The Aharonov-Bohm debate in 3D

Alexander Afriat

July 1, 2018

Abstract

Going from two dimensions (curl and circulation) to three (divergence and flux in electrostatics or 'Newton-Poisson gravity') can shed light on the Aharonov-Bohm debate. The three-dimensional analogy is misleading if taken too literally; it makes sense on a more abstract, formal level (where, for instance, the electromagnetic field is viewed somewhat metaphorically as a 'source'—of electromagnetic turbulence). A slight tweak is enough to produce (a fictitious) gauge freedom in three dimensions.

1 Introduction

In the literature one finds three or even four interpretations (§4) of the Aharonov-Bohm effect[1] (§2); going from two to three dimensions can shed light on all of them. What I propose is an *analogy* and by no means an *exact rendering* of the effect in three dimensions—which in any case would be impossible. More precisely, one has to distinguish between the formalism and (an unduly literal understanding of) the physics: we'll see that the 'overly literal' physics is misleading, whereas the analogy works well on a more abstract, formal level. The scheme depends on differential forms, and is harder to understand in the old vector calculus of curl, divergence *etc*.

Strictly speaking, the Aharonov-Bohm effect isn't even two-dimensional to begin with: it is three- (or even four-)dimensional. But it is rightly discussed in the literature as two-dimensional: one thinks of a horizontal section, an xy plane.[2] Most of the relevant features vary relatively little along the vertical direction z; even if some (such as the wave-function) can vary more, the emphasis on two dimensions remains legitimate. And indeed the version of 'Stokes' that's used is the 'two-dimensional' theorem involving the curl and the circulation around a loop.

Reformulation in three dimensions involves a translation of the main entities; again, some of the correspondences can be surprising, even misleading: the electromagnetic field F in the solenoid (2D) becomes the charge density ρ (3D), the electromagnetic potential A (2D) becomes the electric field E (3D) and so on. To avoid misunderstandings it will be best to have not two but three 'languages' or 'vocabularies': one for

[1]Ehrenberg & Siday (1949), Aharonov & Bohm (1959); see also Franz (1939, 1940, 1965), Olariu & Popescu (1985), Hiley (2013).

[2]The nature of an n-form depends on the size of the environment; so a two-form $\alpha_3 \in \bigwedge^2 \mathbb{R}^3$ corresponds to a vector in $\mathbb{R}^3$ but to a density $\alpha_2 \in \bigwedge^2 \mathbb{R}^2$ in $\mathbb{R}^2$.

two dimensions, one for three, and an abstract vocabulary for both. For instance: loop (2D), membrane (3D), boundary (abstract); electromagnetic field (2D), charge density (3D), source (abstract); electromagnetic potential (2D), electric field (3D), primitive (abstract). All this is in the table below (§3).

The three-dimensional treatment sheds light on a number of philosophical and foundational issues: non-locality, the relationship between empirical accessibility and gauge freedom, the rôle of topology and so forth.

In §2 I briefly describe relevant features of the Aharonov-Bohm effect, in §3 I go into the electrostatic analogy, in §4 I look at the three-dimensional versions of the four interpretations.

2 The Aharonov-Bohm effect

A wave-function is split into two, and these, having enclosed a (simply-connected) region $\mathfrak{L}_2$ containing a solenoid, are made to interfere on a screen. The encircling wave-function is sensitive to any electromagnetism inside inasmuch as the electromagnetic potential $A \in \bigwedge^1 \mathbb{R}^2$, a one-form, contributes a phase

$$\exp i \oint_{\partial \mathfrak{L}_2} A$$

to (the wave-function along) the boundary $\partial \mathfrak{L}_2$ and hence to the interference pattern on the screen. The electromagnetism on $\mathfrak{L}_2$ is related to the circulation around the boundary by Stokes's theorem

$$\mathfrak{F}_2 = \oint_{\partial \mathfrak{L}_2} A = \iint_{\mathfrak{L}_2} dA.$$

One can think of concentric circles: The electromagnetic field[3] $F = dA \in \bigwedge^2 \mathbb{R}^2$ produced by the solenoid is confined to a central disc $\mathfrak{I}_2 \subset \mathfrak{L}_2$ surrounded by an annulus $\mathfrak{D}_2 = \mathfrak{L}_2 - \mathfrak{I}_2$ where F vanishes but not A. Usually there's just one (cylindrical) shield, around the solenoid; but for the clean delimitation of an (arbitrarily large) intermediate annulus it can be useful to think of *two* coaxial cylindrical shields: one keeping the electromagnetic field *in* inside a larger shield keeping the wave-function *out*.[4]

Varying the current through the solenoid changes the arbitrarily distant interference pattern, which is perhaps surprising. The effect is differential, not binary (on/off).

