Explaining the Aharonov-Bohm Effect

John Earman
Dept. HPS University of Pittsburgh

In an often invoked setup used to illustrate the Aharonov-Bohm (AB) effect, charged particles passing on opposite sides of an idealized infinitely long solenoid, with no flux leaks and protected from penetration of the charged particles by an infinitely high potential barrier, experience differential phase changes that result in an observable interference pattern dependent on the magnetic flux in the solenoid, despite the fact that the configuration space of the charged particles is disjoint from regions where the magnetic field B is non-zero. The philosophical literature on the AB effect focuses largely on issues of non-locality, the "reality" of electromagnetic potentials, and the like. While a thorough discussion of the effect must confront these issues, making them the focus risks underappreciating the prior and more fundamental issues about the derivation/explanation of the effect. The mainline explanation is dynamical, relying on Schrödinger evolution using a Hamiltonian operator $\hat{H}_A$ universally cited in standard texts. This operator is derived by following the canonical quantization procedure, starting from a classical Hamiltonian and then substituting for the classical position and momentum variables Hilbert space operators that give a representation of the Heisenberg canonical commutation relations. But $\hat{H}_A$ is not the unambiguous result of canonical quantization since there are representations unitarily inequivalent to the familiar Schrödinger representation used to derive $\hat{H}_A$. This operator contains a gauge variable, the vector potential A of the magnetic field, and the mainline explanation relies on the gauge equivalence of Schrödinger evolution under $\hat{H}_A$ with free evolution multiplied by a phase factor. This equivalence is broken, and the mainline explanation is undermined by an attempt to overcome the apparent nonlocal dependence of the inference pattern on the magnetic field by regarding the vector potential A as a real physical variable. The mainline explanation can be criticized on the grounds that it attempts to explain a gauge-independent effect using gauge-dependent variables. This concern can be answered by showing that the AB phase can be obtained without using the vector potential A but only the gauge-invariant magnetic field B. The extant version of this explanation uses the path integral approach rather than canonical quantization, and it makes manifest the non-local dependence of the interference pattern on the magnetic field over all space. There are also attempts at non-dynamical (a.k.a. "topological") explanations that do not depend on the details of the Hamiltonian operator. While intriguing it is not clear that they address the relevant explanandum. In working through the thicket of issues surrounding the explanation of the AB effect it is natural to look for help from the philosophical literature on scientific explanation. But, with the exception of interventionist accounts of causal explanation, one looks in vain.

1 Introduction

There is a substantial philosophical literature on the Aharonov-Bohm (AB) effect, much of it focused on issues of non-locality, the "reality" of electromagnetic potentials, and the like. But there is a surprising lack of discussion and understanding of the prior issue of how QM explains the AB effect. Standard QM does explain how this surprising and counterintuitive effect is possible, although there seems to be an underappreciation of the subtleties involved. When these subtleties are taken into account one sees that while QM explains how the AB effect is possible, insofar as explanation implies prediction, QM does not explain the occurrence of the AB effect in nature. What it does furnish is a conditional dynamical explanation of the effect, where the explanation is conditioned on the quantum Hamiltonian taking a certain form. That form, however, is not the unequivocal outcome of canonical quantization. Alternative non-dynamical explanations, not dependent on the form of the Hamiltonian, have been offered. While intriguing, these alternatives are underdeveloped and appear to fall short of the mark for various reasons.

It might seem then that the AB effect would serve as a fruitful test bed for various competing accounts of scientific explanation, which include the deductive-nomological model, the statistical relevance model, causal mechanical models, unificationist models, and pragmatic theories (see Woodward and Ross 2021). While aspects of these accounts are marginally helpful, none of them are a good fit for the AB effect. The one exception is the interventionist account of causal explanation which is helpful in focusing a crucial aspect of the AB effect. Given the variety of physical phenomena nature presents it would be surprising if one, or even a small handful, of "models of explanation" suffice to cover the variety of means of comprehending the phenomena found in scientific theorizing about the phenomena.[1]

2 Sketch of the AB set up: idealizations and more idealizations

The magnetic AB effect involves situations where the configuration space of a charged particle is strictly disjoint from regions where the magnetic field is non-zero.[2] Physical realizations of such a scenario would seem to involve severe idealizations; a commonly used one invokes an infinitely long solenoid which does not leak magnetic flux and which is surrounded by an infinitely high potential barrier preventing charged particles from penetrating into the interior of the solenoid. A beam of charged particles emitted from a source is split into two parts, "each going on opposite sides of the solenoid, but avoiding it. (The solenoid can be shielded from the electron beam by a thin plate which casts a shadow.)" (Aharonov and Bohm 1959, p. 486; see Fig. 1).

In the treatment given below an additional idealization, shrinking the radius of the solenoid to a point while the total flux remains the same, is imposed for technical reasons. For a charged particle confined to move in the x-y plane (shown in Fig. 1), the configuration space of the particle is $\Omega_R := \mathbb{R}^2 \backslash D_R$ where $D_R := \{(x, y) : x^2 + y^2 \leq R\}$ is a disk of radius $R \geq 0$ centered at the origin $(0, 0)$ of orthogonal coordinates x, y for $\mathbb{R}^2$, and the Hilbert space is $L^2_{\mathbb{C}}(\Omega_R)$. (The non-simple connectedness of the configuration space Ω_R will play an important role in what follows.) Recalling that the elements of a Hilbert space $L^2_{\mathbb{C}}(X)$ are not square integrable complex functions on X but equivalence classes of such functions, where the equivalence relation is equality up to a set of Lebesque measure 0, there is a natural identification of $L^2_{\mathbb{C}}(\Omega_{R=0})$ and $L^2_{\mathbb{C}}(\mathbb{R}^2)$. And the identification can be used to prove that

[1] This sentiment resonates with that of Kitcher, "The Theory of Scientific Explanation: An Obituary" (2021).

[2] There is also an electric AB effect whose theoretical analysis and experimental verification remain controversial. It will not be discussed here. Henceforth, I will speak simply of the AB effect with the understanding that there is a silent "magnetic" modifier.

the Schrödinger momentum operators[3] $-i\dfrac{d}{dx}$, $-i\dfrac{d}{dy}$ acting on $L^2_{\mathbb{C}}(\Omega_{R=0})$ are essentially selfadjoint on a common dense domain[4] and that their unique selfadjoint extensions together with position operators acting by multiplication give an irreducible representation of the Heisenberg canonical commutation relations. However, when the solenoid has a radius greater than 0 the Schrödinger momentum operators are not essentially selfadjoint on $L^2_{\mathbb{C}}(\Omega_{R>0})$ and, further, among the selfadjoint extensions none satisfy the Weyl form of the canonical commutation relations (Hirokowa 1997, 2000).[5] These problems can be addressed by using non-orthogonal coordinates adapted to the streamlines of an incompressible fluid flowing around the solenoid (Hirokowa 1997, 2000). But unless otherwise specified, attention will focus mainly on the $R = 0$ idealization.

Given the need for such severe idealizations and the need for a deep dive into functional analysis required for a rigorous mathematical treatment of the idealizations, a working physicist could be forgiven for thinking that AB effect is a useless plaything of mathematical physicists who have nothing better to do. While there is some validity to such an attitude it would be a mistake to ignore the AB effect. Because of the insights it gives into the nature of the quantization of classical systems and the conceptual differences between the classical and the quantum that result from quantization, it deserves serious study.

3 The AB effect: the usual (expurgated) story

The most common explanation of the AB effect is dynamical, tracing out the dynamical evolution of the wave functions of the charged particles in the beams as they pass by the solenoid on the way to the registration screen. In order to implement the dynamics we need the quantum Hamiltonian for the AB set up. Standard text books tell us that it is

$$\hat{H}_{\mathbf{A}} = \frac{(-i\boldsymbol{\nabla} - q\ \hat{\mathbf{A}})^2}{2m} \tag{1}$$

[3] Units are chosen so that $\hbar = c = 1$.

[4] A suitable domain is $C_0^\infty(\Omega_{R=0})$, infinitely differentiable functions of compact support on $\mathbb{R}^2\backslash(0,0)$, which is dense in $L^2_{\mathbb{C}}(\mathbb{R}^2)$.

[5] See below for a discussion of the Heisenberg and Weyl forms of the commutation relations.

where q and m are respectively the charge and mass of the particle and $\hat{\mathbf{A}}$ is the operator that acts by multiplication by the function $\mathbf{A}$ of coordinates where $\mathbf{B} = \nabla$ x $\mathbf{A}$ (i.e. $\mathbf{A}$ is the vector potential of the magnetic field $\mathbf{B}$).

