

Topology, holes and sources

Alexander Afriat

August 6, 2013

Abstract

The Aharonov-Bohm effect is often called “topological.” But it seems no more topological than magnetostatics, electrostatics or Newton-Poisson gravity (or just about any radiation, propagation from a source). I distinguish between two senses of “topological.”

1 The Aharonov-Bohm effect

A wavefunction is split into two, and these, having enclosed a (simply-connected) region ω containing a solenoid, are made to interfere on a screen. The enclosing wavefunction is sensitive to any enclosed electromagnetism inasmuch as the electromagnetic potential¹ A , a one-form, contributes a phase

$$\exp i \oint_{\partial\omega} A$$

to (the wavefunction along) the boundary $\partial\omega$ and hence to the interference pattern on the screen. The electromagnetism on ω is related to the circulation around the boundary by Stokes’ theorem

$$(1) \quad C = \oint_{\partial\omega} A = \iint_{\omega} dA.$$

The electromagnetic field² $F = dA$ produced by the solenoid is circumscribed to a middle region $\lambda \subset \omega$ surrounded by an isolating region³ $\lambda' = \omega - \lambda$ where F vanishes but not A . The full *Aharonov-Bohm effect*⁴ can be considered the ‘differential’ or ‘incremental’ sensitivity of the interference pattern to variations in the current through the solenoid.⁵

¹By “potential” I just mean *primitive*: the potential of the electromagnetic two-form $F = dA$ is its primitive $A \leftrightarrow (\mathbf{A}, \varphi)$, the potential of the magnetic two-form $\mathbf{B} = d\mathbf{A}$ is its primitive \mathbf{A} , the potential of the electric one-form $*\mathbf{E} = d\varphi$ is its primitive φ (the Hodge dual $*$ being taken in three dimensions), the potential of the three-form density $\rho = d\mathbf{E}$ is its primitive \mathbf{E} .

²It is perhaps easiest to think of F as a purely *magnetic* field \mathbf{B} produced by the current density $\mathbf{J} = d*\mathbf{B}$ in the solenoid.

³It will be convenient to view λ and ω as concentric disks.

⁴Ehrenberg & Siday (1949), Aharonov & Bohm (1959)

⁵*Cf.* Batterman (2003, p. 555): “Similarly, in the AB effect, it appears that we will need to refer to different nonseparable holonomy properties for each case in which there is a different flux running through the solenoid.”

2 The topological₁ interpretation

I will distinguish between two (related but) different senses of “topological”:

1. *Topological*₁: related to the presence of a hole (containing a source, a solenoid or charge for instance).
2. *Topological*₂: invariant under appropriate continuous deformations.

The *topological*₁ interpretation⁶ of the Aharonov-Bohm effect can be formulated as follows: If A were closed throughout a simply-connected region ω it would also be

⁶Aharonov & Bohm (1959, p. 490): “in a field-free multiply-connected region of space, the physical properties of the system still depend on the potentials.” Wu & Yang (1975, p. 3845): “The famous Bohm-Aharonov experiment [...] showed that in a multiply connected region where $f_{\mu\nu} = 0$ everywhere there are physical experiments for which the outcome depends on the loop integral [...] around an unshrinkable loop.” And p. 3856: “ $f_{\mu\nu}$ underdescribes electromagnetism because of the Bohm-Aharonov experiment which involves a doubly connected space region.” Nash & Sen (1983, p. 301): “We [...] consider the consequence of assuming the field \mathbf{F} to be identically zero in some region Ω . At first one may think that there will be no physically measurable electromagnetic effects in such a region Ω . This is not so, effects may arise if the topology of Ω is non-trivial, e.g. if Ω is not simply connected. [...] In terms of parallel transport one says that zero curvature does not imply trivial parallel transport if the region in which the curvature is zero is not simply connected. This underlies the fact that there is a sense in which the connection is a more fundamental object than the curvature, even though a connection is gauge dependent and not directly measurable.” Ryder (1996, p. 101-4): “the Bohm-Aharonov effect owes its existence to the non-trivial topology of the vacuum [...]. The Bohm-Aharonov effect is the simplest illustration of the importance of topology in this branch of physics. [...] The relevant space in this problem is the space of the vacuum, i.e. the space outside the solenoid, and that space is not simply connected. [...] It is thus an essential condition for the Bohm-Aharonov effect to occur that the configuration space of the vacuum is not simply connected. [...] in other words, it is because the gauge group of electromagnetism, U_1 , is not simply connected that the Bohm-Aharonov effect is possible. [...] The configuration space of the Bohm-Aharonov experiment is the plane \mathbb{R}^2 [...] with a hole in, and this is, topologically, the direct product of the line \mathbb{R}^1 and the circle [...]. There is, nevertheless, a positive effect on the interference fringes. The mathematical reason for this is that the configuration space of the null field (vacuum) is the plane with a hole in [...].” 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.” Martin (2003, p. 48): “in the case of non-trivial spatial topologies, the gauge-invariant interpretation runs into potential complications. [...] So-called holonomies [...] encode physically significant information about the global features of the gauge field.” See also Lyre (2001, pp. S377-80), Nounou (2003), Lyre (2004, p. 659), Agricola & Friedrich (2010, p. 275).

