_r | U_r | q_r) (q_1 q_2 \dots q_n|) + \\ &\quad + \sum_{q'_r \neq q_r} (q_r | U_r | q'_r) (q_1 q_2 \dots q_{r-1} q'_r q_{r+1} \dots q_n|)], \end{aligned} \quad (4)$$

the terms arising from the diagonal matrix elements of H being separated from the non-diagonal ones for convenience later.

If we now make the transformation to the n -representation, using equation (2), equation (4) becomes

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (n_1 n_2 \dots |) &= \sum_r (q_r | U_r | q_r) (n_1 n_2 \dots |) + \\ &\quad + \sum_r \sum_{q'_r \neq q_r} [(n_{q'_r} + 1)/n_{q_r}]^{\frac{1}{2}} (q_r | U_r | q'_r) (n_1 n_2 \dots n_{q_r} - 1 \dots n_{q'_r} + 1 \dots |), \end{aligned} \quad (5)$$

after removal of the factor $[n_1! n_2! n_3! \dots / n!]^{\frac{1}{2}}$ throughout. The sum

$\Sigma_r(q_r|U_r|q_r)$ in (5) means a sum of terms each of the type $(q^{(a)}|U|q^{(a)})$ or U_{aa} , the number of times this type occurs being the number of q 's that equal $q^{(a)}$, which is just n_a . Thus this sum is equal to $\Sigma_a n_a U_{aa}$. Again, the double sum $\Sigma_r \Sigma_{q'_r \neq q_r}$ in (5) consists of terms each of the type $[(n_b+1)/n_a]^\frac{1}{2} U_{ab}(n_1 n_2 \dots n_a - 1 \dots n_b + 1 \dots |)$ with $b \neq a$. The number of times this type occurs is equal to the number of ways of choosing r and q'_r such that $q_r = q^{(a)}$ and $q'_r = q^{(b)}$. This is just n_a , the number of ways of choosing r such that $q_r = q^{(a)}$, since there is always just one way of choosing $q'_r = q^{(b)}$. Equation (5) thus reduces to

$$i\hbar \frac{\partial}{\partial t} (n_1 n_2 \dots |) = \Sigma_a n_a U_{aa} (n_1 n_2 \dots |) + \\ + \Sigma_a \Sigma_{b \neq a} n_a^\frac{1}{2} (n_b + 1)^\frac{1}{2} U_{ab} (n_1 n_2 \dots n_a - 1 \dots n_b + 1 \dots |),$$

which may be written

$$i\hbar \frac{\partial}{\partial t} (n_1 n_2 \dots |) = \\ = \Sigma_{ab} n_a^\frac{1}{2} (n_b + 1 - \delta_{ab})^\frac{1}{2} U_{ab} (n_1 n_2 \dots n_a - 1 \dots n_b + 1 \dots |) \quad (6)$$

if by $(n_1 n_2 \dots n_a - 1 \dots n_b + 1 \dots |)$ when $b = a$ we understand simply $(n_1 n_2 \dots n_a \dots |)$.

The eigenvalues of each of our new dynamical variables n_1, n_2, \dots are the integers 0, 1, 2, 3.... They are thus the same, apart from the factor \hbar , as those of the action variable J in the problem of the simple harmonic oscillator, when the arbitrary additive constant in this action variable is chosen as in equation (22) of § 41. Hence each n_a is a dynamical variable of the same nature as the action variable of a simple harmonic oscillator and we can introduce an angle variable w_a canonically conjugate to it, or rather we can introduce e^{iw_a} and e^{-iw_a} . Corresponding to equations (24) of § 41 we shall have

$$\left. \begin{aligned} e^{iw_a} n_a &= (n_a - 1) e^{iw_a} \\ e^{-iw_a} n_a &= (n_a + 1) e^{-iw_a}. \end{aligned} \right\} \quad (7)$$

Also we have that e^{iw_a}, e^{-iw_a} , and n_a commute with e^{iw_b}, e^{-iw_b} , and n_b for $b \neq a$.

The new observables e^{iw_a}, e^{-iw_a} are defined by their matrix representatives in a representation in which n_a is diagonal, like the e^{iw}, e^{-iw} of § 41. From the form of these matrix representatives it follows that when e^{-iw_a} is multiplied into a ψ -symbol whose representative is $(n_1 n_2 \dots n_a \dots |)$, the representative of the product is

$$(n_1 n_2 \dots n_a - 1 \dots |),$$

and when e^{iw_a} is multiplied into this ψ -symbol, the representative of the product is

$$(n_1 n_2 \dots n_a - 1 \dots) \quad \text{for } n_a \geq 1 \\ 0 \quad \text{for } n_a = 0.$$

This means that when e^{-iw_a} and e^{iw_a} are multiplied into ψ -symbols, they are equivalent to the operations of substitution of $n_a + 1$ and $n_a - 1$ for n_a respectively, the second substitution being understood to give the result zero for $n_a = 0$.

We can now express the operator on the right-hand side of (6) explicitly in terms of the n_a and their canonical conjugates w_a . It is, in fact, just

$$\begin{aligned} & \sum_{ab} n_a^{\frac{1}{2}} (n_b + 1 - \delta_{ab})^{\frac{1}{2}} U_{ab} e^{iw_a} e^{-iw_b} \\ &= \sum_{ab} n_a^{\frac{1}{2}} e^{iw_a} U_{ab} (n_b + 1)^{\frac{1}{2}} e^{-iw_b} \end{aligned} \quad (8)$$

with the help of (7). This quantity (8) is our Hamiltonian expressed in terms of the new dynamical variables n_a and w_a . The U_{ab} are, of course, just numerical coefficients.

We can easily generalize this result to apply to a more general type of Hamiltonian, namely, that describing the perturbation of the assembly, not by an external field of force, but by some other atomic system, which we shall call for definiteness the perturber, the reaction of the assembly on the perturber being taken into account. We now have to introduce some more dynamical variables, β say, to describe the perturber. Our Hamiltonian will be of the form

$$H = H_P + \sum_r U_r, \quad (9)$$

where H_P is the Hamiltonian that describes the perturber alone and U_r is the energy associated with the r -th system of the assembly, consisting of its proper energy plus its interaction energy with the perturber. H_P will be a function of the β 's only and U_r will be a function of the variables describing the r -th system and also the β 's. We can express the new sum $\sum_r U_r$ in terms of the n_a , w_a variables by the same method as before and the result will be of the same form (8), with the difference that the U_{ab} 's will no longer be numbers but will be functions of the β 's. The definition of U_{ab} will now be that its representative in the ζ -representation, the ζ 's being any complete set of commuting observables taken out of the β 's, is

$$(\zeta' | U_{ab} | \zeta'') = (\zeta' q^{(a)} | U | \zeta'' q^{(b)}), \quad (10)$$

the matrix on the right being the representative of U_r in the repre-

sentation in which q_r and ζ are diagonal. We shall still have U_{ab} commuting with the n 's and w 's.

It is possible to express any function of the dynamical variables that is symmetrical between all the particles in terms of the new variables n_a and w_a . The transformation may be conveniently carried out by considering the function of the dynamical variables to be the Hamiltonian for some dynamical system and then writing down the Schrödinger equation and transforming that. The general case has been considered by Jordan.*

§ 68. Discussion of Einstein-Bose Assemblies

In the preceding section we saw how the Hamiltonian describing an Einstein-Bose assembly, or more generally any symmetrical function of the dynamical variables of all the systems of the assembly, can be expressed in terms of variables n_a , w_a , analogous to the action and angle variables of a simple harmonic oscillator. This shows that *an Einstein-Bose assembly is dynamically equivalent to a set of simple harmonic oscillators, there being one oscillator corresponding to each of a complete set of independent states of a system of the assembly, the quantum number of the oscillator corresponding to the number of systems in the state.*

We may replace the set of simple harmonic oscillators by a train of waves, each Fourier component of the waves being dynamically equivalent to a simple harmonic oscillator. Thus our Einstein-Bose assembly is dynamically equivalent to a system of waves. This provides us with a complete reconciliation between the corpuscular and wave theories of radiation. We may regard radiation either as an assembly of photons satisfying the Einstein-Bose statistics or as a system of waves, the two points of view being consistent and mathematically equivalent.

We can gain a greater insight into the connexion between the systems of an Einstein-Bose assembly by considering the limiting case when the number of systems in each state is large, *i.e.* when the n 's are large. We introduce the observable

$$\xi_a = (n_a + 1)^{\frac{1}{2}} e^{-iw_a} = e^{-iw_a} n_a,$$

whose conjugate complex is

$$\bar{\xi}_a = e^{iw_a} (n_a + 1)^{\frac{1}{2}} = n_a^{\frac{1}{2}} e^{iw_a}.$$

* Jordan, *Zeits. f. Phys.*, vol. xlvi, p. 774 (1927).

This ξ_a is the analogue of $p - iq$ for the harmonic oscillator, apart from numerical coefficients. We have

$$\begin{aligned}\xi_a \bar{\xi}_a &= n_a + 1 \\ \bar{\xi}_a \xi_a &= n_a\end{aligned}\quad (11)$$

and thus

$$\xi_a \bar{\xi}_a - \bar{\xi}_a \xi_a = 1. \quad (12)$$

We can now express the Hamiltonian (8), describing the perturbation of the assembly by an external field of force, in terms of the ξ_a 's and their conjugate complexes, the result being

$$H = \sum_{ab} \bar{\xi}_a U_{ab} \xi_b.$$

The equations of motion for the ξ_a 's are

$$i\hbar \dot{\xi}_a = \xi_a H - H \xi_a = \sum_b U_{ab} \xi_b, \quad (13)$$

with the help of (12) and the condition that ξ_a commutes with ξ_b and $\bar{\xi}_b$ when $b \neq a$.

When the n_a 's are large, the ξ_a 's are also large and we may neglect the unity on the right-hand side of (12). With this approximation our observables ξ_a , $\bar{\xi}_a$ all commute with each other and may be counted as numbers. The equations of motion (13) now become ordinary differential equations between numbers. These equations are identical to the Schrödinger equation for a single one of the systems perturbed by the external field of force, the set of numbers ξ_a playing the part of the Schrödinger function ($q^{(a)}$) and U_{ab} being the representative of the Hamiltonian. If this Schrödinger function is normalized to n , it may be considered to represent an assembly of n independent systems in the way discussed in § 56. The interpretation of the Schrödinger function, namely the interpretation of $|q^{(a)}|$ as the number of systems in state $q^{(a)}$, now corresponds exactly to the interpretation of the ξ_a 's provided by equation (11). We thus have the result that *an assembly of a large number of similar systems is described by the same equations, whose solutions are to be interpreted in the same way, whether the systems are independent or satisfy the Einstein-Bose statistics.*

Since an assembly of independent systems and an assembly satisfying the Einstein-Bose statistics are two physically different things, it may seem strange that they are both to be described by the same set of equations, even though we are restricting ourselves to the limiting case of a large number of systems in the assembly. The

solution of this paradox lies in the fact that there is an essential difference between the mathematical treatments of the two assemblies, in spite of the similarities pointed out above, as may be seen from the following discussion. An assembly of independent systems is described as completely as quantum mechanics allows when we are given the number of systems in each state. The modulus of the Schrödinger function ($|q^{(a)}|$) is then determined for each state $q^{(a)}$, but not its phase. *This phase has no physical meaning.* We must average over all values of this phase if it appears in the result of any calculation. On the other hand, for an assembly satisfying the Einstein-Bose statistics, the ξ_a 's are *observables* and their phases as well as their moduli are of physical importance. An Einstein-Bose assembly is not described as completely as it might be unless the phases of the ξ_a 's are given as well as their moduli.

When we do not take the limiting case of a large number of systems, the differences between the Einstein-Bose assembly and independent assembly are greater. To obtain the equations for the Einstein-Bose assembly from those for the independent assembly we must apply a sort of quantization to the Schrödinger function, *i.e.* we must replace the numbers composing the Schrödinger function by observables satisfying definite commutability relations.