To justify (1) appeal is made to the process of canonical quantization. The first step is to find the classical Hamiltonian for a charged particle moving in a magnetic field. Newton's $\mathbf{F} = m\ddot{\mathbf{x}}$ law and the Lorentz force law $\mathbf{F} = q\boldsymbol{\nu} \times \mathbf{B}$ yield the classical equation of motion

$$m\ddot{\mathbf{x}} = q\boldsymbol{\nu} \ \times \ \mathbf{B}. \tag{2}$$

Next we find the Lagrangian that yields (2) as the Euler-Lagrange equations. An appropriate Lagrangian is

$$L = \frac{1}{2}m\boldsymbol{\nu}^2 + q\mathbf{A} \cdot \boldsymbol{\nu}. \tag{3}$$

The momenta conjugate to the position variables are given by

$$p_x := \frac{\partial L}{\partial \nu_x} = m\nu_x + qA_x,\, p_y := \frac{\partial L}{\partial \nu_y} = m\nu_y + qA_y. \tag{4}$$

Substituting (3) and (4) into the definition of the classical Hamiltonian

$$H := \boldsymbol{\nu} \cdot \mathbf{p} - L \tag{5}$$

results in

$$H_{\mathbf{A}} = \frac{(\mathbf{p} - \ q\mathbf{A})^2}{2m}. \tag{6}$$

The final step in canonical quantization is to replace the classical $\mathbf{p}$ and $\mathbf{A}$ by Hilbert space operators that give a representation of the Heisenberg CCR (HCCR)[6]

$$\begin{aligned} \hat{\mathbf{p}}_j\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_k\hat{\mathbf{p}}_j \ &= \ -i\hat{I}\delta_{jk}, \ \ j,k = x,y,z & \text{(7a)} \\ \hat{\mathbf{p}}_j\hat{\mathbf{p}}_k - \hat{\mathbf{p}}_k\hat{\mathbf{p}}_j \ &= \ 0, \ \ \hat{\mathbf{x}}_j\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_k\hat{\mathbf{x}}_j = 0. & \text{(7b)} \end{aligned}$$

[6]The mathematically minded will want to know precisely what is meant by a representation of (HCCR) especially since some of the operators involved must be unbounded and, thus, are defined at most on a dense domain of the Hilbert space. The common answer is that a representation of (HCCR) consists of operators $\hat{\mathbf{p}}_j$, $\hat{\mathbf{x}}_k$ that are essentially selfadjoint on a common invariant and dense domain on which (7a)-(7b) hold.

The conventional choice is to use the Schrödinger representation whereby $\mathbf{p}$ is replaced by $\hat{\mathbf{p}} = -i\boldsymbol{\nabla}$ and functions of x, y, z are replaced by operators that act on elements of $L^2_{\mathbb{C}}(\mathbb{R}^2)$ by multiplication. This choice is supposedly justified by the Stone-von Neumann's uniqueness theorem which is often glossed as showing that, for systems with a finite number of degrees of freedom (which is certainly the case here), any irreducible representation of the CCR is unitarily equivalent to the Schrödinger representation, implying that any other choice of representation would yield physically equivalent results.

With (1) in hand the dynamical explanation can be implemented by solving the Schrödinger equation for (1) and showing that the wave packets of charged particles passing the solenoid on opposite sides suffer a differential phase change that accounts for the dependency on the magnitude of the flux in the solenoid of the interference pattern on the detection screen.[7] In fact, this is never actually done; rather one uses an *Ansätz* of Aharonov and Bohm that provides a short cut to an approximate solution exhibiting the appropriate phase change (see Sec. 5 below). Amazingly, despite all the controversy generated by the AB effect, the validity of this approximation was not rigorously studied until comparatively recently (fifty years after the publication of the seminal Aharonov-Bohm paper). Before turning to an analysis of the dynamical explanation I will remark on some dubious moves used in the above attempt to justify $\hat{H}_{\mathbf{A}}$.

4 Exposing some dubious methodology and a misunderstanding: the unexpurgated story of canonical quantization

The dubious methodology, which is rarely remarked, should be obvious. There seems to be some flimflam in the initial move of starting from the classical equations of motion (2) of a particle moving in a magnetic field because in the AB effect the particle is *not* moving in a magnetic field since, *by definition* of the effect, the configuration space of the charged particle is disjoint from regions where $\mathbf{B} \neq \mathbf{0}$. Thus, the Lorentz force on the classical charged particle is everywhere 0, the Newtonian equation of motion is

[7] Typically, Dirichlet boundary conditions are assumed, by which wavefunctions vanish on the boundary $\partial\Omega_R$. More on this below.

$$m\ddot{\mathbf{x}} = \mathbf{0} \tag{2'}$$

and the corresponding classical Hamiltonian is the free Hamiltonian

$$H_0 = \frac{\mathbf{p}^2}{2m}. \tag{6'}$$

If the Schrödinger $\hat{\mathbf{p}} = -i\boldsymbol{\nabla}$ replaces $\mathbf{p}$ in (6′) the resulting quantum Hamiltonian is

$$\hat{H}_0 = -\frac{\boldsymbol{\nabla}^2}{2m}, \tag{1'}$$

and the motion of the quantum particle shows no dependence on the amount of magnetic flux in the solenoid.

An interesting argument for using (6) and (1) is found in the results of de Oliveira and Pereira (2008). Start with a non-idealized solenoid of length $\ell < \infty$ and radius $R > 0$, shielded by a finite potential barrier of height $h < \infty$. Since in this scenario the charged particle assuredly *is* moving in a magnetic field there is no dispute about the correct form of the classical Hamiltonian $H_{\ell,h}$. Replace the classical momentum and position variables in $H_{\ell,h}$ by their Schrödinger operators to obtain the non-idealized quantum Hamiltonian $\hat{H}_{\ell,h}$. Then take the limit $\hat{H}_\infty := \lim_{\ell,h\to\infty} \hat{H}_{\ell,h}$. Show that the limits $\ell \to \infty$ and $h \to \infty$ exist (in the resolvent sense) and that they commute so that it doesn't matter in which order they are taken. Finally show that $\hat{H}_\infty$ is the Hamiltonian $\hat{H}_{\mathbf{A}}$ of (1).[8] A very pretty result indeed!

But how does the result apply to a situation where we imagine being presented with an idealized infinitely long solenoid shielded by an infinitely high barrier? What should we expect when we turn on the beam of charged particles? If we suppose that the idealized solenoid has been constructed by superhero engineers who start small but manage to achieve the $\ell \to \infty$ and $h \to \infty$ limits at some finite time at which we perform our experiments then, presumably, our expectations should be governed by the Hamiltonian (1). But if we suppose that the idealized infinitely long and perfectly shielded solenoid has existed since time immemorial then the Hamiltonian (1′) remains a plausible candidate for guiding our expectations. And in any case the

[8] Furthermore, Dirichlet boundary conditions show up since wavefunctions in the domain of $\hat{H}_\infty$ vanish at the boundary of the solenoid.

justification for (1) assumes the use of the Schrödinger representation of the (HCCR); and it is just here that an important misunderstanding arises.

The final step in the usual argument for the AB effect involves a misunderstanding of the Stone-von Neumann uniqueness theorem.[9] The theorem concerns not the (HCCR) but an exponentiated form of the CCR called the Weyl CCR (WCCR). The trouble with (7a) is that the operator on the rhs is defined for all elements of the Hilbert space whereas the operators on the lhs are not since either the $\hat{\mathbf{p}}$'s or the $\hat{\mathbf{x}}$'s (or both) must be unbounded and are defined at best on a dense domain of the Hilbert space (see Reed and Simon 1980, p. 274). The problem disappears when the operators are exponentiated to $U(s_j) := \exp(-is_j\hat{\mathbf{p}}_j)$ and $V(t_k) := \exp(-it_k\hat{\mathbf{x}}_k)$, $s_j, t_k \in \mathbb{R}$. Assuming that the $\hat{\mathbf{p}}_j$ and $\hat{\mathbf{x}}_k$ are selfadjoint, the $U(s_j)$ and $V(t_k)$ form unitary groups whose elements are defined on the entire Hilbert space. The (WCCR) give the commutation relations for the $U(s_j)$ and $V(t_k)$:

$$\begin{aligned} U(s_j)V(t_k) &= \exp(is_jt_k)V(t_k)U(s_j), \;\; s_j, t_k \in \mathbb{R} & \text{(8a)} \\ U(s_j)U(s_k) - U(s_k)U(s_j) &= 0, \;\; V(t_j)V(t_k) - V(t_k)V(t_j) = 0. & \text{(8b)} \end{aligned}$$

What the Stone-von Neumann theorem shows is that, for a system with a finite number of degrees of freedom, all irreducible and strongly continuous unitary representations of the (WCCR) are unitarily equivalent to the Schrödinger representation.

