Is the sign change of spinors under 2π rotations observable?

1969 J. Phys. A: Gen. Phys. 2 267

(http://iopscience.iop.org/0022-3689/2/3/004)

J. PHYS. A (GEN. PHYS.), 1969, SER. 2, VOL. 2. PRINTED IN GREAT BRITAIN

Is the sign change of spinors under 2π rotations observable?

J. S. DOWKER
Department of Theoretical Physics, University of Manchester
MS. received 21*st November* 1968

Abstract. Hegerfeldt and Kraus's criticisms of Aharonov and Susskind's discussion of the above question are investigated. It is shown that the rotation operator applied to a wave function can be equivalent to solving Schrödinger's equation if gravitational fields are present. The specific example of quantum mechanics on a cone is considered in detail.

1. Introduction

Hegerfeldt and Kraus (1968) have criticized the arguments of Aharonov and Susskind (1967) which led to an affirmative answer to the question posed in the title to this paper. The result of the latter pair of authors contradicts the standard attitude which seems to be founded on the fact that observables are bilinear functions of the spinor. In the present work we wish to make a few comments on this interesting question.

2. The basic idea

Hegerfeldt and Kraus summarized the calculation of Aharonov and Susskind in the following gedanken experiment.

A wave function $\psi(0)$ is divided, by some means, into two parts ψ_1 and ψ_2,

$$\psi(0) = \psi_1 + \psi_2$$

which are then spatially separated. After this separation ψ_1 is subjected to a rotation through angle θ,

$$\psi_1 \rightarrow \mathscr{D}(\theta)\psi_1.$$

The wave functions are now recombined to give $\psi(\theta)$ where

$$\psi(\theta) = \mathscr{D}(\theta)\psi_1 + \psi_2.$$

In particular we have, in the case of half odd integer spins,

$$\psi(0) = \psi_1 + \psi_2, \qquad \psi(2\pi) = \psi_2 - \psi_1.$$

Thus we might expect to see different things in these two cases. A typical effect would be of the interference type which depends on $|\psi(\theta)|^2$.

Hegerfeldt and Kraus's main criticism of this argument is that one cannot perform a rotation on just one part of the wave function or, as they say, $\mathscr{D}(\theta)$ is a 'universal' operation, i.e. it acts everywhere. They rightly point out that the experiment is really a dynamical, as opposed to a kinematical, situation. What we really want is $\exp(\mathrm{i}Ht)\psi_1$ and this depends strongly on the experimental arrangement. The application of $\mathscr{D}(\theta)$ to ψ_1 is not, they say, the correct recipe for solving Schrödinger's equation. We wish to quarrel with this last statement.†

It will be appreciated that the situation envisaged here is, in fact, identical with that of the 'usual' Aharonov–Bohm effect (Aharonov and Bohm 1959) except that the multicomponent wave function now undergoes a non-diagonal change of phase rather than the simple electromagnetic phase change:

$$\psi_1 \rightarrow \exp(\mathrm{i}\alpha_1)\psi_1, \qquad \psi_2 \rightarrow \exp(\mathrm{i}\alpha_2)\psi_2. \tag{1}$$

Now this latter phase change *is* the result of a dynamical propagation. We could use, for

† It is unclear how general Hegerfeldt and Kraus intend their statement to be.

example, the WKB or the phase-integral method of solving Schrödinger's equation (e.g. Furry 1962). However we can also just say "subject ψ_1 and ψ_2 to phase transformations (1)". These phase changes are functions of the experimental situation and are different, in general. The particular fact which makes, or allows, them to be different is the existence of a non-trivial (non-zero) electromagnetic field. If no field is present α_1 must equal α_2. We refer the reader to the original paper of Aharonov and Bohm (1959) for some particular experimental arrangements.

ψ_1 and ψ_2 are spatially separated and we may describe the fact that α_1 is not equal to α_2 by saying that the phase change is a function of position. This ties in with the Yang and Mills (1954) method of generating a gauge (here, electromagnetic) field. The essential point of this method is the requirement of invariance under a transformation group whose parameters are position dependent.

