AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 30, NUMBER 4 APRIL, 1962

Single Valuedness of Wave Functions

E. MERZBACHER
Oak Ridge National Laboratory, Oak Ridge, Tennessee, and University of North Carolina, Chapel Hill, North Carolina†*

(Received August 21, 1961)

The requirement that the quantal wave function be single-valued is examined in the light of two recent developments: The effect of a magnetic vector potential on a particle moving in a multiply connected field-free region (Aharonov-Bohm effect), and the flux quantization inside a superconducting ring. Both of these effects are new applications of conventional elementary quantum mechanics. The single-valuedness problem is considered, with its historical background, for particles with no spin, particles with spin ½, and rigid bodies. It is shown that the single-valuedness condition, if properly adapted to the particular physical model, is deeply rooted in the foundations of quantum mechanics. However, wave functions which superficially appear to be double-valued are useful in an helicity representation of spinors and in the construction of nuclear wave functions.

1. INTRODUCTION

IN the discussion of the eigenvalue problem for the z component of the orbital angular momentum of a particle, $\mathbf{L}=\mathbf{r}\times\mathbf{p}$, one is in quantum mechanics confronted with the differential equation

$$L_z\psi\equiv\frac{\hbar}{i}\frac{\partial\psi}{\partial\varphi}=m\hbar\psi \tag{1}$$

which has the solution

$$\psi=Ce^{im\varphi}. \tag{2}$$

It is then argued that ψ must be a single-valued function of position, hence that it must have the same value for $\varphi=2\pi$ as for $\varphi=0$. It follows that m must be an integer: $m=0, \pm1, \pm2, \cdots$.

Most students of quantum mechanics have on occasion felt uneasy about this argument because it produces one of the most fundamental results of quantum physics, the quantization of angular momentum, by recourse to a requirement which does not appear to be at all obvious. To be sure, the demand that ψ shall be single-valued is merely a boundary condition for Eq. (1), and we are used to the appearance of discrete eigenstates and eigenvalues as a consequence of the enforcement of boundary conditions, but this particular condition seems somewhat less natural than, for instance, the requirement that the wave function must be finite at large distances.

The doubts concerning the single-valuedness condition have been expressed most concisely by Blatt and Weisskopf in a famous footnote of their book[1]: "The . . . argument is fallacious since multiple-valued wave functions cannot be excluded *a priori*. Only physically measurable

* ORINS Summer Research Participant.

† Partly supported by the U. S. Atomic Energy Commission.

[1] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952), see footnotes on pp. 783 and 787.

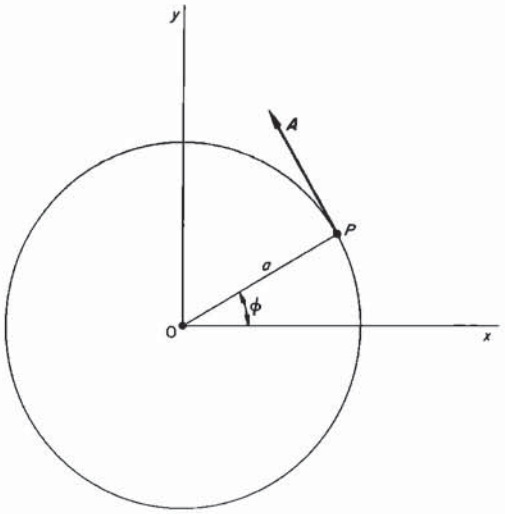

FIG. 1. Particle moving in circular track of radius a around a magnetic flux Φ. The flux is concentrated in a line through 0 perpendicular to the figure. The vector potential (4) is tangent to the circle.

quantities, such as probability densities and expectation values of operators, must be single-valued. Double-valued wave functions are used in the theory of particles with intrinsic spin."

It is the aim of this paper to clarify the problem of the single valuedness of ψ. Recently, the question of whether the wave function of an electron must be a single-valued function of position has been raised in connection with an interesting quantum effect predicted by Aharonov and Bohm.[2] The point made by these authors can be understood in an elementary way by considering an infinitely long, thin solenoid perpendicular to the plane of the paper (Fig. 1). Under idealized conditions the magnetic field is entirely confined to the interior of the solenoid and vanishes outside. However, the vector potential $\mathbf{A}$, related to the magnetic field by $\mathbf{B}=\nabla\times\mathbf{A}$, cannot vanish everywhere outside the coil because the line integral of $\mathbf{A}$ along a curve encircling the coil is equal to the flux Φ in the solenoid:

$$\oint \mathbf{A}\cdot\mathbf{dr}=\int(\nabla\times\mathbf{A})\cdot\mathbf{dS}=\int\mathbf{B}\cdot\mathbf{dS}=\Phi. \quad (3)$$

[2] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959); see also W. Ehrenberg and R. E. Siday, Proc. Phys. Soc. (London) **B62**, 8 (1949); H. Wegener, Z. Physik **159**, 243 (1960); Y. Aharonov and D. Bohm, Phys. Rev. **123**, 1511 (1961).

A possible representation of $\mathbf{A}$ in cylindrical coordinates is given by

$$A_\varphi=\Phi g(\varphi)/\rho, \quad A_z=A_\rho=0, \quad (4)$$

$\rho=(x^2+y^2)^{\frac{1}{2}}$ being the radial coordinate in the xy plane, and φ the azimuthal angle, $\varphi=\arctan y/x$. $g(\varphi)$ is an arbitrary function of φ but normalized so that

$$\int_0^{2\pi} g(\varphi)d\varphi=1. \quad (5)$$

It is easy to verify that $\mathbf{B}=\nabla\times\mathbf{A}=0$ except at the origin.