In many instances satisfaction of the (HCCR) entails satisfaction of the (WCCR). Heuristically the implication seems valid: use a power series expansion of the exponentials $U(s_j)$ and $V(t_k)$, while ignoring issues of convergence, and plug in the (HCCR) at appropriate points in the expansion to arrive at the (WCCR). But for unbounded operators issues of convergence cannot be ignored, and failures of the entailment should not be unexpected. And, in fact, the entailment does fail in the AB setup, making available representations of the (HCCR) unitarily inequivalent to the Schrödinger representation when the configuration space of the charged particle is $\Omega_{R=0} = \mathbb{R}^2 \backslash (0,0)$.[10]

Operators $\hat{A}$ and $\hat{B}$ acting on a Hilbert space $\mathcal{H}$ are said to weakly commute iff $[\hat{A}, \hat{B}]\psi = 0$ for all $\psi \in \mathcal{H}$ such that $\psi \in dom([\hat{A}, \hat{B}])$, and are said to strongly commute iff $\exp(is\hat{A})\exp(it\hat{B}) = \exp(it\hat{B})\exp(is\hat{A})$ for all

[9] For reasons given in Earman (2023) one should speak of Stone's conjecture and von Neumann's theorem. But herein I will use the standard parlance.

[10] For more on the von Neumann uniqueness theorem see Earman (2023).

$s, t \in \mathbb{R}$.[11] The (HCCR) require only that the $\hat{\mathbf{p}}_j$ weakly commute amongst themselves as do the $\hat{\mathbf{x}}_k$, whereas the (WCCR) require that they strongly commute amongst themselves. Strong commutativity can fail in the AB setup. For particles moving in the x-y plane, which is taken to be orthogonal to the symmetry axis of the solenoid, the operators $\hat{\mathbf{P}}_x = -i\dfrac{d}{dx} - qA_x$, $\hat{\mathbf{P}}_y = -i\dfrac{d}{dy} - qA_y$ and the usual position operators $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ acting by multiplication satisfy the (HCCR). $\hat{\mathbf{P}}_x$ and $\hat{\mathbf{P}}_y$ weakly commute but may fail to strongly commute and, thus, fail to satisfy the (WCCR). For any rectangular closed curve with sides lying along the x- and y-axes and having lengths respectively s_x and s_y we have

$$\exp(is_x\hat{\mathbf{P}}_x)\exp(is_x\hat{\mathbf{P}}_y) = \exp(iq\Phi^A_{s_x,s_y})\exp(is_x\hat{\mathbf{P}}_y)\exp(is_x\hat{\mathbf{P}}_x) \tag{9}$$

where $\Phi^A_{s_x,s_y}$ is the line integral of $\mathbf{A}$ around the closed curve (Arai 2020, Theorem 3.3). For a closed curve surrounding the solenoid $\Phi^A_{s_x,s_y}$ is equal to the amount of magnetic flux through a two-surface bounded by the curve. Thus, $\hat{\mathbf{P}}_x$ and $\hat{\mathbf{P}}_y$ strongly commute, as required by the (WCCR) iff q x magnetic flux in the solenoid is an integer multiple of 2π for each $s_x, t_y \in \mathbb{R}$. Otherwise we obtain an irreducible and strongly continuous non-Schrödinger representation of the (HCCR).

Now let's assess the implications for the AB effect of a correct understanding of the Stone-von Neumann theorem. We saw that there are two plausible starting points for canonical quantization: the classical Hamiltonian $H_{\mathbf{A}}$ of (6) and the classical Hamiltonian H_0 of (6′). Suppose we start with (6). Quantum Nature might produce the quantum Hamiltonian $\hat{H}_{\mathbf{A}}$ of (1) by spurning the Schrödinger-inequivalent representations of the (HCCR) and choosing the standard Schrödinger representation. But there is nothing in standard QM to say that Nature will or must make this choice, so there is nothing in the theory as it stands that allows us to predict that the AB effect will be realized. Alternatively suppose we start instead with the classical Hamiltonian H_0 of (6′). Quantum Nature can still produce a quantum Hamiltonian of the form $\hat{H}_{\mathbf{A}}$ by availing herself of Schrödinger-inequivalent representations and, in particular, by choosing the one in which (per usual) the position operators act by multiplication but the momentum operators

[11] For selfadjoint $\hat{A}$ and $\hat{B}$ strong commutativity is equivalent to requiring that the spectral projections of $\hat{A}$ and $\hat{B}$ commute (see Simon and Reed 1980, Theorem VIII.3).

are $\hat{\mathbf{P}}_x = -i\dfrac{d}{dx} - qA_x$, $\hat{\mathbf{P}}_y = -i\dfrac{d}{dy} - qA_y$. But if this choice means that She chooses a particular Schrödinger-inequivalent representation, namely, the one in which $\oint \mathbf{A} \cdot \mathbf{dx}$ is the actual amount of flux now in the solenoid, then an important aspect of the AB effect remains unexplained. The experimenter running the AB setup can control the amount of flux in the solenoid by controlling the current fed to the solenoid, and changing the flux changes the Schrödinger-inequivalent representation unless the flux values differ by an integer multiple of $2\pi/q$. To accommodate this aspect of the AB effect either Nature must anticipate what flux value will be dialed up and change her choice of Schrödinger-inequivalent representation accordingly, or else her choice has to be understood as a choice of a generic Schrödinger-inequivalent representation where the value of the path integral $\oint \mathbf{A} \cdot \mathbf{dx}$ around a closed loop encircling the solenoid is a variable whose value is equal to whatever the value the flux takes in contingently variable circumstances. This might seem artificial. But the $\mathbf{A}$ in the $\hat{H}_{\mathbf{A}}$ of eq. (1) has to be interpreted in this generic manner if $\hat{H}_{\mathbf{A}}$ is to serve as the basis of a dynamical explanation of the AB effect in contingently variable circumstances; and, as the proverb goes, what's sauce for the goose is sauce for the gander. But again there is nothing in standard QM to say that Nature will or must act in this way.

In sum, if "QM" denotes predictions obtained via canonical quantization of classical systems then QM does not predict the AB effect, although it certainly does accommodate the effect. There are, however, non-canonical quantization schemes that appear to yield an unequivocal prediction of the AB effect, one of which will be examined below. There are also non-dynamical (a.k.a. "topological") treatments of the AB effect that are independent of the details of the Hamiltonian and that appear to yield an unequivocal prediction of the AB effect; again one of them will be discussed below.

5 Explaining the effect

5.1 *How-possible*

The first explanatory query that comes to mind for someone whose physical intuitions are trained on classical physics is apt to be "How is the AB effect possible?" since the space of physical possibilities allowed by classical physics appears to have no room for the effect. The answer takes the form of explaining how QM expands the space of possibilities by understanding the process of quantizing classical systems, involving representations of the canonical commutation relations as outlined above. However, it is doubtful that even a detailed filling in of this outline would be accepted as a fully satisfying how-possible explanation until it is explained how the AB is consistent with relativity theory, which is widely thought to rule out or, at least, to be in tension with non-local action. This task will not be attempted here, although a few comments will be offered anon.

It seems pointless to try to develop a taxonomy of how-possible explanations since they are theory specific, and it is a mug's game to try to anticipate the ways in which yet unborn future theories will expand the possibility space.

5.2 *Why does the observed interference pattern emerge in the AB setup and, in particular, why the does the pattern exhibit a certain dependency on the flux in the solenoid?*

5.2.1 The Aharonov-Bohm *dynamical explanation*

The setup described in Section 2 seems to beg for a dynamical explanation. To recapitulate: a beam of charged particles is emitted from a source; the beam is split, and the two parts pass through the apparatus, around the solenoid, and register on a screen. What is it about the temporal evolution of the system that results in the observed interference pattern and its dependency on the flux in the solenoid? QM obliges with a dynamical explanation. The evolution of the state ψ of the charged particle is governed by unitary dynamics generated by the system's Hamiltonian. Waving the qualms expressed in Section 3, the Hamiltonian operator is the $\hat{H}_{\mathbf{A}}$ of eq. (1); and the state $\psi(t)$ at a time $t > 0$ is related to the state $\psi(0)$ at the time $t = 0$ of emission by a one-parameter group of unitary transformations whose generator is $\hat{H}_{\mathbf{A}}$, i.e. $\psi(t) = e^{-it\hat{H}_{\mathbf{A}}}\psi(0)$, the Schrödinger equation being the infinitesimal version. The interference pattern and its flux dependence are features of the solutions of the dynamical equation of motion. That is the aspirational form of the explanation; now it is high time to see the details of how the aspiration is fulfilled.

Before leaving the station, however, the train needs to be put on the right track. The above synopsis would be correct if the $\hat{H}_{\mathbf{A}}$ of eq. (1) were essentially self adjoint, for then with a slight abuse of notation we could use $\hat{H}_{\mathbf{A}}$

to denote the unique selfadjoint extension, and its exponentiation $e^{-it\hat{H}_{\mathbf{A}}}$ defines a one-parameter unitary group. But $\hat{H}_{\mathbf{A}}$ is not an essentially selfadjoint operator on the Hilbert space $L^2_{\mathbb{C}}(\Omega_{R\geq 0})$. It does have selfadjoint extensions, the choice of which correspond to the choice of boundary conditions on the wavefunction ψ at the solenoid boundary $\partial\Omega_R$. Often Dirichlet boundary conditions are assumed, either explicitly or implicitly, whereby $\psi = 0$ at $\partial\Omega_R$. So for the nonce let $\hat{H}_{\mathbf{A}}$ denote the selfadjoint operator corresponding to this choice. Now the train is ready to depart—all aboard.

