# Further Considerations on Electromagnetic Potentials in the Quantum Theory*

Y. AHARONOV
*Department of Physics, Brandeis University, Waltham, Massachusetts*

AND

D. BOHM
*H. H. Wills Physics Laboratory, University of Bristol, Bristol, England*

(Received April 6, 1961)

In this article, we discuss in further detail the significance of potentials in the quantum theory, and in so doing, we answer a number of arguments that have been raised against the conclusions of our first paper on the same subject. We then proceed to extend our treatment to include the sources of potentials quantum-mechanically, and we show that when this is done, the same results are obtained as those of our first paper, in which the potential was taken to be a specified function of space and time. In this way, we not only answer certain additional criticisms that have been made of the original treatment, but we also bring out more clearly the importance of the potential in the expression of the local character of the interaction of charged particles and the electromagnetic field.

## 1. INTRODUCTION

IN a previous paper,[1,2] we have given several examples showing that in the quantum theory, electromagnetic potentials have a further kind of significance that they do not possess in classical theory; viz., in certain kinds of multiply-connected field free regions of space and time, the results of interference and scattering experiments depend on integrals of the potentials, having the form

$$I=\oint A_\mu dx^\mu \tag{1}$$

(where the integration is carried out over a circuit in space and time). *This dependence is present even when the electrons are prevented by a barrier from entering the regions, in which the fields have nonzero values.* On the other hand, according to classical theory, no such dependence of physical results on the potentials is possible, if the electrons are confined to a field-free multiply-connected regions of the type described above.

Since the above paper was published, several experimental confirmations of the predicted dependence of electron interference on potentials have been obtained. First, it was shown by Werner and Brill[3] that, in order to explain the absence of fringe shifts in certain experiments that had been carried out under conditions where there were appreciable 60-cycle stray magnetic fields, one had to take into account the effects of the potentials, which just compensated those of the fields. Secondly, an experiment has been carried out by Chambers,[4] in which the flux was supplied by a very fine magnetized iron "whisker" (about 0.75 $\mu$ in diameter). An electrical bi-prism was used to separate the beam into two parts, which passed on the two sides of the whisker without contact. The resulting interference observations confirmed the existence of a fringe shift, as predicted by the theory. Thirdly, Marton and his collaborators[5] have reported an experiment similar to that of Chambers in its essential points, and they too obtained fringe shifts, as predicted. Finally, Boersch *et al.*[6] have studied the interference patterns of fast electrons passing through thin ferromagnetic layers, and have likewise confirmed that, as predicted, vector potentials have a direct effect on the fringes.

Although all of the above experiments are in agreement with the theory, none of them constitutes an ideal confirmation. For, in each case, the effect of vector potential was mixed up with that of magnetic fields, so that the theory was confirmed only insofar as it was seen that in order to account for the total effect it is necessary to take the influence of the potentials into account. The experiments with whiskers are, however, potentially capable of providing an ideal test, provided that the magnetization of the whiskers can be made sufficiently uniform, so that all stray fields in the region of the beam may be reduced to negligible values.

The existence of effects of potentials on electrons confined to field-free multiply-connected regions of space and time seems to have been regarded with surprise by some physicists. If one reflects on this problem for a while, however, one will see that there is in reality no reason whatsoever to be surprised at this possibility. For a similar effect arises in the much more common case of the stationary states of an electron in an atom. As is well known, according to classical mechanics, any orbit should be possible in such an atom. According to quantum theory, however, the energy levels are restricted in such a way that (at least in the correspondence principle limit of high

* This work was partially supported by the Office of Scientific Research, U. S. Air Force.

[1] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[2] See also, W. Ehrenburg and R. E. Siday, Proc. Phys. Soc. (London) **B62**, 8 (1949), who, on the basis of a semiclassical treatment, obtained some of our results; viz., the prediction of a fringe shift due to magnetic vector potentials in a field-free multiply-connected region.

[3] F. G. Werner and D. R. Brill, Phys. Rev. Letters **4**, 349 (1960).

[4] R. G. Chambers, Phys. Rev. Letters **5**, 3 (1960).

[5] L. Marton *et al.* (private communication).

[6] H. Boersch, H. Hamisch, D. Wohlleben, and K. Grohmann, Z. Phys. **159**, 397 (1960).

quantum numbers) the Bohr-Sommerfeld rule holds, viz.,

$$\oint \mathbf{p}\cdot d\mathbf{q}=nh, \tag{2}$$

where the integral is taken around a closed orbit and $n$ is an integer. If we were to take the classical point of view and to require that any such restriction be explained by a force, then we would be presented with an incomprehensible problem. For the forces known to be present in an atom simply would not constrain an electron, moving in a certain position in its orbit at a given moment of time, to one of a set of possible orbits that depends on an integral of its momentum over the entire orbit in question. If we note, however, that the electron also has a wavelike aspect, the reason for this constraint is quite evident, since an integral number of waves must fit in a circuit (or, in other words, the wave function must be single valued). And, of course, it is this requirement that is really at the root of the Bohr-Sommerfeld condition.[7]

If then, we are ready to accept the fact that there exist characteristically quantum-mechanical phenomena such as discrete energy levels (as well as interference and diffractive scattering), we are, in effect, admitting that the concept of force is not adequate for treating the basic properties of an atom. It is not very much of a further step to add that the concept of force is also not adequate for treating electromagnetic interactions. Or to put the same argument in more precise terms, we note that as the behavior of the electron depends on integrals of the action, $\oint \mathbf{p}\cdot d\mathbf{q}$, in a way that would not occur according to classical mechanics, so it depends on integrals of potential, which would likewise have no such implications classically. And indeed, in both cases, this dependence has basically the same origin, viz., the quantum conditions as given in terms of the "canonical" momentum, $\mathbf{p}=m\mathbf{v}+(e/c)\mathbf{A}(\mathbf{x})$, which are (in the correspondence limit of high quantum numbers)

$$\oint \mathbf{p}\cdot d\mathbf{q}=\oint [m\mathbf{v}+(e/c)\mathbf{A}(\mathbf{x})]\cdot d\mathbf{x}. \tag{4}$$

We see then that the integral of potential, $\oint \mathbf{A}\cdot d\mathbf{x}$, plays a part in the quantum condition, which supplements the corresponding integral, $\oint m\mathbf{v}\cdot d\mathbf{x}$ of the "physical" momentum, $m\mathbf{v}$. This means that the very existence of quantum conditions demands that potential integrals, $\oint \mathbf{A}\cdot d\mathbf{x}$, must have a physical significance which they do not have in classical mechanics. {For example, they influence the eigenvalues of the Hamiltonian which contains the kinetic energy operator $T=\frac{1}{2}mv^2=[\mathbf{p}-(e/c)\mathbf{A}]^2/2m$}.

The notion of force has in the quantum theory at best a very indirect meaning. Thus, one can define *an average force*, $\mathbf{F}$, by means of Ehrenfert's theorem[8]:

$$\frac{dm\bar{\mathbf{v}}}{dt}=\mathbf{F}=\int \psi^*[-\nabla V+e\nabla\phi-(e/c)(\mathbf{v}\times\mathfrak{H})]\psi d\mathbf{x}, \tag{5}$$

where $\psi$ is the electron wave function, $V$ is the non-electrostatic part of the potential, $\mathcal{E}=-\nabla\phi$ is the electric field, $\mathfrak{H}=\nabla\times\mathbf{A}$ is the magnetic field, and $\mathbf{v}$ is the velocity operator of the electron. In spite of the formal similarity of the above equation to the classical Lorentz equation, $dm\mathbf{v}/dt=-\nabla V-e\mathcal{E}-(e/c)(\mathbf{v}\times\mathfrak{H})$, nevertheless there is a very fundamental difference in its physical significance. For all the quantities entering into the classical equation ($\mathcal{E}$, $\mathfrak{H}$, $\mathbf{v}$, etc.) can be determined experimentally and are defined mathematically in a way that does not require the introduction of the potential (which is, in fact, only a mathematically convenient procedure in classical theory). In the quantum theory, however, the average force depends, as can be seen from Eq. (5), on the precise form of the wave function which, in general, can be obtained only by solving Schrödinger's equation (e.g., if one is given the fact that the system is in a certain stationary state of energy, $E$, one must solve for the corresponding eigenfunction, $\psi_E(\mathbf{x})$, of the Hamiltonian operator). Now it is well-known that the potentials must appear in Schrödinger's equation, because there is no way in quantum mechanics to express the interaction of the electron with the electromagnetic field solely in terms of field quantities. *The wave function entering into Eq. (5) for the average force therefore cannot, in general, be known unless one first knows the potentials.* Thus, in quantum mechanics, force is an extremely abstract concept, having at best a highly indirect significance, which is of only secondary importance.[9]

It is clear that at least in the mathematical theory of the quantum mechanics, the electromagnetic potentials (and not the fields) are what play a fundamental role in the expression of the laws of physics. Nevertheless, it seems that physicists have generally been reluctant to accept the notion that potentials also have a more fundamental physical significance than that of the fields. This reluctance is grounded in part on a tendency to regard force as a fundamental concept in quantum theory, a tendency that, as we have seen, cannot be justified. It is also grounded in part, however, on the invariance of all physical quantities to gauge transformation, $A_\mu=A_\mu'+\partial f/\partial x^\mu$, where $f$ is an arbitrary continuous and single-valued scalar function. This invariance implies that even when the physical state of the system is completely specified, the potentials are still arbitrary to within such a gauge transformation. It is therefore argued that the potentials do not have a

[7] D. Bohm, *Quantum Theory* (Prentice-Hall, Englewood Cliffs, New Jersey, 1951), see Chap. 2.