exact, and hence expressible as the gradient $A = d\mu$ of a zero-form μ ; the flux

$$\oint_{\partial\omega} d\mu = \iint_{\omega} d^2\mu$$

through the boundary $\partial\omega$ would then vanish, as d^2 does. But here A is closed on λ' ; from $dA = 0|_{\lambda'}$ it does not follow that A is exact, nor that the flux through the enclosing loop vanishes: it may or may not.

The existence of the source responsible for the effect is therefore ruled out by one topology (A closed throughout a simply-connected region) but not another.

The same applies to a simply-connected three-dimensional region Ω enclosed by a two-dimensional boundary $\partial\Omega$. If the two-form \mathbf{E} were closed throughout Ω it would also be exact, and hence expressible as the curl $\mathbf{E} = d\zeta$ of a one-form ζ ; the flux

$$\iint_{\partial\Omega} d\zeta = \iiint_{\Omega} d^2\zeta$$

through the boundary would then vanish. But if the region on which \mathbf{E} is closed has a hole in it, the flux through the enclosing surface may or may not vanish.

This is precisely what we have in electrostatics, where the electric field $\mathbf{E} = *d\varphi$ is (Hodge-dual to) the gradient $d\varphi$ of the scalar potential φ . The vanishing divergence $d\mathbf{E}$ expresses the conservation of electricity where none is created, away from the charges that produce \mathbf{E} according to the Maxwell-Poisson equation $d\mathbf{E} = d*d\varphi = \rho$, ρ being the charge density. If the divergence $d\mathbf{E}$ vanished throughout the volume Ω , there would be no electricity produced and hence none radiated through the enclosing surface.⁷ But a charge in Ω —say in a region⁸ $\Lambda \subset \Omega$ isolated by $\Lambda' = \Omega - \Lambda$ —would prevent electricity from being conserved throughout Ω .

We have the same formalism in Newton-Poisson gravity, where φ is the gravitational potential, $d\varphi$ and \mathbf{E} both represent gravitational force, and ρ is the mass density. Gravity⁹ would therefore be another topological₁ effect.

Again, the topology of the region where the ‘potential’ (A or \mathbf{E} or whatever) is closed tells us relatively little: if the region were simply-connected, conservation would be *general* within the enclosing surface since there could be no holes containing sources; and if nothing were created inside the enclosing surface, the total radiation through it would vanish. But if the topology does *not* allow the presence of holes to be ruled out, the presence of sources in them cannot either; and sources would produce a flux through the enclosing surface.

A non-trivial topology cannot, on its own at any rate, rule out the *absence* of a source either. Nor does it provide the ‘amount’ or ‘intensity’ of the possible source (which would tell us the intensity of the effect—the flux through the enclosing surface). So the full Aharonov-Bohm effect, which can be considered ‘incremental,’ is hardly topological₁, or at any rate no more so than electrostatics or Newton-Poisson gravity.

The implications

$$([dA = 0|_{\omega}] \ \& \ [\omega \text{ simply-connected}]) \Rightarrow [A \text{ exact}] \Rightarrow \left[\oint_{\partial\omega} A = 0 \right]$$

⁷Over and above any divergence-free electrical background that may or may not be present.