Returning to the particular case under consideration the point we wish to make is that we *can* subject ψ_1 and ψ_2 to different rotations, i.e. subject ψ to a position-dependent rotation, provided there is a non-trivial gravitational field present. Utiyama (1956) and Kibble (1961) have shown that the gravitational field can be introduced as a Yang–Mills gauge field if the parameters of the Lorentz transformation are made functions of position.† The reason for this can, perhaps, be appreciated if we look upon the generalized invariance as allowing independent homogeneous Lorentz transformations at each point of space–time. These transformations can be considered to form the symmetry groups of 'ordinary' four-dimensional spaces attached to each space-time point. Since Riemannian spaces can be defined as spaces which are pointwise flat (i.e. pointwise Euclidean or 'ordinary'), we can see how naturally they enter the picture. In other words the fact that the Lorentz transformations are now associated with each space–time point independently allows the underlying space–time manifold the freedom to be Riemannian, i.e. non-flat in general. Gravitation comes in when, with Einstein, we associate curved space–time with a true gravitational field.

The above considerations make it highly likely, at least, that $\mathscr{D}(\theta)$ applied to ψ_1 *is*, or can be, equivalent to solving Schrödinger's equation and that the words "apply a rotation to ψ_1" is simply a short way of describing the result of some *dynamical* calculation involving gravitational fields.

3. Particular examples

To justify these remarks we should really consider some concrete examples. This we proceed to do.

The first example is one discussed independently by the present author (Dowker 1967) and by Wisnivevsky and Aharonov (1967). It may be termed "quantum mechanics on a cone".

As we know, a cone is locally flat. This is clear because the cone is a developable surface, i.e. it can be unrolled onto a plane. If a cone is so unrolled we see that the difference between a cone and a plane can be thought of as a topological, and therefore global, one. A cone is topologically equivalent to a plane which has had a sector cut out, the two sides of the sector being identified.

The Riemann–Christoffel curvature tensor is zero everywhere on the surface of a cone except at the apex, a singularity of the cone, where it is infinite. It is therefore an ideal case in which to discuss a gravitational analogue of the Aharonov–Bohm effect (Dowker 1967), which, we recall, is concerned with the effects of a confined electromagnetic field, for example that of an infinite solenoid.

Of course actual space is not two dimensional but we can easily incorporate the cone idea simply by constructing a space–time which has a cone-like two-dimensional cross section. Such a space–time has already been discussed by Marder (1959, 1962) and was the starting point of the present author's previous work (Dowker 1967). The metric of this space–time is given by

$$ds^2 = dt^2 - \rho^2\, d\phi^2 - A^2(d\rho^2 + dz^2) \tag{2}$$

† The germ of this idea can be found in Weyl (1929).

where A is a constant. When A is unity, space–time is flat everywhere (ρ, ϕ and z are cylindrical coordinates) but if A differs from one a non-trivial gravitational field is present. This is easily seen from the form of the geodesics,

$$\rho \cos\left(\frac{\phi}{A}+\beta\right) = \rho_0 \cos\beta$$

which does not describe a straight line, unless A is unity.

The cross sections z = const., t = const. are topologically equivalent to a cone. This is best seen by changing variables so that ds^2 'looks flat'. Thus define

$$\rho' = A\rho, \qquad z' = Az, \qquad \phi' = \frac{\phi}{A}$$

whence

$$ds^2 = dt^2 - \rho'^2\, d\phi'^2 - d\rho'^2 - dz'^2.$$

However we must recognize the fact that the whole of the (ρ, ϕ) plane does not map onto the whole of the (ρ', ϕ') plane but rather onto that part of this plane that is left when a sector of angle $2\pi/A$ has been removed (we choose $A \geqslant 1$) and the sides of the sector identified. Thus the (ρ, ϕ) plane is topologically equivalent to a cone of angle $\sin^{-1}\{(A-1)/A\}$. We emphasize that it is the points (t, ϕ, ρ, z), (ϕ modulo 2π), that are in one to one correspondence with the points of *physical* space–time.