Aharonov and Bohm write:

> In singly connected regions, where $\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A} = 0$, we can always obtain a solution for the above Hamiltonian [$\hat{H}_{\mathbf{A}} = \dfrac{(-i\boldsymbol{\nabla} - q\,\hat{\mathbf{A}})^2}{2m}$ in our notation] by taking $\psi = \psi_0 e^{-iS}$, where ψ_0 is a solution when $\mathbf{A} = 0$ and where $\nabla S = q\mathbf{A}$ (Aharonov and Bohm 2009, p. 486).[12]

The latter solution,$\psi = \psi_0 e^{-iS}$, is gauge equivalent to the former; indeed, in a simply connected region any solution for the Hamiltonian $\hat{H}_{\mathbf{A}}$ can be gauge transformed to another consisting of solution for the free Hamiltonian $\hat{H}_0$ multiplied by an appropriate phase factor. Aharonov and Bohm do not prove this, perhaps because they thought it too obvious.[13] But is instructive to see explicitly how it works.

In simply connected regions of the charged particle's configuration space, where $\boldsymbol{\nabla}\times\mathbf{A} = \mathbf{B} = 0$, the vector potential $\mathbf{A}$ can be gauge transformed away. In such a region $\mathbf{A}$ is the gradient of a scalar; and the gauge freedom allows the change $\mathbf{A} \to \mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\chi$, where χ is an arbitrary smooth scalar field, without changing the physical state of the system. So let $\mathcal{R}(L)$ and $\mathcal{R}(R)$ be simply connected regions containing respectively the left and right beams. By choosing $\chi = -\int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{A}\cdot d\mathbf{x}$, with $\mathbf{x}_0$ centered at the location of the beam splitter and $\mathbf{x} \in \mathcal{R}(L)$ or $\mathbf{x} \in \mathcal{R}(R)$ as the case may be, $\mathbf{A}$ can be gauge transformed to 0 along the paths of the beams, in which case the

[12]Notation has been altered to conform to that being used here. In our notation $S = \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{A}\cdot d\mathbf{x}$.

[13]Nor do they emphasize the gauge equivalence of the solutions, which is a crucial part of the derivation. And as we will see shortly, this equivalence undermines their suggestion that the puzzling non-locality in the AB effect can be avoided by positing that the electromagnetic potentials are physically effective.

charged particle propagates freely under the unitary evolution generated by the Hamiltonian $\hat{H}_0$ of $(1')$.

To continue, we seek a unitary operator $\hat{U}(\chi)$ to represent the Hilbert space action of gauge transformations $\mathbf{A} \to \mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\chi$ in simply connected regions. From the requirements that the position operator is invariant under $\hat{U}(\chi)$ and that $\hat{U}(\chi)$ acts on gauge dependent quantities such as magnetic momentum by $\hat{U}(\chi)(-i\boldsymbol{\nabla} - q\ \hat{\mathbf{A}})\hat{U}(\chi)^\dagger = -i\boldsymbol{\nabla} - q(\hat{\mathbf{A}} + \boldsymbol{\nabla}\chi)$ we can infer that $\hat{U}(\chi) = e^{iq\chi}$. And specifically in the case of interest where we want to gauge away $\mathbf{A}$, define $\chi_0 := -\int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{A}\cdot d\mathbf{x}$. Then $\hat{U}(\chi_0)\hat{H}_{\mathbf{A}}\hat{U}^\dagger(\chi_0) = \hat{H}_0$ where $\hat{U}(\chi_0) = e^{iq\chi_0}$. Of course we cannot gauge away $\mathbf{A}$ everywhere, but we can do it in a simply connected region. So let $\Lambda(L)$ (respectively, $\Lambda(R)$) stand for $\int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{A}\cdot d\mathbf{x}$ with the integral taken along a path in a simply region $\mathcal{R}(L)$ containing the left-hand beam (respectively, in a simply connected region $\mathcal{R}(R)$ containing the right-hand beam). Then in $\mathcal{R}(L)$, $\exp(-iq\Lambda(L))\hat{H}_{\mathbf{A}}\exp(iq\Lambda(L)) = \hat{H}_0$ and similarly for $\mathcal{R}(R)$.

To complete the missing proof let $\psi_L(\mathbf{x},t) = \exp(-it\hat{H}_{\mathbf{A}})\psi_L(\mathbf{x},0)$ be a solution for the Hamiltonian $\hat{H}_{\mathbf{A}}$ in the simply connected region $\mathcal{R}(L)$; and let $\psi_{0,L}(\mathbf{x},t) = \exp(-it\widehat{H}_0)\psi_{0,L}(\mathbf{x},0)$ be a solution for the free Hamiltonian $\hat{H}_0$ also in the simply connected region $\mathcal{R}(L)$. And assume that in $\mathcal{R}(L)$ we can validly go from $\exp(iq\Lambda(L))\hat{H}_0\exp(-iq\Lambda(L)) = \hat{H}_{\mathbf{A}}$ to
$\exp(iq\Lambda(L))\exp(-it\widehat{H}_0)\exp(-iq\Lambda(L)) = \exp(-it\widehat{H}_{\mathbf{A}})$. Then

$$\begin{aligned}\exp(iq\Lambda(L))\psi_{0,L}(\mathbf{x},t) &= (\exp(iq\Lambda(L))\exp(-it\widehat{H}_0)\exp(-iq\Lambda(L)))(\exp(iq\Lambda(L))\psi_{0,L}(\mathbf{x},0)) \\ &= \exp(-it\hat{H}_{\mathbf{A}})(\exp(iq\Lambda(L))\psi_{0,L}(\mathbf{x},0)). \qquad (10)\end{aligned}$$

And choosing $\psi_{0,L}(\mathbf{x},0) = \exp(-iq\Lambda(L))\psi_L(\mathbf{x},0)$ we have

$$\psi_L(\mathbf{x},t) = \exp(iq\Lambda(L))\psi_{0,L}(\mathbf{x},t). \qquad (11)$$

And of course a similar result holds for $\mathcal{R}(R)$.

Aharonov and Bohm continue:

> [I]n the experiment discussed above, in which we have a multiply connected region (the region outside the solenoid), $\psi_0 e^{-iS}$ is not a single-valued function and, therefore is not a permissible solution of Schrödinger's equation. Nevertheless, in our problem it is still possible to use such solutions because the wave functions splits into two parts $\psi = \psi_1 + \psi_2$, where ψ_1 represents the beam on

one side of the solenoid and ψ_2 on the opposite side. Each of the beams stays in a simply connected region. We can therefore write $\psi_1 = \psi_1^0 e^{-iS_1}$, $\psi_2 = \psi_1^0 e^{-iS_2}$ where S_1 and S_2 are equal to $q\int \mathbf{A}\cdot d\mathbf{x}$ along the paths of the first and second beams, respectively ... The interference between the two beams will evidently depend on the phase difference" $S_1 - S_2$. (Aharonov and Bohm 1959, pp. 486-487)[14]

In our treatment, this works out as follows. The unitary evolution $\Psi(\mathbf{x},t) = \exp(-it\hat{H}_{\mathbf{A}})\Psi(\mathbf{x},0)$ generated by $\hat{H}_{\mathbf{A}}$ of an initial superposition $\Psi(\mathbf{x},0) = \psi_L(\mathbf{x}) + \psi_R(\mathbf{x})$ of wavepackets $\psi_L(\mathbf{x})$ and $\psi_R(\mathbf{x})$ centered at $\mathbf{x}_0$ and directed to the left and right beams respectively will be given by free evolution in each of the regions $\mathcal{R}(L)$ and $\mathcal{R}(R)$ multiplied respectively by the phase factors $\exp(iq\Lambda(L))$ and $\exp(iq\Lambda(R))$. So for a point $\mathbf{x}$ lying in the overlap of $\mathcal{R}(L)$ and $\mathcal{R}(R)$ the unitary evolution of $\Psi(\mathbf{x},0)$ is given by

$$\begin{aligned}\Psi_{AB}(\mathbf{x},t) &= [\exp(iq\Lambda(L))\exp(-it\widehat{H}_0)(\exp(-iq\Lambda(L))\psi_L(\mathbf{x})] \\ &\quad + \exp(iq\Lambda(R))\exp(-it\widehat{H}_0)[\exp(-iq\Lambda(R))\psi_R(\mathbf{x})] \\ &= [\exp(iq\Lambda(L))\exp(-it\widehat{H}_0)\psi_{0,L}(\mathbf{x},0)) \\ &\quad + \exp(iq\Lambda(R))\exp(-it\widehat{H}_0)\psi_{0,R}(\mathbf{x},0))] \\ &= \exp(iq\Lambda(R))[\exp(i(q\Lambda(L) - q\Lambda(R))(\psi_{0,L}(\mathbf{x},t) + \psi_{0,R}(\mathbf{x},t))]\end{aligned} \tag{12}$$

assuming that the left and right wavepackets remain in $\mathcal{R}(L)$ and $\mathcal{R}(R)$ respectively. The interference between the two beams depends on the phase difference $q\Lambda(L) - q\Lambda(R)$, and when $\mathbf{x}$ lies on the screen $q\Lambda(L) - q\Lambda(R)$ is equal to $q\oint \mathbf{A}\cdot d\mathbf{x}$, the integral of $\mathbf{A}$ around a closed path starting at $\mathbf{x}_0$, then going clockwise to the screen, and returning clockwise to $\mathbf{x}_0$ while encircling the solenoid. And this integral is equal to q x flux in the solenoid—exactly the result needed to explain the interference pattern's dependency on the magnetic flux.