If we place a particle in a narrow circular tube of radius $\rho=a$ around the origin, the particle (of mass μ) is constrained to move on a circle, subject to the Hamiltonian

$$H=\tfrac{1}{2}\mu a^2\dot{\varphi}^2.$$

The generalized momentum corresponding to the coordinate $a\varphi$ is the tangential momentum component

$$p_\varphi=\mu a\dot{\varphi}+QA_\varphi, \quad (6)$$

if Q is the charge of the particle. Hence,

$$H=\tfrac{1}{2}\mu^{-1}(p_\varphi-QA_\varphi)^2. \quad (7)$$

In wave mechanics we use the representation

$$p_\varphi\rightarrow\frac{\hbar}{ia}\frac{\partial}{\partial\varphi}$$

and obtain the Schrödinger equation

$$-\frac{\hbar^2}{2\mu a^2}\left(\frac{\partial}{\partial\varphi}-\frac{iQa}{\hbar}A_\varphi\right)^2\psi=E\psi. \quad (8)$$

For the field (4) this becomes

$$-\frac{\hbar^2}{2\mu a^2}\left[\frac{\partial}{\partial\varphi}-\frac{iQ}{\hbar}\Phi g(\varphi)\right]^2\psi=E\psi. \quad (9)$$

This differential equation is solved by the function

$$\psi(\varphi)=C\exp\left[im'\varphi+i\frac{Q}{\hbar}\Phi\int_0^\varphi g(\varphi')d\varphi'\right] \quad (10)$$

with the energy eigenvalue

$$E=\hbar^2m'^2/(2\mu a^2). \quad (11)$$

Invoking the *single-valuedness condition*

$$\psi(2\pi)=\psi(0), \tag{12}$$

we learn from (10) that

$$2\pi m'+\frac{Q}{\hbar}\Phi\int_0^{2\pi} g(\varphi)d\varphi=2\pi m, \tag{13}$$

where m must be an integer. Hence, by (5),

$$E=\frac{\hbar^2}{2\mu a^2}\left(m-\frac{Q\Phi}{2\pi\hbar}\right)^2, \tag{14}$$

and the wave function can be written as

$$\psi(\varphi)=Ce^{im\varphi}\exp\left\{i\frac{Q\Phi}{\hbar}\left[\int_0^{\varphi} g(\varphi')d\varphi'-\frac{\varphi}{2\pi}\right]\right\}.$$

If $g(\varphi)$ is discontinuous, but finite, the matching conditions

$$\psi \quad \text{continuous}$$

and

$$\left(\nabla-\frac{iQ}{\hbar}\mathbf{A}\right) \quad \text{continuous}$$

must be used.

It is seen that, independent of the choice of the arbitrary gauge function $g(\varphi)$, the energy eigenvalues E given by (14) depend on the flux of the magnetic field.

Indeed, as has been pointed out by Peshkin *et al.*,[3] if the flux is turned on slowly, Faraday's law of induction implies that the rate of change of the energy of the particle is

$$dE/d\Phi=-Q\omega/(2\pi), \tag{15}$$

where ω is the classical circular frequency, $\omega=[2E/(\mu a^2)]^{\frac{1}{2}}$. Hence,

$$\frac{dE}{d\Phi}=-\frac{Q}{\pi}\left(\frac{E}{2\mu a^2}\right)^{\frac{1}{2}},$$

in agreement with the result obtained by differentiating Eq. (14).

Although this argument makes it clear "where the energy comes from," the Zeeman effect of Eq. (14) has seemed paradoxical to many since an electron, brought in from the outside after the field has been turned on, can apparently be used to measure the flux of a field which is zero in all parts of space accessible to the electron. In classical physics no physical feature of the electron motion could be influenced by a static magnetic field which the particle never "sees"; e.g., the Lorentz force on the electron would be zero. However, Eq. (14) provides the example that, although $\mathbf{B}=0$ in the circular track, the nonvanishing of the vector potential $\mathbf{A}$ gives rise to a physical effect. We shall refer to this nonclassical effect briefly as the *Aharonov-Bohm effect*. Such an effect has been observed in electron diffraction experiments.[4] Fringes were shown to be affected by the passage of electrons around an excluded region containing the magnetic flux.

FIG. 2. The superconducting ring P is a multiply connected region. The currents flow only at the surfaces S_1 and S_2. (From Byers and Yang, reference 6.)

It is to be noted that the energy spectrum (14) would coincide with that obtained for $\Phi=0$ if

$$\Phi=\text{integer}\times h/Q. \tag{16}$$

This is a *flux quantization* condition.

In two very remarkable experiments it was recently found that the magnetic flux in the interior of a hollow superconducting ring is in fact quantized in units of $h/2e$, where e is the electronic charge.[5] A superconductor is characterized by the condition that within it $\mathbf{B}=0$. If there is no magnetic field, no body current can flow inside the superconductor in the steady state, and all currents are confined to the surface. A superconducting ring which encloses a hollow region containing a magnetic flux Φ is thus a multiply connected region in which $\mathbf{B}=0$ but $\mathbf{A}\neq 0$. (Fig. 2). The same considerations can be applied as to the Aharonov-Bohm effect, but with the additional restriction that the body current in the ring must be independent of the enclosed flux—and, in fact, equal to zero—im-

[3] M. Peshkin, I. Talmi, and L. J. Tassie, Ann. Phys. 12, 426 (1961).

[4] R. G. Chambers, Phys. Rev. Letters 5, 3 (1960); H. Boersch, H. Hamisch, D. Wohlleben, and K. Grohmann, Z. Physik 159, 397 (1960). For an interpretation of the effect see also W. H. Furry and N. F. Ramsey, Phys. Rev. 118, 623 (1960), and F. G. Werner and D. R. Brill, Phys. Rev. Letters 4, 344 (1960).