§ 69. Application to Photons

In applications of the above theory it is convenient to take the q 's to be constants of the motion for an unperturbed system, so that the $q^{(a)}$'s label the *stationary* states of the unperturbed systems and the n_a 's are the numbers of systems in the stationary states. In the case of photons this means we must take the q 's to be the three Cartesian components of momentum together with a variable specifying the polarization, which variable may be taken to be the direction of the electric vector for a linearly polarized photon. The polarization variable will now continually occur in our calculations along with the momentum. For brevity this polarization variable will usually not be explicitly mentioned but will be understood. Thus when we say a certain photon has a definite momentum, it is to be understood that it has also a definite polarization, and the set of three variables p_x, p_y, p_z (which may be abridged to p) specifying this momentum is to be understood as containing a fourth variable specifying the direction of the electric vector. Again, when it is said

that an integration is made over all values of the variables p_x, p_y, p_z , a summation over the two independent states of polarization is implied as well.

We can apply the theory at the end of § 67 to the interaction of a number of photons with an atom, the atom being the perturber. The energy U for a photon will consist of its proper energy $\hbar\nu$ together with its interaction energy with the atom, V say. Hence

$$U_{ab} = \hbar\nu_a \delta_{ab} + V_{ab},$$

ν_a being the frequency of a photon in the stationary state a . The V_{ab} 's, like the U_{ab} 's, will be functions of the dynamical variables of the atom. The total Hamiltonian, given by (9) and (8), may now be written

$$\begin{aligned} H &= H_P + \sum_{ab} n_a^\dagger e^{iw_a} U_{ab} (n_b + 1)^\dagger e^{-iw_b} \\ &= H_P + \sum_a n_a \hbar\nu_a + \sum_{ab} n_a^\dagger e^{iw_a} V_{ab} (n_b + 1)^\dagger e^{-iw_b} \\ &= H_P + H_R + H_Q, \end{aligned} \quad (14)$$

H_R being the total proper energy of the radiation and H_Q the total interaction energy.

Now photons have the peculiarity that they can be created and annihilated, as happens whenever one of them is emitted or absorbed by an atom, while our theory of the Einstein-Bose assembly has been built up on the basis of the conservation of the total number of systems. We can, however, reconcile our theory with this peculiarity of the photons by assuming a zero state for the photons, in which they have no momentum and energy and are not physically in evidence. We can now say that when a photon is absorbed or emitted, it jumps into or out of this zero state respectively, and can in this way preserve the constancy of the total number of photons. Since there is no limit to the number of photons that may be emitted, we must assume the number in the zero state to be infinite, *i.e.* $n_0 = \infty$. This makes the angle variable conjugate to n_0 a constant of the motion, since

$$\begin{aligned} i\hbar \frac{d}{dt} e^{iw_0} &= e^{iw_0} H - H e^{iw_0} \\ &= (e^{iw_0} n_0 - n_0 e^{iw_0}) (\hbar\nu_0 + V_{00}) + \\ &\quad + [e^{iw_0} n_0^\dagger - n_0^\dagger e^{iw_0}] \sum_{b \neq 0} e^{iw_0} V_{0b} (n_b + 1)^\dagger e^{-iw_b} + \\ &\quad + \sum_{a \neq 0} n_a^\dagger e^{iw_a} V_{a0} [e^{iw_0} (n_0 + 1)^\dagger - (n_0 + 1)^\dagger e^{iw_0}] e^{-iw_a} \\ &= 0, \end{aligned}$$

since ν_0 and V_{00} vanish and the quantities in square brackets [] are of order $n_0^{-\frac{1}{2}}$.

In order that the Hamiltonian (14) may remain finite when n_0 is infinite, V_{a0} and V_{0a} must be infinitely small. We shall suppose that they are infinitely small in such a way that their products with $n_0^{\frac{1}{2}}$ are finite and we shall put

$$\left. \begin{aligned} V_{a0}(n_0+1)^{\frac{1}{2}}e^{-iw_0} &= V_a \\ V_{0a} n_0^{\frac{1}{2}}e^{iw_0} &= \bar{V}_a, \end{aligned} \right\} (15)$$

V_a and \bar{V}_a being two new conjugate complex dynamical variables. We may count V_a and \bar{V}_a as functions only of the dynamical variables describing the atom, like V_{a0} and V_{0a} , since the other factors on the left-hand sides of (15) are constants of the motion (n_0 being effectively constant since changes in n_0 are small compared with n_0) and have no physical significance. The interaction energy H_Q in (14) may now be written

$$H_Q = \sum_a \{ V_a n_a^{\frac{1}{2}} e^{iw_a} + \bar{V}_a (n_a+1)^{\frac{1}{2}} e^{-iw_a} \} + \sum_{ab} V_{ab} n_a^{\frac{1}{2}} e^{iw_a} (n_b+1)^{\frac{1}{2}} e^{-iw_b}, \dots (16)$$

the values $a = 0, b = 0$ being understood to be excluded from the summations here.

A photon has a continuous range of stationary states and not a discrete set, since its components of momentum may have any values from $-\infty$ to ∞ . We therefore have to change the sums in (16) into integrals. To do this accurately would not be very easy, since it would mean dealing according to quantum mechanics with a dynamical system with continuously many degrees of freedom, which would require a new scheme of notation and a new mathematical technique. We are, however, interested in the interaction energy (16) mainly with regard to the limiting case of large n 's, when classical mechanics may be assumed to apply for the radiation, since we wish to compare the interaction energy in this case with that provided by classical electromagnetic theory and thus obtain expressions for the V_a 's and V_{ab} 's. In this limiting case the passage from sums to integrals is quite easy.

Let σ_a denote the number of states of the photon (with a particular polarization) per unit of momentum space about the momentum \mathbf{p}_a . We assume σ_a to be large, but an arbitrary function of \mathbf{p}_a , and investigate the limit of (16) when σ_a is made infinite. The number

of photons (with a particular polarization) per unit of momentum space about the momentum \mathbf{p}_a is

$$\eta_a = n_a \sigma_a,$$

provided n_a varies in some roughly continuous way from one state to the next. Let $(p'|V|p'')$ be the matrix* representing the interaction energy V for one photon in the ordinary p -representation for that photon. This ordinary p -representation differs from the one we have used up to the present in this chapter, in which V is represented by V_{ab} , only through the weight function. In the former representation the weight attached to a small domain δp_a of momentum space is just δp_a , while in the latter it is the number of discrete states in this domain, which is $\sigma_a \delta p_a$. The weight function is thus changed by a factor σ_a . The rule at the end of § 24 now shows that the matrix elements in the two representations are connected by

$$(p^{(a)}|V|p^{(b)}) = V_{ab}(\sigma_a \sigma_b)^{\frac{1}{2}}. \quad (17)$$

Similarly the matrix elements $(p'|V|0)$, $(0|V|p')$, referring to transitions into or out of the zero state, are connected with V_a and \bar{V}_a by

$$(p^{(a)}|V|0) = V_a \sigma_a^{\frac{1}{2}} \quad (0|V|p^{(a)}) = \bar{V}_a \sigma_a^{\frac{1}{2}}.$$

We can now express the interaction energy (16) in the limiting case of large n 's, when the n 's may be assumed to commute with the w 's, in the form

$$\begin{aligned} H_Q &= \sum_a \{(p^{(a)}|V|0)\eta_a^{\frac{1}{2}}e^{iw_a} + (0|V|p^{(a)})\eta_a^{\frac{1}{2}}e^{-iw_a}\} \sigma_a^{-1} + \\ &\quad + \sum_{ab} (p^{(a)}|V|p^{(b)})\eta_a^{\frac{1}{2}}\eta_b^{\frac{1}{2}}e^{i(w_a-w_b)}\sigma_a^{-1}\sigma_b^{-1} \\ &= \int \{(p^{(a)}|V|0)\eta_a^{\frac{1}{2}}e^{iw_a} + (0|V|p^{(a)})\eta_a^{\frac{1}{2}}e^{-iw_a}\} dp_a + \\ &\quad + \iint (p^{(a)}|V|p^{(b)})\eta_a^{\frac{1}{2}}\eta_b^{\frac{1}{2}}e^{i(w_a-w_b)} dp_a dp_b \quad (18) \end{aligned}$$

in the limit $\sigma \rightarrow \infty$. The fact that the σ 's have disappeared from this result justifies our method of dealing with a continuous range of states as a limiting case of a discrete set.

§ 70. Determination of the Interaction Energy between a Photon and Atom

We shall now determine the matrix elements $(p^{(a)}|V|0)$, $(0|V|p^{(a)})$, and $(p^{(a)}|V|p^{(b)})$ by comparing (18) with the classical expression for the interaction energy between an atom and a field of radiation. For

* The matrix elements of this matrix are actually functions of the dynamical variables describing the atom, like the V_{ab} 's, and not numbers, but this does not invalidate the argument. The representation is an 'incomplete' one, the representatives being defined in terms of those of a complete one by equations like (10).

simplicity we shall suppose the atom to consist of a single electron moving in an electrostatic field of force. The field of radiation may be described by the 4-vector potential, which is to a certain extent arbitrary and may be chosen so that its time component vanishes. The field is then completely described by the magnetic potential A_x, A_y, A_z or \mathbf{A} . The change that the field causes in the Hamiltonian describing the atom is now, as explained at the beginning of § 48,

$$\begin{aligned} \frac{1}{2m} \left\{ \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 - \mathbf{p}^2 \right\} &= \frac{e}{mc} (\mathbf{p}, \mathbf{A}) + \frac{e^2}{2mc^2} \mathbf{A}^2 \\ &= \frac{e}{c} (\dot{\mathbf{x}}, \mathbf{A}) + \frac{e^2}{2mc^2} \mathbf{A}^2. \end{aligned} \quad (19)$$

This is the classical interaction energy, which is to be compared with (18). The \mathbf{A} that occurs here ought really to be the value of the magnetic potential at the point where the electron is momentarily situated. It is, however, a good enough approximation if we take this \mathbf{A} to be the magnetic potential at some fixed point in the atom, such as the nucleus, provided we are not dealing with radiation whose wave-length is small compared with the dimensions of the atom.

To make the comparison between (18) and (19) we must first resolve the field of radiation into plane progressive waves. The electric and magnetic fields of one of these waves, whose frequency is ν and whose direction is specified by the momentum \mathbf{p} of the associated photons, are of the form

$$\mathcal{E}_p \cos[(\mathbf{x}, \mathbf{p})/\hbar + 2\pi\nu t + \gamma_p], \quad \mathcal{H}_p \cos[(\mathbf{x}, \mathbf{p})/\hbar + 2\pi\nu t + \gamma_p],$$

the amplitudes \mathcal{E}_p and \mathcal{H}_p being vectors of equal length that are perpendicular to the direction of motion and to each other. The total electric and magnetic fields are expressible as Fourier integrals of the form

$$\begin{aligned} \mathcal{E} &= \int \mathcal{E}_p \cos[(\mathbf{x}, \mathbf{p})/\hbar + 2\pi\nu t + \gamma_p] d\mathbf{p} \\ \mathcal{H} &= \int \mathcal{H}_p \cos[(\mathbf{x}, \mathbf{p})/\hbar + 2\pi\nu t + \gamma_p] d\mathbf{p}, \end{aligned}$$

$\mathcal{E}_p, \mathcal{H}_p$, and γ_p being definite functions of the momentum \mathbf{p} .

