

THE CALCULUS OF VARIATIONS AND THE QUANTUM THEORY*

BY G. A. BLISS

1. *Introduction.* Several years ago an insistent curiosity caused me to forsake temporarily fields of mathematical interest more legitimate for me in order to find out if possible the character of the new quantum theories in which my physicist neighbors and some of my mathematical friends seemed to find so much of scientific interest and excitement. I was entirely ignorant of the theory at that time and was greatly surprised to find that my own specialty, the calculus of variations, through the media of mechanical and optical conceptions, had played an important role in its development. The quantum theory was then in an exceedingly fluent stage, and though the unanswered mathematical questions which presented themselves were numerous and significant it seemed to me useless for mathematicians to spend large amounts of time in endeavoring to perfect mathematical details when some new paper might at short notice cause fundamental changes in the whole structure of the theory. Every one who has followed the development of the quantum theory would agree, I think, that this impression was justified. At the present time the situation seems somewhat different. I have heard assurances from various reliable sources that the leaders in the promotion of quantum mechanics, men of wide physical experience and of daring in mathematical exploration distinctly beyond that of the average pure mathematician who takes his rules so seriously, have reached something like agreement. These statements should perhaps not be accepted too readily, as they sound like others that have been made in the past and which have afterward needed modification. But to the seeker after knowledge, less experienced in this field like myself, it also seems that an equilibrium has approx-

* Retiring Address of the Vice-President of Section A of the American Association for the Advancement of Science, delivered at New Orleans, December 29, 1931, at a joint meeting of this Society, the American Physical Society, and Section A of the A.A.A.S.

imately been reached in the fundamentals of the quantum theory, and that pure mathematicians could therefore reasonably endeavor to contribute in a valuable way to the development and perfection of its mathematical structure.

In the following pages I do not intend to attempt to point out mathematical defects in the quantum theory, though it seems clear that such defects exist. They appear for the most part in the generalities and less often in the special problems which have so far been successfully studied. My endeavor here will rather be to present some of the mathematical aspects of the theory to those of my mathematical colleagues who have not yet delved in the field. They will agree with me, I think, that there is no other applied mathematical domain whose development has involved a wider variety of mathematical techniques.

This paper was written originally as a retiring address for a chairman of the mathematicians in Section A of the Association. I was somewhat perturbed when I found that it had been allotted to a joint meeting of mathematicians and physicists. To those of the latter class who are here present I can only express my sympathy. It will be impossible for me to enlighten them with regard to quantum mechanics, and my mathematics may seem to them too technical. My only comfort is that it may possibly not seem more so than that of their own colleagues who specialize in the quantum theory.

2. *Normal Congruences and Optical Fields.* One of the simplest problems of the calculus of variations is that of finding in the class of arcs

$$y = y(x), \quad z = z(x), \quad (x_1 \leq x \leq x_2),$$

joining two fixed points in xyz -space, one which minimizes the length integral

$$I = \int_{x_1}^{x_2} (1 + y'^2 + z'^2)^{1/2} dx.$$

This is a relatively simple problem concerning whose solution we know much intuitively. The possible minimizing curves are straight lines

$$(1) \quad y = ax + b, \quad z = cx + d,$$

the totality of which depends upon four constants a, b, c, d . A *congruence of lines* is a two-parameter family selected out of this four-parameter family by setting a, b, c, d equal to functions of two other parameters u, v . Geometers have long known that the lines of such a congruence will not in general have a surface which they all cut at right angles. If, however, there is one surface cutting all the lines of the congruence orthogonally, then there is a whole family of such surfaces given by an equation of the form

$$(2) \quad S(x, y, z) = \text{constant}.$$

The congruence is in that case called a *normal congruence*. Such congruences of lines have many interesting properties. For example the segments of lines of the congruence intercepted by two of the surfaces (2) are all equal in length, and the congruence has so-called focal surfaces to which all of the lines of the congruence are tangent. But the property which interests us most here is that a normal congruence may be determined by its family (2) of orthogonal surfaces as well as by its lines. Not every family (2) has a congruence of straight lines cutting it at right angles. The characteristic property of those families of surfaces which do belong to a normal congruence is that for such families the function $S(x, y, z)$ can always be selected so that it satisfies the differential equation

$$S_x^2 + S_y^2 + S_z^2 = 1,$$

where S_x, S_y, S_z are the partial derivatives of S with respect to x, y , and z .

All this will seem rather familiar to my physicist hearers because the principal business of the theory of geometrical optics is with normal congruences. The straight lines are the light rays, the orthogonal surfaces are the wave fronts, and the focal surfaces are the caustic surfaces of optical theory. An interesting property of an optical field of rays for the physicist is of course that it retains its character after a reflection or a refraction. Sir William Hamilton devoted one of his early published papers* to the study of the properties of such optical fields. It was the

* *Theory of systems of rays*, Transactions of the Royal Irish Academy, vol. 15 (1824), pp. 69–173; vol. 16, Part I, pp. 3–61, Part II, pp. 93–125; vol. 17, pp. 1–144.

more than 300-page effort of a youthful writer, and probably unnecessarily long, but it contains the source of the ideas which later influenced profoundly the development of the calculus of variations and the theory of analytical mechanics.

