



**HAL**  
open science

## The reality of loops

Alexander Afriat

► **To cite this version:**

| Alexander Afriat. The reality of loops. 2013. hal-00769689v1

**HAL Id: hal-00769689**

**<https://hal.science/hal-00769689v1>**

Preprint submitted on 3 Jan 2013 (v1), last revised 3 Jan 2014 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The reality of loops

Alexander Afriat

January 3, 2013

## Abstract

The transformation properties of the electromagnetic four-potential can make one wonder whether it has any physical reality. But it is felt (with the Aharonov-Bohm effect in mind) that the reality in question has to be *somewhere*, ‘thereabouts’; if it is not in the potential itself, might it not be in the loops around which it is integrated? I argue that no ontological solidity is gained by retreating to the loops, which are just as flimsy as the potential.

## 1 Introduction

Physical legitimacy, reality, even centrality have been attributed to loops by Lyre (2001),<sup>1</sup> Belot (2003),<sup>2</sup> Lyre (2004) pp. 665-7 and especially Healey (2007).<sup>3</sup> In a nutshell,  $\alpha$ :

---

<sup>1</sup>P. S380: “an ontological universe consisting of matter-fields, gauge field strengths, and holonomies.”

<sup>2</sup>P. 216: “holonomies [...] are well-defined quantities on the spaces of states of the standard formulations of Yang-Mills theories. If it is accepted that these theories describe reality, does not it follow that the quantities in question are as real as any others?”

<sup>3</sup>P. xviii: “In the simplest case (classical electromagnetism interacting with quantum particles) such an account ascribes properties to (or on) a loop of empty space that are not fixed by properties of anything located at points around the loop [...]” P. 30: “Suppose instead that one takes the holonomies themselves directly to represent electromagnetism and its effects on quantum particles.” P. 31: “But if the holonomies directly represent electromagnetism and its effects, then there is still a sense in which the action of electromagnetism on the electrons is not completely local, since holonomies attach to extended curves rather than points.” P. 56: “only gauge-invariant functions of these mathematically localized fields directly represent new electromagnetic properties; and these are predicated of, or at, arbitrarily small neighborhoods of *loops* in space-time—i.e. oriented images of closed curves on the space-time manifold.” P. 106: “we arrive at the view that non-localized EM potential properties in a region are represented by the holonomies [...] of all closed curves in the region [...]. This is the interpretation of classical electromagnetism I shall defend.” P. 118: “One can reformulate the theory as a theory of holonomy properties, so that it does not even appear to mention localized gauge potential properties.” P. 220: “[...] the Aharonov-Bohm effect and other related effects provide vivid examples of physical processes that seem best accounted for in terms of non-localized holonomy properties [...]” P. 221: “Should we believe that non-separable processes involving non-localized holonomy properties are responsible for phenomena like the Aharonov-Bohm effect? This belief may be encouraged by the predictive successes consequent upon introducing classical electromagnetism into the quantum mechanics of particles.” P. 225: “This reinforces the conclusion that the evidence for contemporary gauge theories lends credence to the belief that these describe non-separable processes, while nothing in the world corresponds to or is represented by a locally defined gauge potential.” And the last paragraph of the book, pp. 227-8: “Consider instead the claim that there are physical processes involving properties that are neither localized at or near a point nor determined by properties localized at or near a point, and that some of our theories of fundamental “forces” succeed in capturing significant features of

‘potentials are unreal because they’re *not* gauge invariant, loops are real because they *are*.’<sup>4</sup> In §2.2 I’ll try to understand whether indeed  $\beta$ : ‘potentials are unreal because they’re not gauge invariant,’<sup>5</sup> and will argue that the formulation is misleading: the point is rather that *potentials have the wrong transformation behaviour* (§2.1). It is admittedly easier to attribute ontological solidity to entities with the right transformation properties, but entities that can be transformed away shouldn’t necessarily be dismissed as fictions.

*The reality of potentials* will no doubt remain an open issue for quite a while; in any case it is not settled by  $\beta$ . I’ll say something about this, but my main point (§4) is that whatever reservations one may have or not have about potentials, loops are no better.