[5] B. S. Deaver, Jr., and W. F. Fairbank, Phys. Rev. Letters 7, 43 (1961); R. Doll and M. Näbauer, Phys. Rev. Letters 7, 51 (1961).

plying that a charged particle moving in the superconductor does not "see" the effect of the flux.[6] This can be true only if the flux is quantized according to Eq. (16). Q is the effective charge of the particles whose motion is responsible for the properties of the superconductor. The experiments yield $Q=2e$, and this observation is very important for an understanding of the superconducting state.[6]

The effect predicted by Aharonov and Bohm and the interpretation of the recent flux quantization experiments depend essentially on the assumption that the wave function be single-valued. It may thus be of interest to examine the grounds on which this requirement rests.

The founders of quantum mechanics were aware of the peculiar role which the condition that ψ be single-valued plays. Eddington seems to have given some thought to the possibility of using multivalued wave functions,[7] and Schrödinger showed that, under very general assumptions, one needs to consider only the alternatives that ψ is either a single-valued function of position or a double-valued function, the two values at a given point differing only by a sign.[8] Schrödinger assumed that the probability density $|\psi(P)|^2$ should be single-valued, that together with any state ψ the "time-reversed state" ψ^* (complex conjugate of ψ) should also be a possible state, and that certain continuity requirements had to be fulfilled.

The most important contribution to our subject was made by Pauli: He was dissatisfied with his formulation of the problem in the 1932 edition of Vol. 24.1 of the *Handbuch der Physik*,[9] examined the question in detail in a beautiful paper written in 1939,[10] and finally was led to rewrite the relevant section of his article for the new 1958 edition of what is now known as the *Encyclopedia of Physics*.[11] The discussion of this paper leans heavily on Pauli's 1939 paper in which he examined the double-valued eigensolutions of the orbital angular momentum eigenvalue problem.[10]

2. PARTICLE WITHOUT SPIN

Let us, for the time being, restrict ourselves to the nonrelativistic quantum mechanics of a particle without spin. The wave function ψ is then a function of the position coordinates of the particle, and the question of its single or multivaluedness can be simply phrased: Does or does not ψ resume its value $\psi(P)$ if, starting at some point P, we move along a closed path in space returning to the initial point (Fig. 3)?

The orbital angular momentum operator $\mathbf{L}=\mathbf{r}\times\mathbf{p}$ satisfies the commutation relations

$$[L_x,L_y]=i\hbar L_z \quad \textit{et cycl.}$$

and

$$[\mathbf{L},\mathbf{L}^2]=0.$$

Consider again the eigenfunctions of orbital angular momentum about an arbitrarily chosen origin. Introducing spherical polar coordinates, the simultaneous eigenfunctions of L_z and $\mathbf{L}^2$ obey the differential equations

$$L_z Y_l^m \equiv \frac{\hbar}{i}\frac{\partial}{\partial\varphi} Y_l^m = m\hbar Y_l^m, \tag{17}$$

$$\mathbf{L}^2 Y_l^m \equiv -\hbar^2\left[\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)\right] Y_l^m = \hbar^2 l(l+1)\, Y_l^m. \tag{18}$$

A familiar "ladder" argument made in quantum mechanics[12] and based on the commutation relations alone shows that the eigenvalues are $m=-l, -l+1, \cdots, l$, and l can only be an integer or a half-integral number. For integral l the regular solutions of (17) and (18) are the usual spherical harmonics $e^{im\varphi}P_l^m(\cos\theta)$, and these are single-valued functions of the particle position. For half-integral l the φ dependence of the solutions causes the wave function to change its sign as we

[6] For a complete discussion see N. Byers and C. N. Yang, Phys. Rev. Letters **7**, 46 (1961); L. Onsager, Phys. Rev. Letters **7**, 50 (1961); J. M. Blatt, Phys. Rev. Letters **7**, 82 (1961); and W. Brenig, Phys. Rev. Letters **7**, 337 (1961). Also the original suggestion by F. London, *Superfluids* (John Wiley & Sons, Inc., New York, 1950), Vol. 1, p. 152.

[7] A. S. Eddington, *Relativity Theory of Protons and Electrons* (Cambridge University Press, New York, 1936).

[8] E. Schrödinger, Ann. Physik (5) **32**, 49 (1938).

[9] W. Pauli, *Handbuch der Physik*, edited by H. Geiger and K. Scheel (Springer-Verlag, Berlin, Germany, 1933), Vol. 24, part 1, p. 126; A. Nordsieck, "Quantum Theory Lecture Notes" (University of Illinois, 1949, unpublished).

[10] W. Pauli, Helv. Phys. Acta **12**, 147 (1939).

[11] W. Pauli, *Encyclopedia of Physics*, edited by S. Flügge (Springer-Verlag, Berlin, Germany, 1958), Vol. 5, part 1, p. 45–46.

[12] P. A. M. Dirac, *Quantum Mechanics* (Oxford University Press, Oxford, 1958), 4th ed., Sec. 36.