[8] See, for example, reference 7, Chapter 10.

[9] In the Appendix, we shall discuss a particular example, showing, in more detail, the comparatively indirect and secondary significance of force in the quantum theory.

direct physical significance, but that they are significant only insofar as they determine the field quantities, $F_{\mu\nu}=\partial A_\mu/\partial x^\nu-\partial A_\nu/\partial x^\mu$; which latter are invariant to such a transformation.

Although we must accept that gauge invariance implies that the value of a potential at a given point has by itself no direct physical significance, it does not necessarily follow that the physical significance of the potentials is always exhausted by that of the fields which they define. For, as we have seen, it is possible to confine the electron to multiply connected regions of space by means of suitable potential barriers, and the behavior of electrons thus confined depends on integrals of the potential $\oint A_\mu dx^\mu$, which are physically significant even when all fields in the regions in question vanish. These integrals are gauge invariant so that they are not subject to the arbitrariness, in relation to the physical state, which the potentials themselves have.

It is true, of course, that *in a simply connected region* the integral $\oint A_\mu dx^\mu$ is identically equal to the integral $\int F_{\mu\nu}dS^{\mu\nu}$ of the field quantities $F_{\mu\nu}$, taken over the surface (whose elements are $dS^{\mu\nu}$), which the circuit of the potential integrals encloses. One might therefore be led to conclude that there is no additional physical content in the potential integrals that is not already in the field variables. However, we must keep in mind that the quantum theory as it is now formulated requires that the interaction of electron with electromagnetic field must be a *local* one (i.e., the field can operate only where the charge is). Therefore, in the description of this interaction, only those quantities which differ from zero in the region accessible to the electron can account for observable physical effects on the electron. As a result, when the electron is thus confined to a multiply connected region, the fields in the excluded region (which appear in the above identity, between field and potential integrals) cease to be relevant for the problem under discussion. The observable physical effects in question must therefore be attributed to the potential integrals themselves. Such integrals, being not only gauge invariant, but also Hermitian operators, are perfectly legitimate examples of quantum-mechanical observables. They represent extended (nonlocal) properties of the field, which are evidently directly measurable in the region in question with the aid of the observable properties (interference, diffractive scattering, and energy levels) of electrons confined to this region.

Although the above point of view concerning potentials seems to be called for in the quantum theory of the electromagnetic field, it must be admitted that it is rather unfamiliar. Various of its aspects are often, therefore, not very clearly understood, and as a result, a great many objections have been raised against it (some of them in the published literature, and some of them in private communications to the authors). Such objections have appeared so frequently that we feel that it would be useful to answer them systematically, and in so doing, to present certain further developments concerning the theory of the effects of potentials in quantum mechanics.

The objections mentioned above fall roughly into two types. In the first type, it is accepted that the potentials can be expressed in the one-body Schrödinger's equation as definite functions of space and time, as we did in our first article. On this foundation, however, various points are raised which call some of our conclusions into question. These points will be discussed and answered in Sec. 2 of the present paper.

In the second type of objection, it is not accepted that the potentials can be written as specified functions of time and space, but instead, questions are raised which would suggest that there would be a breakdown of some of our conclusions if we took into account the distribution of charges and currents (e.g., in a solenoid) which are the sources of potentials. To discuss these questions, we begin in Sec. 3 by giving a theory, which treats the source of the electric potential by means of a many-body Schrödinger's equation, and in Sec. 4, the same is done for magnetic potentials. In all cases, we show that the results are precisely the same as those given in our first paper.

In our detailed treatment of the effect of the sources of the potentials, the fact that potentials possess a physical meaning beyond that of the fields emerges with even greater clarity than before. Thus it will be shown in Sec. 3, that whereas the electron does actually exert force on the various parts of the source in experiments of the type that we have described, the total force of reaction of the source back on the electron vanishes. Nevertheless, the electronic interference effects remain, thus confirming our conclusion that in the quantum theory, force does not have the fundamental role that it has in classical physics. In Sec. 4, where we treat the electromagnetic field quantities as dynamical variables, it will be seen that the potentials constitute an intermediary link between the electron and the charges and currents in the source variable. As in the one-body theory (in which the variables of the electromagnetic field are taken as specified functions of space and time), it is only with the aid of the potentials that this link can be established by means of a localized interaction between charged particles and field (the field quantities themselves being, in general, inadequate for this purpose). Thus, we demonstrate the fundamental role of potentials for this problem in still another way.

Finally, in an Appendix, we shall discuss a recent article by Peshkin, Talmi, and Tassie[10] on the subject of potentials in the quantum theory.

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

## 2. FURTHER CLARIFICATION OF EFFECTS OF POTENTIAL IN ONE-BODY SCHRÖDINGER'S EQUATION

In this section, we shall attempt mainly to clarify various questions that have been raised concerning the effects of potentials in the one-body Schrödinger's equation.

First of all, Furry and Ramsey[11] have discussed the relationship of the uncertainty principle to interference experiments such as those suggested in our first article (e.g., an electron beam is split coherently in two, each is allowed to pass through tubes in which there is a different time dependent potential, after which the beams are allowed to come together and interfere). Although the above authors did not intend their article to be regarded as an objection to our conclusions, it seems that it has been so regarded by a number of physicists.[12] It is therefore worthwhile here to make a few remarks about this point.

Now, as long as no observation can be made from which one could tell through which beam the electron actually passes, then there will, of course, be interference as predicted in our paper. If, for any reason, however, an observation as to which beam the electron actually passes through can be carried out, then as is well known, the apparatus that makes this observation possible must introduce a disturbance that destroys the interference pattern.[13] Furry and Ramsey treated this point in some detail, considering a special example of a measuring device (a charge), and showing that as a result of its effects, interference will be destroyed, as is to be expected. Of course, this demonstration does not invalidate our conclusions in any way whatsoever, since by hypothesis, we are considering a case in which the experiment is done under conditions in which no such detailed observation of the path of the electron can be made.

The above discussion indirectly answers a large number of further objections of a certain general type to our conclusions. For example, if the electron passes through one of the condensers, then when that condenser is charged up, the amount of work done by the charging generator will be different from what it would be in the absence of this electron. By measuring this work, one could, in principle, tell which beam the electron passed through, so that interference would be impossible. In accordance with the preceding discussion, however, it is clear that in order to be sure that interference will take place, it is necessary to arrange conditions so that no such measurement can be carried out. We shall treat this problem in more detail in Sec. 3, where we shall show that if the generator is properly constructed (so that its behavior is adiabatic), then no energy measurement permitting us to tell which beam the electron passed through will be possible, and the usual interference pattern will be obtained.

The second general kind of question that has been raised concerns the problem of the single-valuedness of the wave function. In connection with this problem, the magnetic example given in our first paper (a very narrow solenoid with flux inside but no flux outside) is the easiest to discuss, although the conclusions that we shall give here are true in general.

Since the wave function is being solved in a non-simply-connected region (which excludes the solenoid), it is argued that the usual considerations leading to the requirement of the single valuedness of the wave function may not be valid here. For example, if the wave function were to be multiplied by a constant factor, $e^{i\alpha}$, when the polar angle is increased by $2\pi$, then all physical predictions, which depend only on functions like $\psi^* O\psi$ (where $O$ is a Hermitean operator), will still be single-valued.[14] It is proposed then that for this case, the boundary conditions on the wave function might be altered. For example, the vector potential in this case can be chosen (in a certain gauge) as

$$\mathbf{A}=q\mathbf{u}_\phi/2\pi r, \tag{6}$$

where $q$ is the total flux inside the solenoid, and $\mathbf{u}_\phi$ is a unit vector perpendicular to the radius. Then by a certain gauge transformation which is regular in the multiply-connected region under discussion and singular only in the excluded region, viz., $\mathbf{A}\rightarrow\mathbf{A}'-(q/2\pi)\nabla\phi$ (where $\phi$ is the polar angle), one can eliminate the vector potential altogether, reducing the Hamiltonian to that of a free particle. If now we regard $\psi'$ as a proper representation of the wave function, we would obtain solutions corresponding to a free particle. Since such solutions are single valued in the $\psi'$ representation, they would have to be multiple valued in the original representation ($\psi$ would be multiplied by $e^{ieq}/c$ when $\phi$ was increased by $2\pi$). If such a procedure were legitimate, then all effects of potentials in field-free multiply connected regions could be transformed away, and the conclusions of our previous paper would be invalidated.