3. *Generalizations in the Calculus of Variations.* It has been found that properties analogous to those of normal congruences are possessed by the minimizing curves of far more general problems of the calculus of variations than that of determining the shortest distance. When we try to minimize an integral

$$I = \int_{x_1}^{x_2} f(x, y_1, y_2, y_1', y_2') dx$$

in the class of arcs

$$y_1 = y_1(x), \quad y_2 = y_2(x), \quad (x_1 \leq x \leq x_2)$$

joining two fixed points in xy_1y_2 -space we find that the only possible minimizing curves are the solutions of the differential equations

$$(3) \quad \frac{df_{y'_k}}{dx} = f_{y_k}, \quad (k = 1, 2).$$

I use the notations x, y_1, y_2 instead of x, y, z because then all of our equations can be readily interpreted for the $(r+1)$ -dimensional case, instead of the 3-dimensional case, by simply changing the range $k=1, 2$ to $k=1, \dots, r$. Meanwhile, for the space of points (x, y_1, y_2) we can use the language of ordinary 3-dimensional geometry. The differential equations (3) are of the second order and their solutions, which are called *extremals* in the calculus of variations, form a family of curves whose equations, like those of the straight lines (1), contain four arbitrary constants.

It has been found, following Hamilton and probably earlier writers also, that the equations (3) take a very pleasing form when new variables x, y_k, z_k are introduced in place of the original variables x, y_k, y_k' by means of the equations

$$z_k = f_{y'_k}(x, y, y'), \quad (k = 1, 2).$$

Let us define a Hamiltonian function H by means of the equation

$$H(x, y, z) = y'_k f_{y'_k} - f,$$

it being understood here as elsewhere in this paper that a term containing the same subscript k twice stands for the sum of the terms of that form for the various special values in the range of k , as in tensor analysis. It can be proved that the equations of the extremals now take the form

$$(4) \quad \frac{dy_k}{dx} = H_{z_k}, \quad \frac{dz_k}{dx} = -H_{y_k}, \quad (k = 1, 2).$$

The variables x, y_k, z_k are called *canonical variables*, and the equations (4) are the *canonical equations* of the extremals. Equations (4) may also be written in the form

$$(5) \quad \frac{dy_k}{dx} = (y_k, H), \quad \frac{dz_k}{dx} = (z_k, H), \quad (k = 1, 2),$$

where the brackets on the right are so-called *Poisson brackets* defined for two arbitrary functions f and g of x, y_k, z_k by the equation

$$(6) \quad (f, g) = \frac{\partial f}{\partial y_k} \frac{\partial g}{\partial z_k} - \frac{\partial f}{\partial z_k} \frac{\partial g}{\partial y_k}.$$

For the fields of geometrical optics the equation which is satisfied when the direction $dx:dy:dz$ in xyz -space is at right angles with a direction $1:y':z'$ is the well known orthogonality condition

$$dx + y' dy + z' dz = 0.$$

In the calculus of variations it is found that this equation must be replaced by the equation

$$(f - y'_k f_{y'_k}) dx + f_{y'_k} dy_k = 0.$$

A direction $dx:dy_1:dy_2$ is transversal to the direction $1:y'_1:y'_2$ at the point (x, y) if it satisfies this equation; and a surface is transversal to $1:y'_1:y'_2$ at (x, y) if all the directions $dx:dy_1:dy_2$ tangent to the surface at (x, y) satisfy the equation.

We are now in possession of data sufficient to justify a generalization of the notion of a normal congruence of straight lines. Instead of a two-parameter family of straight lines we consider a two-parameter family of the extremal curves which satisfy the differential equations (3) or (4). Such a congruence

of curves is called a Mayer family, after the name of its discoverer, if there is a single surface which cuts every member of the family transversally. It is provable, then, that there is a single infinity of such surfaces, the so-called transversal surfaces of the Mayer family, defined by an equation of the form

$$S(x, y_1, y_2) = \text{constant},$$

and analogous to (2). Not every such set of surfaces is the set of transversal surfaces of a Mayer family. In order that this property may be possessed by the surfaces it is necessary and sufficient that they should be definable by an equation $S(x, y) = \text{constant}$ for which S satisfies the partial differential equation

$$S_x + H(x, y, S_y) = 0.$$

This is the famous Hamilton-Jacobi partial differential equation of the calculus of variations. It has many interesting relationships with the extremals besides the one which has just been described.

We need one further notion from the calculus of variations in order to complete the list of those upon which the development of the quantum theory has depended. The variables x, y_k, z_k introduced above are not the only ones for which the differential equations of the extremals have the canonical form. Transformations

$$(7) \quad x = x(X, Y, Z), \quad y_k = y_k(X, Y, Z), \quad z_k = z_k(X, Y, Z),$$

which take x, y_k, z_k into new variables X, Y_k, Z_k and which transform the canonical differential equations into others of the same form are called *canonical transformations*. Such transformations can be constructed in many different ways but the most important ones for the quantum theory are those which replace (x, y_k, z_k) by new variables $(X, Y_k, Z_k) = (x, w_k, J_k)$ by means of equations of the form

$$(8) \quad x = x, \quad z_k = S_{y_k}(x, y, J), \quad w_k = S_{J_k}(x, y, J),$$

where $S(x, y, J)$ is a solution of the differential equation

$$S_x + H(x, y, S_y) = E(J),$$

involving the variables J_i as parameters. The last set of equations (8) must be solved for the y_k and substituted in the second

in order to secure the transformation in the form (7). The important property of the transformation is that in terms of the variables x , w_k , J_k the differential equations of the extremals take the very simple form

$$\frac{dw_k}{dx} = E_{J_k}, \quad \frac{dJ_k}{dx} = 0,$$

and the equations of the extremals are

$$(9) \quad w_k = E_{J_k}(J)t + \alpha_k, \quad J_k = \beta_k,$$

where the α_k 's and β_k 's are constants.

