PHYSICAL REVIEW D VOLUME 23, NUMBER 6 15 MARCH 1981

Comment on "Condition for nonexistence of Aharonov-Bohm effect"

U. Klein
Institute for Theoretical Physics, University of Linz, Linz, Austria
(Received 7 March 1980)

A recent attempt to deny the crucial role of inaccessible fields in experiments performed to test the predictions of Aharonov and Bohm is critically discussed. It is shown that these experiments may well be considered to show the reality of the Aharonov-Bohm effect even if it is correct to say that the results are completely determined by the accessible field.

It has been generally accepted for some time that electromagnetic potentials acquire a special role in quantum mechanics by transmitting an influence of inaccessible fields to charged particles in multiply connected regions of space and time. There is also a considerable amount of experimental data confirming this concept, the importance of which was first stressed by Aharonov and Bohm[1] (AB). Despite this, there have recently been attempts[2] to show that the Aharonov-Bohm effect (AB effect) is a purely mathematical artifact not leading to observable results when proper use is made of multivalued wave functions. A similar conclusion was recently reached by Roy[3] arguing from a different point of view. Roy claims to show that physical effects of inaccessible fields cannot exist if the electromagnetic potentials fulfill a certain set of mathematical conditions. In particular, he asserts—although he does not completely rule out the possibility of quantum effects in field-free regions (see short remark at end of paper)—that previously observed interference fringe shifts cannot be interpreted as being caused by the inaccessible magentic field.

The present paper is mainly concerned with a discussion of the fundamental questions raised by Roy. His position as compared to that of other authors[2] asserting the nonexistence of the AB effect is clarified. It is shown that a proper interpretation of his mathematical results leads to agreement with Aharonov and Bohm's point of view rather than indicating nonexistence of the AB effect in previously performed experiments.

As a starting point let us recall DeWitt's observation[4] that potentials may be completely eliminated from quantum electrodynamics by means of a gauge transformation

$$\psi' = \psi e^{i\chi}, \quad A'_\mu = A_\mu + \frac{\partial \chi}{\partial x_\mu}, \tag{1}$$

where

$$\chi = \int_{-\infty}^{0} A_\mu(z)\frac{\partial z_\mu}{\partial \xi} d\xi \tag{2}$$

and

$$A'_\mu = \int_{-\infty}^{0} F_{\nu\rho}(z)\frac{\partial z_\nu}{\partial \xi}\frac{\partial z_\rho}{\partial x_\mu} d\xi . \tag{3}$$

Here $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and $z_\mu(x,\xi)$ ($\mu = 1-4$, $z_4 = iz_0$) are single-valued differentiable functions of the space-time coordinates x_μ and a real parameter ξ, defined for all x and $-\infty < \xi \leq 0$, obeying

$$z_\mu(x,0) = x_\mu , \qquad \lim_{\xi\to-\infty} z_\mu(x,\xi) = \text{spatial infinity}. \tag{4}$$

Roy points out that the transformation defined by (1) to (4) is, in fact, only feasible from a physical point of view provided A_μ obeys the additional conditions

$$\int_{-\infty}^{0} A_\mu(z)\frac{\partial z_\mu}{\partial \xi} d\xi < \infty, \quad \lim_{\xi\to-\infty} A_\nu(z)\frac{\partial z_\nu}{\partial x_\mu} = 0 , \qquad \int_{-\infty}^{0} F_{\rho\nu}(z)\frac{\partial z_\rho}{\partial \xi}\frac{\partial z_\nu}{\partial x_\mu} d\xi < \infty . \tag{5}$$

Equation (5) guarantees the single-valuedness of the wave function.[5] The mathematical condition[6] for this reads

$$\frac{\partial^2 \chi}{\partial x_\mu \partial x_\nu} - \frac{\partial^2 \chi}{\partial x_\nu \partial x_\mu} = 0 \text{ everywhere}, \tag{6}$$

which expresses, on the other hand, the physical equivalence of A_μ and A'_μ. Topological singularities of χ show up in δ-like singularities[7] of the left-hand side of (6). There can be no doubt that DeWitt's original framework has to be supplemented by condition (5). Let us discuss its physical implications.

Let a particle be confined to a multiply connected region R where $F_{\mu\nu}$ is either small (exterior of finite solenoid in actual experiments on magnetic Aharonov-Bohm effect) or exactly zero (exterior of idealized infinite solenoid or of toroidal solenoid).

We will consider first the pure case $F_{\mu\nu} = 0$ in R and ask if DeWitt's gauge can be used to eliminate the AB effect. The answer is provided by the following:

Proposition: Let $z_\mu(x, \xi)$ be a single-valued differentiable path obeying condition (4) and lying in R for every x_μ in the multiply connected region R. If $F_{\mu\nu} \to 0$ in R in such a way as to make the right-hand side of Eq. (3) vanish, then the transformation defined by (1) to (5) is inadmissible as a result of the violation of condition (5).