It is easy to see, however, that such non-single-valued wave functions in the original representation are not compatible with the basic principles of quantum mechanics. For they do not take into account the fact that the magnetic flux can be turned off adiabatically and that any potential barriers that surround this flux can, in principle, likewise be decreased adiabatically to zero. From the fact that the Hamiltonian is always a single-valued operator (even when it is thus changing in time), it is easy to show that if the wave function is initially single-valued, it remains single-valued for all

[11] W. Furry and N. Ramsey, Phys. Rev. **118**, 623 (1960).
[12] Private communications.
[13] See, for example, reference (7), Chapter 6, where it is shown that this behavior is, in fact, quite general, and not just restricted to the experiment under discussion.
[14] The requirement that $\psi$ itself be single-valued stems basically from the demand for three-dimensional invariance (see, for example, reference 7, Chapters 14 and 17). If the region is not simply connected, we cease to require this invariance.

time, while if it is originally multiple-valued, it retains the same kind of multiple-valuedness. Hence, if (in the original representation) multiple-valued wave functions were allowed while the barrier was present and the flux was turned on, they would also have to be allowed when the barrier had disappeared and the flux was turned off. This would evidently lead to new quantum conditions on the particle, which depended on its past history (i.e., as to whether it had once been in a multiply connected space with flux in the excluded region). But it is a basic postulate of the quantum theory that the quantum states of a given system allowed in a specified physical situation are independent of the past history of that system (i.e., of how the state was prepared). Therefore, it is not possible to transform away the effects of potentials in field-free multiply-connected regions by giving up the condition of single-valuedness of the wave function.

A third type of question that has been raised is concerned with the electric field which arises when source of the magnetic vector potential (e.g., the current in the solenoid) is turned on. To treat this problem, let us suppose, for example, that in the absence of the flux there is an electron in a stationary orbit going around the solenoid. In the correspondence limit its angular momentum is determined by the Bohr-Sommerfeld quantum conditions [Eq. (2)]. When the flux is turned on, the resulting electric field acting on the electron will alter its "physical" momentum, $m\mathbf{v}$, and as a simple calculation shows, this alteration is equal in magnitude to $(e/c)\mathbf{A}$, where $\mathbf{A}$ is the final vector potential due to the source.

At first sight, one might then suppose that the effects of a vector potential have been explained, as the result of the action of an electric field, by means of the above argument. However, it is possible to begin the experiment with the electron screened by a Faraday cage, so that it experiences no electric field whatsoever. If the electron is subsequently released from the cage and then captured into a stationary orbit, the quantum conditions will be precisely the same as those which would be operative if the electron had initially been in this orbit [and are, in fact, given by Eq. (4), in terms of the vector potential]. This is just a special case of the general rule of the quantum theory that we have cited in connection with the problem of the single valuedness of the wave function; viz., that the possible quantum states are independent of the past history of the system. It is therefore clear that the change of quantum state cannot, in general, be ascribed in this way to the action of electromagnetic fields on the electron.

A similar question has been raised by Pryce,[15] except that he has discussed the problem of the shift of interference fringes, and has tried to explain them as due to the static magnetic field. This explanation starts from the circumstance that in the Chambers experiment,[4] the flux in the whisker varied somewhat in its longitudinal direction (which we shall call $z$). As a result, the displacement of the fringes was a function of $z$, so that the fringes consisted of tilted, (and in general, curved) lines. The $z$ dependence of the flux implies, of course, that there is a magnetic field outside the solenoid. If one assumes that the flux, $q(z)$, does not vary too rapidly as a function of $z$, the vector potential $\mathbf{A}=q(z)\mathbf{u}_\phi/2\pi r$ will still be the correct solution of Maxwell's equations, to a good order of approximation. From this, one can calculate the magnetic field, $\mathfrak{IC}=\nabla\times\mathbf{A}=\hat{r}(dq/dz)/2\pi r$, where $\hat{r}$ is a unit vector in the radial direction. This field implies a force on the electron, $\mathbf{F}=(e/c)(\mathbf{v}\times\mathfrak{IC})=(e/c)(\mathbf{v}\times\hat{r})(dq/dz)/2\pi r$, which is in the $z$ direction. There will be a resulting momentum transfer to the electron of $\Delta\mathbf{p}=\int\mathbf{F}dt=(e/c)\int(\mathbf{v}\times\mathfrak{IC})dt$, where the integration is carried out over the path of the electron. (This transfer will be oppositely directed in accordance with the side of the solenoid on which the electron passes.)

Pryce then noted that the above-described momentum transfer can be used to calculate the slope of the fringe. Evidently this slope is determined by $\partial\Phi/\partial z$, where $\Phi$ is the phase difference of the beams which have passed on opposite sides of the solenoid. Now, as we saw in the discussion of the Bohr-Sommerfeld quantum conditions, this phase shift (the number of wavelengths) is equal to $\oint\mathbf{p}\cdot d\mathbf{x}/h$, where $\mathbf{p}$ is the canonical momentum, $m\mathbf{v}+(e/c)\mathbf{A}$. Now, consider the dependence of this phase shift on $z$, *at the location of the screen, where interference is being detected.* Here $\mathbf{A}$ can be neglected (because $\mathbf{r}$ is large). The $z$ dependence of $\Phi$ will then arise only because the two beams have different $z$ components of the "physical" momentum, $m\mathbf{v}$. From this difference, as calculated in the previous paragraph, one obtains

$$\frac{\partial\Phi}{\partial z}=\frac{e}{ch}\oint\hat{z}\cdot(\mathbf{v}\times\mathfrak{IC})dt, \tag{7}$$

where $\hat{z}$ is a unit vector in the $z$ direction.

We see then that the *slope* of the fringe line can be obtained from the momentum transferred to the electron by the magnetic force due to the stray field outside the solenoid. However, Pryce then went on to consider a case in which the flux in the "whisker" has a value of zero at some point, say $z=z_0$, and in which $q(z)$ rises continuously to its actual value at the altitude $z$. The total phase shift can then be obtained by integrating Eq. (7) from $z_0$ to $z$; viz., $\Phi=\int_{z_0}^{z}(\partial\Phi/\partial z)dz$. Before doing this, however, we first transform the integral in Eq. (7) into

$$\frac{e}{ch}\oint(\mathfrak{IC}\times\hat{z})\cdot\mathbf{v}dt=\frac{e}{ch}\oint(\mathfrak{IC}\times\hat{z})\cdot d\mathbf{x}.$$

We then note from our expression for $\mathfrak{IC}$ that $\mathfrak{IC}\times\hat{z}$

[15] Pryce's arguments have been discussed in reference 4.

$=\partial\mathbf{A}/\partial z$, so that we finally obtain

$$\Phi=\frac{e}{ch}\int_{z_0}^{z}dz\oint(\mathfrak{K}\times\hat{z})\cdot d\mathbf{x}$$

$$=\frac{e}{ch}\int_{z_0}^{z}dz\oint\frac{\partial\mathbf{A}}{\partial z}\cdot d\mathbf{x}=\frac{e}{ch}\oint\mathbf{A}(z)\cdot d\mathbf{x}, \quad (8)$$

using the fact that $\int\mathbf{A}(z_0)\cdot d\mathbf{x}=0$, because $q(z_0)=0$. In this way, it would seem that the whole effect can be explained as a result of forces exerted by the magnetic field on the electron, so that potentials are after all not playing any more fundamental role than they play in classical physics.

One can show the inadequacy of the above argument by noting the electron beam can be limited in the $z$ direction to a region, $\Delta z$, in which the change of flux, $(\partial q/\partial z)\Delta z$, is negligible in comparison to $q(z)$ itself. Moreover, it is always possible, in principle, to find conditions in which this limitation of the beam will have a negligible effect on the interference pattern (it is necessary only that $\Delta z$ be sufficiently large in comparison to a wavelength). In practice, such a limitation could be achieved by suitable slits, but for theoretical purposes, it is more convenient to discuss infinitely high potential barriers which confine the electron to the region in question.

It is clear that the fringe line in the interference experiment described above is, in effect, a map of $\Phi(z)=\oint\mathbf{A}(z)\cdot d\mathbf{x}$ onto the coordinate $[Y(z)]$ perpendicular to $z$, on the screen. To obtain such a map empirically, one could begin by doing such an interference experiment at a certain altitude $z$, with a slit of width $\Delta z$. Then it could be done at a series of altitudes, $z+\Delta z$, $z+2\Delta z$, etc. It must be remembered, however, that in a shift of $n+\delta$ fringes (where $n$ is an integer), only $\delta$ is observable with the aid of measurements made at a definite value of $z$. Nevertheless, if $\Delta z$ is limited in the way described above (so that there is much less than a whole fringe shift in the interval $\Delta z$), then one can make an effectively continuous map, which can be extended over many fringe shifts, and which permits the integer, $n$, to be obtained by counting the fringe shifts down to a point of zero flux. In this way, the function $\Phi(z)$ could be obtained in measurements.