But the assumption that the left and right wavepackets remain in $\mathcal{R}(L)$ and $\mathcal{R}(R)$ respectively is false, for tails of the wave packets will quickly extend over the entire configuration space available to the charged particles. Thus, the analysis comes with the implicit *Ansätz* that $\Psi_{AB}(\mathbf{x},t)$ is a good approximation to evolution for the Hamiltonian $\hat{H}_{\mathbf{A}}$ to the extent that the

[14]Again notation has been altered to conform to that being used here.

wave packets in the two beams remain localized in their respective simply connected regions $\mathcal{R}(L)$ and $\mathcal{R}(R)$.

It is only relatively recently that the *Ansätz* has been rigorously shown to provide a good approximation to the exact Schrödinger solution for high velocity Gaussian wavepackets (see Ballesteros and Weder 2009).[15] The publication occupies over fifty pages of dense argumentation in *Journal of Mathematical Physics*, which helps to explain why, despite its importance, the AB *Ansätz* did not receive a rigorous justification until the lapse of fifty years from the publication of Aharonov and Bohm's seminal paper.

The explanation reviewed above may be seen as an instance of the DN model, although what is important here is not the logical deduction in some formal system of a sentence describing the explanandum events from the laws together with sentences describing initial and boundary conditions but rather the proof that solutions to the equations governing dynamical evolution exhibit the features to be explained. The explanation may also be deemed causal on the productive conception of causation, whereby causes generate, engender or produce their effects,[16] at least if the dynamical unfolding of events is regarded as generating later events from their temporal predecessors. But it does not fit comfortably with some of the leading philosophical accounts of causal explanation such as Salmon's causal mechanical model with its emphasis on causal processes characterized as the ability to transmit a "mark" (Salmon 1984) or Dowe's conserved process theory of causation (Dowe 2000). However, interventionist accounts of causation whereby interventions in a system are employed to discern causal dependencies among the variables used to describe the behavior of the system (see, for example, Woodward 2003) seem to be on the mark.[17] In our case the experimenter can intervene by adjusting the current supplied to the solenoid, thereby changing the amount of flux in the solenoid which, in turn, changes the interference pattern. These dependencies are a crucial part of the AB effect, and their seeming nonlocal character constitutes its most puzzling and controversial aspect.

[15] Ballesteros and Weder study a variant of the AB setup described above, using toroidal magnets. Such an arrangement avoids having to use the idealization of an infinitely long solenoid. For the experiments confirming the AB effect for such magnets see Osakabe et al. (1986) and Tonomura et al. (1986).

[16] See Hall (2004) "Two Concepts of Causation."

[17] Here I am indebted to Gordon Belot for emphasizing this point.

5.2.2 Gauge-independent vs. gauge-dependent

In its present form the dynamical explanation might be deemed unsatisfactory on the grounds that a proper explanation should be couched in terms of observables.[18] The explanation does use self-adjoint operators; but while observables are represented by self-adjoint operators not all such operators represent observables, as we know from the existence of superselection rules. The magnetic momentum operators $\hat{\mathbf{P}}_x = -i\dfrac{d}{dx} - qA_x$ and $\hat{\mathbf{P}}_y = -i\dfrac{d}{dy} - qA_y$ and the Hamiltonian $\hat{H}_{\mathbf{A}}$ represent gauge-dependent quantities, and the phases $\Lambda(L)$ and $\Lambda(R)$ are gauge-dependent as well. (But, of course, the phase difference $\Lambda(L) - \Lambda(R)$ is a gauge-invariant. And this difference *is* observable in the interference effects.) Changes in the electromagnetic potentials need not correspond to change in the physical state of the system (gauge freedom); to hold otherwise implies an abandonment of determinism at the classical level for the electromagnetic field since, when posed in terms of the electromagnetic potentials, the Cauchy problem for Maxwell's equations does not have a unique solution.

To fulfill the demand for an explanation couched in terms of observables it is necessary to provide a gauge-independent description of the AB effect. That it should be possible to meet the demand is indicated by that fact that the Schrödinger equation using $\hat{H}_{\mathbf{A}}$ is gauge-invariant (see Aitchison and Hey 2001, Sec. 2.4). Nevertheless, it takes some work to reach the goal. Li et al. (2022) reach it by utilizing the path integral approach to QM, an example of non-canonical quantization. In this approach the transition amplitude from $\psi(\mathbf{x}_0, 0)$ to $\psi(\mathbf{x}, t)$ is calculated by summing the contributions associated with all the paths from $\mathbf{x}_0$ to $\mathbf{x}$. The contribution of a path is given by $\exp(i\int_0^t Ldt)$ where the action $\int_0^t Ldt$ is computed from the classical Lagrangian L. For present purposes assume that L is given by eq. (3). The contribution of the interaction portion $L_{int} = q\mathbf{A}\cdot\boldsymbol{\nu}$ of this Lagrangian is $\exp(i\int_0^t q\mathbf{A}\cdot\dfrac{\mathbf{dx}}{dt}dt) = \exp(i\int_{\mathbf{x}_0}^{\mathbf{x}} q\mathbf{A}\cdot\mathbf{dx})$. A point $\mathbf{x}$ on the screen can be reached by a path that passes the solenoid either on the left or on the right, and the interference

[18] As used in QM "observable" is a term of art referring not to a quantity whose value can be ascertained by direct observation but (something like) 'a quantity that is in principle measurable by using (perhaps complicated) instruments where the connection between the instrument reading and the value of said quantity may require the use of auxiliary theories.' It is something of a scandal that this important notion does not have a more precise definition.

between the two is given by $\exp(iq(\int_{\mathbf{x}_0}^{\mathbf{x}} q\mathbf{A}\cdot\mathbf{dx}_L - \int_{\mathbf{x}_0}^{\mathbf{x}} q\mathbf{A}\cdot\mathbf{dx}_R))$, where $\mathbf{dx}_L$ and $\mathbf{dx}_R$ indicate that integration is taken along the left and right paths respectively (see Li et al. 2022), and this quantity is equal to $\exp(i\oint q\mathbf{A}\cdot\mathbf{dx})$. Note that this approach is innocent of the complications arising in canonical quantization due to the existence of Schrödinger-inequivalent representations of the CCR, and assuming that the classical Lagrangian (3) is the starting point, the path integral approach gives an unequivocal prediction of the AB effect.[19]

Now the trick is to show that the contribution to the action of the interaction term $L_{int} = q\mathbf{A}\cdot\boldsymbol{\nu}$ can be computed in terms of gauge invariants. This is achieved by including the electromagnetic fields as part of the quantum system and, specifically, including in the action the energy stored in the electromagnetic field generated by both the solenoid and the charged particle.[20] L_{int} is shown to equal δL_F, electromagnetic field generated by the charged particle, and as a result the Aharonov-Bohm phase can be obtained "without using a vector potential, by expressing the Lagrangian in terms of the gauge-invariant magnetic field in all space" (Li et al. 2022). Not surprisingly, however, the price to be paid for the gauge-invariant description is to make manifest that the magnetic field $\mathbf{B}$ has non-local effects.

5.2.3 Non-locality and the status of the electromagnetic potentials

There is pushback in the literature against the demand for a gauge-independent explanation. Aharonov et al. (2016) claim that a gauge-dependent description is required to handle instantaneous aspects of the AB effect. A less controversial and more widespread motivation for a gauge-dependent explanation starts from the premise that relativity theory and locality are in tension if not outright contradiction; and then it concludes that if we do not want to run afoul of relativity theory the AB effect shows that we must

[19]However, the ultimate goal of a quantum theory is to compute the expectation values of observables, and this requires representing observables as linear operators acting on a Hilbert space. The path integral approach offers little help here.

By contrast the program of geometric quantization has the goal of solving the representation problem and, moreover, it seeks to do so when the classical configuration and phase spaces are topologically non-trivial, as in the case of the AB effect. Some brief comments on geometric quantization in relation to our subject are found in the Appendix.

[20]The charged particle is assumed to be moving slowly enough that its radiation effects can be neglected.

recognize the electromagnetic potentials as real or as "physically effective, even when there are no fields acting on the charged particles" (Aharonov and Bohm 1959, p. 490).