We must obtain the distribution of energy of this field over the various Fourier components. At time $t = 0$ we have

$$\begin{aligned} \int \mathcal{E}^2 dx &= \iint (\mathcal{E}_p, \mathcal{E}_{p'}) dp dp' \int \cos[(\mathbf{x}, \mathbf{p})/\hbar + \gamma_p] \cos[(\mathbf{x}, \mathbf{p}')/\hbar + \gamma_{p'}] dx \\ &= \iint (\mathcal{E}_p, \mathcal{E}_{p'}) dp dp'. \frac{1}{2} h^3 \{ \cos(\gamma_p + \gamma_{p'}) \delta(p + p') + \\ &\quad + \cos(\gamma_p - \gamma_{p'}) \delta(p - p') \}, \end{aligned}$$

the integration with respect to x here being similar to that with respect to q performed in § 35. Thus

$$\int \mathcal{E}^2 dx = \frac{1}{2}h^3 \int \{(\mathcal{E}_p, \mathcal{E}_{-p})\cos(\gamma_p + \gamma_{-p}) + \mathcal{E}_p^2\} dp.$$

Similarly

$$\int \mathcal{H}^2 dx = \frac{1}{2}h^3 \int \{(\mathcal{H}_p, \mathcal{H}_{-p})\cos(\gamma_p + \gamma_{-p}) + \mathcal{H}_p^2\} dp.$$

On account of the connexion between the vectors \mathcal{E}_p and \mathcal{H}_p we have $\mathcal{E}_p^2 = \mathcal{H}_p^2$ and also $(\mathcal{E}_p, \mathcal{E}_{-p}) = -(\mathcal{H}_p, \mathcal{H}_{-p})$. Hence the total energy is

$$1/8\pi \cdot \int (\mathcal{E}^2 + \mathcal{H}^2) dx = h^3/8\pi \cdot \int \mathcal{E}_p^2 dp,$$

and the energy per unit of momentum space is $h^3/8\pi \cdot \mathcal{E}_p^2$. This may be equated to $h\nu_p \eta_p$, the η having the same meaning as in the preceding section. Thus

$$\mathcal{E}_p^2 = 8\pi h^{-2} \nu_p \eta_p.$$

The vector potential \mathbf{A} may be expressed as a Fourier integral in the same way as \mathcal{E} and \mathcal{H} . We shall have

$$\mathbf{A} = -\int \mathbf{A}_p \sin[(\mathbf{x}, \mathbf{p})/\hbar + 2\pi\nu t + \gamma_p] dp, \quad (20)$$

the vector \mathbf{A}_p being in the same direction as \mathcal{E}_p and having its length given by

$$\mathbf{A}_p^2 = \left(\frac{c}{2\pi\nu_p}\right)^2 \mathcal{E}_p^2 = \frac{2c^2}{\pi h^2 \nu_p} \eta_p. \quad (21)$$

At the origin \mathbf{A} will have the value

$$\mathbf{A} = -\int \mathbf{A}_p \sin[2\pi\nu t + \gamma_p] dp = \int \mathbf{A}_p \cos w_p dp,$$

w_p being an angle variable of the same nature as those occurring in (18). This value for \mathbf{A} substituted in expression (19) for the interaction energy gives

$$\begin{aligned} & e/c \cdot \int (\dot{\mathbf{x}}, \mathbf{A}_p) \cos w_p dp + e^2/2mc^2 \cdot \iint (\mathbf{A}_p, \mathbf{A}_{p'}) \cos w_p \cos w_{p'} dp dp' \\ &= \frac{e}{\hbar} \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int \frac{1}{\nu_p^{\frac{1}{2}}} \dot{x}_p \eta_p^{\frac{1}{2}} \cos w_p dp + \\ & \quad + \frac{e^2}{\pi m h^2} \iint \frac{1}{\nu_p^{\frac{1}{2}} \nu_{p'}^{\frac{1}{2}}} \cos \theta_{pp'} \eta_p^{\frac{1}{2}} \eta_{p'}^{\frac{1}{2}} \cos w_p \cos w_{p'} dp dp', \end{aligned} \quad (22)$$

with the help of (21), where \dot{x}_p is the component of $\dot{\mathbf{x}}$ in the direction of \mathbf{A}_p or \mathcal{E}_p and $\theta_{pp'}$ is the angle between the vectors \mathbf{A}_p and $\mathbf{A}_{p'}$.

If we write (22) in terms of e^{iw} and e^{-iw} instead of $\cos w$ and compare it with (18), we obtain

$$\left. \begin{aligned} (p|V|0) &= (0|V|p) = \frac{e}{\hbar} \frac{1}{(2\pi\nu_p)^{\frac{1}{2}}} \dot{x}_p \\ (p|V|p') &= \frac{e^2}{2\pi m h^2} \frac{1}{\nu_p^{\frac{1}{2}} \nu_{p'}^{\frac{1}{2}}} \cos \theta_{pp'} \end{aligned} \right\} (23)$$

We also find that there are certain terms in (22), namely those involving $\exp i(w_p + w_{p'})$ or $\exp -i(w_p + w_{p'})$, which have no corresponding terms in (18). This discrepancy shows the inadequacy of the assumption that the Hamiltonian describing the interaction of an assembly of photons with an atom is of the form (9). The extra terms in (22) would correspond to transitions in which two photons are simultaneously absorbed or emitted and the possibility of such transitions requires a more complicated interaction energy than that assumed in (9). The physical effects of these terms are, however, small and unimportant, and so we shall neglect them.

Equations (23) now give the interaction energy V between a single photon and the atom. This interaction energy cannot conveniently be expressed explicitly in terms of dynamical variables. We can get a complete representation of V by introducing a Heisenberg representation for the variables describing the atom. If the different stationary states of the atom alone are denoted by α', α'', \dots , we shall have

$$\left. \begin{aligned} (p\alpha' | V | 0\alpha'') &= (0\alpha' | V | p\alpha'') = \frac{e}{\hbar} \frac{1}{(2\pi\nu_p)^{\frac{1}{2}}} (\alpha' | \dot{x}_p | \alpha'') \\ (p'\alpha' | V | p''\alpha'') &= \frac{e^2}{2\pi m\hbar^2} \frac{1}{\nu_{p'}^{\frac{1}{2}} \nu_{p''}^{\frac{1}{2}}} \cos \theta_{p'p''} \delta_{\alpha'\alpha''}. \end{aligned} \right\} (24)$$

Each p here is, as before mentioned, to be understood as including not only the three Cartesian components of momentum of the photon but also a polarization variable specifying a direction of electric force. The matrix element $(\alpha' | \dot{x}_p | \alpha'')$ is the component of the vector $(\alpha' | \dot{\mathbf{x}} | \alpha'')$ in the direction of the electric force specified by p and similarly $\theta_{p'p''}$ is the angle between the directions of electric force specified by p' and p'' .

§ 71. Emission, Absorption, and Scattering of Radiation

We can now determine directly the coefficients of emission, absorption, and scattering of radiation by substituting in the formulas of Chapter X the values for the matrix elements given by (24). For the case of emission we can use formula (56) of Chapter X. This shows that for an atom in a state α' the probability per unit time per unit solid angle of its spontaneously emitting a photon and dropping to a state α'' of lower energy is

$$\frac{4\pi^2}{\hbar} \frac{WP}{c^2} \left| \frac{e}{\hbar} \frac{1}{(2\pi\nu)^{\frac{1}{2}}} (\alpha' | \dot{x}_p | \alpha'') \right|^2. \quad (25)$$

Now the energy and momentum of a photon of frequency ν are

$$W = h\nu \quad P = h\nu/c.$$

Again from the Heisenberg law (48) of Chapter VI

$$(\alpha'|\dot{x}_p|\alpha'') = 2\pi i\nu(\alpha'\alpha'')(\alpha'|\dot{x}_p|\alpha''),$$

$\nu(\alpha'\alpha'')$ being the frequency connected with transitions from state α' to state α'' , which in the present case is just the frequency ν of the emitted radiation. These results substituted in (25) make the emission coefficient reduce to

$$\frac{(2\pi\nu)^3}{hc^3} |(\alpha'|\dot{e}\dot{x}_p|\alpha'')|^2. \quad (26)$$

To obtain the rate of emission of energy per unit solid angle we must multiply this by $h\nu$. If we now integrate over all solid angles, we shall obtain for the total rate of emission of energy

$$\frac{4}{3} \frac{(2\pi\nu)^4}{c^3} |(\alpha'|\dot{e}\dot{x}|\alpha'')|^2. \quad (27)$$

This is in agreement with expression (50) of Chapter VI and justifies Heisenberg's assumption for the interpretation of his matrix elements.

In the same way the absorption coefficient, given by formula (59) of Chapter X, becomes for photons

$$\frac{4\pi^2 h^2 W}{c^2 P} \left| \frac{1}{\hbar(2\pi\nu)^{\frac{1}{2}}} (\alpha'|\dot{x}_p|\alpha'') \right|^2 = \frac{8\pi^3 \nu}{c} |(\alpha'|\dot{e}\dot{x}_p|\alpha'')|^2.$$

This absorption coefficient refers to an incident beam of one photon crossing unit area per unit time per unit energy range. If we take one per unit frequency range instead of energy range, as is usual when dealing with radiation, the absorption coefficient becomes

$$\frac{8\pi^3 \nu}{hc} |(\alpha'|\dot{e}\dot{x}_p|\alpha'')|^2.$$

This result is the same as (24) of § 53, if we substitute for the E , there the energy $h\nu$ of a single photon. Thus *the elementary theory of § 53, in which the radiation field is treated as an external perturbation, gives the correct value for the absorption coefficient*. The average absorption for all directions of motion and of polarization of the incident beam is

$$\frac{8\pi^3}{3} \frac{\nu}{ch} |(\alpha'|\dot{e}\dot{x}|\alpha'')|^2,$$

which is just equal to the emission coefficient (27) divided by the factor $8\pi h\nu^3/c^2$. This ratio for the absorption and emission coefficients may be verified by elementary statistical arguments.

Let us now consider scattering. The true scattering coefficient is given by formula (38) of Chapter X. Such scattering of photons will not be accompanied by any change of state of the atom on account of the factor $\delta_{\alpha'\alpha''}$ in the expression for the matrix element $(p'\alpha'|V|p''\alpha'')$ in (24). Thus the final energy W' of the photon will equal its initial energy W^0 . The scattering coefficient now reduces to

$$e^4/m^2 c^4 \cdot \cos^2 \theta_{p^0 p'}.$$

This is the same as that given by classical mechanics for the scattering of radiation by a free electron. We thus see that the true scattering of radiation by an electron in an atom is independent of the atom and is correctly given by the classical theory. This result, it should be remembered, holds only provided the wave-length of the radiation is large compared with the dimensions of the atom.

The true scattering is a mathematical concept and cannot be separated out experimentally from the total scattering, given by formula (44) of Chapter X. Let us see what this total scattering is in the case of photons. A modification must now be made in the application of formula (44) of Chapter X. The summation Σ_k in this formula may be considered as representing the contribution to the scattering of double transitions consisting of transitions firstly from the initial state to state k and secondly from state k to the final state. The first transition may be an absorption of the incident photon and the second an emission of the required scattered photon, but it is also possible for the first transition to be the emission and the second the absorption. It is clear from the general nature of the method used for deriving formula (44) of Chapter X that both these kinds of double transitions must be included in the summation Σ_k when this formula is applied to photons, although only the first of them was taken into account in the actual derivation given in Chapter X.