It is clear that the argument of Pryce applies only to the calculation of the shift of the fringe line in the interval $\Delta z$ [which is proportional to $(\partial\Phi/\partial z)\Delta z$]. The main part of the deviation, $\delta$, of the fringe shift from an integer (which can be measured directly in observations taken in the interval $\Delta z$) is, however, not explained by this argument at all, since the electron cannot penetrate into the regions over which the integration of $\partial\Phi/\partial z$ [in Eq. (8)] was carried out. This deviation, $\delta$, is, in fact, determined directly by the potential integrals, $\oint\mathbf{A}\cdot d\mathbf{x}$, while (as has already been pointed out in Sec. 1), the concept of the force exerted by the fields acting on the charges is seen to be a comparatively abstract one, having at best, a secondary importance in comparison with that of the potential integrals themselves.

Finally, another point of interest that has been raised is in connection with the possibility that in the experiments described here, the flux is actually quantized in units of "fluxons" (1 fluxon$=ch/e$). In those cases where stray fields are present (e.g., in the Chambers experiment[4]) such a suggestion implies that there is, in reality, always an integral number of fluxons at any given altitude, $z$, and that this number changes abruptly at certain altitudes. The stray field would then be present only at these altitudes where the number of fluxons suffers an abrupt change of the type described above.

Of course, as pointed out in our first article, all interference and scattering experiments must vanish in a field-free multiply-connected region containing an integral number of fluxons. If the flux were quantized, one would first sight, therefore, expect no observable fringe shifts except at those altitudes, $z$, where the flux changes abruptly, (and where the resulting magnetic field might perhaps be expected to deviate the fringe pattern in the manner indicated by the argument given by Pryce). Such a discontinuous pattern would contradict the observed results which, as we have already pointed out, show a continuous tilted and, in general, curved fringe line. In order to answer this objection it could, however, further be suggested that the electron is not fully localized in the $z$ direction, so that it effectively experiences a magnetic field averaged over a certain range of $z$. In this way, one could perhaps hope to explain the observed continuity of the fringe lines, while still holding on to the notion that the flux is quantized, and that all observable effects are really due to the fields.

In order to settle this question of quantization of flux finally, it would suffice if an experiment were done in which $\Delta z$ were small enough so that the fringe shift along its length would be negligible in comparison to the deviation, $\delta$, from an integral number of fringe shifts. In this way, one could demonstrate that the average field experienced by the electron (which is proportional to the slope of the fringe line) is too small to account for the observed fringe shift, $\delta$, so that the assumption of quantized flux lines with discrete changes in intensity would have to be given up.

Finally, it must be pointed out that the quantization of flux is not compatible with the quantum theory of the electromagnetic field as it stands now. Some arguments have been given with aim of deducing the quantization of flux from the present theory, but these arguments are erroneous. For example, consider an electron moving in a uniform magnetic field of strength, $H$, in the $z$ direction. The vector potential can be taken as $A=-Hr\mathbf{u}_\phi/2$. The Bohr-Sommerfeld condition is $\oint[m\mathbf{v}+(e/c)\mathbf{A}]\cdot d\mathbf{x}=nh$. But for a circular orbit in a uniform field, $m\mathbf{v}=(Her/c)\mathbf{u}_\phi$. We thus obtain $(\pi er^2/c)H=nh$, so that

the flux is $q=\pi r^2H=nch/e$, which is just a whole number of "fluxons." If flux were always confined by charges moving in uniform magnetic fields, then the effects of potentials would vanish in the way described above. In general, however, the electrons are confined to a given region by other kinds of forces (e.g., electric) so that the above conclusion of quantized flux does not hold.

Additional arguments in favor of the assumption of quantized flux arise in the theory of superconductivity.[16] Even if these arguments are accepted, however, they would imply at most that flux was quantized for superconductors, and therefore would not hold for the experiments that we have considered.

Of course, there is a possibility that current electromagnetic field theories should be modified in such a way as to introduce quantization of flux as a general property of the field. In this connection, the experiments that we have cited furnish strong evidence against such an assumption. However, in order that this evidence shall be made conclusive, it is desirable that the experiment be done with a suitably limited slit width, $\Delta z$, in the manner described earlier (so that the fringe shift along $\Delta z$ should be much less than the main deviation, $\delta$, from an integer).

### 3. EFFECT OF SOURCES OF POTENTIAL IN INTERFERENCE EXPERIMENTS (CASE OF AN ELECTRICAL POTENTIAL)

We have thus far been treating the interference experiment under the assumption that the one-body Schrödinger equation, with the potentials given as specified functions of space and time, is adequate. We shall now show that the same results are obtained, when we take into account quantum-mechanically the fact that all potentials originate in some kind of source (or set of sources). In this section we shall discuss only the case of electrical potentials (i.e., no magnetic fields), for which the problem is simplified by the fact that these potentials satisfy Poisson's equation (in the gauge in which $\mathrm{div}\mathbf{A}=0$).

$$\nabla^2\phi(\mathbf{x})=-4\pi\rho(\mathbf{x}), \tag{9}$$

so that

$$\phi(\mathbf{x})=\int\frac{\rho(\mathbf{x}')d\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|}. \tag{10}$$

The above equation shows that the value of $\phi(\mathbf{x})$ at a given time is determined completely by the distribution of charged particles at that same moment of time. (In the next section, we shall see that there is no analogous complete determination of magnetic vector potential by the distribution of currents.)

Our procedure will then be to include in a many-body Schrödinger equation, not only the electronic coordinates, $\mathbf{x}$, but also the coordinates, $y_i$, of the various parts of the apparatus, which are used to generate the potential. In general, the potential energy of interaction of the electron with the source will be a function, $V(\mathbf{x},\cdots y_i\cdots)$, which depends on the $y_i$ as well as on $\mathbf{x}$, because the $y_i$ determine how the various charges in the source are placed. Thus, the system will have to be described by the wave equation

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{x},\cdots y_i\cdots,t)$$
$$=[H_e+H_S+V(\mathbf{x},\cdots y_i)]\Psi(\mathbf{x},\cdots y_i\cdots,t), \tag{11}$$

where $H_e$ is the Hamiltonian of the electron by itself. The Hamiltonian of the source can be expressed in more detail as

$$H_S=\sum_i\frac{p_i^2}{2M_i}+W(\cdots y_i\cdots), \tag{12}$$

where $M_i$ is the mass associated with the $i$th coordinate and $W(\cdots y_i\cdots)$ is the potential energy of interaction of all parts of the source with each other, while $p_i$ is the momentum conjugate to $y_i$.

We now take advantage of the fact that the source consists of a macroscopic piece of apparatus. Thus, all its parts will be very heavy, so that it can be treated with the aid of the WKB approximation. As is well known,[17] the approximation leads to an expression for the wave function of the source by itself, having the form

$$\phi(\cdots y_i\cdots,t)=R(\cdots y_i\cdots,t)e^{iS(\cdots y_i\cdots,t)/\hbar}, \tag{13}$$

where $S$ is a solution of the classical Hamilton-Jacobi equation

$$\frac{\partial S}{\partial t}+\sum_i\left(\frac{\partial S}{\partial y_i}\right)^2\Big/2M_i+V(\cdots y_i\cdots)=0, \tag{14}$$

and the momenta, $p_i$, are given by

$$mv_i=\partial S/\partial y_i. \tag{15}$$

The probability density, $P=R^2$, satisfies the conservation equation

$$\frac{\partial P}{\partial t}+\sum_i\frac{\partial}{\partial y_i}\left(\frac{p_i}{M_i}P\right)=0. \tag{16}$$

In a typical state, $P(\cdots y_i\cdots,t)$ takes the form of a packet function (in the configuration space), such that the probability density is appreciable only in a small region of width $\Delta y_i$, near a point $y_i=y_i(t)$, which follows the classical orbit. In view of the fact that such a wave packet spreads and otherwise changes its shape negligibly, it can be approximated as a function $P(\cdots, y_i-y_i(t)\cdots)$, which depends only on the difference, $y_i-y_i(t)$.

[16] See F. London, *Superfluids* (John Wiley & Sons, Inc., New York, 1950–54).

[17] See, for example, reference 7, page 270.

Since $y_i$ is very close to $y_i(t)$, wherever the probability density is appreciable, the interaction potential, $V(\cdots y_i\cdots,\mathbf{x})$ can be approximated as $V(\cdots y_i(t)\cdots,\mathbf{x})$. Thus, we obtain a time-dependent potential. Our problem is then to show that the equation for the electron wave function factors out of (11), to yield the time-dependent Schrödinger equation for the one-body problem

$$i\hbar\frac{\partial\psi}{\partial t}=[H_e+V(\cdots y_i(t)\cdots,\mathbf{x})]\psi. \tag{17}$$

If we succeed in doing this, we will have shown that a complete quantum-mechanical treatment which includes the source of the potential leads (in a suitable approximation) to the same result as does the treatment given in our first paper.