The results from the calculus of variations mentioned in the preceding paragraphs have been described for the 3-dimensional space of points (x, y_1, y_2) . But the equations which have been written are valid for the $(r+1)$ -dimensional space of points (x, y_1, \dots, y_r) if we change the range of the indices k from 1, 2 to 1, \dots , r . The number of parameters in the equations of the family of extremals is of course $2r$ instead of 4, and the Mayer families for $(r+1)$ -dimensional space contain r parameters instead of only 2.

4. *Applications of the Calculus of Variations.* The calculus of variations appears in the quantum theory through its applications in mechanics. The equations of motion of a particle of mass m moving in a field of force with the potential function $V(x, y, z)$ are the equations

$$(10) \quad \frac{dmx'}{dt} = -V_x, \quad \frac{dmy'}{dt} = -V_y, \quad \frac{dmz'}{dt} = -V_z,$$

where the primes indicate derivatives with respect to t . According to the well known principle of Hamilton they are the differential equations of the extremals of the integral

$$(11) \quad I = \int_{t_1}^{t_2} (T - V)dt,$$

where $T = m(x'^2 + y'^2 + z'^2)/2$ is the kinetic energy of the moving particle. Thus all of the results of the calculus of variations described in the preceding section, with suitable changes in notation, find their applications in mechanics. It is not neces-

sary to review them in detail for this special case. It should be mentioned, however, that a similar formulation of Hamilton's principle holds for n particles instead of one, and for any auxiliary variables q_i in terms of which the coordinates of the particles may be expressed. The variables t, q_k, q_k' replace the x, y_k, y_k' of the preceding section. The kinetic energy T is a homogeneous quadratic form in the derivatives q_k' with coefficients functions of the q_k , and the canonical variables are t, q_k, p_k , where the momenta p_k are defined by the equations $p_k = T_{q_k'}$. The Hamiltonian function $H(t, p, q)$ turns out to be the total energy $H = T + V$ of the system and is equal to a constant E for the motion of every system.

This is a familiar basis for the theory of Newtonian analytical mechanics, but the modification of equations (10) which arises in the theory of relativity should also be mentioned. In the non-relativistic case the mass m is a constant, but in the relativity theory it varies in accordance with the law $m = m_0/(1 - v^2/c^2)^{1/2}$, where m_0 is the so-called rest-mass, v is the velocity of the particle, and c the velocity of light. The equations (10) are no longer the differential equations of the extremals of the integral (11) when T has the value described above. They will retain this property, however, if we replace T by the function

$$T_1 = m_0 c^2 [1 - (1 - v^2/c^2)^{1/2}],$$

as can readily be verified. The Hamiltonian function for this relativity case turns out to have the value $H = T_2 + V$, where the kinetic energy T_2 is now given by the formula

$$T_2 = m_0 c^2 [(1 - v^2/c^2)^{-1/2} - 1].$$

There are two further principles, important for the purposes of this paper, which are phrased in terms of concepts of the calculus of variations. The principle of least action in mechanics says that every trajectory of a particle of mass m , moving with total energy E in a field of force with the potential function V , is an extremal of the integral

$$\begin{aligned} I &= \int_{\tau_1}^{\tau_2} [2m(E - V)(x_\tau^2 + y_\tau^2 + z_\tau^2)]^{1/2} d\tau \\ (12) \quad &= \int_{s_1}^{s_2} [2m(E - V)]^{1/2} ds, \end{aligned}$$

where τ is a parameter in terms of which the coordinates x , y , z of the trajectory are to be expressed and s is the length of arc. The time t at which the particle arrives at the place defined on the trajectory by a value of τ is expressible by the equation

$$t = t_1 + \int_{\tau_1}^{\tau} [m(x_\tau^2 + y_\tau^2 + z_\tau^2)/(2(E - V))]^{1/2} d\tau.$$

Fermat's principle in geometrical optics has a form much like that of this mechanical principle of least action. It says that rays of light in a medium in which the velocity of light at a point (x, y, z) is $u(x, y, z)$ are extremals of the integral

$$(13) \quad I = C \int_{\tau_1}^{\tau_2} (1/u)(x_\tau^2 + y_\tau^2 + z_\tau^2)^{1/2} d\tau = C \int_{s_1}^{s_2} (1/u) ds,$$

where C is an arbitrary constant whose value does not affect the determination of the extremals in any way. The two integrals (12) and (13) are identical if

$$(14) \quad u = C/[2m(E - V)]^{1/2}.$$

Hence the dynamical trajectories of a particle moving with total energy E are the same as the paths of light in a medium where the velocity of light at each point is given by the equation (14).

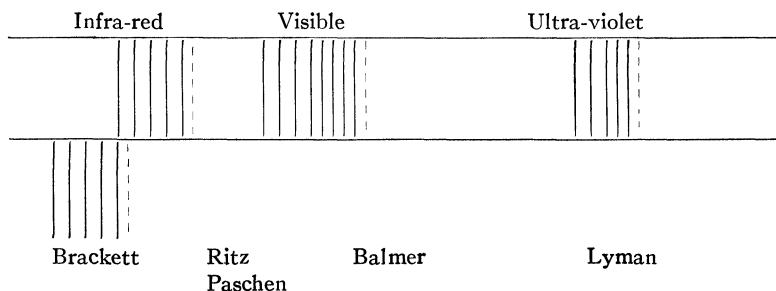
John Bernoulli used an analogy similar to this in 1697 when he determined the curve down which a particle would fall under the action of gravity from one fixed point to another in the shortest time. He found that the time of descent was given, except for a constant factor, by the integral

$$\int_{s_1}^{s_2} z^{-1/2} ds.$$

Hence the curve of quickest descent must be a light-ray in a medium where $u = Cz^{1/2}$. By dividing the medium into horizontal layers and applying the law of refraction he was able to find the differential equation of the minimizing curve and to show that the curve of quickest descent is an inverted cycloid. The paper in which he found this result is often designated as the origin of the calculus of variations.