For the double transition of absorption followed by emission we must take, using zero, single prime and double prime to refer to the initial, final, and intermediate k state respectively,

$$(k|V|p^0\alpha^0) = (0\alpha''|V|p^0\alpha^0) \quad (p'\alpha'|V|k) = (p'\alpha'|V|0\alpha'')$$

$$E - E_k = h\nu^0 + H_P(\alpha^0) - H_P(\alpha'') = h[\nu^0 - \nu(\alpha''\alpha^0)],$$

where ν^0 is the frequency of the incident photon and

$$\hbar\nu(\alpha''\alpha^0) = H_P(\alpha'') - H_P(\alpha^0).$$

Similarly for the double transition of emission followed by absorption we must take

$$(k|V|p^0\alpha^0) = (p'\alpha''|V|0\alpha^0) \quad (p'\alpha'|V|k) = (0\alpha'|V|p^0\alpha'')$$

$$E - E_k = h\nu^0 + H_P(\alpha^0) - H_P(\alpha'') - h\nu^0 - h\nu' = -\hbar[\nu' + \nu(\alpha''\alpha^0)],$$

where ν' is the frequency of the scattered photon, there being now two photons, of frequencies ν^0 and ν' , in existence for the intermediate state k . The expression for the scattering coefficient now reduces to

$$\frac{e^4}{h^2 c^4} \frac{\nu'}{\nu^0} \left| \frac{\hbar}{m} \cos \theta_{01} \delta_{\alpha^0 \alpha'} + \sum_{\alpha''} \left\{ \frac{(\alpha'|\dot{x}_1|\alpha'')(\alpha''|\dot{x}_0|\alpha^0)}{\nu^0 - \nu(\alpha''\alpha^0)} - \frac{(\alpha'|\dot{x}_0|\alpha'')(\alpha''|\dot{x}_1|\alpha^0)}{\nu' + \nu(\alpha''\alpha^0)} \right\} \right|^2, \quad (28)$$

where x_0 and x_1 have been written for x_{p^0} and $x_{p'}$, the components of \mathbf{x} in the directions of the electric vectors of the incident and scattered photons, and θ_{01} has been written for $\theta_{p^0 p'}$, the angle between these electric vectors. If we write (28) in terms of x instead of \dot{x} , we get

$$\frac{(2\pi e)^4}{h^2 c^4} \frac{\nu'}{\nu^0} \left| \frac{\hbar}{2\pi m} \cos \theta_{01} \delta_{\alpha^0 \alpha'} - \sum_{\alpha''} \nu(\alpha'\alpha'')\nu(\alpha''\alpha^0) \left\{ \frac{(\alpha'|\dot{x}_1|\alpha'')(\alpha''|\dot{x}_0|\alpha^0)}{\nu^0 - \nu(\alpha''\alpha^0)} - \frac{(\alpha'|\dot{x}_0|\alpha'')(\alpha''|\dot{x}_1|\alpha^0)}{\nu' + \nu(\alpha''\alpha^0)} \right\} \right|^2. \quad (29)$$

We can simplify (29) with the help of the quantum conditions. We have

$$x_1 x_0 - x_0 x_1 = 0,$$

which gives

$$\sum_{\alpha''} \{(\alpha'|\dot{x}_1|\alpha'')(\alpha''|\dot{x}_0|\alpha^0) - (\alpha'|\dot{x}_0|\alpha'')(\alpha''|\dot{x}_1|\alpha^0)\} = 0, \quad (30)$$

and also

$$x_1 \dot{x}_0 - \dot{x}_0 x_1 = 1/m \cdot (x_1 p_0 - p_0 x_1) = i\hbar/m \cdot \cos \theta_{01},$$

which gives

$$\begin{aligned} \sum_{\alpha''} \{(\alpha'|\dot{x}_1|\alpha'') \cdot \nu(\alpha''\alpha^0) (\alpha''|\dot{x}_0|\alpha^0) - \nu(\alpha'\alpha'') (\alpha'|\dot{x}_0|\alpha'') \cdot (\alpha''|\dot{x}_1|\alpha^0)\} &= \\ &= \frac{1}{2\pi i} \cdot \frac{i\hbar}{m} \cos \theta_{01} \delta_{\alpha^0 \alpha'} = \frac{\hbar}{2\pi m} \cos \theta_{01} \delta_{\alpha^0 \alpha'}. \end{aligned} \quad (31)$$

Multiplying (30) by ν' and adding to (31), we obtain

$$\begin{aligned} \sum_{\alpha''} \{(\alpha'|\dot{x}_1|\alpha'') (\alpha''|\dot{x}_0|\alpha^0) [\nu' + \nu(\alpha''\alpha^0)] - (\alpha'|\dot{x}_0|\alpha'') (\alpha''|\dot{x}_1|\alpha^0) [\nu' + \nu(\alpha'\alpha'')]\} &= \\ &= \hbar/2\pi m \cdot \cos \theta_{01} \delta_{\alpha^0 \alpha'}. \end{aligned}$$

If we substitute this expression for $\hbar/2\pi m \cdot \cos \theta_{01} \delta_{\alpha^0 \alpha'}$ in (29), we

obtain, after a straightforward reduction making use of identical relations between the ν 's,

$$\frac{(2\pi e)^4}{h^2 c^4} \nu^0 \nu'^3 \left| \sum_{\alpha''} \left\{ \frac{(\alpha' | x_1 | \alpha'') (\alpha'' | x_0 | \alpha^0)}{\nu^0 - \nu(\alpha'' \alpha^0)} - \frac{(\alpha' | x_0 | \alpha'') (\alpha'' | x_1 | \alpha^0)}{\nu' + \nu(\alpha'' \alpha^0)} \right\} \right|^2. \quad (32)$$

This gives the scattering coefficient in the form of the effective area that a photon has to hit per unit solid angle of scattering. It is known as the *Kramers-Heisenberg dispersion formula*, having been first obtained by these authors from analogies with the classical theory of dispersion.

The fact that the various terms in (29) can be combined to give the result (32) justifies the assumption made in deriving formula (44) of Chapter X, that the matrix elements $(p'\alpha'|V|p''\alpha'')$ of the interaction energy are of the second order of smallness compared with the $(p'\alpha'|V|k)$ ones, at any rate when the scattered particles are photons.

§ 72. Einstein's Laws of Radiation

In the preceding section we determined the probability coefficients for absorption, emission, and scattering of a photon by an atom. We were there concerned with only a single photon interacting with the atom (or at most with two), the interaction energy being given by (24). To complete our theory of radiation we require to know the laws governing the interaction of a number of photons with the atom. If the atom is exposed to an incident beam of radiation containing many photons, how do the absorption, emission, and scattering probabilities depend on the intensity of this beam?

This question cannot, of course, be answered simply from a consideration of the interaction energy, defined by (24), for a single photon. We have to fall back on the general interaction energy (16) for a number of photons, and this requires incidentally that we must perform the passage from sums to integrals once again. We make use of the general result (28) of § 54, according to which a transition probability is proportional to the square of the modulus of the matrix element of the perturbing energy that refers to this transition.

Let us consider an absorption process in which the number of photons in state a is reduced from n_a to $n_a - 1$, the atom simultaneously jumping from state α^0 to state α' . The probability of such a process will be proportional to the square of the modulus of the matrix element

$$(n_1 n_2 \dots n_a \dots \alpha^0 | H_Q | n_1 n_2 \dots n_a - 1 \dots \alpha')$$

of the total interaction energy H_Q . The only term in the expression (16) for H_Q which can contribute to this matrix element is $V_a n_a^{\frac{1}{2}} e^{iw_a}$. This matrix element is thus proportional to $n_a^{\frac{1}{2}}$ and the transition probability is proportional to n_a . The passage from sums to integrals is now quite trivial, the final result being that the probability of an absorption process is proportional to the intensity of the incident radiation.

Similarly for an emission process, in which the number of photons in state a is increased from n_a to $n_a + 1$, we must consider the matrix element

$$(n_1 n_2 \dots n_a \dots \alpha^0 | H_Q | n_1 n_2 \dots n_a + 1 \dots \alpha').$$

The only term in expression (16) that contributes to this is $\bar{V}_a (n_a + 1)^{\frac{1}{2}} e^{-iw_a}$. This matrix element is thus proportional to $(n_a + 1)^{\frac{1}{2}}$ and the transition probability to $n_a + 1$. In the same way a scattering process, in which the number of photons in state a is decreased from n_a to $n_a - 1$ and that in state b is increased from n_b to $n_b + 1$, is due to the term $V_{ab} n_a^{\frac{1}{2}} e^{iw_a} (n_b + 1)^{\frac{1}{2}} e^{-iw_b}$, if it is a true scattering process, and to the product of the two terms $V_a n_a^{\frac{1}{2}} e^{iw_a}$ and $\bar{V}_b (n_b + 1)^{\frac{1}{2}} e^{-iw_b}$, if otherwise. The scattering probability is thus in any case proportional to $n_a(n_b + 1)$. To interpret these results we must now make an accurate passage from the discrete to the continuous ranges of stationary states for the photons.

Suppose we have a distribution n_a of photons over the discrete states a . To obtain the density of these photons (in ordinary space) we may suppose them to be represented by a Schrödinger function $(p^{(a)}) = n_a^{\frac{1}{2}}$, and transform this Schrödinger function to the (x, y, z) -representation by means of the transformation function $(x|p^{(a)})$. This transformation function must now have the value

$$(x|p^{(a)}) = h^{-\frac{1}{2}} e^{i(\mathbf{x}, \mathbf{p}^{(a)})/\hbar} \sigma_a^{-\frac{1}{2}},$$

differing from the value given by (36) of Chapter VI by the factor $\sigma_a^{-\frac{1}{2}}$, on account of the weight function of our present p -representation differing from that of the usual one by the factor σ_a , as was discussed in obtaining equation (17). Thus

$$(x|) = \sum_a (x|p^{(a)})(p^{(a)}|) = h^{-\frac{1}{2}} \sum_a e^{i(\mathbf{x}, \mathbf{p}^{(a)})/\hbar} n_a^{\frac{1}{2}} \sigma_a^{-\frac{1}{2}}.$$

Suppose n_a has the value unity for one state \mathbf{p} and zero for all the others. We shall then have

$$(x|) = h^{-\frac{1}{2}} e^{i(\mathbf{x}, \mathbf{p})/\hbar} \sigma^{-\frac{1}{2}},$$

and the density of photons will be

$$|(x^{\dagger})|^2 = h^{-3} \sigma^{-1}.$$

For an arbitrary distribution n_a of the photons over the discrete states a , the photon density will be given by addition of the contributions from each state and will therefore be

$$h^{-3} \sum_a n_a \sigma_a^{-1} = h^{-3} \int n_a dp_a.$$

Thus the number of photons per unit volume per unit of momentum space is $h^{-3}n_a$, corresponding to an energy $h^{-2}\nu_a n_a$ per unit volume per unit of momentum space. The intensity per unit frequency range, equal to c times the energy density per unit solid angle per unit frequency range, is therefore

$$I_a = h\nu_a^3/c^2 \cdot n_a.$$

The probability for an emission process, which we found was proportional to $n_a + 1$, is thus proportional to $I_a + h\nu_a^3/c^2$. This means that with no incident radiation there is still a certain amount of emission (which is, in fact, given by expression (26)), but that the emission is increased or stimulated by incident radiation in the same direction and having the same frequency (and state of polarization) as the emitted radiation under consideration. Our present theory of radiation thus completes the imperfect one of § 53, and gives a ratio for the stimulated and spontaneous emissions in agreement with Einstein's laws of radiation discussed at the end of § 53.

The probability for a scattering process from state a to state b , which we found was proportional to $n_a(n_b + 1)$, is in the same way proportional to $I_a(I_b + h\nu_b^3/c^2)$. Thus the scattering of radiation is also stimulated by incident radiation in the same direction and having the same frequency as the scattered radiation. The stimulation phenomenon is, in fact, a general one, as has been shown by Einstein and Ehrenfest * from general statistical arguments.

* Einstein and Ehrenfest, *Zeits. f. Phys.*, vol. xix, p. 301 (1923). See also Pauli, *Zeits. f. Phys.*, vol. xviii, p. 272 (1923).

XIII

RELATIVITY THEORY OF THE ELECTRON

§ 73. Relativity Treatment of a Single Particle

Our theory of special dynamical systems from Chapter VI onwards was essentially non-relativistic. We worked all the time with one particular Lorentz frame of reference and did not make it an essential requirement of the theory that its results should be independent of this frame. Let us now inquire into what sort of modifications we may expect relativity to introduce.