The careless invocation of the notion that relativity theory is in conflict with non-local action without supplying a precise specification of what "non-locality" means and without giving a proof of the alleged inconsistency has done untold mischief. To give one example of the problematic nature of this notion: It might be thought that it is impossible to have non-trivial Lorentz invariant equations of motion for particles acting instantaneously-at-a-distance (e.g. equations that have solutions in which the particle world-lines are not timelike geodesics of Minkowski spacetime). This is demonstrably false. Additionally, the original platform for the AB effect is not suited to fruitful discussion of issues of locality and relativity theory. Besides some bits and pieces of electromagnetic theory, the mainstay of that platform is ordinary non-relativistic QM. The spacetime setting for ordinary QM is Newtonian spacetime (or more properly neo-Newtonian spacetime whose symmetry group is the inhomogeneous Galilean group) which, of course, is compatible with instantaneous action-at-a-distance; and, not surprisingly, ordinary QM countenances the instantaneous spread of effects; e.g. a particle initially trapped in some small neighborhood of the origin of spatial coordinates and released at $t = 0$ spreads infinitely fast—for any $t > 0$ and any $\mathbf{x} > \mathbf{0}$ no matter how far from the origin there is a non-zero probability for the particle to be detected in some small neighborhood $\mathbf{x}$. The proper setting for treating issues of locality and compatibility of relativity theory raised by the AB effect is relativistic quantum field theory, set in Minkowski space, permitting the formulation both of various requirements of relativistic invariance and of locality conditions whose satisfaction or violation can be studied with precision. This will not be attempted here.

What can be said here—and what needs to be underscored because, though obvious, is often overlooked—is that the quest to secure locality by promoting the electromagnetic potentials to physically effective fields undermines Aharonov's and and Bohm's explanation of the AB effect. Recall a key move in the explanation: the Schrödinger evolution under the Hamiltonian $\hat{H}_{\mathbf{A}}$ in each of the regions $\mathcal{R}(L)$ and $\mathcal{R}(R)$ is replaced by free evolution multiplied by phase factors appropriate to $\mathcal{R}(L)$ and $\mathcal{R}(R)$. The replacement is justified *if* the transformation $\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\chi$, where χ is an arbitrary smooth scalar field, is a gauge transformation in the sense that it represents a change in the description of a physical state of affairs without changing the physical state itself. But this is precisely what is denied by promoting the vector potential $\mathbf{A}$ to a real physical field.

Aharonov and Bohm were, of course, aware of the tension, and their struggle with it is palpable. After presenting the AB effect they opine that

> It would therefore seem natural at this point to propose that, in quantum mechanics, the fundamental physical entities are the potentials, while the fields are derived from them by differentiations. (1959, p. 490)

But then they immediately add that

> The main objection that could be raised against the above suggestion is grounded in the gauge invariance of the theory. In other words, if the potentials are subject to the transformation $\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \boldsymbol{\nabla}\chi$, then all the known physical quantities are left unchanged. As a result, the same physical behavior is obtained from any two potentials, $\mathbf{A}(\mathbf{x})$ and $\mathbf{A}'(\mathbf{x})$, related by the above transformation. This means that insofar as the potentials are richer in properties than the fields, there is no way to reveal this additional richness. It was therefore concluded that the potentials cannot have any meaning, except insofar as they are used mathematically, to calculate the fields. (1959, p. 490. Notation changed to conform to ours)

To this they add a disclaimer: "We have seen from the examples described in this paper that the above point of view cannot be maintained for the general case" (1959, p. 490). What is clear from the context is that what they mean is that the electromagnetic potentials cannot be mere gauge quantities *if* all interactions are to be treated as local.

Two possible options for further development are offered: The first is to formulate a nonlocal theory in which the electron could interact with a field that was a finite distance away. But this option, they say, entails "severe difficulties." The second option is "to regard $\mathbf{A}(\mathbf{x})$ as a physical variable. This means that we must be able to define the physical difference between two quantum states which differ only by gauge transformation." This they promise to do in a future paper which, to my knowledge, never appeared in print. What gets lost in this internal dialectic is that the second option is incompatible with their own derivation of the AB effect which depends on the gauge interpretation of $\mathbf{A}(\mathbf{x})$. This tension is relieved in alternative derivations of the effect, one of which was considered above and the other of which is considered in the following section. But neither of them promotes $\mathbf{A}(\mathbf{x})$ to the status of a physical variable; and, indeed, one of them—using the path integral approach—shows that the effect can be explained in terms of the magnetic field without invoking $\mathbf{A}(\mathbf{x})$. What I am suggesting is that if Aharonov and Bohm had followed the logic of their own derivation of thee AB effect they would have come to the conclusion that the resulting non-locality is something that needs to be acknowledged and come to terms with rather than being run away from.

5.2.4 *Non-dynamical explanation*?

Arai (2020) presents what is termed a "purely topological" interpretation of the AB effect. The AB effect is most surely a topological effect in that it derives from the topology of the configuration space of the charged particle and, in particular it derives from the non-simple connectedness of this space which opens the door to inequivalent representations of the CCR. But for present purposes the important point about Arai's interpretation is that it is non-dynamical in the sense that it is not dependent on the details of the Hamiltonian:

> [I]n this interpretation, the AB effect in the present context is purely topological and independent of the details of dynamics of the system (forms of Hamiltonian). (p. 165)

The result reported in eq. (9) is suggestive of the AB effect. The suggestion is parlayed into an explicit topological/non-dynamical interpretation of the AB effect as follows. The Schrödinger momenta operators $\hat{\mathbf{p}}_x = -i\dfrac{d}{dx}$ and $\hat{\mathbf{p}}_y = -i\dfrac{d}{dy}$ are respectively the generators of translation of the wave function ψ along the x- and y-axes. The "physical momenta" $\hat{\mathbf{P}}_x = \hat{\mathbf{p}}_x - qA_x$ and $\hat{\mathbf{P}}_y = \hat{\mathbf{p}}_y - qA_y$ are interpreted as generators of translation of ψ under the existence of a magnetic field. Hence $e^{is_y\hat{\mathbf{P}}_y}e^{is_x\hat{\mathbf{P}}_x}\psi$ gives the result of first translating ψ under the influence of a magnetic field a distance s_x along the straight line $(x, y) \rightarrow (x + s_x, y)$ and thence a distance s_y along the straight line $(x+s_x, y) \to (x+s_x, y+s_y)$, while $\exp(is_x\hat{\mathbf{P}}_x)\exp(is_y\hat{\mathbf{P}}_y)\psi$ gives the result of first translating ψ under the influence of a magnetic field a distance s_y along the straight line $(x, y) \to (x, y+s_y)$ and thence a distance s_x along the straight line $(x, y+s_y) \to (x+s_x, y+s_y)$ (see Fig. 2). The value of $(\exp(is_x\hat{\mathbf{P}}_x)\psi)(\mathbf{x})$ is equal to $\psi(\mathbf{x}+ s_x\mathbf{e}_x)$, where $\mathbf{e}_x$ is a unit vector in the x-direction, multiplied by a phase factor of $-iq\int_0^{s_x} A_x(\mathbf{x}+ s\mathbf{e}_x)ds$, and similarly for the other terms (Arai 2020, Theorem 3.2). The upshot is that $\exp(is_x\hat{\mathbf{P}}_x)\exp(is_y\hat{\mathbf{P}}_y)\psi$ and $\exp(is_y\hat{\mathbf{P}}_y)\exp(is_x\hat{\mathbf{P}}_x)\psi$ differ by a phase factor $\exp(-iq\Phi^A_{s_x,s_y})$, where (recall) that $\Phi^A_{s_x,s_y}$ is the integral of $\mathbf{A}$ around the closed curve; and if the curve surrounds the solenoid $\Phi^A_{s_x,s_y}$ is equal to the amount of magnetic flux through a two-surface bounded by the curve.

> Physically this means that the phase of the state function of the charged particle starting at (x, y) and arriving at $(x+s_x, y+s_y)$ through C_- $((x, y) \to (x+s_x, y) \to (x+s_x, y+s_y))$ does not coincide with that through C_+ $((x, y) \to (x, y+s_y) \to (x+s_x, y+s_y))$, and a phase shift occurs. This is exactly interpreted as the AB effect in the present context. (Arai 2020, p. 165).[21]

And notably, the phase factor $\exp(-iq\Phi^A_{s_x,s_y})$ is non-trivial ($\neq 1$) iff the physical momenta $\hat{\mathbf{P}}_x$, $\hat{\mathbf{P}}_y$ along with the usual position operators $\hat{x}$, $\hat{y}$ give a representation of the (HCCR) inequivalent to the Schrödinger representation. Thus, on this interpretation, the AB effect invokes inequivalent representations in the sense that it occurs in exactly the circumstances that inequivalent representations exist.