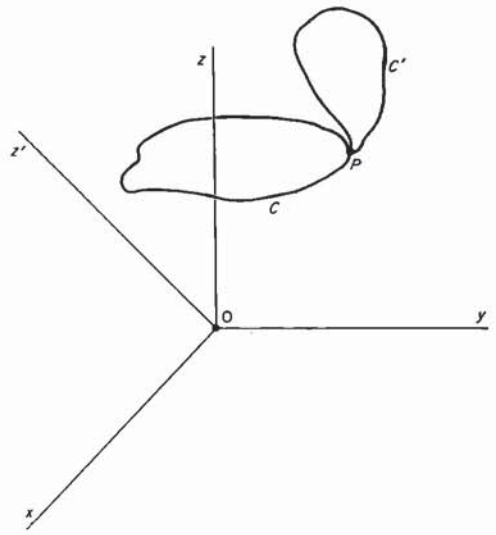

FIG. 3. Closed paths through point P. If the polar axis is arbitrary, the "inside" of a closed loop has no unambiguous meaning. The loop C can be continuously deformed into C', and both can be shrunk to a point.

change φ from 0 to 2π. These wave functions are thus double-valued in agreement with Schrödinger's assumptions. Indeed, if one wants the time-reversed state $e^{-im\varphi}$ to be an eigensolution together with $e^{im\varphi}$, both belonging to the same angular momentum l, then $2m$ must be an integer, since the probability density in a state like $e^{im\varphi}+e^{-im\varphi}=2\cos m\varphi$, obtained by superposition, is single-valued only if $2m$ is an integer. Besides, $-m$ can be reached from m in an integral number of steps of ± 1 only if m is integral or half-integral.

Some typical examples of double-valued solutions of (17) and (18) are

$$Y_{\frac{1}{2}}^{\frac{1}{2}} \propto (\sin\theta)^{\frac{1}{2}} e^{i\varphi/2}, \qquad (l=\tfrac{1}{2},\ m=\tfrac{1}{2}) \quad (19)$$

$$Y_{\frac{3}{2}}^{\frac{1}{2}} \propto (\sin\theta)^{\frac{1}{2}} \cos\theta e^{i\varphi/2}, \quad (l=\tfrac{3}{2},\ m=\tfrac{1}{2}). \quad (20)$$

These wave functions are finite everywhere, but they are not sufficiently well behaved to provide a basis for the angular momentum operators. For example, we have the strange property that the lowering operator $L_-=L_x-iL_y$ applied to $Y_{\frac{1}{2}}^{\frac{1}{2}}$ does *not* yield $Y_{\frac{1}{2}}^{-\frac{1}{2}}$, as the general operator theory would lead us to expect. Instead,

$$L_- Y_{\frac{1}{2}}^{\frac{1}{2}} = e^{-i\varphi}\hbar\left(-\frac{\partial}{\partial\theta}+i\cot\theta\frac{\partial}{\partial\varphi}\right)Y_{\frac{1}{2}}^{\frac{1}{2}} \propto \frac{\cos\theta}{(\sin\theta)^{\frac{1}{2}}}e^{-i\phi/2},$$

giving a function which is a solution of (18) with $l=\frac{1}{2}$, but which is singular at $\theta=0$ and π, although still quadratically integrable. It follows that the scalar product

$$(Y_{\frac{1}{2}}^{-\frac{1}{2}}, L_- Y_{\frac{1}{2}}^{\frac{1}{2}}) = \int Y_{\frac{1}{2}}^{-\frac{1}{2}*} L_- Y_{\frac{1}{2}}^{\frac{1}{2}} d\Omega \neq 0.$$

Hence, if we write the matrix element of, say, L_x as $\langle l'm'|L_x|lm\rangle$, we find that

$$\langle \tfrac{3}{2}, -\tfrac{1}{2}|L_x|\tfrac{1}{2},\tfrac{1}{2}\rangle \neq 0,$$

although the commutation relation

$$\mathbf{L}^2 L_x - L_x \mathbf{L}^2 = 0$$

leads to the vanishing of any matrix element of L_x which does not obey the selection rule $l'=l$. The trouble arises because

$$(\mathbf{L}^2 Y_{\frac{1}{2}}^{-\frac{1}{2}}, L_x Y_{\frac{1}{2}}^{-\frac{1}{2}}) \neq (Y_{\frac{1}{2}}^{-\frac{1}{2}}, \mathbf{L}^2 L_x Y_{\frac{1}{2}}^{-\frac{1}{2}}),$$

i.e., $\mathbf{L}^2$ is not Hermitian with respect to the functions generated from double-valued eigenfunctions by repeated application of L_- (or L_+). Nor does the relation $(L_-)^2 Y_{\frac{1}{2}}^{\frac{1}{2}}=0$ hold. Hence, the "ladder" does not terminate as it does for single-valued Y_l^m. The ladder algorithm, which seems so unexceptionable, since it appears to be based only on the use of the commutation relations for angular momentum, breaks down in these examples, as Pauli recognized, because repeated application of the "raising" or "lowering" operators here ultimately produces singular functions which lie outside the class of functions with respect to which $\mathbf{L}$ is Hermitian. Such an eventuality is usually not contemplated in the formal theory of angular momentum.

Spherical harmonics with half-integral l have further undesirable properties. Since L_x is seen to connect states with different l values, it follows that the rotation operator $e^{-iL_x\alpha/\hbar}$ acting on Y_l (l half-integral) produces functions with different l values; hence the angular momentum of a state can be changed by merely rotating the state in space. In the language of group theory we may say that the half-integral spherical harmonics do not carry a representation of the rotation group. Thus a theorem like

$$Y_l^m(\theta',\varphi') = \sum_{m'=-l}^{l} Y_l^{m'}(\theta,\varphi) D_{m'm}^{(l)}(R), \quad (21)$$

describing the behavior of spherical harmonics under a rotation R, cannot hold for half-integral l, since on circling the pole of $Y_l^m(\theta,\varphi)$ we may not be circling that of $Y_l^m(\theta',\varphi')$, so that one side of Eq. (21) may change sign while the other side does not.[13] (See Fig. 3.)

We conclude that in the nonrelativistic quantum mechanics of a spinless mass point the half-integral values of l must be excluded as eigenvalues of $\mathbf{L}^2$ because the corresponding eigenfunctions do not provide a suitable basis for a representation.