Now, in a typical interference experiment (e.g., the first case treated in our previous paper,[1] with a split electron beam passing through a pair of drift tubes), the generator of the potential is so arranged that $V(\cdots y_i(t)\cdots,\mathbf{x})$ is zero before the experiment begins, then rises continuously to a maximum, finally falling again to zero when the experiment is over. Since the probability density is negligible when $y_i$ is appreciably different from $y_i(t)$, it follows that $V(\cdots y_i\cdots\mathbf{x})$ also satisfies the same conditions in the domain in which the wave function of the source is appreciable. When the experiment begins, there is therefore no interaction between the electron and the source of potential, so that we can write the solution of the wave equation for the combined system as a simple product function

$$\Psi_0=R(\cdots,y_i-y_i(t),\cdots)e^{iS(\cdots y_i\cdots,t)/\hbar}\psi_0(\mathbf{x},t), \tag{18}$$

where $\psi_0(\mathbf{x},t)$ is the initial electronic wave function (which also takes the form of a suitable packet).

After the experiment is over (and the interaction vanishes again), the wave function will, in general, take the form of a sum of products

$$\Psi=\sum_n\phi_n(\cdots y_i\cdots,t)\psi_n(\mathbf{x},t), \tag{19}$$

where the $\psi_n(\mathbf{x},t)$ represents a set of solutions of the wave equation for the electron alone, and $\phi_n(\cdots y_i\cdots,t)$ a corresponding set for the source variables. If such a sum of products is necessary, then it is clear that it will be impossible to factor out a one-body Schrödinger equation applying to the electronic variables alone. As we shall see, however, because the parts of the source are so heavy, only a single such product is actually needed.

To treat this problem, let us first tentatively write the solution as

$$\Psi=R(\cdots,y_i-y_i(t),\cdots)\times e^{iS(\cdots y_i\cdots,t)/\hbar}\psi(\mathbf{x},\cdots y_i\cdots,t). \tag{20}$$

If we substitute this function in Schrödinger's Eq. (11), we obtain

$$i\hbar\frac{\partial\psi}{\partial t}=\Big[H_e+V(\cdots y_i\cdots,\mathbf{x})-\sum_i\frac{\hbar}{M_i}\Big(i\frac{\partial S}{\partial y_i}+\hbar\frac{\partial\ln R}{\partial y_i}\Big)\frac{\partial}{\partial y_i}-\frac{\hbar^2}{2}\sum_i\frac{1}{M_i}\frac{\partial^2}{\partial y_i^2}\Big]\psi. \tag{21}$$

We now apply the adiabatic approximation,[18] which is based on the large value of the $M_i$ plus the fact that $V(\cdots y_i\cdots,\mathbf{x})$ and $R(\cdots y_i-y_i(t)\cdots)$ are fairly regular and slowly varying functions of the $y_i$. Because these conditions are satisfied, we can neglect $\hbar\partial\ln R/\partial y_i$, and the term on the right-hand side of (21) involving $(1/M_1)\partial^2/\partial y_i^2$ in comparison to the term containing $\partial S/\partial y_i$. This leaves us with

$$i\hbar\frac{\partial\psi}{\partial t}=\Big[H_e+V(\mathbf{x},\cdots y_i\cdots)-i\hbar\sum_i\bar{v}_i(t)\frac{\partial}{\partial y_i}\Big]\psi. \tag{22}$$

(Note that we have also replaced

$$v_i=p_i/M_i=(1/M_i)\partial S/\partial x_i,$$

by its average, $\bar{v}_i(t)$, because the probability of an appreciable difference between $v_i$ and $\bar{v}_i(t)$ is negligible.) We then make the substitution

$$y_i=y_i(t)+u_i;\quad\psi(\mathbf{x},\cdots y_i\cdots,t)=\psi'(\mathbf{x},\cdots,y_i(t)+u_i,t).$$

Equation (22) becomes

$$i\hbar\frac{\partial\psi'}{\partial t}=[H_e+V(\mathbf{x},\cdots,y_i(t)+u_i)]\psi'. \tag{23}$$

The complete wave function is obtained by multiplying $\psi'$ with

$$\phi(\mathbf{x},\cdots y_i\cdots,t)=R(\cdots u_i\cdots)\times\exp[iS(\cdots,y_i(t)+u_i\cdots,t)/\hbar].$$

This is

$$\Psi=R(\cdots u_i\cdots)\exp[iS(\cdots,y_i(t)+u_i,\cdots,t)]\times\psi'(\mathbf{x},\cdots u_i\cdots,t). \tag{24}$$

Since Eq. (22) does not contain derivatives of $u_i$, these variables can be set equal to any specified set of values. But from (24), we see that the wave function is appreciable, only in a small range, near $u_i=0$. Thus, to a good approximation, we can write $\psi'(\mathbf{x},\cdots u_i\cdots,t)=\psi(\mathbf{x},t)$, where $\psi(\mathbf{x},t)$ is the value of $\psi'$ when all the $u_i$ are set equal to zero. Equation (24) becomes (after setting $u_i=0$)

$$i\hbar\frac{\partial\psi}{\partial t}=[H_e+V(\mathbf{x},\cdots y_i(t)\cdots)]\psi, \tag{25}$$

which is just the one-body Schrödinger equation with the appropriate time-dependent potential [i.e., the same

[18] See reference 7, Chapter 12, for more details.

as Eq. (17)]. We have thus accomplished our objective of showing that when the source of potential is taken into account quantum-mechanically, we obtain the same result as that given in our first paper, where the potential was assumed to be a specified function of space and time.

We can now obtain directly from the above treatment the same conclusion that was drawn by Furry and Ramsey[11] in terms of an illustrative example of a measurement process (see Sec. 2). Thus, if, in the case of the drift tube experiment, the interaction with the source were such as to make a measurement of which tube the electron actually passed through possible, then after this interaction is over, the wave function would take the form of a sum of products, in which the electron wave function is correlated to the wave function of the apparatus.[19] Such a sum would be a special case of the expansion given in Eq. (19). As is well known, however, *when the wave function of the combined system takes the form of a sum of products of the kind described above, then there is no interference between the different parts of the electronic wave function.* In order that there shall be such interference, it is necessary that the wave function of the combined system take the form of a simple product (18). But if this happens, then by observing the apparatus, we will obtain no further information about the electron. Therefore, the adiabaticity of the interaction, which guarantees that the source shall not destroy the interference properties of the electron, also guarantees that no measurement can be made as to which partial beam the electron actually passed through.

We shall now illustrate the equivalence of the one-body treatment to that in which the apparatus is treated quantum mechanically, in terms of some examples. We shall begin with our first case of a split electron beam passing through a pair of drift tubes. Suppose that one of these tubes is attached by a wire to a generator of electric potential, while the other is grounded (at zero potential). We now suggest a simplified model of such a generator. Consider a sphere of capacity $C$ (much greater than that of the drift tube $C_0$), with two small holes in it, at opposite ends of a diameter. We then take an insulating rod of mass $M$, with some charge distribution fixed near its center. Let this rod move in such a way that it passes through the two holes in the sphere. If we let $y$ be the coordinate of the center of the rod, then the potential on the sphere will evidently rise from essentially zero to some maximum value, as the charge enters the sphere, after which it will fall again to zero when the charge leaves. This potential will therefore have a form, $V(y)$, which resembles a localized packet-like function in $y$ space. If the rod moves (by its own inertia) with some velocity $\dot{y}(t)$, this movement will produce a potential $V(y(t))$, which is time dependent, in such a way that it starts at $V=0$ at $t=0$, rises to a maximum of some time, $t=t_m$, and then falls back to zero as $t\rightarrow\infty$. Thus, we are able to produce the kind of time dependent potential required in this experiment. And if the adiabaticity conditions are satisfied, then, as we have shown, the one-body Schrödinger equation with this time dependent potential will yield the same results as would the exact quantum mechanical treatment, based on solving for the wave function $\psi(\mathbf{x},y,t)$ for the combined system.

Another example that is interesting to study is afforded by the consideration of a condenser, consisting of two large insulating flat sheets of mass $M$, charged uniformly and oppositely with a surface density, $\sigma$, which is attached without possibility of moving relative to the sheets. Let $y_1$ be the coordinate of the first sheet, $y_2$ that of the second (in a direction perpendicular to the sheets). The energy of interaction with an electron outside the sheets is then

$$W=\pm 4\pi\sigma e(y_1-y_2), \tag{26}$$

the sign being opposite, in accordance with the side of the condenser on which the electron is.

Let us suppose that initially the condenser sheets are practically in contact, so that the above potential energy is essentially zero. At this time, a pair of wave packets corresponding to a split electron beam is allowed to pass on opposite sides of the condenser. Then, when the packets have gone far enough so that edge fields can be neglected, the condensers are allowed to separate with some relative momentum, $p_1-p_2$, thus generating a potential difference between the two beams given by (26). After some time, the attraction of the two sheets for each other overcomes their initial relative momentum, and they turn around to approach each other. After they touch, so that $W$ is zero again, the electron beams are allowed to pass over the edge of the condenser, and are brought together to interfere.