3 In three dimensions

We can begin with a table outlining the correspondences, which should in due course become intelligible.

[3] It is best to view F as a purely *magnetic* field B produced by the current density $J = d{*}B$ in the solenoid, where the star indicates Hodge duality.

[4] In fact there are not two but three 'discs' or 'circles' or 'loops': first, the support of F; then, the shield delimiting the annulus on the outside; and finally, the loop running through the wave-function. To simplify, I have conflated the last two. If one would rather distinguish, there is an 'integration loop' outside a circle 'keeping the wave-function out' (outside the support of F).

Abstract	Two dimensions	Three dimensions
source $\mathfrak{S}$	EM field $\mathfrak{S}_2 = F$	charge density $\mathfrak{S}_3 = \rho$
primitive $\mathfrak{P}$	EM potential $\mathfrak{P}_2 = A$	electric field $\mathfrak{P}_3 = E$
equiv. class $[\mathfrak{P}] = d^{-1}\mathfrak{S}$	equ. class $[A] = d^{-1}F$	equ. class $[E] = d^{-1}\rho$
support $\mathfrak{I}$ (of $\mathfrak{S}$)	inner disc $\mathfrak{I}_2$	inner sphere $\mathfrak{I}_3$
larger region $\mathfrak{L}$	larger disc $\mathfrak{L}_2$	larger sphere $\mathfrak{L}_3$
difference $\mathfrak{D}$	annulus $\mathfrak{D}_2$	difference $\mathfrak{D}_3$
boundary $\mathfrak{B} = \partial\mathfrak{L}$	loop $\mathfrak{B}_2 = \partial\mathfrak{L}_2$	membrane $\mathfrak{B}_3 = \partial\mathfrak{L}_3$
homotopy class $\mathfrak{H} = [\mathfrak{B}]$	hoop $\mathfrak{H}_2 = [\mathfrak{B}_2]$	homotopy class $\mathfrak{H}_3 = [\mathfrak{B}_3]$
flux $\mathfrak{F}$	circulation $\mathfrak{F}_2$	flux $\mathfrak{F}_3$
exterior derivative d	curl $\nabla\times$	divergence $\nabla\cdot$
source interpretation	EM field interpretation	charge interpretation
primitive interpretation	potential interpretation	electric field interpretation
boundary interpretation	holonomy interpretation	membrane interpretation
topology interpretation	topology interpretation	topology interpretation

In three dimensions we have a charge $\rho = dE \in \bigwedge^3 \mathbb{R}^3$ radiating an electric field $E \in \bigwedge^2 \mathbb{R}^3$ which is then caught by an enclosing membrane $\mathfrak{B}_3 = \partial\mathfrak{L}_3$; Stokes's theorem becomes

$$\mathfrak{F}_3 = \iint_{\partial\mathfrak{L}_3} E = \iiint_{\mathfrak{L}_3} dE = \iiint_{\mathfrak{L}_3} \rho.$$

Here we can think of the (nonsimply-connected) isolating region $\mathfrak{D}$ as the difference $\mathfrak{D}_3 = \mathfrak{L}_3 - \mathfrak{I}_3$ between the region $\mathfrak{L}_3$ bounded by the membrane $\mathfrak{B}_3$ and the support $\mathfrak{I}_3$ of ρ. To obtain a satisfactory analogy, it won't even be necessary to *modify* standard electrostatics: a slight tweak will be enough. What we need is a three-dimensional (electrostatic) freedom analogous to