Based on this analysis, Pauli suggested that the single-valuedness requirement for ψ should be replaced by the demand that the repeated application of the operators "belonging" to the particular eigenvalue problem under consideration must produce a representation of the pertinent transformation group, such as the rotation group. Fierz applied Pauli's criterion to the interesting example of a charged particle moving in the field of a (fictitious) magnetic pole.[14]

In retrospect it seems, however, that the simple single-valuedness requirement is altogether reasonable, and, if properly applied, leads in all examples to the correct answer. In the case of a particle with no spin we note that it is a fundamental assumption of quantum mechanics for nonrelativistic particles that a state Ψ can be expanded, in Dirac's bracket notation, as

$$|\Psi\rangle = \int |xyz\rangle dxdydz\langle xyz|\Psi\rangle, \tag{22}$$

since x, y, z are assumed to constitute a complete commuting set of observables. The wave function is identified as the probability amplitude

$$\langle xyz|\Psi\rangle = \psi(x,y,z). \tag{23}$$

A double-valued ψ in the sense considered in this paper could, in the general framework, be accommodated only by introducing an additional physical observable. This new observable would measure whether a certain coordinate point has been reached from some standard position by an even or by an odd number of circulations around the z axis. But there is no unambiguous way to distinguish between those closed loops which have encircled the polar axis once, say, and those which have not encircled it at all (Fig. 3). The first can be deformed continuously into the second, and, generally, any two closed loops can be continuously transformed into each other without leaving the space.

From the point of view of an observer on the polar axis the distinction between the two varieties of closed loops is not dictated by physical conditions but merely by the accidental choice of the polar axis. The axis may go through the "inside" of a particular loop, but this designation is quite arbitrary, and an equally valid polar axis would be on the "outside" of the same loop. The connection of this simple argument with Pauli's criterion can be understood by noting that precisely the same reasoning was used to demonstrate that for half-integral l an equation like (21) could not be valid and that such values of l must therefore be excluded. Hence, the kind of double valuedness we have postulated cannot be given a physical meaning for a spinless point particle moving in ordinary three-dimensional space. From this point of view the usual argument that ψ must be a single-valued function of position is entirely sound. It may even be said that the strange double-valued eigenfunctions of angular momentum have appeared only because we have changed from *Cartesian* coordinates, which are adapted to the homogeneity and isotropy of ordinary space, to *polar* coordinates, which are singular at the coordinate origin and distinguish a particular direction in space. It is then perhaps not surprising that we must subsequently seek to establish criteria for excluding unwanted solutions to our equations which were brought in through the introduction of an awkward (but, of course, eminently useful) coordinate system.

In the light of these observations it might appear as if single valuedness of ψ need *not* be required in the discussion of the Aharonov-Bohm effect. For, in the example of the particle moving in a fixed circular tube (and in all other similar examples of this effect), the particle is excluded from the portion of space where $\mathbf{B}\neq 0$, and the remaining space in which the particle may travel is by no means the simple Euclidean space of our ordinary experience. Rather, it appears to be a space "with a hole in it," the hole being the line

[13] M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957).
[14] M. Fierz, Helv. Phys. Acta **17**, 27 (1943).

of magnetic flux, concentrated on the z axis in our elementary example. In such a space it is entirely possible to distinguish between closed loops which encircle the excluded z axis, and those which do not. No longer can we deform one kind of loop into the other continuously without leaving the allowed space, since the loop must be pulled through the z axis. Therefore, it is now possible that a physical observable may be defined which "counts" the number of revolutions as we circle the axis. The axis appears to be no longer a purely arbitrary mathematical construct but is a part of the physical environment in which the particle moves. Multivalued basis functions appear to be acceptable for a particle moving in such a restricted space which is not simply connected. We claim that the single valuedness of ψ must nevertheless be demanded and that the paradox is resolved by observing that the "space with a hole in it" is an idealization whose nature and limitations must not be ignored.

For it should be clear from the preceding discussion that such questions as "Is ψ single-valued?" or "What are the boundary conditions on ψ?" have no meaning *per se*. They can only be answered in terms of a *model* of the physical situation at hand. We have seen that there is no room for a double-valued ψ in the framework of a model which assumes that the coordinates of the particle form a complete set of observables and that the particle can be located anywhere in ordinary space. The existence of even very high potential barriers tending to keep the particle out of certain portions of space is entirely compatible with this model. Such potential barriers, or the equivalent condition that $\psi=0$ on the boundaries of the allowed region, simulate the effect on the particle of other parts of the physical system and are an indispensable shorthand way of including the extremely complicated interactions which a particle experiences with its surroundings. We are not, in this paper, discussing the merits of this model, but we do wish to point out that if its premises are applied in the description of a physical process, ψ *cannot be other than single-valued.* If observation calls for the use of a model in which a portion of space is actually and permanently "off limits" to a particle, the possibility of multivalued wave functions would have to be examined, but as long as we think of such restrictions merely as limiting cases of high, finite, but "in principle" penetrable barriers, the wave function must be taken to be single-valued in the discussion of the Aharonov-Bohm effect.

The same reasoning applies to the idealized picture of a multiply-connected superconductor which encloses quantized flux.