The mathematics of the topological/non-dynamical interpretation of the AB effect is certainly interesting, but how does it explain the behavior of charged particles in the AB setup? It doesn't, at least not if what is to be explained are properties that are exhibited in a temporal process—the transiting of charged particles through the apparatus and their registration on the screen. It does suggest an alternative to the AB *Ansätz*; namely, the state function $\psi(\mathbf{x}, t_?)$ at some appropriate $t_? > 0$ of a charged particle whose state $\psi(\mathbf{x}, 0)$ at $t = 0$ is a Gaussian wave packet $\psi(\mathbf{x})$ which is centered at the location $\mathbf{x}_0$ of the beam splitter and which traverses the path C_- is given by

[21] Notation has been changed to conform to conventions used herein. The "does not coincide with" should be read as containing the proviso that $\exp(-iq\Phi^A_{s_x,s_y}) \neq 1$, which means that $q\Phi^A_{s_x,s_y}$ is not an integer multiple of 2π.

$(\exp(is_y\hat{\mathbf{P}}_y)\exp(is_x\hat{\mathbf{P}}_x)\psi)(\mathbf{x})$, and is given by $(\exp(is_x\hat{\mathbf{P}}_x)\exp(is_y\hat{\mathbf{P}}_y)\psi)(\mathbf{x})$ for a charged particle that traverses C_+. But what time is $t_?$? The "arriving at" phrase in the above quotation suggests that $t_?$ should be the time of arrival of the wave packets at the location of the recombining of the two beams at the screen. But this suggestion means that, after all, the explanation is really a dynamical one; and the justification of the alternative *Ansätz* must consist, as does the justification for the AB *Ansätz*, in showing that it holds to good approximation as a result of the exact Schrödinger evolution.

The desire for an explanation of the AB effect that is independent of the details of the Hamiltonian is understandable, but it has to be tempered by the fact that the details of the dynamics do matter. As mentioned above the quantum Hamiltonian operator (1) is not essentially selfadjoint when the configuration space of the charged particle is $\Omega_{R\geq 0}$ (finite or zero radius solenoid); it does have selfadjoint extensions, indeed, infinitely many. Until some selfadjoint extension is singled out, dynamical explanations of the behavior of charged particles in the AB setup do not get off the ground. The choice of a selfadjoint extension of the Hamiltonian amounts to a choice of boundary conditions at the surface of the solenoid. For wave packets whose trajectories intersect minimally with the solenoid the choice of boundary conditions should not matter much if at all. (This was apparently Aharanov's and Bohm's intent in their seminal (1959) as indicated by their diagram, whose main features are reproduced in Fig. 1, showing the solenoid lying in the shadow of a metal plate.) But for wave packets that scatter off the solenoid the choice does matter for predicting scattering amplitudes (see de Oliveira and Pereira 2010). The topological interpretation as it stands seems to have no way to address this matter and other matters where the details of the dynamics matter.

6 Conclusion

How then does QM explain the AB effect? The most common explanation is a dynamical one: The AB setup, as illustrated in Fig. 1, leads to the quantum Hamiltonian $\hat{H}_{\mathbf{A}}$ of eq. (1) that generates unitary dynamics in which the wavepackets of charged particles passing on opposite sides of the solenoid experience differential phase changes that result in an observable interference pattern that is dependent on the magnitude of magnetic flux in the solenoid, even though the charged particles never encounter the magnetic field. But the explanation/justification of $\hat{H}_{\mathbf{A}}$ remains elusive. The most direct route to $\hat{H}_{\mathbf{A}}$ starts from the classical Lagrangian (3) and Hamiltonian (6) for a charged particle moving in a magnetic field and then quantizes (6) by substituting for the classical momentum and position variables Hilbert space operators that give a Schrödinger representation of the Heisenberg CCR. There are two problems here. First, why start from the classical Lagrangian (3) and Hamiltonian (6) for a charged particle moving in a magnetic field when, by construction, the charged particles in an AB setup do not encounter a magnetic field? Second, since representations unitarily inequivalent to the Schrödinger representation are available in the AB setup, why use the Schrödinger representation? There is nothing in standard QM that implies that Nature will or must make this choice and, thus, no way to know in advance of doing the experiment that the interference pattern characteristic of the AB effect will be detected. The upshot is that on the canonical approach to quantization, while QM accommodates the AB effect it does not predict it.

The path integral formulation of QM is an example of non-canonical quantization that bypasses issues about inequivalent representations of the CCR and predicts the AB phase, at least if the classical Lagrangian (3) is used to define the action associated with a path in the charged particle's configuration space. The approach can also be used to overcome an awkwardness in the most common explanation of the AB effect; namely, the explanans uses gauge-dependent variable to while the explanandum is a gauge-independent effect. By including the magnetic field as part of the quantum system the path integral derivation of the AB phase can be couched in terms of the gauge-independent magnetic field rather than the gauge-dependent electromagnetic potentials. But this derivation makes manifest the non-local dependence of the interference effect on the magnetic field. A number of commentators find such non-locality intolerable, and propose to restore locality by recognizing the electromagnetic potentials as real physical variables. Their motivation is often based on the debatable premise that relativity theory is inconsistent with non-locality, and in any case a robust sense of the physical effectiveness of the electromagnetic potentials undermines the Aharonov-Bohm dynamical explanation of the AB effect which relies on the gauge equivalence of Schrödinger evolution under $\hat{H}_{\mathbf{A}}$ and free evolution multiplied by a phase factor. Nor do alternative derivations support a non-gauge interpretation of the electromagnetic potentials.

Alternative non-dynamical (a.k.a. "topological") explanations of the AB effect that do not depend on the details of the Hamiltonian offer the promise of avoiding some of these contentious issues. But the extant attempts at a non-dynamical treatment do not provide a well motivated and coherent explanation. And unlike a dynamical explanation they appear to remain silent about observable effects in scenarios that differ from the original AB setup, e.g. where the solenoid is not shielded from the beams of charged particles which are allowed to scatter off the solenoid.

Little help on these matters can be expected from the philosophical literature on scientific explanation which seems consumed with internal disputes, e.g. can the "mark method" distinguish between genuine and pseudo-causal processes; can the DN model cope with perceived explanatory asymmetries, and the like. The one exception is the interventionist account of causal explanation, which substantiates the causal dependency of the observed interference pattern on the amount of flux in the solenoid, i.e. the experimenter can intervene by changing the current supplied to the solenoid, which changes the flux, which results in a change in the interference pattern. Nevertheless, I venture that it will be fruitful to spend more effort on the study of concrete cases of explanation in physics before fashioning and debating philosophical "models of explanation"; and I am skeptical that any one, or even any finite number, of models of explanation will capture, in anything more than broad brush strokes, the variety and nuances of explanations in physics.

Appendix on geometric quantization

Good general references are Woodhouse (1992) and Ali and Engliš (2005). The ambition of the geometric quantization program is to start from an arbitrary classical phase space in the guise of a $2n$ dim manifold Γ equipped with a symplectic form ω.[22] Classical observables are construed as phase functions, real valued functions f on Γ. Quantization means finding a Hilbert space $\mathcal{H}$ and a map $f \mapsto \widehat{f}$ from classical observables f to selfadjoint operators $\widehat{f}$ on $\mathcal{H}$. The major idea of the program is encapsulated in the first axiom

> (A1) For pairs f, g of classical observables the CCR are satisfied in the sense that $[\widehat{f}, \widehat{g}] = i\widehat{\{f, g\}}$ where $\{,\}$ is the Poisson bracket

[22] ω is a closed non-degenerate 2-form defined on all of Γ.

given by $\{f,g\} = -\omega(X_f, X_g)$ and X_f is the Hamiltonian vector field determined by $\iota_{X_f}\omega = -df$.[23]

In the case where $\Gamma = \mathbb{R}^{2n}$ we can choose global coordinates q^i, p_i where $\omega = dq^i \wedge dp_i$, and the Poisson bracket takes the familiar form $\{f,g\} = \sum_{i,j=1}^{n} \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i}$.[24]

Some other seemingly natural restrictions on the association $\mapsto$ suggest themselves:

(A2) The map $f \mapsto \widehat{f}$ is linear, i.e. $\widehat{f+g} = \widehat{f} + \widehat{g}$.

(A3) $\widehat{1} = I$, where 1 is the constant function of value unity.

It might also seem natural to demand that

(A4) For smooth $\phi : \mathbb{R} \to \mathbb{R}$, $\widehat{\phi \circ f} = \phi(\widehat{f})$.

Unfortunately, (A1)-(A4) are inconsistent if they are supposed to hold for arbitrary phase functions. One reaction is to restrict or abandon (A4) and also restrict the domain of the mapping $\mapsto$ to a subset *Obs* of quantizable classical observables which is small enough so as to avoid inconsistency and but which is large enough to cover the system at issue. It is concerning that a surprising amount of mathematical maneuvering is required to satisfy (A1)-(A3) for an *Obs* that accommodates Hamiltonians as bland as that for a harmonic oscillator.