5. *The Quantum Theory of Bohr.* In my effort to give my fellow mathematicians some idea of the role which the calculus of variations has played in the development of the quantum theory I do not hesitate to begin with the early theory of Bohr, even though it has been superseded in recent years by quite different mathematical mechanisms. It still affords a convenient picture by means of which the physicist correlates a considerable part of the data in which the quantum theory had its source, and with the theory of Schrödinger it forms the chapters of the development of the quantum theory in which the calculus of variations had its greatest influence. Furthermore I believe that a knowledge of these earlier stages is essential to an understanding of quantum mechanics in its more recent forms.

To the mathematicians I may explain that one of the first problems of the quantum theory, and one of the least complicated illustrations of it, is the theory of the spectrum of hydrogen. On the one hand we have the spectrum, itself, as shown in the figure which is drawn without reference to scale. The lines fall into four series which have been observed experimentally, and others not indicated, each of which has lines converging to a dotted limiting line. Each line is due to light of a particular wave length λ , and when the figure is drawn to scale the lines



are spaced proportionally to their wave numbers $\nu = 1/\lambda$. It has been found empirically that these wave numbers are given with great accuracy by the famous Balmer formula

$$\nu = R(1/m^2 - 1/n^2),$$

in which R is a constant and m and n are integers. The Lyman series is the sequence of lines corresponding to $m=1$ and arbitrary positive integral values for n , the Balmer series to

$m = 2$, and so on. Associated with this picture was the physicist's conception of a hydrogen atom as a dynamical system consisting of a heavy positively charged nucleus with a much lighter electron carrying an equal negative charge revolving around it. What could be the connection between the two?

To answer this question, Bohr assumed that the electron is not free to move in all of the orbits which would be permitted by ordinary dynamical theory, but only in any one of a denumerable infinity of orbits, the so-called stationary states of motion, selected in a way which will presently be described. Radiation was supposed to be emitted only when the electron fell from one of these orbits to another in which it had a smaller total energy, and the wave number of the radiation emitted was proportional to the difference $E_1 - E_2$ between the total energies of the atomic system in the initial and final states. This somewhat crude dynamical picture corresponds to the fact that hydrogen gas does not emit radiation except when suitably excited, that is, except when the electrons in its atoms have been knocked by a bombardment of some sort into stationary states with larger total energies from which they can fall to states with smaller energy-values. Similar assumptions were made for much more complicated atoms with many electrons revolving around their nuclei.

The calculus of variations was of assistance in the specification of the stationary states and the determination of their orbits. It provided canonical transformations from the variables t, q_k, p_k of the atomic system to new canonical variables t, w_k, J_k which had among other properties those described in equations (8) and the following equations of §3 above. For an atomic system these equations do not involve t and have the somewhat simpler form

$$(15) \quad p_k = S_{q_k}(q, J), \quad w_k = S_{J_k}(q, J),$$

$$(16) \quad H(q, S_q) = E(J).$$

The equations of the trajectories in the new variables are identical with equations (9) which show among other things that the J 's are really constants of integration. In accordance with assumptions which had proved successful in the simplest cases the quantum orbits were determined by values of the J 's of the special form $J_k = n_k h$, where n_k is an integer and h is Planck's

constant, and the frequency $\nu = c/\lambda$ of the emitted radiation when the system fell from a stationary state of total energy E_1 to one of total energy E_2 was to be given by the formula $\nu h = E_1 - E_2$. This last equation, when the extremals for two stationary states of the hydrogen atom were determined and the total energies E_1, E_2 expressed in terms of the integers m, n belonging to the two states, gave the Balmer formula. It was the first great success of the theory.

But the agreement between theory and experimental results was even more striking than this. When hydrogen gas is excited in a constant electric force field each spectrum line is replaced by a symmetrically situated group of neighboring lines, and a similar effect with a different type of group is produced by a constant magnetic field. By modifying the potential function V of the atomic system to take account of these electric or magnetic forces, formulas for the frequencies ν , agreeing well with those observed on the spectrum, were found by the process described in the preceding paragraphs. It was still more interesting perhaps that the fine structure of the lines of the spectrum, detectable by more delicate measurement, was accounted for by replacing Newtonian mechanics by the mechanics of the relativity theory, as described in §3 above. These were indeed impressive agreements between observation and theory.

For more complicated atoms the mathematics became of course much more difficult and was theoretically somewhat disturbing. Besides the properties listed above the canonical transformation (15) was further required to satisfy conditions of the form

$$(17) \quad J_k = \int \dot{p}_k dq_k = \int S_{q_k} dq_k,$$

where the integrals were taken throughout a period of the motion of the atomic system. These could be satisfied by transforming the J 's to other similar constants when the integral $S(q, J)$ of equation (16) was a sum of functions each containing only one of the variables q_i , and progress had been made in the study of more general cases. But I think it must have been a relief to many when at this stage the quantum theory was given an entirely new form by the advent of the methods of Heisenberg and Schrödinger.