This example is useful because it brings out an important point, viz., that whereas *the electron exerts a force on each sheet, there is, nevertheless, no net force in the electron*, because the sheets exert equal and opposite forces on it. Thus, while the electron can be seen to change the relative momentum of the parts of the source, the total momentum of the source is not altered; and it is basically for the reason that the reaction forces of the source on the electron cancel out. (Note that the kinetic energy of the electron is therefore not altered; the change of energy of the parts of the source can be shown to come from cross terms of the electrostatic *field* of the electron with the electrostatic *field* of the source.)

We note also that a similar argument can be applied in the general case. For example, with the drift tubes, we can consider pairs of small elements of charge on opposite diameters of these tubes as a basic unit. To simplify the problem, let us suppose that the electron is at the center of the tube. Then it will exert equal and

[19] See reference 7, Chapter 22, for a more detailed discussion of this aspect of measurement theory.

opposite forces on the elements of the pair, so that the net reaction of this pair on the electron vanishes. Since this happens for every such pair, it follows that the total force on the electron is zero, even though the electron is actually exerting a force in every element of charge in the tube.

We thus verify again that in quantum mechanics, there are experimental situations in which the behavior of an electron can be influenced by interactions, under conditions in which there is no force on it, so that according to classical theory, no effects could occur.

## 4. EFFECTS OF SOURCES FOR CASE OF MAGNETIC VECTOR POTENTIALS

In Sec. 3, we treated the source of electric potential quantum mechanically, and showed that if the source is heavy enough for the adiabatic approximation to hold, the results are the same as if the potential is taken to be a specified function of space and time entering the one-body Schrödinger equation for the electron. We shall now go on to obtain a similar result for the case of magnetic potentials. This case is not so straightforward as is the corresponding electrical problem, because the magnetic vector potentials satisfy d'Alembert's equation,

$$\left(\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{A}_\perp=-\frac{4\pi}{c}\mathbf{j}_\perp, \tag{27}$$

where we are using the gauge in which $\mathrm{div}\mathbf{A}=0$, and where $\mathbf{j}_\perp$ is the divergence free part of $\mathbf{j}$ (i.e., $\mathrm{div}\mathbf{j}_\perp=0$ and $\nabla\times\mathbf{j}_\perp=0$). As a result, there is no simple expression corresponding to the integral (10) for the electric potential, which would, in general, determine the magnetic vector potential at a given time in terms of the current distribution at that time. Rather, as is well known, the field has a "dynamic" character, implied by the fact that even when $\mathbf{j}(\mathbf{x},t)$ is given everywhere, an arbitrary solution of the homogeneous wave equation can be added to $\mathbf{A}$. As a result, $\mathbf{A}$ cannot be eliminated, as was the case with $\phi$; and in quantum mechanics, it must therefore be included in the wave function and in the wave equation.

It follows then that the method given in Sec. 3 will not be adequate for the magnetic case, and that a more general treatment will be needed. This treatment, which we shall give here, will also show how the local character of the interaction between charge and field, which played an essential part in our discussion in Sec. 1, appears in the theory when the fields are treated dynamically, instead of as specified functions of position and time.

We begin by writing down the Hamiltonian for the system, consisting of sources, the electromagnetic field, and the electron under discussion.[20]

[20] For a more detailed discussion, see, for example, W. Heitler, *Quantum Theory of Radiation* (Oxford University Press, New York, 1954), 3rd ed., Chap. 1.

This is

$$H=H_F+H_S+H_e. \tag{27}$$

The Hamiltonian of the field is

$$H_F=\int\frac{(\mathcal{E}^2+\mathfrak{H}^2)d\mathbf{x}}{8\pi}, \tag{28a}$$

with

$$\mathcal{E}=-\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t},\quad \mathfrak{H}=\nabla\times\mathbf{A}.$$

The momentum canonically conjugate to $\mathbf{A}(\mathbf{x})$ is $\mathbf{\Pi}(\mathbf{x})=(1/c^2)\partial\mathbf{A}(\mathbf{x})/\partial t$. The Hamiltonian of the source is

$$H_S=\sum_j\left[\mathbf{P}_j-\frac{e}{c}\mathbf{A}(\mathbf{y}_j)\right]^2\Big/2M_j+V(\cdots\mathbf{y}_j\cdots), \tag{28b}$$

where $\mathbf{y}_j$ is the coordinate of one of the moving particles which constitute the current in the source whose mass is $M_j$. $\mathbf{P}_j$ is the momentum canonically conjugate to $\mathbf{y}_j$ and $V(\cdots y_j\cdots)$ is the potential energy of interaction of the source particles with each other.

The Hamiltonian of the electron is

$$H_e=[\mathbf{p}-(e/c)A(\mathbf{x})]^2/2m, \tag{28c}$$

where $\mathbf{x}$ is the coordinate of the electron, and $\mathbf{p}$ is the conjugate momentum.

The wave function of the system must depend on the above variables, so that it can be expressed as

$$\Psi(\mathbf{x},\cdots y_j\cdots,\cdots\mathbf{A}(\mathbf{z})\cdots,t),$$

where $\mathbf{A}(\mathbf{z})$ is the potential at the point.

As in Sec. 3, we can use the WKB approximation for the source, so that its wave function may be written as $\phi(\cdots y_j\cdots,t)$, where $\phi$ is a narrow packet-like function which is appreciable only in a small region near $y=y_j(t)$, the classical orbit of the particle. We then make the adiabatic approximation, based on the large mass, which we are assuming for the source particles. In analogy with Eq. (20), we write

$$\Psi=\phi(\cdots\mathbf{y}_j\cdots,t)\xi(\cdots\mathbf{y}_j\cdots,\mathbf{x},\cdots\mathbf{A}(\mathbf{z})\cdots,t). \tag{29}$$

By a calculation based on approximations similar to those used in Sec. 3, we obtain

$$i\hbar\frac{\partial\xi}{\partial t}=\left[H_F+H_e-\sum_j\frac{e}{c}\langle\mathbf{v}_j(t)\rangle_{\mathrm{av}}\cdot\mathbf{A}(\mathbf{y}_j)\right]\xi', \tag{30a}$$

where we have set

$$\xi(\cdots,y_j+u_j(t),\cdots,\mathbf{x},\cdots\mathbf{A}(\mathbf{z})\cdots,t)=\xi'(\cdots u_j\cdots,\mathbf{x},\cdots\mathbf{A}(\mathbf{z})\cdots,t),$$

and where $\langle\mathbf{v}_j(t)\rangle_{\mathrm{av}}=\langle\mathbf{p}_j(t)\rangle_{\mathrm{av}}/M_j$, the average velocity of the $j$th source particle in its wave packet. In this expression, we have neglected the terms involving $e^2\mathbf{A}^2(y_j)/2M_jc^2$, in comparison to the sum involving

$\langle \mathbf{v}_j(t)\rangle_{\rm av}\cdot\mathbf{A}(\mathbf{y}_j)$, which is permissible if $M_j$ is large enough. (To do this is equivalent to assuming that the effects of mutual induction between the source particles can be neglected in comparison to those of their own inertia. This simplification evidently does not change the results in any essential way.)

As in Sec. 3, we can set $u_j=0$ in Eq. (30a), because derivatives of $u_j$ do not appear in this equation, and because $\phi(\cdots\mathbf{y}_j\cdots,t)$ is negligible for appreciable values of $u_j$. We shall write

$$\xi'(\cdots 0,\cdots\mathbf{x},\cdots\mathbf{A}(\mathbf{z})\cdots,t)=\eta(\mathbf{x},\cdots\mathbf{A}(\mathbf{z})\cdots,t).$$

Equation (30a) then becomes

$$i\hbar\frac{\partial\eta}{\partial t}=\left[H_F+\frac{[p-(e/c)\mathbf{A}(\mathbf{x})]^2}{2m}-\frac{e}{c}\sum_j\langle\mathbf{v}_j(t)\rangle_{\rm av}\cdot\mathbf{A}(\mathbf{y}_j(t))\right]\eta, \quad (30b)$$

where $\langle\mathbf{y}_j(t)\rangle_{\rm av}$ is the average of $\mathbf{y}_j$ over its wave packet.