The matter can be considered with reference to the Aharonov-Bohm effect<sup>6</sup> (§3), which—if the electromagnetic four-potential  $A$  is just a mathematical fiction—has to be conveyed by something else. What *does* have the right transformation behaviour, what *is* invariant? The circulation  $C$ . But  $C$  is just a number, not enough on its own to convey or account for the effect. The number must surely be related to an ontologically promising—perhaps extended—entity involved in the transmission of the effect: why not the loop around which the circulation is taken?

The circulation, which is indeed invariant under the deformations of one curve, namely the loop, is also invariant under deformations of other curves: the level curves of  $A$ ’s local potential. So why prefer loops? Given the duality (§4) between loops and potentials one can wonder whether loops are any better.

---

these processes. [...] But while it is a very general and very abstract ontological claim that can be related only distantly to observation, the evidence for contemporary Yang-Mills gauge theories does provide some reason to believe this claim. Or so I have argued.”

<sup>4</sup>Cf. Healey (2007) p. 51: “If the value of the vector potential  $A_\mu$  at each space-time point  $x$  in a region does not represent some qualitative intrinsic physical properties in the vicinity of  $x$ , it may be that some function of its integral around each closed curve  $C$  in that region does represent such properties of or at (the image of)  $C$ . [...] Since the gauge dependence of the vector potential made it hard to accept Feynman’s view that it is a real field that acts locally in the Aharonov-Bohm effect, there is reason to hope that a gauge-invariant function of its line integral around closed curves might facilitate a local account of the action of electromagnetism on quantum particles in the Aharonov-Bohm effect and elsewhere.” P. 105: “The non-localized gauge potential properties view is motivated by the idea that the structure of gauge potential properties is given by the gauge-invariant content of a gauge theory. The most direct way to implement this idea would be to require that the gauge potential properties are just those that are represented by gauge-invariant magnitudes. [...] While the vector potential  $A_\mu$  is gauge dependent, its line integral  $S(C) = \oint_C A_\mu dx^\mu$  around a closed curve  $C$  is gauge invariant.”

<sup>5</sup>Lyre (2001) p. S377. Lyre (2004) p. 665: “realists can hardly be satisfied by the gauge dependence of entities as imminent in the A-interpretation [...]” Healey (2007) pp. 25-6; and pp. 55-6: “If there are new localized gauge properties, then neither theory nor experiment gives us a good grasp on them. Theoretically, the best we can do is to represent them either by a mathematical object chosen more or less arbitrarily from a diverse and infinite class of formally similar objects related to one another by gauge transformations, or else by this entire gauge-equivalence class.”

<sup>6</sup>Ehrenberg & Siday (1949), Aharonov & Bohm (1959)

## 2 Transformation behaviour and objective reality

### 2.1 Wegtransformierbarkeit

Physics is full of entities one might call ‘perspectival,’ whose representation depends on the perspective of the subject or observation. Where perspective is unimportant, the objective reality of the entity can be hard to contest; but perspective can also dominate, and there it may be better to speak of *opinion* than of objective reality. In a transformation theory like quantum mechanics or relativity, transformations turn one perspective’s representation into another’s; bad transformation behaviour changes representations more radically than good.

Even if the role of the observer cannot be eliminated, it can be limited by appropriate transformation properties. *Vanishing* is an important criterion: good transformation behaviour will not *annihilate*, or *create out of nothing*. Tensors are defined accordingly: if the components of a tensor all vanish in one basis they will in all others too, so that all observers agree on *presence* or *absence*—the field is there for everyone or for no one. There is less agreement as to the connection components  $\Gamma_{\nu\sigma}^{\mu}$  of general relativity, which transform badly: they come and go with the accelerations of the observer. The connection can be transformed away,<sup>7</sup> the point of tensors is that they can’t.

### 2.2 The objective reality of the electromagnetic potential

The potential  $A$  is a connection too, and behaves accordingly. The curvature  $F = dA = dA'$  is alone observable, and as it is indifferent to the substitution

$$(1) \quad A \mapsto A' = A + d\lambda$$

one is free to add an exact term. Even if  $A$  is not exact, *at a point  $x$*  all covectors are on an equal footing, so one can always choose a  $\lambda$  satisfying  $d\lambda(x) = -A(x)$  that gets rid of the connection at  $x$ . Much as in general relativity, the connection can always be eliminated or engendered at a point, where some will see it but not others.<sup>8</sup>

It is almost as though *logic itself* were violated by *Wegtransformierbarkeit*: how can the connection be both *there* and *not there*? Consistency can of course be restored by longer statements specifying the different gauges, but the inconsistency between the short statements, which leave gauge out, is neither uninteresting nor irrelevant, and is troubling enough to suggest ontological doubts. The subjectivity expressed by the transformation behaviour (1) is therefore sufficiently radical to make one wonder whether the electromagnetic potential is an objectively real field.