$$[A] = [A + d\lambda]_\lambda = d^{-1}F,$$

where $\lambda \in \bigwedge^0 \mathbb{R}^2$ is a function and $[A]$ the equivalence class of all one-forms differing by an exact term $d\lambda$—which corresponds to the kernel[5] of the curl $d : \bigwedge^1 \mathbb{R}^2 \to \bigwedge^2 \mathbb{R}^2$. In a sense we already have the right freedom in

$$[E] = [E + d\xi]_\xi = d^{-1}\rho, \tag{1}$$

where $d\xi \in \bigwedge^2 \mathbb{R}^3$ is the curl of a one-form $\xi \in \bigwedge^1 \mathbb{R}^3$. The difference being that the electric field E is empirically accessible, whereas the potential A isn't. So we'll have to assume that the individual electric field E (as opposed to the class $[E]$) is just as unmeasurable as A in two dimensions. The electric charge ρ and flux $\mathfrak{F}_3$, which remain measurable, determine the class $[E]$, not an individual electric field E.

In what sense is F a *source* $\mathfrak{S}$? Just as the charge $\rho = \mathfrak{S}_3$ is the source of the electric field $E = \mathfrak{P}_3$ (a perturbation of the surrounding medium) which is then caught

[5]Differential operators are destructive, their kernels are not trivial; but the loss can be overcome, or rather reversed, by an appropriate specification (along the lines of a 'constant of integration')—here a zero-form. Each element of $[A] = d^{-1}dA$ corresponds to a different zero-form; a particular λ restores the original A.

by the boundary $\mathfrak{B}_3$, the electromagnetic field $F = \mathfrak{S}_2$ (a curl, after all) is the source of the surrounding electromagnetic 'turbulence' which, carried by the potential $A = \mathfrak{P}_2$, is likewise caught by the boundary $\mathfrak{B}_2$.

Newton-Poisson gravity[6] (where ρ is the gravitational source density, E the gravitational field and so on) is isomorphic to electrostatics and would be equivalent for our purposes. I've concentrated on electrostatics which is no doubt more familiar.

4 The four interpretations in three dimensions

In two dimensions the interpretations can be called:

(1-2) *electromagnetic field interpretation*[7]

(2-2) *potential interpretation*[8]

(3-2) *holonomy interpretation*[9]

(4-2) *topology interpretation.*[10]

This is not the place to summarise the whole two-dimensional debate, details can be found in the cited literature; the two-dimensional interpretations can moreover be largely inferred from the three-dimensional treatment, for which I propose the following terms:

(1-3) *electric charge interpretation*

[6]By this I just mean Newtonian gravity with Poisson's equation.

[7]Aharonov & Bohm (1959) p. 490: "we might try to formulate a nonlocal theory in which, for example, the electron could interact with a field that was a finite distance away. Then there would be no trouble in interpreting these results, but, as is well known, there are severe difficulties in the way of doing this." See also Healey (1997).

[8]Aharonov & Bohm (1959) pp. 490-1: "we may retain the present local theory and [...] try to give a further new interpretation to the potentials. In other words, we are led to regard $A_\mu(x)$ as a physical variable." See also Feynman, Leighton, Sands (1964) §15-5.

[9]See Wu & Yang (1975), Healey (1997, 2001, 2004, 2007), Belot (1998), Lyre (2001, 2002, 2004a,b), Myrvold (2011).

[10]See Afriat (2013), especially footnote 6, especially Batterman (2003, p. 544): "We now have a $U(1)$ bundle over a nonsimply connected base space: $\mathbb{R}^2 - \{\text{origin}\}$. This fact is responsible for the AB effect." *Ibid.* pp. 552-3: "most discussions of the AB effect very quickly idealize the solenoid to an infinite line in space or spacetime. The flux, in this idealization, just is the abstract topological property of having space or spacetime be nonsimply connected. [...] The issue is whether the idealizations—[...] and nonsimply connected space in the AB effect—do better explanatory work than some less idealized description. I believe that the idealized descriptions do, in fact, do a better job." *Ibid.* p. 554: "It seems to me that for a full understanding of these anholonomies, one needs to appeal to the topology and geometry of the base space. [...] If we take seriously the idea that topological features of various spaces [...] can play an explanatory role [...]." Footnote 29, same page: "it is most fruitful to treat the AB solenoid as an idealization that results in the multiple connectedness of the base space of a fiber bundle." *Ibid.* p. 555: "The different cases are unified by the topological idealization of the solenoid as a string absent from spacetime which renders spacetime nonsimply connected. [...] This topological feature enables us to understand the common behaviour in different AB experiments [...]. [...] how can it possibly be the case that appeal to an idealization such as the AB solenoid as a line missing from spacetime, provides a better explanation of genuine physical phenomena than can a less idealized, more "realistic" account where one does not idealize so severely? [...] quite often [...] appeal to highly idealized models does, in fact, provide better explanations."