Let us now consider the equations of motion in a curved space–time. For our purposes here the exact form of these equations is not essential. The point is that in them the derivative of the multicomponent field or wave function ψ occurs as the covariant derivative $\nabla_\mu \psi$, where

$$\nabla_\mu \psi = (\partial_\mu + \Gamma_\mu)\psi$$

Γ_μ being a matrix connection. The theory of spin in curved space (see e.g. Dowker and Dowker 1966) gives for Γ_μ the expression

$$\Gamma_\mu = \mathrm{i}[\Gamma_{\mu\nu}^{\alpha}\, b_{\alpha}^{\cdot\alpha'} a^{\nu}_{\cdot\beta'} - \partial_\mu b_{\nu}^{\cdot\alpha'} a^{\nu}_{\cdot\beta'}] J_{\alpha'}^{\cdot\beta'} \tag{3}$$

where $\Gamma_{\mu\nu}^{\alpha}$ is the usual Christoffel symbol and the a and b are quantities which locally diagonalize the actual metric $g_{\mu\nu}$ to the $\eta_{\mu\nu}$, $= (-1, -1, -1, 1)$, one thus

$$\begin{aligned} g_{\mu\nu} &= b_{\mu}^{\cdot\mu'} b_{\nu}^{\cdot\nu'} \eta_{\mu'\nu'} \\ \eta_{\mu'\nu'} &= a^{\mu}_{\cdot\mu'} a^{\nu}_{\cdot\nu'} g_{\mu\nu}. \end{aligned} \tag{4}$$

The $J_{\alpha'}^{\cdot\beta'}$ are the generators of the homogeneous Lorentz group in the representation to which ψ belongs. We can think of the $(j, 0)$ representations for definiteness.

The a and b define at each point a set of Cartesian axes—the local 'tetrad' and the formalism defined by the equations (3) and (4) is called the tetrad formalism.

The argument now proceeds as in the electromagnetic calculation. We rapidly sketch it here. Consider a wave function ψ to be split into two beams ψ_1, ψ_2 at the point A, $\phi = 0$ and $\rho = \rho_0$. These beams then pass on opposite sides of the z axis and recombine at, say, point B, $\phi = \pm\pi$, $\rho = \rho_0'$. We want the value of ψ at B. To obtain this we can either solve the equations of motion directly, perhaps using an approximation scheme or Feynman's action principle, or we can divide the region into two field-free parts—those above and below the x axis ($\phi = 0$) and treat these separately.

It should be tolerably clear that, by analogy with the Aharonov–Bohm calculation, from A to B, ψ_1 will undergo the change

$$\psi_1(\mathrm{A}) \to \psi_1(\mathrm{B}) = T \exp\left(-\int_{\mathrm{A}}^{\mathrm{B}} \Gamma_\mu\, dx^\mu + \delta\beta_1\right) \psi_1(\mathrm{A}) \tag{5}$$

and similarly for ψ_2

$$\psi_2(\mathrm{A}) \to \psi_2(\mathrm{B}) = T\exp\left(-\int_{\mathrm{A}}^{\mathrm{B}} \Gamma_\mu\, \mathrm{d}x^\mu + \delta\beta_2\right)\psi_2(\mathrm{A}).$$

The paths of integration are the paths of the corresponding beams. Since A is the common starting point we can take $\psi_1(\mathrm{A}) = \psi_2(\mathrm{A}) = \psi(\mathrm{A})$.

The only difference from the electromagnetic case is the necessity for the ordering operator T which acts with respect to the parameter labelling the points on the paths (Bergmann 1962, Feynman 1951).

The quantities $\delta\beta_1$ and $\delta\beta_2$ are diagonal phase changes and are composed of the ordinary optical phase change and an effect due to the replacement of $\eta_{\mu\nu}$ by $g_{\mu\nu}$ in the equation of motion. We are not interested in $\delta\beta_1$ and $\delta\beta_2$ and shall henceforth ignore them.