⁸It will be convenient to view Λ and Ω as concentric spheres.

⁹Or rather the *total* gravitational attraction radiated by a mass.

are legitimate, unlike

$$([dA = 0|_{\omega}] \& \neg [\omega \text{ simply-connected}]) \Rightarrow \neg [A \text{ exact}] \Rightarrow \neg \left[\oint_{\partial\omega} A = 0 \right],$$

or

$$([dA = 0|_{\omega}] \& [\omega \text{ multiply-connected}]) \Rightarrow [A \text{ not exact}] \Rightarrow \left[\oint_{\partial\omega} A \neq 0 \right].$$

3 Topological₂ effects

3.1 Aharonov-Bohm

The Aharonov-Bohm effect is topological₂ in the sense that certain basic quantities (say the circulation C and resulting interference pattern) are invariant under appropriate continuous deformations; but it seems that (fundamental aspects of) electrostatics and Newton-Poisson gravity are just as topological₂. Magnetostatics may be even more topological₂.

Since one measures the curvature $F = dA = dA'$ rather than A itself, the freedom expressed by the substitution

$$(2) \quad A \mapsto A' = A + d\xi$$

(ξ being a zero-form) is invisible. Such transformations deform the level sets of A 's local potential¹⁰ γ . One can first imagine a purely 'radial' or 'angular' γ (with values running from zero to $2\pi k = C$),¹¹ whose level lines are straight rays radiating through the annulus λ' from the inner disk λ to the edge $\partial\omega$. A gauge transformation (2) would then deform the level rays, bending them without making them cross. The circle $\partial\omega$ can likewise be deformed into any loop going around the solenoid once. The Aharonov-Bohm effect is topological₂ in the sense that neither deformation affects the circulation C (or the resulting interference pattern).

It is perhaps easier to picture the denumerable set $\{\gamma_1, \dots, \gamma_N\}$ of level curves at intervals of C/N than all of them; they will each be cut once¹² by any loop going around the solenoid once.¹³

¹⁰For wherever A is closed it can be written locally as the gradient $A = d\gamma$ of a zero-form γ —just as \mathbf{E} can be written locally, wherever it is closed, as the curl $\mathbf{E} = d\zeta$ of a one-form ζ .

¹¹Such a γ cannot be continuous everywhere; we can imagine a single discontinuity, say on the ray with values $\gamma = 2\pi nk$, where the integer n is zero then one, $k = C/2\pi$ being a constant.

¹²One should really say *an odd number of times*, as Jean-Philippe Nicolas has pointed out to me. Crossings in opposite directions cancel, and add nothing to the integral.

¹³Batterman (2003, pp. 545-6): "The phase or anholonomy depends continuously on the flux in the solenoid, but [...] it depends discontinuously upon the shape of the circuit. For example, two loops around gives an anholonomy twice that of one loop around for constant magnetic flux." And 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. In this way we can understand why, for a given fixed magnetic flux, a loop that goes n times around the solenoid will have an anholonomy that is n times that of a loop that goes around once." A loop going around twice will cross each level curve twice. Alternatively, two different loops will also catch the same flux twice. Similar double-counting can be arranged in electrostatics too: a membrane enclosing the source twice, or two different membranes.

Whereas here the deformations are allowed by (and part of) theory, in the next cases they will be counterfactual.

3.2 Electrostatics

The basic law here, the Gauß-Maxwell equation

$$(3) \quad \iint_{\partial\Omega} \mathbf{E} = \iiint_{\Omega} d\mathbf{E} = \iiint_{\Omega} \rho$$

(or $d\mathbf{E} = \rho$), is topological₂ inasmuch as the boundary $\partial\Omega$ and electric field lines can be continuously deformed without affecting the integral (3)—which corresponds to the (circulation giving rise to the) Aharonov-Bohm effect. We can imagine a spherically symmetric charge distribution ρ contained in $\Lambda \subset \Omega$ (everything concentric): the electrical field lines radiated by the charge in Λ correspond to the level rays radiating from the solenoid in λ . Nothing in electrostatics prevents the deformation of $\partial\Omega$. Admittedly the electric rays cannot be bent without violating $\mathbf{E} = *d\varphi$; despite preserving the divergence $d\mathbf{E} = d\mathbf{E}'$, the transformation

$$(4) \quad \mathbf{E} \mapsto \mathbf{E}' = \mathbf{E} + d\alpha$$

(the three-dimensional version of (2), α being a one-form) is counterfactual—which does not prevent a conditional characterisation of the effect as topological₂: “the integral would remain the same even if the field lines were bent.”¹⁴

Most of this applies, *mutatis mutandis*, to Newton-Poisson gravity.