(2-3) *electric field interpretation*

(3-3) *membrane interpretation*

(4-3) *topology interpretation.*

4.1 The electric charge interpretation

This interpretation can be expressed as follows: Since (1) is a class, full of individuals, it has to be a physically meaningless[11] mathematical fiction; which means there's nothing at all between the source and the boundary $\mathfrak{B}_3$. The flux $\mathfrak{F}_3$ through the boundary is therefore a non-local effect, in the sense that it isn't conveyed by a 'carrier' E.

4.2 The electric field interpretation

The flux $\mathfrak{F}_3$ is carried from the source to the boundary $\mathfrak{B}_3$ by the electric field $E \in [E]$. To the question "which $E \in [E]$ in particular?" there are at least three answers:

a) It really doesn't matter, any $E \in [E]$ will do—all elements of $[E]$ are on an equal footing.

b) The elements of $[E]$ are not all on an equal footing; only one of them, E, is the *right one*. But since the distinguished element E is assumed to be empirically inaccessible, any element of $[E]$ will do.

c) The elements of $[E]$ are all on an equal footing, *empirically*. But measurement isn't the only way of selecting or ruling out elements of $[E]$: some could be æsthetically or pragmatically (or even historically[12]) privileged; simplicity, elegance, beauty, economy, convenience or even computational considerations could be relevant.[13]

A wave-function is in fact an equivalence class $[\psi]$ of functions that differ only on a set of vanishing measure; is the cardinality of $[\psi]$ enough to rule out the physical relevance of wave-functions and indeed the objects (*e.g.* electrons, the universe) they describe, consigning them all to a shady realm of mathematical fictions? Is *embarras de richesses* so embarrassing that the riches should all be altogether renounced?

4.3 The membrane interpretation

Since (1) is a class, full of individuals, it has to be a physically meaningless mathematical fiction. But to avoid the non-locality of the electric charge interpretation, *something* in $\mathfrak{D}_3$ has to carry the flux $\mathfrak{F}_3$ from the source to the boundary. Since the flux is the

[11] See Nguyen, Teh, Wells (2018) on the status of structure often called "superfluous."

[12] Duhem (1989) p. 388ff

[13] For instance, one could favour a primitive $\mathfrak{P}$ characterised by *purely radial* lines; in standard electrostatics, with measurable E, this geometrical criterion would (with a spherically symmetric charge) give the right electric field lines. In two dimensions, where the lines are level sets, the criterion would correspond to a very natural gauge choice. Why *wouldn't* the lines be purely radial? Must they really be allowed to bend and wiggle?

same for the whole homotopy class $\mathfrak{H}_3$, we can replace E with $\mathfrak{H}_3$; so a class of boundaries somehow conveys the electric field, or rather $[E]$ (or whatever it is that manifests itself on the boundary as an electric flux $\mathfrak{F}_3$). In other words: There has to be *something* in $\mathfrak{D}$; if it can't be the class $[\mathfrak{P}]$, it has to be the class $\mathfrak{H}$.

So far we've assumed that the individual $\mathfrak{P} \in [\mathfrak{P}]$ is unmeasurable. The homotopy class $\mathfrak{H}$ owes its physical legitimacy to the empirical inaccessibility of $\mathfrak{P}$. Now suppose that $\mathfrak{P}$ becomes measurable—a new experiment is devised to pick a single $\mathfrak{P}$ out of the class $[\mathfrak{P}]$. The new experiment dissolves $\mathfrak{H}$, degrading it into a mere mathematical fiction. But how can an experiment concerning $[\mathfrak{P}]$ change the ontological status of $\mathfrak{H}$?