Because the position $\langle\mathbf{y}_j(t)\rangle_{\rm av}$ and the velocity $\langle\mathbf{v}_j(t)\rangle_{\rm av}$ appear in Eq. (30b) only as average quantities, which are $c$ numbers and not operators, it is now possible to divide $\mathbf{A}(\mathbf{z},t)$ into two parts, one of which $\langle\mathbf{A}(\mathbf{z},t)\rangle_{\rm av}$, is a $C$ number associated with the average movement of the source, while the other, $\mathbf{A}'(\mathbf{z},t)$, is an operator associated with the "zero point" quantum fluctuations of the electromagnetic field, plus whatever field is generated by the electrons. Thus,

$$\mathbf{A}(\mathbf{z},t)=\langle\mathbf{A}(\mathbf{z},t)\rangle_{\rm av}+\mathbf{A}'(\mathbf{z},t). \quad (31a)$$

Rigorously, the average potential $\langle\mathbf{A}(\mathbf{z},t)\rangle_{\rm av}$ should satisfy d'Alembert's equation, corresponding to the average current density $\mathbf{j}(\mathbf{z},t)=(e/c)\sum_j\delta(\mathbf{z}-\mathbf{z}_j')\langle\mathbf{v}_j(t)\rangle_{\rm av}$. However, because the acceleration of the source particles is negligible in typical cases (e.g., the electrons in a solenoid) and because the velocities are small enough for relativistic effects to be neglected, we can replace the exact solution of d'Alembert's equation by

$$\langle\mathbf{A}(\mathbf{z},t)\rangle_{\rm av}=\frac{e}{c}\sum_j\langle\mathbf{v}_j(t)\rangle_{\rm av}/|\mathbf{z}-\mathbf{y}_j(t)|. \quad (31b)$$

This is, of course, just the expression leading to the Biot-Savart law.

We have thus separated out a part $\langle\mathbf{A}(\mathbf{z},t)\rangle_{\rm av}$ of the total vector potential, $\mathbf{A}(z,t)$, which is related to the current distribution, in the same way that the electrostatic potential is related to the charge distribution by the integral (10). Note, however, that there remains another part of the potential, $\mathbf{A}'(\mathbf{z},t)$, so that we have not expressed the *total* potential as a function of the current density at the same moment of time (i.e., we have not eliminated the dynamic character of $\mathbf{A}$).

When the transformation implied by the substitution (31a) and (31b) is used in the wave function, then Eq. (30b) reduces to

$$i\hbar\frac{\partial\eta}{\partial t}=\left\{H_F'+\frac{\{\mathbf{p}-(e/c)[\langle\mathbf{A}(\mathbf{z},t)\rangle_{\rm av}+\mathbf{A}'(\mathbf{x},t)]\}^2}{2m}\right\}\eta, \quad (31)$$

where $H_F'$ represents part of the electromagnetic field energy associated with the potential $\mathbf{A}'(\mathbf{z},t)$, viz.,

$$(1/8\pi)\int[(1/c^2)(\partial\mathbf{A}'/\partial t)^2+(\nabla\times\mathbf{A}')^2]d\mathbf{z}. \quad (32)$$

In the above equation, the term $\mathbf{A}'(\mathbf{x},t)$ (which couples the electron to the part of the electromagnetic field that is not generated by the source) describes, as we pointed out above, the effects of zero point "vacuum" fluctuations of the field, along with associated effects (such as the back reaction of the electron's own field on itself), which are taken into account in standard renormalization theory.[21] However, it is well known that in the first approximation, the effects of the zero-point fluctuations on the electron average out to zero, while in the second approximation they introduce corrections which (along with those of the self-field of the electron) are quite small. We shall neglect these corrections here. This is evidently equivalent to leaving out the term $\mathbf{A}'(\mathbf{x})$ in Eq. (31). The Hamiltonian $H_F'$ then ceases to be coupled to the electron variables. The wave function can therefore be chosen as a simple product

$$\chi=\chi_0(\cdots\mathbf{A}'(\mathbf{z})\cdots)\psi(\mathbf{x},t), \quad (33)$$

where $\chi_0(\cdots A'(\mathbf{z})\cdots)$ represents the ground state of the $\mathbf{A}'(\mathbf{z})$ field (describing therefore the "zero-point" fluctuations) while $\psi(\mathbf{x},t)$ satisfies the equation

$$i\hbar\frac{\partial\psi}{\partial t}=\frac{[p-(e/c)\mathbf{A}(\mathbf{x},t)]^2}{2m}\psi.$$

We have thus achieved our aim of showing that the quantum mechanical treatment of the magnetic source leads to the same results as those of the one-body treatment, in which the vector potential is taken to be a specified function of space and time.

The fundamental role of potentials can now be illustrated in more detail by considering the following simple example, in which a solenoid surrounded by an impenetrable potential barrier is switched on adiabatically. Suppose that, as suggested in our discussion in Sec. 2, the electron originates in a Faraday cage, so that it is not acted on by the electric field resulting from turning on the solenoid. (This electric field is cancelled by the effects of currents induced in the wall of the cage, currents which can, however, be treated in the formalism as just another part of the source variables.) The average initial momentum and position of the electron wave packet are then so arranged that after this electric field vanishes, the electron emerges through a hole in the cage. (The hole is so small that the penetration of the electric field through it can be neglected.) The packet

[21] H. Bethe and S. Schweber, *Mesons and Fields* (Row, Peterson and Company, Evanston, Illinois, 1955), Vol. 1. Chap. 21.

is then split into two parts by a bi-prism, which go around the solenoid on opposite sides, after which they are reunited by another bi-prism, into a single coherent packet. This packet then enters a second Faraday cage, through a small hole, after which the flux in the solenoid is turned off adiabatically. Interference phenomena are then observed inside the second Faraday cage.

The above experiment satisfies the conditions assumed in our treatments in Secs. 3 and 4; viz., that initially there is no interaction between electron and source, while during the course of the experiment, this interaction rises to its full value and then falls back to zero. Moreover, it is evident that at no stage is the electron wave packet in a region containing electromagnetic fields [i.e., $\mathcal{E}(\mathbf{x})$ and $\mathcal{H}(\mathbf{x})$].

It is clear that there is no way to formulate this problem in the quantum theory without considering potentials. Thus, when the initial quantum state of the electron at $t=0$ is determined as represented by a certain packet wave function, this packet by itself contains no reflection whatsoever of the fact that there is a flux in the solenoid (since the electron was screened from electric fields by a Faraday cage). Indeed, we must start with the initial wave function for the combined system of the form (33), where $\psi$ is taken to be the initial electron wave packet $\psi_0(\mathbf{x})$ and $\phi$ the initial packet of the set of particles $\phi_0(\cdots\mathbf{y}_j\cdots,t)$ in the source, whlie the electromagnetic field is represented by $\chi_0(\mathbf{A}'(z,t))$ describing the zero-point "vacuum" fluctuations in the field. By solving Schrödinger's equation for this system, we see from the changing form of $\Psi$ how the quantum fluctuations take place around a changing average $\langle\mathbf{A}(\mathbf{z},t)\rangle_{\text{av}}$ representing the effects of the source. The electron responds mainly, as we have seen, however, to the average $\langle\mathbf{A}(\mathbf{z},t)\rangle_{\text{av}}$, while the quantum fluctuations have effects which can, to a good approximation, be neglected. As a result, the operator $\mathbf{A}(\mathbf{z},t)$ can be replaced by the $c$ number, $\langle\mathbf{A}(\mathbf{z},t)\rangle_{\text{av}}$; and this is how we came back to the one-body Schrödinger equation with specified potentials.

We see then that whether we treat the potentials as specified functions of space and time (as we did in Secs. 1 and 2), or as dynamical variables furnishing a link between the source and the electron (as we did in this section), there is no way in the quantum theory to express the effect of a flux inside the solenoid on an electron outside in terms of a *localized* interaction, except with the aid of potentials. In no case does the theory ever contain any kind of interaction between the electron and the source, which does not go through the intermediary of potentials and, as we have seen, fields are not, in general, adequate for expressing all aspects of this intermediary role.

## ACKNOWLEDGMENTS

We are indebted to Professor M. H. L. Pryce and to Dr. R. Chambers for many helpful discussions.

## APPENDIX

### Some Comments on a Paper by Peshkin, Talmi, and Tassie, Concerning the Role of Potentials in the Quantum Theory

Recently there has appeared a paper by Peshkin, Talmi, and Tassie,[10] on the role of potentials in the quantum theory.[22] This paper seems to have two objectives; firstly, to show that the nonclassical consequences of potentials in the quantum theory should not be regarded as surprising, and secondly, to suggest that these consequences should not be ascribed to the potentials, but rather to the effects of suitable fields on the quantum conditions applying to the electron. With the first of these objectives, we are, of course, in agreement, as we indicated in Sec. 1. We do not, however, regard the second objective as a valid one; and we shall show here the inadequacy of such an approach, in terms of several of the points that were treated in the above article.

The first question considered by these authors is concerned with the problem of stationary states of an electron in a multiply-connected region of space, which contains flux in the excluded region. They begin with an analysis, which leads them to the same conclusions that we gave with regard to this problem in Sec. 2; viz., that the allowed values of the energy of the electron are related to this flux in a way that is independent of the past history of the electron (e.g., of whether the electron was in the orbit in question or not while the flux was being turned on). However, in stating this condition, they assert that "The presence of the field in the excluded region permanently changes the allowed values of the physical angular momentum of every electron in the world; regardless of how the system was actually prepared. Such arguments from the correspondence principle, which are now very old, apply equally to the Zeeman effect." In their Abstract they also state that the observable effects of potentials arise, not from forces exerted by magnetic fields or by their vector potential, but from modifications of the quantum conditions. In view of the above statements, it seems difficult to avoid the condlusion that these authors wish to imply that somehow the role of the potentials can be eliminated, because the fields in the excluded regions are able to account for the change in the quantum conditions and presumably, in a similar way, for all possible physical properties of the electron.