Why bother with obvious facts about integration? Because much is made of the deformable loop $\partial\omega$ (which corresponds to the deformable surface $\partial\Omega$) and gauge transformation (2) (which corresponds to the counterfactual transformation (4)).

3.3 Magnetostatics

The basic law here, Maxwell’s equation

$$\iint_{\partial\Omega} \mathbf{B} = \iiint_{\Omega} d\mathbf{B} = 0$$

(or $d\mathbf{B} = 0$), holds because a magnet has two poles, that act as source and sink of the same field lines, which form loops going from one pole to the other: all magnetism produced is eventually recovered. If a magnetic loop crosses the boundary $\partial\Omega$ it will cross it again¹⁵ on the way back to the magnet, thus erasing whatever it contributed to the integral on the way out. The law is topological₂ in that the boundary $\partial\Omega$ and field lines can be deformed¹⁶ continuously without affecting the integral. Nothing in magnetostatics prevents the deformation of $\partial\Omega$; the deformation of the field lines by $\mathbf{B} \mapsto \mathbf{B}' = \mathbf{B} + d\beta$ is again counterfactual, since the magnetic field is observable.

¹⁴Here I am indebted to Dennis Dieks, Éric Gourgoulhon and Jean-Philippe Nicolas.

¹⁵Here—returning to the point made by Jean-Philippe Nicolas—the number of perforations will be even: 0, 2, 4 etc.

¹⁶Deformations of $\partial\Omega$ can of course lead to the exclusion or inclusion of certain loops.

But if I dwell on these old three-dimensional theories I may give the impression I want to make a point about *them*, whereas my real point concerns the Aharonov-Bohm effect: it is hardly topological, or at any rate no more topological than electrostatics *etc.*

I thank Nazim Bouatta, Dennis Dieks, Éric Gourgoulhon, Marc Lachièze-Rey and Jean-Philippe Nicolas for valuable clarifications and corrections.

References

- Agricola, I. and T. Friedrich (2010) *Vektoranalysis: Differentialformen in Analysis, Geometrie und Physik*, Vieweg+Teubner, Berlin
- Aharonov, Y. and D. Bohm (1959) “Significance of electromagnetic potentials in the quantum theory” *Physical Review* **115**, 485-91
- Batterman, R. W. (2003) “Falling cats, parallel parking and polarized light” *Studies in History and Philosophy of Modern Physics* **34**, 527-57
- Brading, K. and E. Castellani, editors (2003) *Symmetries in physics*, Cambridge University Press
- Ehrenberg, W. and R. E. Siday (1949) “The refractive index in electron optics and the principles of dynamics” *Proceedings of the Physical Society B* **62**, 8-21
- Lyre, H. (2001) “The principles of gauging” *Philosophy of Science* **68**, S371-81
- Lyre, H. (2004) “Holism and structuralism in $U(1)$ gauge theory” *Studies in History and Philosophy of Modern Physics* **35**, 643-70
- Martin, C. (2003) “On continuous symmetries and the foundations of modern physics” pp. 29-60 in Brading & Castellani (2003)
- Nash, C. and S. Sen (1983) *Topology and geometry for physicists*, Academic Press, London
- Nounou, A. (2003) “A fourth way to the Aharonov-Bohm effect” pp. 174-99 in Brading & Castellani (2003)
- Ryder, L. (1996) *Quantum field theory*, Cambridge University Press
- Wu, T. T. and C. N. Yang (1975) “Concept of nonintegrable phase factors and global formulation of gauge fields” *Physical Review D* **12**, 3845-57