In general, because of the T operator, the phase multiplier

$$T\exp\left(-\int_{\mathrm{A}}^{\mathrm{B}} \Gamma_\mu\, \mathrm{d}x^\mu\right)$$

is not equivalent to a 'Lorentz' transformation; however, in our example this is so. Further, in the present instance because $R_{\mu\nu\alpha\beta}$ vanishes in the two regions, the paths of integration in (5) can be any paths homotopic to the 'dynamical' ones.

We now compute the changes (5) explicitly for Marder's metric, and choose the tetrad formalism given by the expressions for a and b

$$\|a^{\mu}_{.\mu'}\| = \begin{pmatrix} A^{-1} & & & 0 \\ & \rho^{-1} & & \\ & & A^{-1} & \\ 0 & & & 1 \end{pmatrix}, \qquad \|b^{.\mu'}_{\mu}\| = \begin{pmatrix} A & & & 0 \\ & \rho & & \\ & & A & \\ 0 & & & 1 \end{pmatrix}$$

$$(x^1 = \rho,\ x^2 = \phi,\ x^3 = z,\ x^0 = t).$$

This means that the local tetrad axes are parallel to the local t, ρ, ϕ, z axes (cf. Brill and Wheeler 1957). Computation yields for the Christoffel symbols and the spinor connection the values

$$\Gamma^2_{12} = \frac{1}{\rho}, \qquad \Gamma^1_{22} = -\frac{\rho}{A^2}$$

$$\Gamma_1 = \Gamma_3 = \Gamma_0 = 0$$

$$\Gamma_2 = \frac{1}{A} J_{1'2'}.$$

Equation (5) now reads

$$\begin{aligned} \psi_1(\mathrm{A}) \to \psi_1(\mathrm{B}) &= \exp\left(-\mathrm{i}\frac{\pi}{A} J_{1'2'} + \delta\beta_1\right)\psi(\mathrm{A}) \\ \psi_2(\mathrm{A}) \to \psi_2(\mathrm{B}) &= \exp\left(\mathrm{i}\frac{\pi}{A} J_{1'2'} + \delta\beta_2\right)\psi(\mathrm{A}). \end{aligned} \qquad (6)$$

$J_{1'2'}$ is the generator of spatial rotations about the z axis

$$J_{1'2'} = J_z$$

and equation (6) represents rotations through angles π/A and $-\pi/A$ about this axis. However we must remember that part of this rotation is due to the rotation of the local tetrad frame and so now, with respect to a tetrad of *fixed* orientation, we have the spinor transformations

$$\psi_1(\mathrm{A}) \to \exp\left\{\mathrm{i}\left(1-\frac{1}{A}\right)\pi J_z + \mathrm{i}\delta\beta_1\right\}\psi(\mathrm{A})$$

$$\psi_2(\mathrm{A}) \to \exp\left\{\mathrm{i}\left(\frac{1}{A}-1\right)\pi J_z + \mathrm{i}\delta\beta_2\right\}\psi(\mathrm{A})$$

whence

$$\begin{aligned}\psi(\mathrm{B}) &= \psi_1(\mathrm{B})+\psi_2(\mathrm{B}) \\ &= \mathscr{D}(-\tfrac{1}{2}\theta)\{\mathscr{D}(\theta)\exp(\mathrm{i}\delta\beta_1)+\exp(\mathrm{i}\delta\beta_2)\}\psi(\mathrm{A})\end{aligned}$$

with

$$\theta = 2\pi\frac{(A-1)}{A}.$$

The overall unitary factor, $\mathscr{D}(-\frac{1}{2}\theta)$, of the *unsplit* wave function is immaterial. The factor in braces is the important one. If A is 1, i.e. space–time is flat everywhere, θ is zero and this factor is the purely 'optical' one. The real question is whether θ ever equals $2\pi n$ i.e. whether