It is fairly certain that the general theory of states and observables developed in Chapters II–V will apply also to relativity treatments of dynamical systems. We are faced with the problem, however, of deciding with what observables we shall now work. There are serious disadvantages in taking these observables to be the values, ξ_t say, of dynamical variables ξ at the time t . If the ξ_t 's occur in our analysis, they would have to appear on the same footing as the ξ_τ 's, the values of the ξ 's at the time τ in some other Lorentz frame. We should therefore require to know the relations between the ξ_t 's and the ξ_τ 's, and these would in general be very complicated and artificial, as they would require us to connect distant parts of space-time. In any case the ξ_t 's are not quantities that could easily be observed and we should not expect them to play any fundamental role in the theory. A possible way out of the difficulty would be to build up a purely field theory and to take as observables the values of the field quantities at definite points in space-time. This appears to be the most straightforward way of dealing with general dynamical systems on relativity lines, but it involves complicated mathematics and appears to be too difficult for practical application.*

The difficulty of a relativity treatment becomes much less severe when one confines one's attention to the problem of a single particle moving in a given field of force. If we now take a representation in which the observables x_t, y_t, z_t specifying the position of the particle at time t are diagonal, we have as the wave function representing a state a function $(x_t y_t z_t |)$ of the three variables x_t, y_t, z_t depending on the parameter t , which is the same as a function $(xyzt |)$ of the four variables x, y, z, t . The domain of our wave function thus becomes

* See Heisenberg and Pauli, *Zeits. f. Phys.*, vol. Ivi, p. 1; vol. lix, p. 168 (1929).

identical with the ordinary space-time continuum, and this circumstance makes possible an elementary treatment of the problem and allows us to use considerations which cannot be extended to more general dynamical systems. We may expect, for instance, the physical conditions at any point in space-time to depend only on the value of the wave function at that point and neighbouring points, and thus the wave function, if not actually invariant under a Lorentz transformation, should transform according to simple laws.

Let us now see how we can bring the momentum of the particle into the theory. The value of a component of momentum at a specified time is an observable of a rather artificial kind, even in the case of a system with a single particle, and we should not expect it to play an important role. This observable, we saw in § 36, is connected with a certain space-displacement operator, which, when it operates on any wave function, produces *at the specified time*, just a spatial displacement, the value of the new wave function at any other time being then determined by the wave equation. It would seem more natural in a relativity theory to deal with an operator which produces *at all times* simply a spatial displacement of the wave function, such an operator being essentially a simple partial differentiation of the type $\partial/\partial x$ of the wave function ($xyzt|$) in four variables. The result of such an operator operating on a wave function is a new wave function which in general does not satisfy the wave equation and hence does not represent a state of the system, so that *this operator is not an observable*. All the same we may expect the operator $-ih\partial/\partial x$ to play the part of a momentum in the theory, in spite of the fact that since it refers to momentum in general and not momentum at a particular time, we can give no precise meaning to an observation of it.

Thus we are led to introduce the operators

$$p_x = -ih \frac{\partial}{\partial x} \quad p_y = -ih \frac{\partial}{\partial y} \quad p_z = -ih \frac{\partial}{\partial z}, \quad (1)$$

and also the corresponding

$$W = ih \frac{\partial}{\partial t}, \quad (2)$$

referring to time displacement, to play the part of momentum and energy. They can operate on any wave function, but since the result of such operation does not satisfy the wave equation and does not represent a state, they are not observables. All the same they may

be used in algebraic analysis like observables and will satisfy all the axioms of ordinary algebra except the commutative law of multiplication. The complete algebraic scheme of Chapter II will not, however, apply, since we cannot interpret $\phi\alpha\psi$ as a number when α is an operator of this more general kind. It will be more convenient in the present chapter to regard the symbols ψ , p , &c., not in the abstract sense of Chapter II, but as wave functions and linear operators in the x , y , z , t representation.

§ 74. The Wave Equation for the Electron

Let us consider first the case of the motion of an electron in the absence of an electromagnetic field, so that the problem is simply that of the free particle, which was discussed in § 39. The Hamiltonian for this system provided by classical mechanics is given by equation (1) of § 39, and this leads to the wave equation (5) of that section. This wave equation may be written

$$\{W/c - (m^2 c^2 + p_x^2 + p_y^2 + p_z^2)^{\frac{1}{2}}\}\psi = 0, \quad (3)$$

where W and the p 's are to be interpreted as operators in accordance with equations (1) and (2). Equation (3), although it takes into account correctly the variation of the mass of the particle with its velocity, is yet unsatisfactory from the point of view of relativity, because it is very unsymmetrical between W and the p 's, so much so that one cannot generalize it in a relativistic way to the case when there is a field present. We must therefore look for a new wave equation for the free particle.

If we multiply the wave equation (3) on the left by the operator $\{W/c + (m^2 c^2 + p_x^2 + p_y^2 + p_z^2)^{\frac{1}{2}}\}$, we obtain the equation

$$\{W^2/c^2 - m^2 c^2 - p_x^2 - p_y^2 - p_z^2\}\psi = 0, \quad (4)$$

which is of a relativistically invariant form and may therefore more conveniently be taken as the basis of a relativity theory. Equation (4) is not completely equivalent to equation (3) since, although every solution of (3) is also a solution of (4), the converse is not true. Only those solutions of (4) belonging to positive values for W are also solutions of (3).

The wave equation (4) is not in agreement with the general laws of the quantum theory on account of its being quadratic in W . In § 37 we deduced from quite general arguments that the wave equation must be linear in the operator $\partial/\partial t$ or W , like equation (43) of

that section. We therefore seek a wave equation that is linear in W and that is roughly equivalent to (4). In order that this wave equation shall transform in a simple way under a Lorentz transformation, we try to arrange that it shall be rational and linear in p_x , p_y , and p_z as well as in W , and thus of the form

$$\{W/c + \alpha_x p_x + \alpha_y p_y + \alpha_z p_z + \beta\} \psi = 0, \quad (5)$$

where the α 's and β are independent of W and the p 's. Since we are considering the case of no field, all points in space-time must be equivalent, so that the operator in the wave equation must not involve x , y , z , or t . Thus the α 's and β must also be independent of x , y , z , and t . They must therefore denote some quite new dynamical variables, which may be pictured as describing some internal motion in the electron. We shall see later that they just describe the spin of the electron. The α 's and β must, of course, commute with W and the p 's and also with x , y , z , and t .

Multiplying (5) by the operator $\{W/c - \alpha_x p_x - \alpha_y p_y - \alpha_z p_z - \beta\}$ on the left, we obtain

$$\{W^2/c^2 - \sum_{xyz} [\alpha_x^2 p_x^2 + (\alpha_x \alpha_y + \alpha_y \alpha_x) p_x p_y + (\alpha_x \beta + \beta \alpha_x) p_x] - \beta^2\} \psi = 0.$$

This is the same as (4) if the α 's and β satisfy the relations

$$\begin{aligned} \alpha_x^2 &= 1, & \alpha_x \alpha_y + \alpha_y \alpha_x &= 0, \\ \beta^2 &= m^2 c^2, & \alpha_x \beta + \beta \alpha_x &= 0, \end{aligned}$$

together with the relations obtained from these by permuting x , y , and z . If we write

$$\beta = \alpha_m mc,$$

these relations may be summed up in the single one,

$$\alpha_\mu \alpha_\nu + \alpha_\nu \alpha_\mu = 2\delta_{\mu\nu} \quad (\mu, \nu = x, y, z, \text{ or } m). \quad (6)$$

The four α 's all anticommute with one another and the square of each is unity.

Thus by giving suitable properties to the α 's and β we can make the wave equation (5) equivalent to (4), in so far as the motion of the electron as a whole is concerned. We may now assume (5) is the correct relativity wave equation for the motion of an electron in the absence of a field. This gives rise to one difficulty, however, owing to the fact that (5), like (4), is not exactly equivalent to (3), but allows solutions corresponding to negative as well as positive values of W . The former do not, of course, correspond to any actually observable motion of an electron. For the present we shall simply

evade the difficulty by ignoring the negative-energy solutions. Their proper physical interpretation will be discussed in § 79.

We can easily obtain a representation of the four α 's. They have similar algebraic properties to the σ 's introduced in § 43 to describe the spin of an electron, which σ 's can be represented by matrices with two rows and columns. So long as we keep to matrices with two rows and columns we cannot get a representation of more than three anticommuting quantities, and we have to go to four rows and columns to get a representation of the four anticommuting α 's. It is convenient first to express the α 's in terms of the σ 's and also of a second similar set of three anticommuting observables whose squares are unity, ρ_1, ρ_2, ρ_3 say, that are independent of and commute with the σ 's. We may take, amongst other possibilities,

$$\alpha_x = \rho_1 \sigma_x, \quad \alpha_y = \rho_1 \sigma_y, \quad \alpha_z = \rho_1 \sigma_z, \quad \alpha_m = \rho_3, \quad (7)$$

and the α 's will then satisfy all the relations (6), as may easily be verified. If we now take a representation with ρ_3 and σ_z diagonal, we shall get the following scheme of matrices:

$$\begin{aligned} \sigma_x &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} & \sigma_y &= \begin{pmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix} & \sigma_z &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \\ \rho_1 &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} & \rho_2 &= \begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \end{pmatrix} & \rho_3 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \end{aligned}$$

Corresponding to the four rows and columns, the wave function must have four components. We saw in § 43 that the spin of the electron requires the wave function to have two components. The fact that our present theory gives four is due to our wave equation (5) having twice as many solutions as it ought to have, half of them corresponding to states of negative energy.

With the help of (7), the wave equation (5) may be written in the vector form

$$\{W/c + \rho_1(\sigma, \mathbf{p}) + \rho_3 mc\} \psi = 0.$$

To generalize this equation to the case when there is an electromagnetic field present, we follow the classical rule of replacing W and \mathbf{p} by $W + eA_0$ and $\mathbf{p} + e/c \cdot \mathbf{A}$, A_0 and \mathbf{A} being the scalar and vector

potentials of the field at the place where the electron is. This gives us the equation

$$\left\{ \frac{W}{c} + \frac{e}{c} A_0 + \rho_1 \left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) + \rho_3 mc \right\} \psi = 0, \quad (8)$$

which is the fundamental wave equation of the relativity theory of the electron. The conjugate imaginary equation

$$\phi \left\{ \frac{W}{c} + \frac{e}{c} A_0 + \rho_1 \left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) + \rho_3 mc \right\} = 0 \quad (9)$$

must be treated on the same footing as (8). The operators W and \mathbf{p} in (9), which operate to the left, must be interpreted, according to §§ 36 and 37, as having the meanings in equations (1) and (2) with the signs reversed.

§ 75. Invariance under a Lorentz Transformation

Before proceeding to discuss the physical consequences of the wave equation (8) or (9), we shall first verify that our theory really is invariant under a Lorentz transformation, or, stated more accurately, that the physical results the theory leads to are independent of the Lorentz frame of reference used. This is not by any means obvious from the form of the wave equation (8). We have to verify that, if we write down the wave equation in a different Lorentz frame, the solutions of the new wave equation may be put into one-one correspondence with those of the original one in such a way that corresponding solutions may be assumed to represent the same state. For either Lorentz frame, the square of the modulus of the wave function, summed for the four components, gives the probability per unit volume of the electron being at any given place in that Lorentz frame. This probability is of the nature of an electric density (and will be called the electric density in future, for brevity), and its values, calculated in different Lorentz frames for wave functions representing the same state, should be connected like the time components in these frames of some 4-vector. Further, the 4-dimensional divergence of this 4-vector should vanish, signifying conservation of charge, or that the electron cannot appear or disappear in any volume without passing through the boundary.