Before we take up the consideration of these later developments we should note that for canonical variables w_k, J_k a consequence of the equations (17) was that each coordinate q of a particle in the periodic atomic system was expressible as a function

$$q = q(w, J) = q(w, nh),$$

which was periodic with period unity in each variable w_k . On account of this periodicity q could be expanded, for a quantum orbit having $J_k = n_k h$, in a Fourier series of the form

$$q = \sum C(n, \tau) e^{2\pi i(\tau_1 w_1 + \dots + \tau_r w_r)},$$

where the sum is taken for all sets $\tau = (\tau_1, \dots, \tau_r)$ of positive and negative integers. This is an ordinary multiple Fourier series with the sines and cosines replaced by their values in terms of exponentials, and it is readily provable that $c(n, -\tau)$ and $c(n, \tau)$ are conjugate imaginaries. The corresponding equation for q in terms of the time is found by substituting the values (9) for the w 's and has the form

$$(18) \quad q = \sum Q(n, \tau) e^{2\pi i N(n, \tau) t},$$

where

$$N(n, \tau) = \tau_k E_{J_k}(nh).$$

For a fall from a state defined by integers n to one defined by m 's the frequency of the emitted radiation is

$$\begin{aligned} \nu(m, n) &= [E(nh) - E(mh)]/h \\ &= (n_k - m_k) \int_0^1 E_{J_k}[m + \theta(n - m)] d\theta, \end{aligned}$$

by Taylor's formula with integral form of remainder term. This is a mean value between the frequencies $N(n, \tau)$ and $N(m, \tau)$ of the two terms having $\tau_k = n_k - m_k$ in the series (18) for the two orbits. The correspondence principle of Bohr resulted in the assumption that the radiation emitted would be like that of a set of simple oscillators whose q -coordinate has the mathematical representation

$$(19) \quad q(m, n) e^{2\pi i \nu(m, n) t},$$

where $q(m, n)$ is the corresponding mean value

$$q(m, n) = \int_0^1 Q[m + \theta(n - m), \tau] d\theta, (\tau_k = n_k - m_k),$$

between the coefficients $Q(n, \tau)$ and $Q(m, \tau)$. By means of this further rather doubtfully justifiable hypothetical specification of the character of the radiation, theoretical polarizations and intensities of spectrum lines were calculated whose agreements with observed data were close enough to be surprising. This situation is, however, a physical one of rather refined delicacy for a mathematician to discuss, and I will emphasize further therefore only the important result that corresponding to every coordinate q of the atomic dynamical system the theory gave a matrix of elements of the form (19) for which elements situated symmetrically with respect to the principal diagonal are conjugate imaginaries. Such a matrix is called by the mathematicians a Hermitian matrix.

6. *The Theory of Schrödinger.* The quantum theory as proposed by Schrödinger is based upon a very interesting extension of the analogy of Bernoulli between optical paths and mechanical trajectories described near the end of §4 above. His argument can be phrased as follows. It has been seen that the extremals for the integral (13) in the optical field are the same as the mechanical trajectories defined by the action integral (12) when the light-velocity u is given by equation (14). In an optical field, however, the theory of geometrical optics is not sufficient to explain minute optical phenomena, such as interference and diffraction. To have a satisfactory theory of these latter phenomena, it has been found necessary to associate with every field of light rays solutions of the wave equation

$$(20) \quad w_{xx} + w_{yy} + w_{zz} = w_{tt}/u^2.$$

It seems reasonable, therefore, in the effort to explain minute atomic mechanical phenomena, to associate with mechanical trajectories solutions of the wave equation with the value of u defined by (14).

The constant C now assumes a more important role. In order to determine its value, Schrödinger made two assumptions. The first was that the frequency ν of the wave to be associated with the moving particle should be related to the total energy E of the particle by the equation $E = h\nu$, as de Broglie had earlier

suggested; the second was that the group velocity of the one-parameter family of waves associated with the particle for different values of E should be equal for each E to the velocity v of the particle itself. It would take too long to explain here the mathematically somewhat vaguely formulated definition of group velocity. For mathematicians who may be interested the best discussion which I know of is that of Fraenkel.† The result of the two assumptions is that $C = E$, so that the final value for u^2 in the wave equation is

$$u^2 = E^2/[2m(E - V)].$$

If we make the substitution

$$(21) \quad w = \psi(x, y, z)e^{-2\pi i(E/h)t}$$

in equation (20), in order to determine the wave functions which have frequency $\nu = E/h$, we find the new equation for ψ

$$(22) \quad \psi_{xx} + \psi_{yy} + \psi_{zz} + \psi 8m\pi^2(E - V)/h^2 = 0.$$

This is the fundamental equation upon which Schrödinger based his theory of the mechanics of an atom.

With each state of motion of a dynamical system having a potential function $V(x, y, z)$ there is to be associated a real or imaginary solution $\psi(x, y, z)$ of the wave equation (22) which has continuous first and second partial derivatives everywhere in xyz -space, which vanishes at infinity, and which furthermore has first partial derivatives ψ_q with respect to the coordinates $q = x, y, z$ such that the products $r^2\psi_q$ are bounded, where $r^2 = x^2 + y^2 + z^2$, as is customarily assumed in the potential theory. The determination of solutions ψ with these properties is what the mathematicians call a boundary value problem, the parameter in equation (22) which they would usually denote by λ being E . For the mechanical system of the hydrogen atom, for example, it turns out that the only negative values of E for which the problem has solutions form a denumerable infinity E_1, E_2, \dots , and these are precisely the energies of the Bohr stationary states. Thus the Schrödinger theory singles out automatically the possible energy states for the electron, which in the Bohr theory had to be specified by the special assumptions $J_i = n_i h$.

† *Einführung in die Wellenmechanik*, p. 24.