$$A = A_n \equiv \frac{1}{1-n}$$

is a physical possibility. Since the sign of A is immaterial A_2 is equivalent to A_0, $= 1$. A_1 is infinite and gives a singular metric which we rule out. The essential point about the remaining A_n, for n greater than two, is that their magnitudes are all less than 1. Now, in this case, classical particles are repelled by the singularity at $\rho = 0$. This can be deduced from the geodesic equations. Thus we might expect something odd to occur inside the source generating the gravitational field when A becomes less than 1. To see this explicitly we shall construct a source metric such that (2) is valid in some exterior region and matches onto the source, or interior, metric at some radius. This was, in fact, done by Marder (1959). He rewrote the exterior metric by introducing a new radial coordinate ρ_1 equal to $\rho' - K$ where K is a positive constant. Then

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}\rho_1{}^2 - A^{-2}(\rho_1+K)^2\,\mathrm{d}\phi^2 - \mathrm{d}z'^2.$$

He found, for $A > 1$, the interior solution

$$\begin{aligned}\mathrm{d}s^2 &= \mathrm{d}t^2 - \mathrm{d}\rho_1{}^2 - f^2(\rho_1)\,\mathrm{d}\phi^2 - \mathrm{d}z'^2 \\ f &= z^{-1}\sin z\rho_1\end{aligned} \qquad (7)$$

where

$$z = \frac{1}{a_1}\cos^{-1}\frac{1}{A} = \frac{(A^2-1)^{1/2}}{a_1+K}$$

the matching radius, or boundary of the source, being $\rho_1 = a_1$.

From Einstein's equations the energy density, T_{00}, follows as the positive value

$$T_{00} = z^2/8\pi.$$

If A is less than 1 it is not difficult to see by inspection that (7) still holds with

$$f = z^{-1}\sinh z\rho_1$$

and

$$z = \frac{1}{a_1}\cosh^{-1}\frac{1}{A} = \frac{(1-A^2)^{1/2}}{a_1+K}$$

but now we find for the energy density the *negative* value

$$T_{00} = -z^2/8\pi.$$

This is as we would expect if the source is to be repulsive. It is, however, unphysical. Negative masses have never been seen.

Thus we can never reach the interesting values $\theta = 2\pi n$ and the unobservability of the spinor sign change is reduced to questions of detailed dynamics and not principle.

To see whether a similar situation holds for other gravitational fields we shall briefly consider another example—that of the Schwarzschild solution. All the relevant expressions have been evaluated by Brill and Wheeler (1957) and we shall just copy their answers. Although they were specifically concerned with spin half their formula for the spinor connection Γ_μ is quite general. Spherical polar coordinates $r, \theta, \phi, t, = (x^1, x^2, x^3, x^0)$, are used and again the local tetrad axes are chosen to be parallel to the local r, θ, ϕ, t axes. For Γ_μ we have

$$\Gamma_1 = 0, \qquad \Gamma_2 = \mathrm{i}J_{3'}\left(1-\frac{2m}{r}\right)^{1/2}$$

$$\Gamma_3 = \mathrm{i}\left[J_{1'}\cos\theta + J_{2'}\sin\theta\left(1-\frac{2m}{r}\right)^{1/2}\right], \qquad \Gamma_0 = \frac{1}{2}\frac{m}{r^2}J_{1'}.$$

The path formed by the two beams is assumed to circulate the z axis and to be for fixed θ, say $\frac{1}{2}\pi$ and $r = a$. Then the relevant phase change is†, allowing for the rotation of the local tetrad frame,

$$2\pi\left\{1-\left(1-\frac{2m}{a}\right)^{1/2}\right\}J_{2'}, \qquad J_{2'} \equiv J_z.$$

This expression equals 2π when $2m$ equals a. This means that the beams must pass where the metric is infinite, as in the previous case. For physical sources this singularity 'occurs' inside the source where the exterior solution does not apply anyway. Again, the observability of the spinor sign change is limited by physical considerations.

Whether it is possible in principle or practice to make the beams undergo multiple circuits of the z axis, which would alter the above situation, the author does not know.