For discussing Lorentz transformations it is convenient to make a slight change in our notation. We shall use the suffixes 1, 2, 3 instead of x, y, z and shall put p_0 for W/c , and we shall also use the

convention that terms containing a repeated suffix are to be summed over the values 0...3 for that suffix. We can now write equation (8) in the form

$$\{\alpha_\mu(p_\mu + e/c \cdot A_\mu) + \alpha_m mc\} \psi = 0, \quad (10)$$

α_0 being equal to unity, and similarly we can write equation (9) in the form

$$\phi \{\alpha_\mu(p_\mu + e/c \cdot A_\mu) + \alpha_m mc\} = 0. \quad (11)$$

We now apply a Lorentz transformation and denote quantities referring to the new frame by a star. The components of the 4-vectors \mathbf{p} and \mathbf{A} will transform according to a linear law of the type

$$p_\mu = a_{\mu\nu} p_\nu^*, \quad A_\mu = a_{\mu\nu} A_\nu^*. \quad (12)$$

Substituting these expressions for p_μ and A_μ in equations (10) and (11), we obtain

$$\text{and } \{\alpha_\mu a_{\mu\nu} (p_\nu^* + e/c \cdot A_\nu^*) + \alpha_m mc\} \psi = 0 \quad (13)$$

$$\phi \{\alpha_\mu a_{\mu\nu} (p_\nu^* + e/c \cdot A_\nu^*) + \alpha_m mc\} = 0. \quad (14)$$

We now try to bring these equations back to the form of the original (10) and (11) by introducing a new wave function ψ^* , whose four components are linear functions (with constant numerical coefficients) of the four components of the original ψ . This means that ψ^* is connected with ψ by an equation of the type

$$\psi^* = \gamma \psi, \quad (15)$$

where γ is an operator, like the α 's, which can be represented as a matrix with four rows and columns. The conjugate imaginary equation to (15) is

$$\phi^* = \bar{\phi} \bar{\gamma}.$$

Equations (13) and (14) will go over into the equations

$$\text{and } \bar{\gamma} \{\alpha_\nu(p_\nu^* + e/c \cdot A_\nu^*) + \alpha_m mc\} \psi^* = 0 \quad (16)$$

$$\phi^* \{\alpha_\nu(p_\nu^* + e/c \cdot A_\nu^*) + \alpha_m mc\} \gamma = 0 \quad (17)$$

provided we can choose γ such that

$$\bar{\gamma} \alpha_\nu \alpha \gamma = \mu a_{\mu\nu}, \quad \bar{\gamma} \alpha_m \gamma = \alpha_m. \quad (18)$$

These equations (16) and (17) are of the same form as (10) and (11), as required, since one can divide out by the extra factors $\bar{\gamma}$ and γ .

In order to verify that we can always choose γ to satisfy equations (18), let us first take the special case when the change of our frame of reference consists simply of a rotation through a hyperbolic

angle θ in the xt plane, so that the transformation equations for the components of a 4-vector are of the type

$$\left. \begin{aligned} p_0 &= p_0^* \cosh \theta + p_1^* \sinh \theta \\ p_1 &= p_0^* \sinh \theta + p_1^* \cosh \theta \\ p_2 &= p_2^*, \quad p_3 = p_3^*. \end{aligned} \right\} \quad (19)$$

The values of the $a_{\mu\nu}$ may be written down at once from a comparison of these equations with (12). With these values for the $a_{\mu\nu}$ it is easy to see that equations (18) hold when we take

$$\gamma = e^{\frac{1}{2}\theta\alpha_1} = \bar{\gamma}. \quad (20)$$

We have, in fact,

$$\begin{aligned} \bar{\gamma}\alpha_0\gamma &= \bar{\gamma}\gamma = e^{\theta\alpha_1} \\ &= 1 + \theta\alpha_1 + \theta^2\alpha_1^2/2! + \theta^3\alpha_1^3/3! + \dots \end{aligned}$$

On account of $\alpha_1^2 = 1$, this reduces to

$$\begin{aligned} \bar{\gamma}\alpha_0\gamma &= \{1 + \theta^2/2! + \dots\} + \alpha_1\{\theta + \theta^3/3! + \dots\} \\ &= \cosh \theta + \alpha_1 \sinh \theta \\ &= \alpha_0 \cosh \theta + \alpha_1 \sinh \theta. \end{aligned}$$

Again, $\bar{\gamma}\alpha_1\gamma = \alpha_1\bar{\gamma}\gamma = \alpha_0 \sinh \theta + \alpha_1 \cosh \theta$.

Further, $\bar{\gamma}\alpha_2\gamma = e^{\frac{1}{2}\theta\alpha_1}\alpha_2e^{\frac{1}{2}\theta\alpha_1} = e^{\frac{1}{2}\theta\alpha_1}e^{-\frac{1}{2}\theta\alpha_1}\alpha_2 = \alpha_2$,

since α_2 anticommutes with α_1 , which results in $\alpha_2 f(\alpha_1) = f(-\alpha_1)\alpha_2$ for any function $f(\alpha_1)$ of α_1 . Similarly

$$\bar{\gamma}\alpha_3\gamma = \alpha_3, \quad \bar{\gamma}\alpha_m\gamma = \alpha_m.$$

Thus the five equations (18) hold with γ given by (20) when the $a_{\mu\nu}$ are given by (19).

As a second typical change of the frame of reference, we may consider a rotation through an angle θ in ordinary space about the x -axis. The transformation equations are now

$$\begin{aligned} p_0 &= p_0^* & p_1 &= p_1^* \\ p_2 &= p_2^* \cos \theta + p_3^* \sin \theta \\ p_3 &= -p_2^* \sin \theta + p_3^* \cos \theta. \end{aligned}$$

With the new values for the $a_{\mu\nu}$ we can easily verify that equations (18) hold with

$$\gamma = e^{-\frac{1}{2}\theta\alpha_2}\alpha_3 \quad \bar{\gamma} = e^{-\frac{1}{2}\theta\alpha_3}\alpha_2 = e^{\frac{1}{2}\theta\alpha_2}\alpha_3,$$

the analysis being very similar to the preceding case.

If two changes of the frame of reference are made consecutively, we simply have to multiply the corresponding γ 's to get the γ for the resultant change. Now any change of the frame of reference may

be built up from two rotations of the types we have considered, and hence there will always be a γ satisfying (18).

In this way we see that the solutions of the wave equation in the new frame of reference, equation (16), can be put into a natural one-one correspondence with those of the original wave equation (10), corresponding solutions being connected by (15), and we may assume that corresponding solutions represent the same state. It remains for us to verify that the electric density transforms like the time component of a 4-vector and that the divergence of this 4-vector vanishes.

We shall introduce the notation $\phi_r \cdot \psi_s$ to denote the sum of the product of each of the four components of ϕ_r with the corresponding component of ψ_s . In the same way $\phi\xi \cdot \eta\psi$, where ξ and η are any linear operators that can operate on the wave functions, will denote the sum of the product of each component of $\phi\xi$ with the corresponding component of $\eta\psi$. Our new symbols of the type $\phi\xi \cdot \eta\psi$ are functions of x, y, z , and t , and are quite distinct from the products $\phi\xi\eta\psi$ of Chapter II, which products, we have seen, have in general no meaning for the more general type of linear operator with which we are now dealing. It should be noted that

$$\phi \cdot \alpha\psi = \phi\alpha \cdot \psi \quad (21)$$

when α is one of the α 's in the wave equation, or more generally when it is any operator which means simply taking four linear functions (whose coefficients are numbers or functions of x, y, z , and t) of the four components of the wave function.

We can now express the electric density as $\phi \cdot \psi$, which is the same as $\phi \cdot \alpha_0 \psi$ or $\phi\alpha_0 \cdot \psi$ since $\alpha_0 = 1$. Let us see how the four quantities $\phi \cdot \alpha_\mu \psi$, with $\mu = 0 \dots 3$, transform under a Lorentz transformation. We have, from (15) and (18),

$$\begin{aligned} \phi^* \cdot \alpha_\nu \psi^* &= \phi \bar{\gamma} \cdot \alpha_\nu \gamma \psi = \phi \cdot \bar{\gamma} \alpha_\nu \gamma \psi \\ &= \phi \cdot \alpha_\mu \alpha_{\mu\nu} \psi = (\phi \cdot \alpha_\mu \psi) \alpha_{\mu\nu}. \end{aligned}$$

Comparing this result with (12), we see that the four quantities $\phi \cdot \alpha_\mu \psi$ transform like the covariant components of a 4-vector. The contravariant components will be

$$\phi \cdot \psi, \quad -\phi \cdot \alpha_1 \psi, \quad -\phi \cdot \alpha_2 \psi, \quad -\phi \cdot \alpha_3 \psi.$$

This verifies that our electric density $\phi \cdot \psi$ is the time component of a 4-vector and that the corresponding space components are $-\phi \cdot \alpha_r \psi$ (with $r = 1, 2, 3$). These space components give the electric current,

or, more accurately, the probability of the electron crossing unit area per unit time.

The divergence of our 4-vector is

$$\Sigma_\mu \pm \frac{\partial}{\partial x_\mu} (\phi \cdot \alpha_\mu \psi), \quad (22)$$

where x_0 denotes ct and the \pm sign means that the $+$ sign is to be taken for $\mu = 0$ and the $-$ sign for $\mu = 1, 2, 3$ before one does the summation. To prove this divergence vanishes, multiply equation (10) by ϕ and (11) by ψ , taking the sum over the four components in each case, and subtract. The result is

$$\phi \cdot \alpha_\mu p_\mu \psi - \phi \alpha_\mu p_\mu \cdot \psi = 0,$$

the other terms cancelling on account of (21). With the help of (1) and (2) this gives

$$\Sigma_\mu \pm \left\{ \phi \cdot \alpha_\mu \frac{\partial \psi}{\partial x_\mu} + \frac{\partial \phi}{\partial x_\mu} \alpha_\mu \cdot \psi \right\} = 0,$$

which just expresses the vanishing of (22). In this way we complete the proof that our theory gives consistent results in whichever frame of reference it is applied.

§ 76. Existence of the Spin

In § 74 we saw that the correct wave equation for the electron in the absence of an electromagnetic field, namely equation (5), is equivalent to the wave equation (4) which is suggested from analogy with the classical theory. This equivalence no longer holds when there is a field. By treating the correct wave equation for this case, namely (8), in the same way as we treated (5) and comparing it with the wave equation to be expected from analogy with the classical theory, namely

$$\left\{ \left(\frac{W}{c} + \frac{e}{c} A_0 \right)^2 - \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 - m^2 c^2 \right\} \psi = 0, \quad (23)$$

in which the operator is just the classical relativity Hamiltonian, we may expect to get an indication of the new physical features of the present theory.

We must multiply (8) by some factor on the left to make it resemble (23) as closely as possible. Taking this factor to be

$$\frac{W}{c} + \frac{e}{c} A_0 - \rho_1 \left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) - \rho_3 mc,$$

we get

$$\left\{ \left(\frac{W}{c} + \frac{e}{c} A_0 \right)^2 - \left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 - m^2 c^2 + \rho_1 \left[\left(\frac{W}{c} + \frac{e}{c} A_0 \right) \left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) - \left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) \left(\frac{W}{c} + \frac{e}{c} A_0 \right) \right] \right\} \psi = 0. \quad (24)$$

We now use the general formula that, if \mathbf{B} and \mathbf{C} are any two vectors that commute with σ ,

$$\begin{aligned} (\sigma, \mathbf{B})(\sigma, \mathbf{C}) &= \sum_{xyz} \{ \sigma_x^2 B_x C_x + \sigma_x \sigma_y B_x C_y + \sigma_y \sigma_x B_y C_x \} \\ &= (\mathbf{B}, \mathbf{C}) + i \sum_{xyz} \sigma_z (B_x C_y - B_y C_x) \\ &= (\mathbf{B}, \mathbf{C}) + i(\sigma, \mathbf{B} \times \mathbf{C}). \end{aligned} \quad (25)$$

Taking $\mathbf{B} = \mathbf{C} = \mathbf{p} + e/c \cdot \mathbf{A}$, we find, since

$$\begin{aligned} \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right) \times \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right) &= \frac{e}{c} \{ \mathbf{p} \times \mathbf{A} + \mathbf{A} \times \mathbf{p} \} \\ &= -i\hbar e/c \cdot \text{curl } \mathbf{A} = -i\hbar e/c \cdot \mathcal{H}, \end{aligned}$$

where \mathcal{H} is the magnetic field, that

$$\left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 = \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 + \frac{\hbar e}{c} (\sigma, \mathcal{H}).$$