To every value E_n there corresponds a finite number of solutions

$$w_n = \psi_n e^{-2\pi i(E_n/h)t}$$

of the wave equation (22), and these solutions can be normed and orthogonalized so that

$$\int \psi_n \psi_n^* dx dy dz = 1, \quad \int \psi_m \psi_n^* dx dy dz = 0, \quad (m \neq n),$$

where ψ_n^* is the conjugate imaginary of ψ_n and the integrals here and in the following equations are taken over the whole of space. In accord with analogies in electro-magnetic theory it was further assumed that the radiation emitted when the system fell from a state of higher energy E_n to one of lower energy E_m would be that of a simple oscillator with coordinates $q = x, y, z$ of the form

$$(23) \quad \int q w_m w_n^* dx dy dz,$$

where w_n^* is the conjugate imaginary of w_n . The expression (23) has the value

$$(24) \quad q(m, n) e^{2\pi i(E_n - E_m)t/h},$$

with

$$q(m, n) = \int q \psi_m \psi_n^* dx dy dz.$$

The assumptions of the theory therefore lead automatically to a matrix of elements (24) of the same form as that which arose from the Bohr theory.

Mathematically, the crucial part of the Schrödinger theory is the determination of the characteristic numbers E_n of the boundary value problem associated with the wave equation (22) and their corresponding characteristic functions ψ_n . The methods used are those of separation of variables. Theoretically there is an interesting contact at this point also with the calculus of variations. For the boundary value problem is what the mathematicians call self-adjoint, and every such problem is closely associated with a problem of the calculus of variations. For the simple 3-dimensional case of the Schrödinger theory dis-

cussed above this latter problem is that of finding in the class of functions ψ satisfying the condition

$$\int \psi^2 dx dy dz = 1$$

one which minimizes the integral

$$I = \int [\psi_x^2 + \psi_y^2 + \psi_z^2 + \psi^2 8m\pi^2 V/h^2] dx dy dz.$$

This is an isoperimetric problem of familiar character whose extremals have the differential equation (22). The relationship between boundary problems and such problems of the calculus of variations has had an elaborate development. It was studied for the special case of minimal surfaces by Schwarz, and for more general cases by Hilbert and many others.

The calculations of the quantum theory of Schrödinger applied to the hydrogen atom and other cases gave theoretical results which for the most part agreed as well or better with observed data than those of the earlier theory of Bohr. Sommerfeld's treatment of the fine structure gave, however, a less satisfactory formula, according to his own statement.† The method used in all cases was to substitute the potential function V of the system in the wave equation, then to determine the energy levels and characteristic functions for the boundary value problem, and finally to analyze the radiation emitted by the oscillators (24). What could not be accomplished directly could sometimes be attained by the use of perturbation methods similar to those used in other mechanical theories.

In closing this section I wish to emphasize again the fact that in the Schrödinger theory, as in the preceding theory of Bohr, the final result was a matrix of elements of the form (24) which characterized the emitted radiation. Furthermore, for solutions of the form (21), the equation (20) or (22) is equivalent to the new equation

$$(25) \quad ihw_i/(2\pi) = -h^2(w_{xx} + w_{yy} + w_{zz})/(8\pi^2m) + Vw,$$

as one readily sees by eliminating E by means of the equation

$$w_i = -2\pi iEu/h.$$

† *Atombau und Spektrallinien*, Wellenmechanischer Ergänzungsband, pp. 132 ff.

If each momentum p in the Hamiltonian function

$$(26) \quad H(q, p) = (p_x^2 + p_y^2 + p_z^2)/(2m) + V$$

is replaced by the operator $p = -(ih)/(2\pi) \cdot \partial/\partial q$, where q is the corresponding coordinate, and if the resulting operator is designated by $H(q, -(ih)/(2\pi) \cdot \partial/\partial q)$, then equation (26) is expressible in the form

$$(27) \quad \frac{ih}{2\pi} \frac{\partial w}{\partial t} = H\left(q, -\frac{ih}{2\pi} \frac{\partial}{\partial q}\right) w.$$

These remarks have significance for the description in the next section of the more recent forms of the quantum theory.

7. *Present Forms of the Quantum Theory.* The recent forms which the quantum theory has taken are not easy to describe in brief fashion because they involve matrix algebra and mathematical theories of probability as well as much of the mathematics which has been discussed in preceding pages. I should like to try nevertheless to give my fellow mathematicians some idea of the contacts which still remain with the calculus of variations, in spite of the fact that my experience with the newer theories is even more restricted than with the older ones.

The first step toward the newer point of view seems to have been taken by Heisenberg. Because of the artificial character of the mechanical mechanisms involved in the Bohr theory, or perhaps one should say the non-mechanical mechanisms, he proposed to cast those aspects of the theory aside completely and to begin with the matrices (24) or (19) themselves which appear as the prime results of the theory of Schrödinger as well as that of Bohr. To the physicist this seemed a very bold proposal, perhaps because the notions of matrices and the matrix calculus are not a part of his usual mathematical kit. But to the mathematician Heisenberg's proposal would, I think, seem a natural one in view of the results of the preceding theories. The lines of the spectrum have a double or matrix array of frequencies which can be regarded as caused by a similar array of oscillators. What then are the properties of the matrices of oscillators which characterize such spectra? It is this question which would have stumped the mathematician, and it is to Heisenberg's unusual combination of physical experience with

mathematical instincts that an answer has been found which seems in many respects to be successful.

Heisenberg's mathematical postulates are in part simple enough when once they have been formulated. He proposed to replace each coordinate q and momentum p of a classical mechanical system by a Hermitian matrix of the kind described in equation (24), and to replace the energy function $H(q, p)$, expressed as a power series in the p 's and q 's, by the matrix defined by the same series when the matrices for p and q are substituted. If a suitable interpretation could be found in matrix theory for a Poisson bracket (6) then the differential equations of the system could be expressed in the form

$$(28) \quad \frac{dq_k}{dt} = (q_k, H), \quad \frac{dp_k}{dt} = (p_k, H),$$

analogous to the equations (5).