There remains then a question of consistency: Our analysis of the Aharonov-Bohm effect and of the flux quantization in a superconducting ring is based on an idealized model in which the magnetic flux is assumed to be strictly confined to a region which is inaccessible to the particle. Yet, both effects were also seen to depend on the single valuedness of ψ, which is assured only if the particle can penetrate the excluded region, with however small a probability. This contradiction is only apparent, and it has been indicated by Aharonov and Bohm (and can presumably also be demonstrated for the case of the superconductor) that the theory of the quantal effects under consideration does not involve any improper limiting process; rather, the effects arise in a continuous manner from a more realistic model which allows for partial penetration. The experimental results lend full support to this view.[4,5]

3. PARTICLE WITH SPIN

We now consider a particle with spin $\frac{1}{2}$. In the usual representation the wave function now has two components and is written as

$$\psi=\begin{pmatrix}\phi_1(x,y,z)\\ \phi_2(x,y,z)\end{pmatrix}, \tag{24}$$

$\phi_1(x,y,z)$ is the amplitude for finding the particle at point $P(x,y,z)$ with spin "up," i.e., in the *positive* z direction. As such it must be a single-valued function of the position coordinates by our previous discussion. Similarly, $\phi_2(x,y,z)$, which denotes the amplitude for spin "down," must return to its original value if we vary x, y, z along a closed loop in space.

However, it may be instructive to see that single valuedness of a function *per se* is not a sacred requirement, since it is possible to construct a representation for the states of a particle

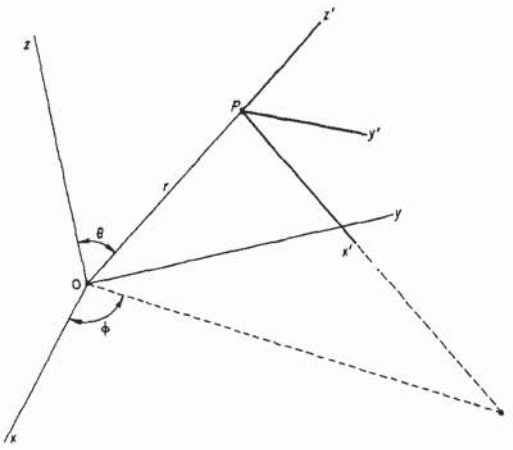

FIG. 4. The coordinate system $Px'y'z'$ underlies the triad representation of spinors.

with spin $\frac{1}{2}$ which utilizes just those double-valued functions that were rejected in the last section. For some purposes this second representation is particularly appropriate. The "new" representation is obtained by choosing again a spherical polar coordinate system with its origin at O and by relating the spin amplitudes at a given point P to the direction OP rather than the axis Oz. To remind ourselves of this change in representation we attach a suffix 0 to the spinors based on this representation. The equation

$$\psi=\begin{pmatrix}\chi_1(r,\theta,\varphi)\\ \chi_2(r,\theta,\varphi)\end{pmatrix}_0 \tag{25}$$

is to be interpreted as follows: $\chi_1(r,\theta,\varphi)$ is the amplitude for finding the particle at position r, θ, φ with spin "out," i.e., away from the origin but along OP. Similarly $\chi_2(r,\theta,\varphi)$ is the amplitude for spin "in," i.e., toward the origin. Thus, while ϕ_1 in Eq. (24) refers to a spin orientation which is the same for all spatial points, χ_1 in Eq. (25) refers to a spin orientation which is defined by the angular coordinates θ, φ and, therefore, depends on the particular point at which the wave function is being evaluated.

The connection between the two representations (24) and (25) is easily established. The change from the x, y, z dependence of ϕ_1 and ϕ_2 to the r, θ, φ dependence of χ_1 and χ_2 is trivial. But the spinor itself undergoes a unitary transformation, since the axis of quantization is rotated from Oz to OP (Fig. 4). One way (among infinitely many) of accomplishing this is to rotate the original Cartesian coordinate system by an angle θ about the y axis, and follow this with a rotation by an angle φ about the original vertical z axis.[13] The result is that the new z' axis points along OP, while the new x' axis lies in the zOP plane. For convenience we shall refer to the "new" coordinate system, with its origin at P and its orientation dependent upon P, as a *triad*. The unitary operator which corresponds to the over-all rotation is

$$S=e^{-i\sigma_z\varphi/2}e^{-i\sigma_y\theta/2}, \tag{26}$$

where σ_y and σ_z are the usual Pauli spin matrices. Explicitly

$$S=\begin{pmatrix}e^{-i\varphi/2}\cos\frac{1}{2}\theta & -e^{-i\varphi/2}\sin\frac{1}{2}\theta\\ e^{i\varphi/2}\sin\frac{1}{2}\theta & e^{i\varphi/2}\cos\frac{1}{2}\theta\end{pmatrix}, \tag{27}$$

and the connection between the two representations is given by

$$\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}=S\begin{pmatrix}\chi_1\\ \chi_2\end{pmatrix}_0. \tag{28}$$

The *total* angular momentum vector operator

$$\mathbf{J}=\mathbf{L}+\mathbf{S} \tag{29}$$

can now be written down in the new as well as in the familiar old representation. In the old representation it is

$$\mathbf{J}=\mathbf{r}\times\frac{\hbar}{i}\nabla+\frac{\hbar}{2}\boldsymbol{\sigma}, \tag{30}$$

but in the new representation it is, instead,

$$\mathbf{J}'=S^{-1}\mathbf{J}S. \tag{31}$$

The Cartesian components of this operator can be worked out easily and are most conveniently exhibited in this form:

$$J_+'=J_x'+iJ_y'$$

$$=\hbar e^{i\varphi}\left(\frac{\partial}{\partial\theta}+i\cot\theta\frac{\partial}{\partial\varphi}+\frac{1}{2\sin\theta}\sigma_z\right), \tag{32a}$$

$$J_-'=J_x'-iJ_y'$$

$$=-\hbar e^{-i\varphi}\left(\frac{\partial}{\partial\theta}-i\cot\theta\frac{\partial}{\partial\varphi}-\frac{1}{2\sin\theta}\sigma_z\right), \tag{32b}$$

$$J_z'=\frac{\hbar}{i}\frac{\partial}{\partial\varphi}. \tag{32c}$$

The operator for the square of the magnitude of total angular momentum is

$$\mathbf{J}^{2\prime}=\mathbf{L}^2+\frac{\hbar^2}{4\sin^2\theta}+\frac{i\hbar^2\cos\theta}{\sin^2\theta}\sigma_z\frac{\partial}{\partial\varphi}, \tag{33}$$

where L^2 is the usual differential operator

$$\mathbf{L}^2=-\hbar^2\left[\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)\right]. \tag{34}$$

It is easily verified that the operators defined by Eqs. (32) and (33) satisfy the usual commutation relations

$$[J_x',J_y']=i\hbar J_z' \quad et\ cycl.$$
$$[\mathbf{J}',\mathbf{J}^{2\prime}]=0$$

as they must since they are the generators of infinitesimal rotations of the system.