Also we have

$$\begin{aligned} \left(\frac{W}{c} + \frac{e}{c} A_0 \right) \left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) - \left(\sigma, \mathbf{p} + \frac{e}{c} \mathbf{A} \right) \left(\frac{W}{c} + \frac{e}{c} A_0 \right) &= \\ = \frac{e}{c} \left(\sigma, \frac{W}{c} \mathbf{A} - \mathbf{A} \frac{W}{c} + A_0 \mathbf{p} - \mathbf{p} A_0 \right) & \\ = \frac{i\hbar e}{c} \left(\sigma, \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} + \text{grad } A_0 \right) &= -i \frac{\hbar e}{c} (\sigma, \mathcal{E}), \end{aligned}$$

where \mathcal{E} is the electric field. Thus (24) becomes

$$\left\{ \left(\frac{W}{c} + \frac{e}{c} A_0 \right)^2 - \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 - m^2 c^2 - \frac{\hbar e}{c} (\sigma, \mathcal{H}) - i\rho_1 \frac{\hbar e}{c} (\sigma, \mathcal{E}) \right\} \psi = 0.$$

This equation differs from (23) through having two extra terms in the operator. The electron according to the present theory is more closely analogous to a classical system with the Hamiltonian function

$$\left(\frac{W}{c} + \frac{e}{c} A_0 \right)^2 - \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 - m^2 c^2 - \frac{\hbar e}{c} (\sigma, \mathcal{H}) - i\rho_1 \frac{\hbar e}{c} (\sigma, \mathcal{E}).$$

If we neglect relativity corrections, so that we can put $W = mc^2 + W_1$

and count W_1 as small, this Hamiltonian reduces, after division throughout by $2m$, to

$$W_1 = \left\{ -eA_0 + \frac{1}{2m} \left(\mathbf{p} + \frac{e}{c} \mathbf{A} \right)^2 + \frac{\hbar e}{2mc} (\sigma, \mathcal{H}) + i\rho_1 \frac{\hbar e}{2mc} (\sigma, \mathcal{E}) \right\}.$$

We can now see that the two extra terms may be considered approximately as due to the electron possessing an additional potential energy of amount

$$\frac{\hbar e}{2mc} (\sigma, \mathcal{H}) + i\rho_1 \frac{\hbar e}{2mc} (\sigma, \mathcal{E}),$$

which may be interpreted as arising from *the electron having a magnetic moment $-\hbar e/2mc \cdot \sigma$ and an electric moment $-i\rho_1 \hbar e/2mc \cdot \sigma$.* This magnetic moment is in agreement with the assumptions of § 43 and is what is required by experiment. The electric moment, on the other hand, is a pure imaginary quantity and thus cannot be considered as having a physical meaning. The Hamiltonian of our original wave equation (8) is real, and the imaginary term has appeared only on account of our having performed a rather artificial operation to get a Hamiltonian that can be compared with the classical one.

The spin angular momentum does not give rise to any potential energy and therefore does not appear in the result of the preceding calculation. The simplest way of showing the existence of the spin angular momentum is to take the case of the motion of an electron in a central field of force and determine the angular momentum integrals. We therefore take $\mathbf{A} = 0$ and A_0 a function of r only, so that the wave equation (8) becomes

$$(W - H)\psi = 0,$$

where

$$H = -eA_0(r) - c\rho_1(\sigma, \mathbf{p}) - \rho_3 mc^2. \quad (26)$$

This H is the Hamiltonian to be used in the equations of motion.

If we take the x -component of orbital angular momentum, $m_x = yp_z - zp_y$, we find for its rate of change, with the help of commutability relations proved in §§ 44 and 45,

$$\begin{aligned} i\hbar \dot{m}_x &= m_x H - H m_x \\ &= -c\rho_1 \{m_x(\sigma, \mathbf{p}) - (\sigma, \mathbf{p})m_x\} \\ &= -c\rho_1 (\sigma, m_x \mathbf{p} - \mathbf{p} m_x) \\ &= -i\hbar c \rho_1 \{ \sigma_y p_z - \sigma_z p_y \}. \end{aligned}$$

$\propto k$

Thus $\dot{m}_x \neq 0$ and the orbital angular momentum is not a constant of the motion. We have further

$$\begin{aligned} i\hbar\dot{\sigma}_x &= \sigma_x H - H\sigma_x \\ &= -c\rho_1\{\sigma_x(\sigma, p) - (\sigma, p)\sigma_x\} \\ &= -c\rho_1(\sigma_x\sigma - \sigma\sigma_x, p) \\ &= -2ic\rho_1\{\sigma_z p_y - \sigma_y p_z\} \end{aligned}$$

with the help of equations (42) of § 43. Hence

$$i\hbar(\dot{m}_x + \frac{1}{2}\hbar\dot{\sigma}_x) = 0,$$

so that the vector $\mathbf{m} + \frac{1}{2}\hbar\sigma$ is a constant of the motion. This result one can interpret by saying *the electron has a spin angular momentum $\frac{1}{2}\hbar\sigma$* , which must be added to the orbital angular momentum \mathbf{m} before one gets a constant of the motion.

§ 77. Transition to Polar Variables

For the further study of the motion of an electron in a central field of force, it is convenient to make a transformation to polar co-ordinates, as was done in § 45 in the non-relativity case. We can introduce r and p_r , as before, but instead of k , the magnitude of the orbital angular momentum \mathbf{m} , which is no longer a constant of the motion, we must now use the magnitude of the total angular momentum $\mathbf{M} = \mathbf{m} + \frac{1}{2}\hbar\sigma$. If j is this magnitude expressed in units of \hbar , we shall have

$$j^2\hbar^2 = M_x^2 + M_y^2 + M_z^2 + \frac{1}{4}\hbar^2. \quad (27)$$

The eigenvalues of m_z are integral multiples of \hbar , those of $\frac{1}{2}\hbar\sigma_z$ are $\pm\frac{1}{2}\hbar$, and hence those of M_z must be half-odd integral multiples of \hbar . It follows from the general result of § 30 that the eigenvalues of j must be integers greater than zero.

If in formula (25) we take $\mathbf{B} = \mathbf{C} = \mathbf{m}$, we get

$$\begin{aligned} (\sigma, \mathbf{m})^2 &= \mathbf{m}^2 + i(\sigma, \mathbf{m} \times \mathbf{m}) \\ &= \mathbf{m}^2 - \hbar(\sigma, \mathbf{m}) \\ &= (\mathbf{m} + \frac{1}{2}\hbar\sigma)^2 - 2\hbar(\sigma, \mathbf{m}) - \frac{3}{4}\hbar^2. \end{aligned}$$

Hence $(\sigma, \mathbf{m}) + \hbar = \mathbf{M}^2 + \frac{1}{4}\hbar^2$.

Thus $(\sigma, \mathbf{m}) + \hbar$ is a quantity whose square is $\mathbf{M}^2 + \frac{1}{4}\hbar^2$ and we could, consistently with equation (27), define $j\hbar$ as $(\sigma, \mathbf{m}) + \hbar$ instead of as the positive square root of $\mathbf{M}^2 + \frac{1}{4}\hbar^2$. This would not be convenient,

however, since we want j to be a constant of the motion and $(\sigma, \mathbf{m}) + \hbar$ is not constant. We have, in fact, from applications of (25),

$$(\sigma, \mathbf{m})(\sigma, \mathbf{p}) = i(\sigma, \mathbf{m} \times \mathbf{p})$$

and

$$(\sigma, \mathbf{p})(\sigma, \mathbf{m}) = i(\sigma, \mathbf{p} \times \mathbf{m}),$$

so that

$$\begin{aligned} (\sigma, \mathbf{m})(\sigma, \mathbf{p}) + (\sigma, \mathbf{p})(\sigma, \mathbf{m}) &= i \sum_{xyz} \sigma_x \{ m_y p_z - m_z p_y + p_y m_z - p_z m_y \} \\ &= i \sum_{xyz} \sigma_x \cdot 2i\hbar p_x = -2\hbar(\sigma, \mathbf{p}), \end{aligned}$$

or $\{(\sigma, \mathbf{m}) + \hbar\}(\sigma, \mathbf{p}) + (\sigma, \mathbf{p})\{(\sigma, \mathbf{m}) + \hbar\} = 0.$

Thus $(\sigma, \mathbf{m}) + \hbar$ anticommutes with one of the terms in the expression (26) for H , namely the term $-c\rho_1(\sigma, \mathbf{p})$, and commutes with the other two. It follows that $\rho_3\{(\sigma, \mathbf{m}) + \hbar\}$ commutes with all the three terms in H and is a constant of the motion. But the square of $\rho_3\{(\sigma, \mathbf{m}) + \hbar\}$ is also $\mathbf{M}^2 + \frac{1}{4}\hbar^2$. We can therefore take

$$j\hbar = \rho_3\{(\sigma, \mathbf{m}) + \hbar\}, \quad (28)$$

which gives us a convenient rational definition for j which is consistent with (27) and makes j a constant of the motion. The eigenvalues of this j are all positive and negative integers, excluding zero.

By a further application of (25), we get

$$\begin{aligned} (\sigma, \mathbf{x})(\sigma, \mathbf{p}) &= (\mathbf{x}, \mathbf{p}) + i(\sigma, \mathbf{m}) \\ &= rp_r + i\rho_3 j\hbar, \end{aligned} \quad (29)$$

with the help of (28) and also of equation (13) of Chapter VIII. We introduce the observable ϵ defined by

$$r\epsilon = \rho_1(\sigma, \mathbf{x}). \quad (30)$$

Since r commutes with ρ_1 , and with (σ, \mathbf{x}) , it must commute with ϵ . We thus have

$$r^2\epsilon^2 = [\rho_1(\sigma, \mathbf{x})]^2 = (\sigma, \mathbf{x})^2 = \mathbf{x}^2 = r^2,$$

or

$$\epsilon^2 = 1.$$

Since there is symmetry between \mathbf{x} and \mathbf{p} so far as angular momentum is concerned, $\rho_1(\sigma, \mathbf{x})$, like $\rho_1(\sigma, \mathbf{p})$, must commute with \mathbf{M} and j . Hence ϵ commutes with \mathbf{M} and j . Further, ϵ must commute with p_r , since we have

$$(\sigma, \mathbf{x})(\mathbf{x}, \mathbf{p}) - (\mathbf{x}, \mathbf{p})(\sigma, \mathbf{x}) = (\sigma, \mathbf{x}(\mathbf{x}, \mathbf{p}) - (\mathbf{x}, \mathbf{p})\mathbf{x}) = i\hbar(\sigma, \mathbf{x}),$$

which gives $r\epsilon(rp_r + i\hbar) - (rp_r + i\hbar)r\epsilon = i\hbar r\epsilon$

or $r\epsilon(p_r r + 2i\hbar) - (rp_r + i\hbar)r\epsilon = i\hbar r\epsilon,$

which reduces to $\epsilon p_r - p_r \epsilon = 0.$

From (29) and (30) we obtain

$$r\epsilon\rho_1(\sigma, \mathbf{p}) = rp_r + i\rho_3 j\hbar$$

or

$$\rho_1(\sigma, \mathbf{p}) = \epsilon p_r + i\epsilon\rho_3 j\hbar/r.$$

Thus

$$H/c = -e/c \cdot A_0 - \epsilon p_r - i\epsilon\rho_3 j\hbar/r - \rho_3 mc.$$

This gives our Hamiltonian expressed in terms of polar variables.

It should be noticed that ϵ and ρ_3 commute with all the other variables occurring in H and anticommute with one another. This means that we can take a representation in which ϵ and ρ_3 are represented respectively by the matrices

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (31)$$

and in which r , say, is diagonal, and the wave function $(r|)$ will then have two components, $(r|)_a$ and $(r|)_b$ say, referring to the two rows and columns of the matrices.