Let us now try to understand  $\beta$ . Consider the following statements:

---

<sup>7</sup>The infinitive *wegtransformieren* means “to transform away;” the noun *Wegtransformierbarkeit* is harder to translate (“away-transformability”?). Schrödinger (1918) p. 7: “daß es wirkliche Gravitationsfelder (d. i. Felder, die sich nicht „wegtransformieren“ lassen) gibt, mit durchaus verschwindenden oder richtiger gesagt „wegtransformierbaren“ Energiekomponenten; Felder, in denen nicht nur Bewegungsgröße und Energiestrom, sondern auch die Energiedichte und die Analoga der Maxwell’schen Spannungen durch geeignete Wahl des Koordinatensystems für endliche Bezirke zum Verschwinden gebracht werden können.” See also Afriat & Caccese (2010).

<sup>8</sup>Of course  $A$  cannot be transformed away by the Lorentz group:  $\bar{A}_\mu = \Lambda_\mu^\nu A_\nu$  only vanishes if  $A_\mu$  does. So the same object can transform well *and* badly.

- (A) *what's unobservable is unreal*<sup>9</sup>
- (B) *what's observable is real*
- (C) *what's subject to an unreal transformation is itself unreal*
- (D)  *$d^2$  vanishes*
- (E) *(1) is unobservable*
- (F)  *$F$  and  $C$  are indifferent to (1)*
- (G)  *$F$  and  $C$  are real*
- (H)  *$F$  and  $C$  are observable*
- (I)  *$A$  is unobservable*<sup>10</sup>
- (J)  *$A$  is unreal.*

(F) is a theorem, (H) is relatively uncontroversial, (F) & (H) seem good reasons to accept (G). But the implication (G)  $\Rightarrow$  (J) is manifestly wrong, since  $F$ ,  $C$  and  $A$  could all be real. The implication [(A) & (I)]  $\Rightarrow$  (J) is right, but why accept (A)? The implications [(D) & (I)]  $\Rightarrow$  (E) and [(A) & (C) & (E)]  $\Rightarrow$  (J) are not uninteresting, but the conclusion (J) again rests on (A); and (C)—to which tensors and even vectors would succumb—is just as questionable.<sup>11</sup> The reality of  $A$  should be investigated by looking at the peculiarities of (1), not by appealing to the indifference of *other* objects to (1). Again, one has to distinguish good transformation behaviour from bad, and *Wegtransformierbarkeit* seems a useful criterion.

To uphold the reality of  $A$  one could insist on the merely *punctual* character of the elimination: wherever the curvature  $F = dA$  doesn't vanish, the annihilation (and the radical subjectivity it suggests) cannot be extended beyond the point to its neighbourhood. And a point isn't much; a more extended domain of disagreement would be better, for otherwise a kind of intersubjective *agreement* would subsist almost everywhere in the neighbourhood. Alternatively one can be less demanding, and put up—like Einstein<sup>12</sup> in 1918—with the radical subjectivity expressed by the wrong transformation properties, extending physical reality to quantities that can be transformed away. And

---

<sup>9</sup>Cf. Healey (2007) p. 49: “Now even if the value of  $A_\mu(x)$  is not observable, it does not automatically follow that the vector potential has no value at  $x$ , nor that this value cannot represent any qualitative intrinsic physical properties at or near  $x$ . Only a positivist or instrumentalist would be prepared to make such an inference without further ado.”

<sup>10</sup>Cf. Lyre (2002) p. 82: “Der Eichsymmetrie zufolge lassen sich Eichpotentiale nicht direkt beobachten – nur eichinvariante Größen können observabel sein.” Healey (2007) p. 49: “Certainly no properties represented by  $A_\mu(x)$  are *observable*, if the combination of quantum mechanics and classical electromagnetism used to account for the Aharonov-Bohm effect exhausts the empirical content of  $A_\mu$ . [...] Only gauge-invariant magnitudes are observable, including the electromagnetic field strength  $F_{\mu\nu}$  and the Dirac phase factor [...]”