Heisenberg determined the interpretation of the Poisson bracket for his matrix theory by a limiting process which had been used in the so-called correspondence principle of Bohr and which I somewhat arbitrarily avoided in the preceding pages. The method of Dirac is more accessible to the mathematician. Dirac listed the rules of computation for Poisson brackets as defined for the classical theory and found an equation pointing to the fact that he could make them valid for matrices also by assigning to the bracket (f, g) for two matrices f and g the value $c(fg - gh)$, where c is a constant. The difference $fg - gh$ is not in general zero for matrices. When it is so the matrices are said to be *permutable*. It turns out that the value of c must be $-2\pi i/\hbar$ in order to make the results of the theory agree with experiment.

In classical mechanics the Poisson brackets for the coordinates and momenta have the simple values

$$(q_k, q_l) = 0, \quad (p_k, p_l) = 0, \quad (q_k, p_l) = \delta_{kl},$$

where $\delta_{kk} = 1$ and $\delta_{kl} = 0$ for $k \neq l$. Hence by analogy, and on account of the definition of the Poisson bracket, the matrices q, p of the quantum theory should satisfy the relations

$$(29) \quad \begin{aligned} q_k q_l - q_l q_k &= 0, & p_k p_l - p_l p_k &= 0, \\ q_k p_l - p_l q_k &= i\hbar \delta_{kl} / (2\pi), \\ i\hbar \dot{q}_k / (2\pi) &= q_k H - H_{q_k}, & i\hbar \dot{p}_k / (2\pi) &= p_k H - H_{p_k}, \end{aligned}$$

where the dots indicate time derivatives and the last two equations are the equations (28) with the values of the brackets substituted. For a particular dynamical system the energy matrix $H(q, p)$ is the classical energy function of the system with matrices q, p in place of the classical coordinates and momenta. These are the *fundamental equations* of the quantum theory in matrix form.

A *canonical transformation* is defined by a matrix S whose inverse S^{-1} is the matrix formed from S by replacing all its elements by their conjugate imaginaries and then turning the matrix through 180° on its principal diagonal as axis. In mathematical language S^{-1} is the conjugate transpose of S . If such a matrix is at hand, then the transformation $P = SpS^{-1}$, $Q = SqS^{-1}$ takes every set of matrices p, q satisfying equations (29) into a new set P, Q satisfying equations of the same form. Such transformations were used by Heisenberg to find solutions of equations (29). When, for the case of the hydrogen atom for example, a canonical transformation had been found which took the energy matrix H into a diagonal matrix E , the elements $E(n, n) = E_n$ in the principal diagonal were the energy levels of Bohr, and the frequencies of the elements of the matrices q, p turned out to be the radiation frequencies

$$\nu(m, n) = (E_n - E_m)/h.$$

This matrix calculus of Heisenberg has been given a different form and has been supplemented by a probability interpretation which is based upon analogs of well known algebraic properties of Hermitian matrices. A vector $\psi = (\psi_1, \dots, \psi_n)$ of complex numbers in n dimensions is said to be *normed* if $\psi_k \psi_k^* = 1$, where ψ_k^* is the conjugate imaginary of ψ_k . It is *orthogonal* to a second such vector ψ' if $\psi_k^* \psi'_k = 0$. An n -dimensional Hermitian matrix m transforms every vector ψ into another $m\psi$ of the same type, and has a set of n normed and orthogonal vectors $\Psi(m')$, each possessing a real number m' such that $m\Psi(m') = m'\Psi(m')$. The vectors $\Psi(m')$ are called the *characteristic vectors* of the matrix, and the numbers m' are its *characteristic numbers*.† An arbitrarily selected vector ψ is expressible linearly in terms of the characteristic vectors in the form

$$\psi = \sum \Psi(m')c(m'),$$

† They are analogous to the “*eigenfunctions*” and “*eigenvalues*” of Dirac.

where the sum is taken for all the characteristic vectors. If ψ is normed, then it is readily provable that

$$\psi^*\psi = \sum c(m')c^*(m') = \sum |c(m')|^2 = 1,$$

an equation which suggests the possibility of interpreting the real numbers $|c(m')|^2$ as probabilities. These results are stated here for n -dimensional vectors and matrices because it is that case which is familiar to most of us. In the quantum theory the range of the subscript k and of the characteristic values m' may be a continuum, or a combination of a discrete and a continuous range, or of still more general character. For these other ranges the matrix theorems are usually not proved but are assumed to hold by analogy.

In the presentation of the theory as given by Dirac the matrices of Heisenberg are replaced by much more general linear operators which operate on symbols ψ representing the possible states of motion of the dynamical system under consideration. Each state is represented not only by its symbol ψ but also by $c\psi$, where c is an arbitrary complex number, and by a symbol ψ^* to which ψ is related as a vector to its conjugate imaginary. All linear combinations of symbols for states, of the form $c_1\psi_1 + c_2\psi_2$, also represent states. For any two states ψ_1 and ψ_2 a scalar product $\psi_1^*\psi_2$ is postulated whose value is a complex number such that $\psi_2^*\psi_1 = (\psi_1^*\psi_2)^*$. If this number is zero the two states are said to be *orthogonal*. A symbol for a state may be *normed* so that $\psi^*\psi = 1$ by multiplying it by a suitable complex constant. One easily sees that the absolute value of this constant is uniquely determined but its amplitude or phase is arbitrary.