There would be ways of making the measurability of $\mathfrak{P}$ a matter of degree: *more or less measurable*, to *some degree or other* (say one-half, or perhaps 0.73, on a scale from zero to one), rather than just *entirely measurable* or *not at all*. Would the ontological status of $\mathfrak{H}$ vary accordingly? *Very real* rather than *somewhat real*? One can imagine an 'ontological intensity lever' which, by controlling the measurability of $\mathfrak{P}$, determines the *ontological intensity* or *degree of reality* of the homotopy class $\mathfrak{H}$: at one end the lever makes $\mathfrak{H}$ a mere mathematical abstraction, at the other it gives $\mathfrak{H}$ full physical legitimacy.

Summing up, these first three interpretations can be distinguished by what they put in the isolating region $\mathfrak{D}$:

(1-3) nothing at all

(2-3) electric field

(3-3) homotopy class $\mathfrak{H}$.

4.4 The topology interpretation

If a differential form $\mathfrak{P}$ is closed on a simply-connected region $\mathfrak{L}$,

$$d\mathfrak{P} = 0|_{\mathfrak{L}},$$ [14]

the integral $\mathfrak{F}$ through the (outer and only) boundary $\mathfrak{B} = \partial\mathfrak{L}$ will vanish. In other words if no source radiates any '$\mathfrak{P}$' *anywhere* inside the outer boundary $\mathfrak{B}$, the flux $\mathfrak{F}$ through $\mathfrak{B}$ has to vanish—the topological condition means there can be no inner boundary that might contain a '$\mathfrak{P}$-producing' source, whose flux through the inner boundary would reach the outer boundary as well.

But if instead of the simply-connected region $\mathfrak{L}$ we have $\mathfrak{D}$, with a hole $\mathfrak{I}$, the logic of the matter gets more complicated: the absence of a source on $\mathfrak{D}$ no longer allows us to conclude that no '$\mathfrak{P}$' is produced *anywhere* inside the outer boundary $\mathfrak{B}$—for there may be an inner boundary containing a '$\mathfrak{P}$-producing' source (whose flux through the inner boundary would reach the outer boundary as well). So even if $d\mathfrak{P} = 0|_{\mathfrak{L}}$ means that $\mathfrak{F} = 0$ through $\mathfrak{B}$, the condition $d\mathfrak{P} = 0|_{\mathfrak{D}}$ is compatible with both $\mathfrak{F} = 0$ and $\mathfrak{F} \neq 0$ through the outer boundary $\mathfrak{B}$.

The topology interpretation in electrostatics can be summarised as follows: If there's no charge contained in $\mathfrak{B}_3$, the flux $\mathfrak{F}_3$ through $\mathfrak{B}_3$ will vanish; if there may

[14]The notation, with the vertical bar, means that the statement holds where indicated (here on $\mathfrak{L}$).

be a charge in $\mathfrak{B}_3$, the flux through $\mathfrak{F}_3$ may not vanish. More generally, if there's no source in $\mathfrak{B}$, the flux $\mathfrak{F}$ through $\mathfrak{B}$ will vanish; if there may be a source in $\mathfrak{B}$, we may have $\mathfrak{F} \neq 0$.

But the flux is produced by the source, not the hole. Without a source, the (desired) implication

$$[d\mathfrak{P} = 0|_{\mathfrak{D}}] \Rightarrow [\mathfrak{F} \neq 0]$$

is groundless, indeed wrong.[15] If one cannot avoid attributing the flux to the source that produced it, why bother with all the supposedly autonomous topological circumlocution? The 'topologist' seems to want to 'shift the explanation' from the source to the topology, replacing the source with an appropriate topological condition. There's nothing wrong with an *emphasis* on the flux over the source; but a mere hole on its own doesn't even provide the flux.

5 Final remarks

The physics of the analogy, if taken too literally, can be misleading: in two dimensions, the electromagnetic field F is viewed as a 'source'; in three dimensions, the electric field E has the same rôle as the potential A in two; and so on. The analogy makes sense on a more abstract level. A single stipulation is enough to make it work: the individual $E \in [E] = d^{-1}\rho$ must be empirically inaccessible.