<sup>11</sup> $\beta$  becomes especially awkward in general relativity, where the gauge group  $\text{diff}M$  induces coordinate transformations—which are observable—and forces a choice between  $\beta$  and (B).

<sup>12</sup>Einstein (1918a) p. 167: “[Levi-Civita] (und mit ihm auch andere Fachgenossen) ist gegen eine Betonung der Gleichung  $[\partial_\nu(\mathfrak{T}_\sigma^\nu + t_\sigma^\nu) = 0]$  und gegen die obige Interpretation, weil die  $t_\sigma^\nu$  keinen T e n s o r bilden. Letzteres ist zuzugeben; aber ich sehe nicht ein, warum nur solchen Größen eine physikalische

so on; but what matters to us for the time being is that there are reasons to question the objective reality of  $A$ .

### 3 The Aharonov-Bohm effect

A wavefunction is split into two, and these, having enclosed a (simply-connected) region  $\omega$  containing a solenoid, are made to interfere on a screen. The enclosing wavefunction is sensitive to any enclosed electromagnetism inasmuch as the electromagnetic potential  $A$  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  $\omega$  is related to the circulation around the boundary by Stokes' theorem

$$C = \oint_{\partial\omega} A = \iint_{\omega} dA.$$

The electromagnetic field<sup>13</sup>  $F = dA$  produced by the solenoid is circumscribed to a middle region  $\lambda \subset \omega$  surrounded by an isolating region<sup>14</sup>  $\lambda' = \omega - \lambda$  where  $F$  vanishes but not  $A$ .

Varying the current through the solenoid changes the arbitrarily distant interference pattern, which is surprising.

### 4 Duality between loops and potentials

Since the objective reality of the electromagnetic potential is questionable—and *something* has to convey the effect—why not attribute reality to the loop  $\partial\omega$  instead, as Lyre, Belot and Healey suggest? Because it seems just as flimsy and deformable as the potential. In fact there is a *duality* between the two: just as a vector  $\dot{\sigma}'_0(x) \in T_x M$  and a covector  $A(x)$  from the dual space  $T_x^* M$  give a number  $\langle A(x), \dot{\sigma}'_0(x) \rangle$ , the loop<sup>15</sup>  $\sigma_0 = \partial\omega$  and potential give a number  $\langle A, \sigma_0 \rangle = C$ . Both  $A$  and  $\sigma_0$  can be deformed without affecting the circulation: the potential according to (1); a loop can be deformed into any other loop going around the solenoid once. Both could be replaced by their equivalence classes  $[A]$  and  $[\sigma_0]$ , one could even write  $([A], [\sigma_0]) = C$ .

---

Bedeutung zugeschrieben werden soll, welche die Transformationseigenschaften von Tensorkomponenten haben." Einstein (1918b) p. 447: "Diese Formulierung stößt bei den Fachgenossen deshalb auf Widerstand, weil  $(\mathcal{M}'_\sigma)$  und  $(t'_\sigma)$  keine Tensoren sind, während sie erwarten, daß alle für die Physik bedeutsamen Größen sich als Skalare und Tensorkomponenten auffassen lassen müssen." He later changed his mind, according to Afriat & Caccese (2010).

<sup>13</sup>It 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.

<sup>14</sup>It will be convenient to view  $\lambda$  and  $\omega$  as concentric disks.

<sup>15</sup>The one-dimensional manifold  $\sigma_0 \subset M$  is the image of the mapping  $\sigma'_0 : I \rightarrow M ; t \mapsto \sigma'_0(t)$ , without its parameter  $t$ , which is not part of the boundary  $\partial\omega$  (where  $I \subset \mathbb{R}$  is an interval and the manifold  $M$  is an appropriate base space).