At each time t there is associated with each coordinate of the dynamical system a linear operator q which transforms each state ψ into another and which has properties like those of a Hermitian matrix. The operator q has a system of normed and orthogonalized characteristic states $\Psi(q')$ with characteristic numbers q' such that every state ψ is expressible in the form

$$\psi = \sum_{q'} \Psi(q')c(q'),$$

by analogy with the matrix theory. If the symbol ψ is normed then one finds readily that

$$(30) \quad \sum_{q'} |c(q')|^2 = 1.$$

The fundamental idea of the probability form of the quantum theory is now that when an observation is made at the time t to determine the value of a coordinate of the mechanical system in the state ψ the result is not necessarily unique. The values which may be found for the coordinate are the characteristic numbers q' of the operator q , and the probability that a particular value q' will be found at the time t for the state ψ , is the term $|c(q')|^2$ of the sum (30). The classical case would be the one for which q has a single characteristic number q' and behaves like a Hermitian matrix with all elements in the principal diagonal equal to q' and all other elements equal to zero. Such an operator has but one characteristic number q' , and all states ψ are characteristic states, so that a measurement of q always gives the value q' as a certainty.

The momenta of the mechanical system are similarly represented by operators p , and the energy by an operator which is a function $H(q, p)$ of the operators q and p . The equations which characterize these operators for a particular mechanical system are the equations (29) of Heisenberg for reasons similar to those described above in deducing them. It is provable that if two operators p and q satisfy the third of equations (29), then every real number q' is a characteristic number for q , and similarly for p . Moreover, the set of permutable operators q_1, \dots, q_r has normed and orthogonal characteristic states $\Psi(q') = \Psi(q'_1, \dots, q'_r)$ in common, one and only one of which corresponds to every set of real numbers (q'_1, \dots, q'_r) , and in terms of which every other state is expressible in the form

$$\psi = \int \Psi(q') \psi(q', t) dq'.$$

The integral is taken over the whole of the r -dimensional q' -space and $\psi(q', t)$ is a function of q'_1, \dots, q'_r, t , which is said to *represent* ψ . The state ψ does not change with the time, but its representative must do so, since the operators q and their characteristic states Ψ are changing. When ψ is normed it follows that

$$1 = \int |\psi(q', t)|^2 dq' = \int \psi^*(q', t) \psi(q', t) dq'.$$

If an observation to determine all of the coordinates q is made at the time t upon a system in the state ψ the possible values for the coordinates are the characteristic sets (q_1', \dots, q_r') , and the probability that a set will be found in the volume element dq' is $|\psi(q', t)|^2 dq'$, where $q' = (q_1', \dots, q_r')$ is a point in dq' .

It is important and interesting to note that the representation $\psi(q', t)$ of the states which are possible for a given dynamical system are precisely the solutions of Schrödinger's wave equation in the form (27). For it is provable from the equations (29) that the results of operating with q_k , p_k , or H upon a state with representative $\psi(q', t)$ are, respectively, the states represented by

$$q'_k \psi(q', t), -\frac{i\hbar}{2\pi} \frac{\partial}{\partial q'_k} \psi(q', t), \frac{i\hbar}{2\pi} \frac{\partial}{\partial t} \psi(q', t),$$

provided that the phases of the normed characteristic states $\Psi(q')$ are suitably chosen. But since H is also expressible in terms of the q 's and p 's it follows then readily that

$$\frac{i\hbar}{2\pi} \frac{\partial}{\partial t} \psi(q', t) = H\left(q', -\frac{i\hbar}{2\pi} \frac{\partial}{\partial q'}\right) \psi(q', t),$$

where the operator on the right is formed by replacing in $H(q, p)$ each of q and p by its value as an operator. Thus by reasoning quite different from that of Schrödinger we are led again to the wave equation (27).

A mathematician should perhaps not venture to discuss the physical reasons for things, but I think I owe to my mathematical hearers an effort to explain why the quantum theory has adopted a probability aspect. There are a number of reasons, but I shall speak of only one. The Schrödinger theory provided functions $w(q, t)$ which belonged to the stationary states of an atom but it did not attempt to describe the mechanical character of the atomic system itself except in its identification of the group velocity of a one-parameter family of waves with what would ordinarily be the velocity of a mechanical particle. Such a family or packet of waves may have a resultant amplitude different from zero only in a limited portion of space whose extent may increase or contract as it moves. The coordinates and momenta of such a packet taking the place of the mechanical particle would be to some extent uncertain. In the general

theory this uncertainty corresponds to the fact that the momenta q are never permutable with the coordinate operators p . Operators have characteristic states in common in general only when they are permutable, and the only states for which the probabilities described above for a particular operator q or p are unity, and the results of observations therefore certain, are the characteristic states of the operator. It follows, therefore, that in general there is no state for which the coordinates and momenta are determinable simultaneously with certainty. The mean square deviations Δq , Δp of a corresponding coordinate and momentum from their mean values can be calculated at least in simpler cases, and it has been shown that $\Delta q \Delta p \geq h/(2\pi)$, which is the inequality expressing Heisenberg's famous uncertainty principle.

This is as far as we shall have time to go in the discussion of the various modifications of quantum mechanics. In the foundations of the operational or matrix form of the theory the influence of the calculus of variations remains only in the preservation of the notions of canonical equations and transformations, and in the boundary value problems connected with the wave equation of Schrödinger. But other equally interesting mathematical domains have been freely called upon to contribute their quotas. An abstract operational calculus, a kind of general analysis of operations, with an associated theory of probability, seem to be requisites, and the details of the applications of these theoretical mechanisms to the special cases so far studied have also great mathematical interest. This can be said in spite of the fact that at the present time the pure mathematician, in his effort to understand the theory, must plough steadily forward oblivious to a multitude of detailed questions of mathematical accuracy. In view of the success which the theory seems to have had it is not too much to hope that in the end all such questions may be satisfactorily answered or avoided, and it is even possible that the pure mathematician may contribute a valuable mite to this happy result.

THE UNIVERSITY OF CHICAGO