The simultaneous eigenfunctions of J_z' and $\mathbf{J}^{2\prime}$ can be found from the usual angular momentum eigenfunctions by application of the unitary transformation S^{-1}, or they can be obtained directly by solving the equations

$$J_z'\psi=\frac{\hbar}{i}\frac{\partial}{\partial\varphi}\psi=m\hbar\psi, \tag{35}$$

$$\mathbf{J}^{2\prime}\psi=j(j+1)\hbar^2\psi. \tag{36}$$

Equation (35) gives, of course,

$$\psi(r,\theta,\varphi)=e^{im\varphi}f(\theta), \tag{37}$$

and (36) then becomes

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)f(\theta)-\frac{m^2+\frac{1}{4}}{\sin^2\theta}f(\theta)+\frac{m\cos\theta}{\sin^2\theta}\sigma_z f(\theta)+j(j+1)f(\theta)=0, \tag{38}$$

where $f(\theta)$ is a spinor with two components

$$f(\theta)=\begin{pmatrix}f_1(\theta)\\ f_2(\theta)\end{pmatrix}. \tag{39}$$

Since σ_z is a diagonal matrix, (38) separates into two uncoupled equations

$$\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)-\frac{m^2+\frac{1}{4}\mp m\cos\theta}{\sin^2\theta}+j(j+1)\right]f_{1,2}(\theta)=0, \tag{40}$$

where the upper sign goes with f_1 and the lower with f_2.

So far nothing has been said about the quantum numbers m and j. But now the usual single-valuedness reasoning can be applied, except that special caution is required since, in the new representation, spin and space coordinates are more intimately coupled together than in the usual representation. In the latter one can vary the position coordinates without changing the axis of spin quantization which is the z axis once and for all. In the new representation, as we change the r, θ, φ coordinates of the particle location we also change the axis of spin quantization and the entire triad which is tied to the line OP. As we move, for instance, on a circle around the z axis, we return to the initial space point P, but the triad has been rotated by 2π about the z axis. It is known from the geometric properties of spinors that under such a full rotation of the coordinate system all spinor components change sign.[13] (Vectors and tensors behave differently, their components being restored without change of sign.) The state ψ cannot be single-valued unless this change of sign is compensated by another change of sign incurred by the transformation of the spatial function $e^{im\varphi}$ in Eq. (37). Hence, we see that m now must be half-integral so that ψ will return to its initial value when we move on a closed loop and simultaneously correct for the rotation of the triad. Having thus deduced from the fundamental postulates of quantum mechanics that, for a particle with spin $\frac{1}{2}$, m must be *half-integral*, it follows from the usual mathematical analysis of Eq. (40) that nonsingular solutions exist only if j is likewise half-integral and $j\geqslant m$, in agreement with the general results derived from the commutation relations for angular momentum. Equation (40) occurs in the quantal description of the symmetric top, and its eigenfunctions are the quantities $d_{\frac{1}{2},\pm m}^{(j)}(\theta)$ which are known from the representations of the

rotation group.[13] Indeed, the eigenfunctions (37) may be written in terms of the rotation matrices

$$D_{m'm}^{(j)}(\alpha,\beta,\gamma)=e^{-im'\alpha}d_{m'm}^{(j)}(\beta)e^{-im\gamma} \quad (41)$$

as

$$\psi(r,\theta,\varphi)=\begin{pmatrix} c_1 D_{\frac{1}{2},m}^{(j)*}(0,\theta,\varphi) \\ c_2 D_{\frac{1}{2},-m}^{(j)}(0,\theta,\varphi)\end{pmatrix}_0. \quad (42)$$

The arbitrary constants c_1 and c_2 may be determined by requiring ψ to be an eigenfunction of some operator which commutes with $\mathbf{J}^{2'}$ and J_z'. Usually one chooses the magnitude of orbital angular momentum, $\mathbf{L}^{2'}$ (not $\mathbf{L}^2$!), as this operator, and assigns the quantum number l to ψ besides j and m. However, it is much more natural to take advantage of the fact that $\mathbf{J}^{2'}$ and J_z' commute with the Pauli matrix σ_z which represents in the new triad representation the component of the spin along the radius vector $\mathbf{r}=\overrightarrow{OP}$. The two eigenfunctions

$$\begin{pmatrix} D_{\frac{1}{2},m}^{(j)*}(0,\theta,\varphi) \\ 0\end{pmatrix}_0, \quad \begin{pmatrix} 0 \\ D_{\frac{1}{2},-m}^{(j)}(0,\theta,\varphi)\end{pmatrix}_0$$

represent two eigenstates of $\mathbf{J}^{2'}$ and J_z' corresponding to "spin in" and "spin out," since $\sigma_z=(\boldsymbol{\sigma}\cdot\mathbf{r})'/r$ has, respectively, the values $+1$ and -1. These two simple states are said to have *helicity* $+1$ and -1. Since both components contain the factor $e^{im\varphi}$, with m being half-integral, we see that they are double-valued functions of position. However, it must be remembered that this happens here because we have chosen a representation in which spin and space coordinates cannot be varied independently.