Although (A4) may need to be restricted, it is certainly desirable to retain special cases. For instance, when Γ may not be $\mathbb{R}^{2n}$ but admits global coordinates q^i, p_i with $\omega = dq^i \wedge dp_i$ (as in the AB setup where $\mathcal{Q} = \mathbb{R}^2 \backslash \Omega_{R \geq 0}$ is the configuration space and Γ is the cotangent space $T^*\mathcal{Q} \subseteq \mathbb{R}^{2n}$), f is q^i or p_i, and $\phi(t) = t^n$, we would particularly like to have $\widehat{(q^i)^n} = (\widehat{q^i})^n$ and $\widehat{(p_i)^n} = (\widehat{p_i})^n$. Then in keeping with the original canonical quantization program we know that the quantization of a classical Schrödinger type Hamiltonian

[23] The interior derivative $\iota_{X_f}\omega$ sends the 2-form ω to the 1-form given by $\omega(X_1) = \omega(X_f, X_1)$.

[24] To conform to standard notation for symplectic geometry I use q^i rather than x^i for spatial coordiantes.

$\dfrac{\mathbf{p}^2}{2m}+V(\mathbf{q})$ is given by $\dfrac{\widehat{\mathbf{p}}^2}{2m}+V(\widehat{\mathbf{q}})$. A quantization scheme without some systematic connection between the quantization of the position and momentum observables on one hand and quantization of the Hamiltonian on the other seems incomplete.

For the AB setup does geometric quantization, with the noted provisos on axioms (A1)-(A4), accommodate the inequivalent representations of the CCR that arise in canonical quantization? If not, why not, and is something amiss with the scheme? If so, then the issues discussed above about the prediction/explanation of AB effect as a dynamical effect arise for geometrical quantization as well.

Acknowledgment: Many thanks to the Club—Laura Ruetsche, Gordon Belot, and David Baker—for helpful suggestions and corrections. Any remaining errors of omission or commission are entirely mine. And my gratitude to my long-suffering proofreader, Frances, who does not understand why her husband would want to spend his time writing such boring papers.

References

[1] Aharonov, Y., Cohen, E., and Rohrlich, D. 2016. "Nonlocality of the Aharonov-Bohm effect," *Physical Review A* 93: 042110.

[2] Aharonov, Y. and Bohm, D. 1959. "Significance of Electromagnetic Potentials in the Quantum Theory," *Physical Review* 115: 485-491.

[3] Ali, S. T. and Engliš, M. 2005. "Quantization Methods: A Guide for Physicists and Analysts," *Reviews in Mathematical Physics* 17: 391-490; https://arxiv.org/abs/math-ph/0405065v1.

[4] Aitchison, I. J .R., and Hey, A .J. G. 2001. *Gauge Theories in Particle Physics* (2nd ed.). Bristol: Institute of Physics Publishing.

[5] Arai, A. 2020. *Inequivalent Representations of Canonical Commutation and Anti-Commutation Relations*. Singapore: Springer Nature.

[6] Ballesteros, M. and Weder, R. 2009. "The Aharonov–Bohm effect and Tonomura et al. experiments: Rigorous results," *Journal of Mathematical Physics*: 50: 122108.

[7] de Oliveira, C. R. and Pereira, M. 2010. "Scattering and self-adjoint extensions of the Aharonov-Bohm Hamiltonian," *Journal of Physics A* 43: 354011-1-29.

[8] Dowe, P. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

[9] Earman, J. 2023. "As Revealing in the Breach as in the Observance: von Neumann's Uniqueness Theorem," preprint http://philsci-archive.pitt.edu/id/eprint/22115.

[10] Hall, N. 2004. "Two Concepts of Causation," J. Collins, N. Hall, and L. Paul (eds.), *Causation and Counterfactuals*. MIT Press. pp. 225-276.

[11] Hirokawa, M. 1997. "Weyl's Relation on a Doubly Connected Space and the Aharonov-Bohm Effect," *Publications of the Research Institute for Mathematical Sciences* 982: 240–257.

[12] ____________ 2000. "Canonical Quantization on a Doubly Connected Space and the Aharonov Bohm Phase," *Journal of Functional Analysis* 174: 322 363.

[13] Kitcher, P. 2021. "The Theory of Scientific Explanation: An Obituary," preprint.

[14] Li, X., Hansson, T. H., and Ku, W. 2022. "Gauge-independent description of the Aharonov-Bohm effect," *Physical Review A* 106: 032217.

[15] Osakabe, N., Matsuda, T., Kawasaki, T., Endo, J., Tonomura, A., Yano, S., and Yamada, H.. 1986. "Experimental confirmation of Aharonov-Bohm effect using a toroidal magnetic field confined by a superconductor," *Physical Review A* 34: 815-.

[16] Reeh, H. 1988. "A remark concerning canonical commutation relations," *Journal of Mathematical Physics* 29: 1535-1536.

[17] Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

[18] Tonomura, A., Osakabe, N., Matsuda, T., Kawasaki, T., Endo, J., Yano, S., and Yamada, H.. 1986. "Evidence for Aharonov-Bohm effect with magnetic field completely shielded from electron wave," *Physical Review Letters* 56: 792.

[19] Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

[20] Woodward, J. and Ross, L. 2021. "Scientific Explanation," *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/

[21] Woodhouse, N. M. J. 1992. *Geometric Quantization*, 2nd ed. Oxford: Clarenden Press.

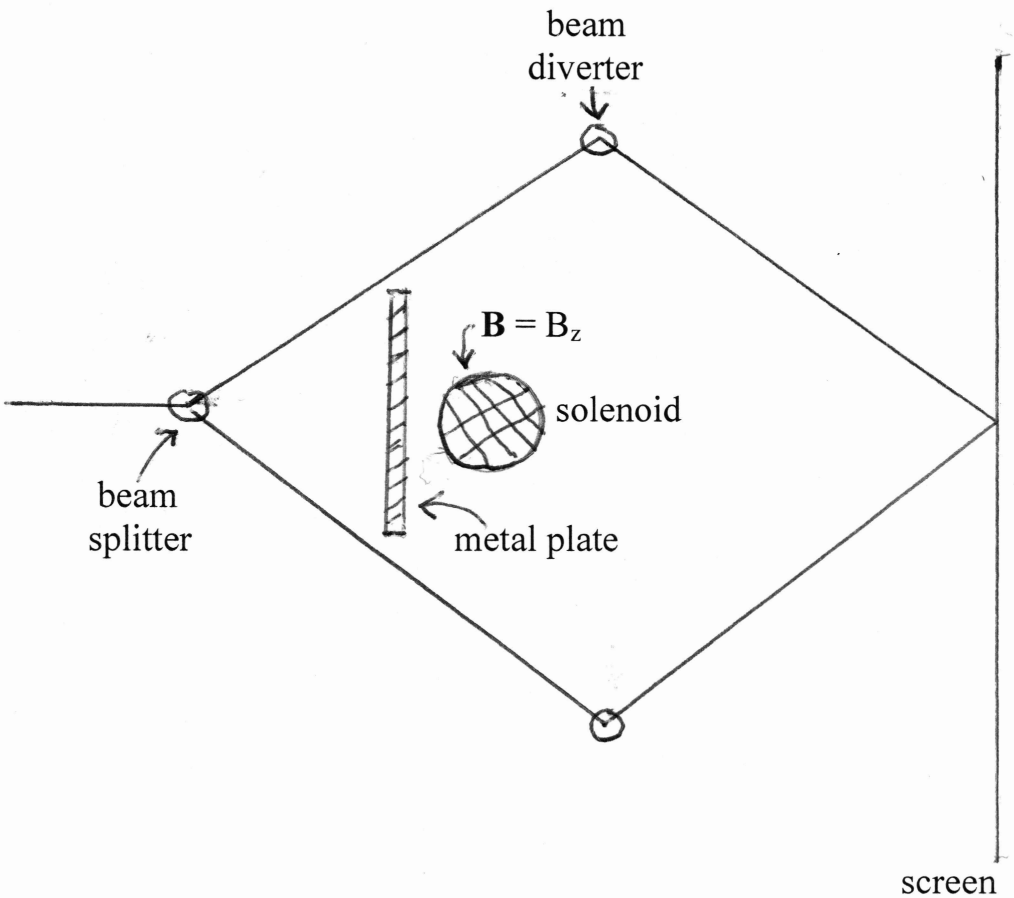
beam
diverter
$\mathbf{B} = B_z$
solenoid
beam
splitter
metal plate
screen

Fig. 1

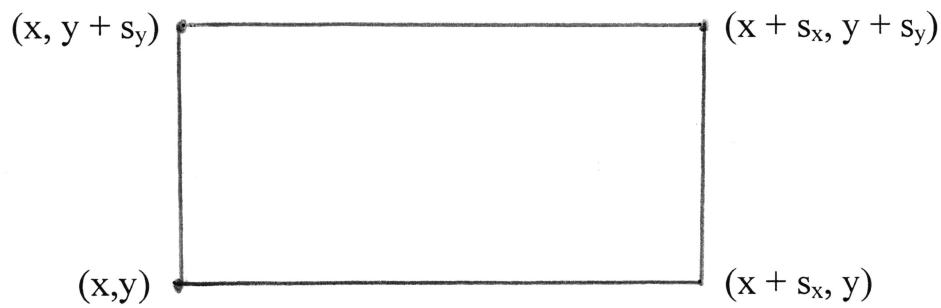

Fig. 2