In accordance with the discussion of this problem given in our paper, however, we see that the fields in the excluded region cannot be regarded as interacting directly with the electron. Indeed, this interaction goes by the intermediary of the potentials; and it is only when this is taken into account that the essential feature of the *locality* of interaction of electromagnetic field

[22] We wish to thank the above authors for sending us a preprint of their work.

with electron can be brought properly into the theory. And since the above applies to the exact form of the quantum theory, it must also apply in the correspondence limit of high quantum numbers (which is, after all, an approximation to the exact theory), so that this conclusion cannot be altered by an appeal to the correspondence principle. It is therefore not sufficient to attribute the change of quantum state of the electron to the field in the excluded region, but in a more nearly complete treatment, one must take into account how this change is brought about by means of the influence of the potentials which link the quantum state of the electron to the current in the source by means of purely local interactions.

The next problem considered by Peshkin, Talmi, and Tassie, was that of the origin of the average force (i.e., the average rate of momentum transfer) in the scattering of an electron beam off a solenoid of negligible radius. In their discussion of this problem, they indicated without, however, giving a proof that (a) when no barrier is present, this force can be accounted for as a result of the possibility that the electron will penetrate into the magnetic field region; and (b) when there is a barrier, the force comes from interaction of the electron with the barrier in question.

Before proceeding further, it is worthwhile here to show that these conclusions can be proved directly from Ehrenfest's theorem [Eq. (5)]. Thus, if there is no barrier $V=0$, and the average force is equal in this case to $(e/c)\int\psi^*(\mathbf{v}\times\mathfrak{H})\psi d\mathbf{x}$ (since $\mathfrak{E}=0$ also). The above is, of course, just the average magnetic force. If there is a barrier, then to simplify the problem, let us suppose that it is very high, but not infinite. (In this way, we will guarantee that the wave function and its derivatives are finite everywhere, so that the conditions necessary for the validity of Ehrenfest's theorem are satisfied.) Then, as is well known, the electron will penetrate with appreciable probability only a short distance into the barrier, so that the wave function, $\psi(\mathbf{x})$ will be essentially zero near the origin, where the magnetic field is not zero. As a result, the average magnetic force vanishes, and the average force will be $\int\psi^*(\nabla V)\psi d\mathbf{x}$, which is just the force coming from the barrier, as was indeed suggested in the above paper.

Peshkin, Talmi, and Tassie then assert that because the force comes from the barrier in the manner described above, there is no reason to ascribe any force to the excluded magnetic field or to the local vector potential. They recognize, however, that the average force exerted by a given barrier depends on the flux inside. For example, in the absence of flux, such a barrier has a cross section and a proportional average force of the order of the radius, while when there is flux within, the cross section for the same barrier can rise to a generally much larger value, of the order of the wavelength of the incident electrons. They ascribe this change in the effectiveness of a given barrier to the "modification of the quantum conditions" (which they have, in turn, ascribed to the effects of the magnetic field inside the barrier). From these statements, it would seem once again that the above authors are giving arguments against the notion that in quantum mechanics the potentials play a role more significant than that which they played classically.

In answer to these arguments, we first point out that from the modification of quantum conditions alone, there is, in general, no way to calculate either the scattering cross section or the average force on the electron, without first specifying the vector potential in the whole region outside the barrier, and then solving Schrödinger's equation in detail. (For example, in the case of a barrier of a radius that is appreciable in comparison to the wavelength of the incident electrons, all physical effects will depend on this detailed solution.) It is only by thus introducing the potentials that we can account for the change of the average force exerted by the same barrier, when no fields of any kind change except those in the region that is not accessible to the electron.

The above discussion illustrates once again that (as we have emphasized throughout this article), the concept of force is, in the quantum theory, an abstraction of secondary importance. Therefore, from the fact that potentials exert no forces, it does not follow that (as seems to be implied by the above authors) these potentials can have no physically significant effects.

Finally, Peshkin, Talmi, and Tassie give a model for the interaction of the electron with the source of the field. To do this, they assume a direct velocity-dependent mechanical interaction between the electron and the source, which gives rise essentially to the Biot-Savart law for the electron. On the basis of this model, they are led to a result analogous to that which we give in Sec. 4; viz., that the interference effects are the same as those obtained on the basis of the one-body Schrödinger equation with the potentials given as specified functions of the time.

In connection with the above model, there are two points that we wish to stress. First, this model involves the assumption of a nonlocal mechanical interaction between electron and source, which is adequate for the purpose of proving the result described above, but which cannot be used for a treatment of the problem of the locality of the interaction, which we have stressed in our article. Secondly, this model has been used (within its proper domain of validity) to make certain inferences, which are, however, misleading for other reasons. These inferences were based on the fact that what appears in the Hamiltonian obtained by the above authors is not the vector potential, but rather, the momentum, $p_\beta$, canonically conjugate to the source variable. From the constancy of $p_\beta$ with time (which follows from their Hamiltonian), it can be seen that quantum fluctuations of the magnetic field in the source have no influence on the interference phenomena under discussion. And since the vector potential fluctuates along with this magnetic field, it would seem at first sight that in a

situation in which quantum fluctuations are important, the interference effects are determined, not by the potentials, but rather by some other variables (in this case, $p_\beta$). Thus, one might be led to conclude that in such cases, the potentials are not of fundamental significance in the theory.

A more careful analysis shows, however, that while the calculation of Peshkin, Talmi, and Tassie showing the dependence of interference effects on the constant of the motion $p_\beta$ is correct, their discussion does not make clear that the vector potential plays an essential part in bringing about this result. For as can be shown quite readily, this dependence of interference effects on $p_\beta$ (with their resulting independence from the quantum fluctuations of the field in the source) is due to a *compensation* of the effects of the fluctuating part of the vector potential by the effects of the electric fields, $\mathcal{E}=-(1/c)(\partial\mathbf{A}/\partial t)$, that inevitably accompanies such a change of vector potential. (In Sec. 1 we treated a similar problem,[23] where we saw that stray 60-cycle magnetic fields compensated the effects of the corresponding fluctuating potentials to produce a constant and stable interference pattern.)

In order to seen in more detail what is happening in this problem, we first note that there is a back reaction of the magnetic field of the (moving) incident electron on the source (see, for example, the latter part of Sec. 3, where a similar reaction was found in the electrostatic case). This (in general, changing) magnetic field induces an electromotive force in the source solenoid; and as a result there is mutual interaction of the electronic variables and the source variable, in which the states of both are altered. Nevertheless, as can be shown by a simple calculation, this interaction is such that it leads to the constancy of $p_\beta$.

In all discussions given until now, the source has been assumed to be so heavy that for practical purposes, the effects of the electronic magnetic field on the source current can be neglected (as indeed also follows in the treatment of the above authors, if the mass of the source is allowed to become very large). If, however, the mass is not large, then one will have to take into account the effects of the changing flux produced by the source, which will, as pointed out in the above discussion, give rise to a further electric field. It is clear, of course, that when such an electric field is present, interference effects will no longer, in general, be given by the formulae of our first paper, in which we, by hypothesis, restricted ourselves to the case in which there were *no fields of any kind* in the region accessible to the electron. In fact when fields are present, the interference properties of the electron are as pointed out in Sec. 1, determined (in the limit of high quantum numbers) by $\oint\mathbf{p}\cdot d\mathbf{x}=\oint[m\mathbf{v}+(e/c)\mathbf{A}']\cdot d\mathbf{x}$, where $\mathbf{p}$ is its *canonical* momentum, and $m\mathbf{v}$ its "physical" momentum [see Eq. (4)]. However, as a result of the change of source strength brough about by the magnetic field of the electron, the vector potential, $\mathbf{A}(\mathbf{x})$, that is actually present on the path of the electron will be slightly different from what it would have been, if this reaction had not occurred. Thus, it will no longer be correct to calculate the integral, $\oint\mathbf{A}\cdot d\mathbf{x}$, under the assumption that no such a reaction takes place. On the other hand, the resulting electric field acting on the electron will change the "physical" momentum, and as can be verified by a simple calculation, this alteration just compensates the effects of the change in $\mathbf{A}$ on the interference pattern. This result shows that the implication of Peshkin, Talmi, and Tassie, that potentials are not playing a fundamental role here, is wrong. For the fluctuating part of the potentials is just what is needed to explain the dependence of the interference pattern on the constant of the motion, $p_\beta$, despite the presence of the fluctuating electric field, which necessarily accompanies this fluctuating potential.

[23] See reference 3.