4. Conclusion and discussion

The two examples we have discussed seem to rule out the possibility of a test of the observability of the spinor sign change using a gravitational field to rotate the wave functions. It may be possible to prove this more generally.

The experimental arrangement we have had in mind has been that described in § 3. However there does not seem to be any objection in principle to the arrangement whereby a beam is made to pass any number of times, say m, around the z axis and then allowed to interfere with itself at the starting point much as in the Sagnac experiment. In this case instead of requiring $A = (1-n)^{-1}$ we would need $A = m(m-n)^{-1}$ which can easily be greater than 1.

Because of the definite dynamical origin of the rotations we do not think that we can say, with Hegerfeldt and Kraus, that the above result (if θ *can* equal 2π) implies that rotation through zero angle produces an effect.‡

We note that a generalization of Marder's space–time easily results if we take a space–time whose constant z cross sections are the three-dimensional space–times discussed by Staruskiewicz (1963). In this way we could discuss the gravitational analogue of the two (or more) solenoids considered by Peshkin *et al.* (1961).

Finally we should just like to say that the 'results' of the present paper raise again the interesting question of what happens to the theories of quantum mechanics and spin on non-orientable surfaces Möbius strips, Clifford–Klein surfaces, etc.

Hegerfeldt and Kraus (private communication, since the preparation of this paper) have explained that the dynamical situation they had in mind was literally that of rotating

† For theoretical convenience we have assumed that t is constant.

‡ Hegerfeldt and Kraus's argument applies only to the *kinematic* situation, as they themselves imply.

the 'box' which contained ψ_1. Clearly, the effect of this is not necessarily to rotate ψ_1. In this case, there is *no* disagreement between Hegerfeldt and Kraus and ourselves.

Concerning the observability of the sign change we should like to say that it is just as observable as the sign change in the Pauli principle. Exchanging identical fermions reverses the sign of the wave function but alters nothing physically, yet no one would say the exclusion principle was unobservable.

References

AHARONOV, Y., and BOHM, D., 1959, *Phys. Rev.*, **115**, 485–91.
AHARONOV, Y., and SUSSKIND, L., 1967, *Phys. Rev.*, **158**, 1237.
BERGMANN, P. G., 1962, *Handb. d. Phys.*, **4**, 203–90 (Berlin: Springer-Verlag).
BRILL, D. R., and WHEELER, J. A., 1957, *Rev. Mod. Phys.*, **29**, 465–79.
DOWKER, J. S., and DOWKER, Y. P., 1966, *Proc. Phys. Soc.*, **87**, 65–78.
DOWKER, J. S., 1967, *Nuovo Cim.*, **52B**, 129–35.
FEYNMAN, R. P., 1951, *Phys. Rev.*, **84**, 108–28.
FURRY, W. H., 1962, *Lectures in Theoretical Physics*, Vol. 5 (New York: Interscience).
HEGERFELDT, G. C., and KRAUS, K., 1968, *Phys. Rev.*, **170**, 1185–6.
KIBBLE, T. W. B., 1961, *J. Math. Phys.*, **2**, 212–21.
MARDER, L., 1959, *Proc. R. Soc.* A, **252**, 45–9.
—— 1962, *Developments in General Relativity* (Warsaw: Pańistwowe Wydawnyctwa Naukowe).
PESHKIN, M., TALMI, I., and TASSIE, L. J., 1961, *Ann. Phys., N.Y.*, **12**, 426–35.
STARUSZKIEWICZ, A., 1963, *Acta. Phys. Polon.*, **24**, 735–40.
UTIYAMA, R., 1956, *Phys. Rev.*, **101**, 1597–607.
WEYL, H., 1929, *Proc. Natn. Acad. Sci.*, **15**, 323–34.
WISNIVEVSKY, D., and AHARONOV, Y., 1967, *Ann. Phys. N.Y.*, **45**, 479.
YANG, C. N., and MILLS, R. L., 1954, *Phys. Rev.*, **96**, 191–5.