To prove this, let us assume that (5) holds. Then $A'_\mu = 0$ in R and $\oint A'_\mu\, dx_\mu = \oint A_\mu dx_\mu$ in the forbidden region S. Consider now an unshrinkable closed path l_R in R[8] and a second closed path l_S in S which together span a surface F lying partly in R and partly in S. Application of Stokes's theorem gives

$$\int_F dS_{\mu\nu} F'_{\mu\nu} = \oint_{l_S} A_\mu dx_\mu \,. \tag{7}$$

If $l_S \to l_R$, the right-hand side of Eq. (7) does not vanish, showing that the new fields $F'_{\mu\nu}$ are singular at the boundary of R and S. This contradicts the validity of condition (5).

The proposition shows that the influence of the inaccessible field cannot be eliminated. In particular, the vector potential of an infinitely long solenoid used by AB (Ref. 1) cannot be replaced by $A'_\mu = 0$. The proposition further asserts the reality of the quantum effects of the enclosed flux for electrons moving outside a toroidal solenoid; here the condition $F_{\mu\nu} = 0$ in R is exactly fulfilled.

As regards recent assertions[2] that the AB effect (in the pure case) does not exist, one concludes from the proposition and from Eqs. (1) and (6) that the choice $A'_\mu = 0$ in R may only be justified if one assumes that (a) the original wave function is multivalued in exactly such a way as to make ψ' single valued, and that (b) the singular field appearing as a result of the violation of (6) is physically irrelevant. We will not discuss these points in detail since they are not directly related to the present analysis but restrict ourselves to quoting several papers[9] which discuss the inadmissibility of these assumptions from a physical point of view.

We proceed to the case of small but finite field strengths in R presenting a more realistic description of the situation actually met in experiments.[10] For a thin solenoid of length L with axis along the z direction the Stokesian vector potential in cylindrical coordinates r, φ, z in the accessible region R near $z = L/2$ is given by $A_0 = A_r = A_z = 0$ and

$$A_\varphi = \phi[|1 + r^2/(L-z)^2|^{-1/2} + |1 + r^2/z^2|^{-1/2}]/4\pi r \,, \tag{8}$$

where ϕ is the flux contained in the solenoid at $z = L/2$. In all previous treatments, the limiting form of (8) for $L \to \infty$ (set $z = L/2$),

$$A_\varphi = \frac{\phi}{2\pi r} \,, \tag{9}$$

has been used. Apart from the Lorentz force contained in (8) there is a crucial difference between (8) and (9). According to our proposition, (9) cannot—and in fact does not, as Roy shows—obey condition (5) (since it yields $\vec{B} = 0$ in R), while (8) does, if a proper path, e.g.,[3]

$$z_0 = x_0, \quad \vec{z} = \vec{x}(1 - \xi) \,, \tag{10}$$

is chosen.

Roy's criticism of the common interpretation of the AB experiments concentrates on the use of (9) which is considered to be an inadmissible oversimplification insofar as it does not allow the above transformation. He points out that for experiments performed up to now, where, in fact, straight solenoids of finite length have been employed, a potential A'_μ which is constructed according to Eq. (3) with a path given by (10) and fields determined by (8), gives a valid description of the physical situation. Finally, his doubts as to the reality of the AB effect stem from the fact that this A'_μ is completely determined by accessible fields.

We are going to investigate if this property implies a breakdown or an essential modification of the long-accepted AB interpretation of the experimental results. Obviously, in quantum mechanics electromagnetic fields may have an influence on charged particles very different from that of the corresponding classical forces. The excitement aroused by the claim of Aharonov and Bohm rather stems from the fact that there may be effects of fields even for particles experiencing nowhere the slightest classical force. In contrast to the idealized situation discussed by AB in actual experimental arrangements the electron beams were always subject to a small Lorentz force. The data, however, showed good agreement with the predictions of Aharonov and Bohm and seemed to indicate that the accessible field need not be taken into consideration since the effect of the Lorentz force is seen to be negligibly small.[11] In this sense the results have been accepted as confirmation of the Aharonov and Bohm prediction even if the ideal conditions were not fulfilled and are, in fact, not realizable for straight solenoids.