§ 78. The Fine-Structure of the Energy-Levels of Hydrogen

We shall now take the case of the hydrogen atom, for which $A_0 = 0$ and work out its energy-levels, given by the eigenvalues H' of H . The equation $(H' - H)\psi = 0$ which defines these eigenvalues, written in terms of representatives in the representation discussed above with ϵ and ρ_3 represented by the matrices (31), gives the equations

$$\left(\frac{H'}{c} + \frac{e^2}{cr} \right) (r|)_a - \hbar \frac{\partial}{\partial r} (r|)_b - \frac{j\hbar}{r} (r|)_b + mc (r|)_a = 0$$

$$\left(\frac{H'}{c} + \frac{e^2}{cr} \right) (r|)_b + \hbar \frac{\partial}{\partial r} (r|)_a - \frac{j\hbar}{r} (r|)_a - mc (r|)_b = 0.$$

If we put

$$\frac{\hbar}{mc + H'/c} = a_1, \quad \frac{\hbar}{mc - H'/c} = a_2, \quad (32)$$

these equations reduce to

$$\left. \begin{aligned} \left(\frac{1}{a_1} + \frac{\alpha}{r} \right) (r|)_a - \left(\frac{\partial}{\partial r} + \frac{j}{r} \right) (r|)_b &= 0 \\ \left(-\frac{1}{a_2} + \frac{\alpha}{r} \right) (r|)_b + \left(\frac{\partial}{\partial r} - \frac{j}{r} \right) (r|)_a &= 0, \end{aligned} \right\} (33)$$

where $\alpha = e^2/\hbar c$, which is a small number. We shall solve these equations by a similar method to that used for equation (20) in § 46.

Put

$$(r|)_a = e^{-r/a} f, \quad (r|)_b = e^{-r/a} g,$$

introducing two new functions, f and g , of r , where

$$a = (a_1 a_2)^{\frac{1}{2}} = \hbar(m^2 c^2 - H'^2/c^2)^{-\frac{1}{2}}. \quad (34)$$

Equations (33) become

$$\left. \begin{aligned} \left(\frac{1}{a_1} + \frac{\alpha}{r} \right) f - \left(\frac{\partial}{\partial r} - \frac{1}{a} + \frac{j}{r} \right) g &= 0 \\ \left(-\frac{1}{a_2} + \frac{\alpha}{r} \right) g + \left(\frac{\partial}{\partial r} - \frac{1}{a} - \frac{j}{r} \right) f &= 0. \end{aligned} \right\} \quad (35)$$

We now try for a solution in which f and g are in the form of power series,

$$f = \sum_s c_s r^s, \quad g = \sum_s c'_s r^s, \quad (36)$$

in which consecutive values of s differ by unity though these values need not be integers. Substituting these expressions for f and g in (35) and picking out coefficients of r^{s-1} , we obtain

$$\left. \begin{aligned} c_{s-1}/a_1 + \alpha c_s - (s+j)c'_s + c'_{s-1}/a &= 0 \\ -c'_{s-1}/a_2 + \alpha c'_s + (s-j)c_s - c_{s-1}/a &= 0. \end{aligned} \right\} \quad (37)$$

By multiplying the first of these equations by a and the second by a_2 and adding, we can eliminate both c_{s-1} and c'_{s-1} , since from (34) $a/a_1 = a_2/a$. This gives

$$c_s[a\alpha + a_2(s-j)] + c'_s[a_2\alpha - a(s+j)] = 0, \quad (38)$$

a relation which shows the connexion between the primed and unprimed c 's.

The boundary condition at $r = 0$ requires that the series (36) shall terminate on the side of small s . If s_0 is the minimum value of s for which c_s and c'_s do not both vanish, we obtain from (37), by putting $s = s_0$ and $c_{s_0-1} = c'_{s_0-1} = 0$,

$$\begin{aligned} \alpha c_{s_0} - (s_0 + j)c'_{s_0} &= 0 \\ \alpha c'_{s_0} + (s_0 - j)c_{s_0} &= 0, \end{aligned}$$

which give

$$\alpha^2 = -s_0^2 + j^2.$$

Since the boundary condition requires that the minimum value of s shall be greater than zero, we must take

$$s_0 = +\sqrt{j^2 - \alpha^2}.$$

To investigate the convergence of the series (36) we shall determine the ratio c_s/c_{s-1} for large s . Equation (38) and the second of equations (37) give approximately, when s is large,

$$c_s a_2 = c'_s a$$

and

$$\varepsilon c_s = c_{s-1}/a + c'_{s-1}/a_2.$$

Hence

$$c_s/c_{s-1} = 2/as.$$

The series (36) will therefore converge like

$$\sum_s \frac{1}{s!} \left(\frac{2r}{a}\right)^s$$

or $e^{2r/a}$. This result is similar to that obtained in § 46 and allows us to infer, as before, that all values of H' are permissible for which a is pure imaginary, *i.e.* for which, from (34), $H' > mc^2$, but of those values of H' for which a is real, only those are permissible for which the series (36) terminate on the side of large s .

If the series (36) terminate with the terms c_s and c'_s , so that $c_{s+1} = c'_{s+1} = 0$, we obtain from (37) with $s+1$ substituted for s

$$\begin{aligned} c_s/a_1 + c'_s/a &= 0 \\ -c'_s/a_2 - c_s/a &= 0. \end{aligned}$$

These two equations are equivalent on account of (34). When combined with (38), they give

$$a_1[a\alpha + a_2(s-j)] = a[a_2\alpha - a(s+j)],$$

which reduces to

$$2a_1a_2s = a(a_2 - a_1)\alpha$$

or

$$\frac{s}{a} = \frac{1}{2} \left(\frac{1}{a_1} - \frac{1}{a_2} \right) \alpha = \frac{H'}{ch} \alpha,$$

with the help of (32). Squaring and using (34), we obtain

$$s^2(m^2c^2 - H'^2/c^2) = \alpha^2 H'^2/c^2.$$

Hence

$$\frac{H'}{mc^2} = \left(1 + \frac{\alpha^2}{s^2} \right)^{-\frac{1}{2}}.$$

The s here, which specifies the last term in the series, must be greater than s_0 by some integer not less than zero. Calling this integer n , we have

$$s = n + \sqrt{j^2 - \alpha^2}$$

and thus

$$\frac{H'}{mc^2} = \left\{ 1 + \frac{\alpha^2}{(n + \sqrt{j^2 - \alpha^2})^2} \right\}^{-\frac{1}{2}}.$$

This formula gives the discrete energy-levels of the hydrogen spectrum and was first obtained by Sommerfeld working with Bohr's orbit theory. There are two quantum numbers n and j involved, but owing to α^2 being very small the energy depends almost entirely on

$n+|j|$. Values of n and $|j|$ that give the same $n+|j|$ give rise to a set of energy-levels lying very closely to one another, and to the energy-level given by the non-relativistic formula (27) of § 46 with $s = n+|j|$.

For a general value of n , j can have any integral value except zero. The value $n=0$ is, however, exceptional as it makes equation (38) vanish identically. A closer investigation shows that in this case only negative values for j are allowed.*

§ 79. Physical Meaning of the Negative-Energy Solutions

It has been mentioned before that the wave equation for the electron admits of twice as many solutions as it ought to, half of them referring to states with negative values for the kinetic energy $W+eA_0$. This difficulty was introduced as soon as we passed from equation (3) to equation (4) and is inherent in any relativity theory. It occurs also in classical relativity theory, but is not then serious since, owing to the continuity in the variation of all classical dynamical variables, if the kinetic energy $W+eA_0$ is initially positive (when it must be greater than or equal to mc^2), it cannot subsequently be negative (when it would have to be less than or equal to $-mc^2$). In the quantum theory, however, discontinuous transitions may take place, so that if the electron is initially in a state of positive kinetic energy it may make a transition to a state of negative kinetic energy. It is therefore no longer permissible simply to ignore the negative-energy states, as one can do in the classical theory.

Let us examine the negative-energy solutions of the equation

$$\left\{ \left(\frac{W}{c} + \frac{e}{c} A_0 \right) + \alpha_x \left(p_x + \frac{e}{c} A_x \right) + \alpha_y \left(p_y + \frac{e}{c} A_y \right) + \alpha_z \left(p_z + \frac{e}{c} A_z \right) + \alpha_m mc \right\} \psi = 0 \quad (39)$$

a little more closely. For this purpose it is convenient to use a representation of the α 's in which all the elements of the matrices representing α_x , α_y , and α_z are real and all those of the matrix representing α_m are pure imaginary. Such a representation may be obtained, for instance, from that of § 74 by interchanging the expressions for α_y and α_m in (7). With such a representation, if we write

* See W. Gordon, *Zeits. f. Phys.*, vol. xlvi, p. 11 (1928).

$-i$ for i in the operator of equation (39), we get, remembering (1) and (2),

$$\left(-\frac{W}{c} + \frac{e}{c} A_0 \right) + \alpha_x \left(-p_x + \frac{e}{c} A_x \right) + \alpha_y \left(-p_y + \frac{e}{c} A_y \right) + \left. \alpha_z \left(-p_z + \frac{e}{c} A_z \right) - \alpha_m mc \right\} \psi = 0. \quad (40)$$

Thus the conjugate complex of any wave function that is a solution of (39) is a solution of (40). Further, if the solution of (39) belongs to a negative value for $W + eA_0$, the conjugate complex solution of (40) will belong to a positive value for $W - eA_0$. But equation (40) is just what one would get if one substituted $-e$ for e in (39). It follows that the conjugate complex of any solution of (39) belonging to a negative value for $W + eA_0$ is a solution, belonging to a positive value for $W - eA_0$, of the wave equation obtained from (39) by substitution of $-e$ for e , and therefore represents an electron of charge $+e$, instead of $-e$, moving through the given electromagnetic field. Thus the unwanted solutions of (39) are connected with the motion of an electron with a charge $+e$. (It is not possible, of course, with an arbitrary electromagnetic field, to separate the solutions of (39) definitely into those referring to positive and those referring to negative values for $W + eA_0$, as such a separation would imply that transitions from one kind to the other do not occur. The preceding discussion is therefore only a rough one, applying to the case when such a separation is approximately possible.)

In this way we are led to infer that the negative-energy solutions of (39) refer to the motion of protons or hydrogen nuclei, although there remains the difficulty of the great difference in the masses. We cannot, however, simply assert that the negative-energy solutions represent protons, as this would make the dynamical relations all wrong. For instance, it is certainly not true that a proton has a negative kinetic energy. We must therefore establish the protons on a somewhat different footing. We assume that *nearly all the negative-energy states are occupied*, with one electron in each state in accordance with the exclusion principle of Pauli. An unoccupied negative-energy state will now appear as something with a positive energy, since to make it disappear, *i.e.* to fill it up, we should have to add to it an electron with negative energy. We assume that *these unoccupied negative-energy states are the protons*.

These assumptions require there to be a distribution of electrons of infinite density everywhere in the world. A perfect vacuum is a region where all the states of positive energy are unoccupied and all those of negative energy are occupied. In a perfect vacuum Maxwell's equation

$$\operatorname{div} \mathbf{E} = 0$$

must, of course, be valid. This means that the infinite distribution of negative-energy electrons does not contribute to the electric field. Only departures from the distribution in a vacuum will contribute to the electric density ρ in Maxwell's equation

$$\operatorname{div} \mathbf{E} = -4\pi\rho.$$

Thus there will be a contribution $-e$ for each occupied state of positive energy and a contribution $+e$ for each unoccupied state of negative energy.

The exclusion principle will operate to prevent a positive-energy electron ordinarily from making transitions to states of negative energy. It will still be possible, however, for such an electron to drop into an unoccupied state of negative energy. In this case we should have an electron and proton disappearing simultaneously, their energy being emitted in the form of radiation. Such processes probably actually occur in nature.

The present theory is very symmetrical between the electrons and protons. The symmetry is not mathematically perfect, as may easily be verified, when one takes interaction between the electrons into account. This cause, however, hardly appears to be sufficient, according to present ideas, to account for the very considerable observed differences between electrons and protons, in particular their different masses. Possibly the solution of this difficulty will be found in a better understanding of the nature of interaction.

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRESS, OXFORD
BY JOHN JOHNSON, PRINTER TO THE UNIVERSITY