Indeed we have seen that the problem, the paradox, the whole debate is produced by the *unmeasurability of a primitive* (A in two dimensions, which corresponds to E in three): if the *individual* primitive $\mathfrak{P}$ (rather than the whole inverse image $[\mathfrak{P}] = d^{-1}\mathfrak{S}$ of the source) were measurable, *it* would carry the effect from the source to the boundary, without non-locality—or the need to look, beyond E, for an 'invariant' medium (a homotopy class of boundaries) to carry the effect in its place.

As to the topology interpretation, one wonders how a flux can be due to a mere hole, that may or may not contain a source; it is clearly produced by the source itself.

Again, I am only claiming that the three-dimensional analogy *sheds light* on the debate, not that it *captures absolutely everything*—indeed it is easy to find features of the two-dimensional case that get 'lost in translation.'

References

Afriat, A. (2013) "Topology, holes and sources" *International Journal of Theoretical Physics* **52**, 1007-12

Aharonov, Y. and D. Bohm (1959) "Significance of electromagnetic potentials in the quantum theory" *Physical Review* **115**, 485-91

Batterman, R. (2003) "Falling cats, parallel parking and polarized light" *Studies in History and Philosophy of Modern Physics* **34**, 527-57

[15] See Afriat (2013).

Belot, G. (1998) “Understanding electromagnetism” *The British Journal for the Philosophy of Science* **49**, 531-55

Duhem, P. (1989) *La théorie physique: son objet – sa structure*, Vrin, Paris

Ehrenberg, W. and R. Siday (1949) “The refractive index in electron optics and the principles of dynamics” *Proceedings of the Physical Society B* **62**, 8-21

Feynman, R., R. Leighton and M. Sands (1964) *The Feynman lectures on physics, volume 2*, Addison-Wesley, Reading

Franz, W. (1939) “Elektroneninterferenzen im Magnetfeld” *Verhandlungen der Deutschen Physikalischen Gesellschaft* **20**, 65-6

Franz, W. (1940) “Elektroneninterferenzen im Magnetfeld” *Physikalische Berichte* **21**, 686

Franz, W. (1965) “Elektroneninterferenzen im Magnetfeld” *Zeitschrift für Physik* **184**, 85-91

Healey, R. (1997) “Nonlocality and the Aharonov-Bohm effect” *Philosophy of Science* **64**, 18-41

Healey, R. (2001) “On the reality of gauge potentials” *Philosophy of Science* **68**, 432-55

Healey, R. (2004) “Gauge theories and holisms” *Studies in History and Philosophy of Modern Physics* **35**, 619-42

Healey, R. (2007) *Gauging what’s real: the conceptual foundations of contemporary gauge theories*, Oxford University Press, New York

Hiley, B. (2013) “The early history of the Aharonov-Bohm effect” arXiv:1304.4736v1 [physics.hist-ph]

Lyre, H. (2001) “The principles of gauging” *Philosophy of Science* **68**, S371-81

Lyre, H. (2002) “Zur Wissenschaftstheorie moderner Eichfeldtheorien” in A. Beckermann and C. Nimtz (editors) *Argument & Analyse – Sektionsvorträge*, Mentis, Paderborn

Lyre, H. (2004a) *Lokale Symmetrien und Wirklichkeit: eine Naturphilosophische Studie über Eichtheorien und Strukturenrealismus*, Mentis, Paderborn

Lyre, H. (2004b) “Holism and structuralism in $U(1)$ gauge theory” *Studies in History and Philosophy of Modern Physics* **35**, 643-70

Myrvold, W. (2011) “Nonseparability, classical and quantum” *The British Journal for the Philosophy of Science* **62**, 417-32

Olariu, S. and I. Popescu (1985) “The quantum effects of electromagnetic fluxes” *Reviews of Modern Physics* **57**, 339-436

Nguyen, J., N. Teh and L. Wells (2018) “Why surplus structure is not superfluous” *British Journal for the Philosophy of Science*

Wu, T. and C. Yang (1975) “Concept of nonintegrable phase factors and global formulation of gauge fields” *Physical Review D* **12**, 3845-57