The authors[2] who deny the reality of the AB effect in the pure case try to explain the observed interference fringe shifts as due to Lorentz forces. According to their point of view, no AB shift will be seen for nearly ideal (L very long) conditions. Like these authors, Roy considers the accessible field as being responsible for the observed phenomena but, in contrast to them, does not offer an explanation in terms of classical forces. His justification for using the phrase "nonexistence of AB effect" stems from the fact that observable results may be logically deduced from values of accessible fields for any finite solenoid, no matter how long. Although his treatment partly suggests that there would be no AB shift if the accessible field were exactly zero, he introduces, in fact, a different meaning of this phrase rather than giving a logically independent proof of the nonexistence of the AB effect as asserted by other authors.[2]

We doubt, however, that Roy's findings imply the nonexistence of the AB effect even in a restricted sense of the word. Of course, we continue to associate with this phrase the absence of any inaccessible field effect. Let us recall that the path integral $\oint A'_\mu dx_\mu$ which is responsible for the AB shift is invariant under the transformation (1) and equal to the enclosed flux which stems mainly from fields in inaccessible regions. One cannot conclude that no effect of the inaccessible field has been seen, since the relevant path integral of the four potential (3) determined by the accessible field is itself almost entirely determined by the inaccessible field. Thus, there is a correlation between accessible and inaccessible fields which holds until the field in R approaches zero and condition (5) breaks down. Then, there is no longer the possibility of referring to the accessible field since the above proposition prevents the use of DeWitt's gauge for paths lying entirely in R and states that the data can only be explained in terms of the inaccessible field. The quantity entering the physical results is again the enclosed flux $\oint A_\mu dx_\mu$.

If the principle of localizability of physical effects is assumed to hold in quantum mechanics then the electromagnetic potentials must be interpreted as causal agents of the AB effects. While Roy's treatment shows that such a special significance of the vector potential[4,12] cannot be asserted for solenoids of finite length, this significance remains valid in the idealized situation considered by AB. Note also that there are no limitations in principle which would prevent the realizability of the condition: $\vec{B}=0$ in R.

We have shown that the original predictions of Aharonov and Bohm are fully confirmed in the framework of DeWitt's line-dependent gauge, provided Roy's single-valuedness condition for the wave function is taken into consideration. The fact that for experiments on solenoids of finite length the physical results are completely determined by accessible fields should not be termed as "nonexistence of AB effect" since these results may equally well be deduced from inaccessible fields. Our analysis supports the common belief that these experiments have indeed shown the reality of the quantum effects of the enclosed flux in accordance with the ideas of Aharonov and Bohm.

---

[1]Y. Aharonov and D. Bohm, Phys. Rev. 115, 485 (1959); W. Ehrenburg and R. E. Siday, Proc. Phys. Soc. London B62, 8 (1949).

[2]P. Bocchieri and A. Loinger, Nuovo Cimento 47A, 475 (1978); P. Bocchieri, A. Loinger, and G. Siragusa, *ibid.* 51A, 1 (1979).

[3]S. M. Roy, Phys. Rev. Lett. 44, 111 (1980).

[4]B. S. DeWitt, Phys. Rev. 125, 2189 (1962).

[5]Condition (5) also excludes transformations where χ is multivalued in exactly such a way as to make the wave function single valued. The significance of these transformations has been discussed by T. T. Wu and C. N. Yang, Phys. Rev. D 12, 3845 (1975).

[6]M. Wadati, H. Matsumoto, and H. Umezawa, Phys. Rev. D 18, 520 (1978).

[7]C. G. Bollini and J. J. Giambiagi, Nucl. Phys. B123, 311 (1977).

[8]For the first example of Aharonov and Bohm (see Ref. 1) the region R consists of (1) all space for $t < t_0$, (2) the interior of the two Faraday cages for $t_0 < t < t_1$ while potentials are turned on, and (3) all space for $t > t_1$.

[9]U. Klein, Lett. Nuovo Cimento 25, 33 (1979); A. Zeilinger, *ibid.* 25, 333 (1979); D. Bohm and B. J. Hiley, Nuovo Cimento 52A, 295 (1979); J. A. Mignaco and C. A. Novaes, Lett. Nuovo Cimento 26, 453 (1979).

[10]R. G. Chambers, Phys. Rev. Lett. 5, 3 (1960); H. Boersch *et al.*, Z. Phys. 159, 397 (1960); G. Möllenstedt and W. Bayh, Naturwissenschaften 48, 400 (1961); W. Bayh, Z. Phys. 169, 492 (1962); R. C. Jaklevic *et al.*, Phys. Rev. Lett. 12, 274 (1964); Phys. Rev. 140, A 1628 (1965); G. Matteucci and G. Pozzi, Am. J. Phys. 46, 619 (1978).

[11]Interference fringe shifts due to Lorentz forces will show very distinct features as compared to those caused by quantum effects of the enclosed flux. See T. H. Boyer, Phys. Rev. D 8, 1679 (1973); D. H. Kobe, Ann. Phys. (N. Y.) 123, 381 (1979).

[12]Y. Aharonov and D. Bohm, Phys. Rev. 125, 2192 (1962); 130, 1625 (1963); A. I. Vainshtein and V. V. Sokolov, Yad. Fiz. 22, 618 (1975) [Sov. J. Nucl. Phys. 22, 319 (1976)].