Pauli showed,[10] for the case of the Dirac equation of the relativistic electron, that wave functions containing half-integral values of m are the only admissible ones if we insist that repeated application of "raising" and "lowering" operators, J_+ and J_-, must not yield singular functions.[15] We have seen, however, that the geometric transformation properties of spinors together with a properly interpreted single-valuedness condition give the same result very naturally, and these considerations can be extended to the relativistic case.[16]

The triad representation, in which the spin of a particle is related to its radius vector rather than to an axis of quantization fixed in space, has useful applications in many problems, especially in the description of scattering of a particle with spin $\frac{1}{2}$. For such a particle it is evidently simplest to describe its polarization by referring its spin to an axis of quantization along the direction of the momentum vector of the particle. Jacob and Wick have shown how such a helicity description can be utilized in the analysis of collision processes.[17]

4. RIGID BODY

Finally, we may consider the quantum mechanics of a rigid body rotating about a fixed point, the coordinate origin.[18] In this model of a physical system one may introduce the Euler angles α, β, γ as suitable "orientation coordinates" on which the wave function depends. Again we inquire into the change of $\psi(\alpha,\beta,\gamma)$ as the coordinates α, β, γ are varied, and we ask if ψ resumes its initial value when the original spatial orientation of the rigid body is restored. This happens, for example, if β and γ are kept fixed but α is varied from 0 to 2π. If it is assumed that the spatial orientation of the body gives a complete set of observables, then ψ must be a single-valued function of the orientation coordinates and, hence, we must have

$$\psi(2\pi,\beta,\gamma)=\psi(0,\beta,\gamma). \quad (43)$$

If, for example, a rotating molecule is represented by the model of a rigid body, the wave function must satisfy the single-valuedness condition (43) because the rigidity of the structure is an idealization, approximating the high but nevertheless finite potential barriers which hold the atoms in their equilibrium configuration. Since, furthermore, the nuclear spins interact only weakly with the orbital motion of the nuclei, the appropriate wave functions depend only on the position coordinates and involve the addition of orbital angular momenta (L–S coupling). Hence, the molecular rotations are described by single-valued rigid body wave func-

[15] E. Schrödinger, Commentationes Pontificia Academia Scientiarium **2**, 231 (1938).

[16] F. Bakke and M. Wergeland, Physics Seminar, Trondheim, have also considered the double-valued solutions of the Dirac electron in a Coulomb field (unpublished).

[17] M. Jacob and G. C. Wick, Ann. Phys. **7**, 404 (1959).

[18] F. Bopp and R. Haag discuss the rigid body and the quantal many-body problem from a similar point of view in Z. Naturforsch. **5a**, 644 (1950).

tions, and the corresponding angular momentum can only have integral values.

However, for the idealized model of a truly rigid body there exists also the possibility of double valuedness of ψ as a function of orientation, since the rotations of a rigid body which restore it to its original orientation can be divided into two separate classes: Any such complete rotation is equivalent either to the identity operation in which no displacement occurs at all or to the rotation by 2π about an arbitrary axis. The two values which a double-valued function of orientation takes on can be assigned to these two classes in a physically unambiguous fashion. Hence, for a rigid body it is possible to supplement the observables describing the orientation by an additional observable which allows for double-valued wave functions in the framework of the general principles of quantum mechanics. The basic reason for this possibility is that the group space for the three-dimensional rotation group, in which every rotation is assigned a point, subject to the continuity properties of the rotation group, is a doubly connected space.[19] This means that the operation "rotation by 2π about an axis" cannot be continuously deformed into the identity operation "no rotation at all." Rotations by 4π, on the other hand, can be deformed into "no rotation." These statements are easily proved if the rotations are appropriately parametrized (e.g., in terms of Cayley-Klein parameters), but it seems difficult to visualize these features of three-dimensional rotations directly.

In any event, the topological structure of the orientation coordinate space of a rigid body permits us to distinguish "odd" from "even" rotations, when we have brought a rigid body back to its initial orientation. It is even possible to imagine an approximate physical model of a rigid body whose wave function has the property

$$\psi(2\pi,\beta,\gamma) = -\psi(0,\beta,\gamma). \tag{44}$$

One only needs to couple strongly a spin $\frac{1}{2}$ particle to a rigid body which has a single-valued wave function $\psi_1(\alpha,\beta,\gamma)$. The total wave function will be of the form

$$\psi = \omega_{(\alpha\beta\gamma)}(r')\psi_1(\alpha,\beta,\gamma), \tag{45}$$

where r' designates a set of intrinsic coordinates describing the motion of the extra particle with respect to a coordinate system fixed in the rigid body and dependent on α, β, γ. As explained in Sec. 3, ω changes sign as α goes from 0 to 2π. Hence, the total wave function also has this property.

As an example, it might be mentioned that in the collective model of nuclear structure one encounters symmetric top wave functions[20]

$$D_{MK}^{(I)}(\alpha\beta\gamma),$$

where I is the nuclear spin, so that for odd A nuclei, which have half-integral spin, the model wave function is a double-valued function of the orientation coordinates. The usefulness of such an approximation depends, of course, on the nature of the nuclear forces. These are strongly spin dependent and favor j–j coupling, so that I results from adding the individual half-integral j values.

ACKNOWLEDGMENTS

I wish to thank Y. Aharonov, L. C. Biedenharn, A. Bohr, and W. H. Furry for several illuminating discussions.

[19] E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959), see p. 89 and Chap. 15.

[20] A. Bohr, Kgl. Danske Videnskab Selskab, Mat.-fys. Medd 26, No. 14, 27 (1952).