It will be useful to understand the transformation (1) more geometrically, as a deformation of the level sets of  $A$ 's local primitive<sup>16</sup>  $\sigma$ . One can first imagine a purely 'angular' or 'radial'  $\sigma$  (with values running from zero to  $2\pi k = C$ ),<sup>17</sup> whose level lines are straight rays radiating through the annulus  $\lambda'$  from the inner disk  $\lambda$  to the edge  $\partial\omega$ . A gauge transformation (1) would then deform the level rays, bending them without making them cross. It is easier to picture the denumerable set  $\{\sigma_1, \dots, \sigma_N\}$  of level curves at intervals of  $C/N$  than all of them; they will each be cut once<sup>18</sup> by any loop  $\sigma_0$  going around the solenoid once.

In this construction we have  $N + 1$  deformable curves  $\{\sigma_0, \dots, \sigma_N\}$ , which all seem pretty much on the same footing;  $\alpha$  amounts to the surprising claim that

$\alpha'$ :  $\sigma_0$  is ontologically superior to the other curves  $\sigma_1, \dots, \sigma_N$  since they can be deformed.

Why should one curve  $\sigma_k$  be any better than the others? How about  $\sigma_7$ ? It remains true that  $\sigma_0, \dots, \sigma_6, \sigma_8, \dots, \sigma_N$  can be deformed.

To emphasise that loops are no better than  $A$ , we can even arrange for a gauge transformation to *induce* a loop deformation (thus strengthening the duality): Level rays of unit length determine a unit circle, which will then be 'deflated' into another loop by a gauge transformation (1); to every gauge transformation there corresponds a different loop  $\sigma_0^\lambda$ . If a potential subject to (1) is too flimsy to exist, how can loops also subject to (1) be any better? Are vectors any more real than the covectors dual to them? Is a (nondegenerate) Lagrangian any less real than the Hamiltonian dual to it?

## 5 Final remarks

It seems that statements like  $\beta$  are not enough to condemn potentials, whose reality remains an open issue despite their *Wegtransformierbarkeit*, which is the real problem. But my main point (§4) has been that whatever the status of potentials, loops are no better. Potentials—which I am not here to defend—even have advantages: if the transmission of the Aharonov-Bohm effect is to be accounted for,  $A$  is a traditional field (of a kind that physics is used to and more or less knows how to handle) connecting the solenoid to the screen; one can even imagine, as there's nothing else on  $\lambda'$ , a propagation of the effect *through*  $A$ .

I thank Jean-Philippe Nicolas for valuable clarifications.

## References

Afriat, A. and E. Caccese (2010) "The relativity of inertia and reality of nothing" *Studies in History and Philosophy of Modern Physics* **41**, 9-26

<sup>16</sup>For wherever  $A$  is closed it can be written locally as the gradient  $A = d\sigma$  of a zero-form  $\sigma$ .

<sup>17</sup>Such a  $\sigma$  cannot be continuous everywhere; we can imagine a single discontinuity, say on the ray with values  $\sigma = 2\pi nk$ , where the integer  $n$  is zero then one,  $k = C/2\pi$  being a constant.

<sup>18</sup>One should really say *an odd number of times*, as Jean-Philippe Nicolas has pointed out. Crossings in opposite directions cancel, and add nothing to the integral.

- Aharonov, Y. and D. Bohm (1959) "Significance of electromagnetic potentials in the quantum theory" *Physical Review* **115**, 485-91
- Belot, G. (2003) "Symmetry and gauge freedom" *Studies in History and Philosophy of Modern Physics* **34**, 189-225
- 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
- Einstein, A. (1918a) "Über Gravitationswellen" *Sitzungsberichte der Preußischen Akademie der Wissenschaften, Berlin (Mitteilung der 31. Januar)* 154-67
- Einstein, A. (1918b) "Der Energiesatz in der allgemeinen Relativitätstheorie" *Sitzungsberichte der Preußischen Akademie der Wissenschaften, Berlin (Sitzung der physikalisch-mathematischen Klasse vom 16. Mai)* 448-59
- Healey, R. (2007) *Gauging what's real: the conceptual foundations of contemporary gauge theories*, Oxford University Press, New York
- 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. (2004) "Holism and structuralism in  $U(1)$  gauge theory" *Studies in History and Philosophy of Modern Physics* **35**, 643-70
- Schrödinger, E. (1918) "Die Energiekomponenten des Gravitationsfeldes" *Physikalische Zeitschrift* **19**, 4-7