PROCEEDINGS OF SCIENCE

Electricity, gravity and matter

Alexander Afriat
Université de Bretagne Occidentale & SPHÈRE (UMR 7219), Université Paris Diderot
E-mail: afriat@gmail.com

Weyl uses a $\mathbb{U}(1)$ freedom caught between matter and gravity to produce an electromagnetic potential A and field $F = dA$. As his potential is curved the electromagnetic field doesn't vanish, which is noteworthy—for an exact potential $A = d\lambda$ is often used to produce *nothing at all*: $F = d^2\lambda = 0$.

Frontiers of Fundamental Physics 14
15-18 July 2014
Aix Marseille University (AMU) Saint-Charles Campus, Marseille, France

1. Introduction

Weyl's first gauge theory[1] was a generalization of Einstein's general relativity; his second,[2] which grew out of the first, remained a relativistic theory of curved spacetime, but with a matter field of two-spinors governed by 'half' of the (massless) Dirac equation. The proper orthochronous Lorentz group $\mathbb{SO}^+(1,3)$ changes neither the length, origin, spatial parity nor temporal orientation of spacetime four-vectors, which are accordingly propagated in Weyl's second theory by a connection $\mathscr{A} = \mathscr{A}^a_\mu dx^\mu \otimes \mathbf{T}_a$ with values in the Lie algebra $\mathfrak{o}(1,3) = \mathrm{Lie}\,\mathbb{SO}^+(1,3)$. The parallel transport of Weyl's two-spinors, which are subject to a group we can call $\mathbb{W}(2,\mathbb{C}) = \{g \in \mathbb{GL}(2,\mathbb{C}) : |\det g| = 1\}$, is given by a connection $\mathfrak{A}$ with values in $\mathfrak{w}(2,\mathbb{C}) = \mathrm{Lie}\,\mathbb{W}(2,\mathbb{C})$; the homomorphism $h : \mathbb{W}(2,\mathbb{C}) \to \mathbb{SO}^+(1,3)$ is indeed at the core of Weyl's theory. We'll see how he exploits the angular freedom $e^{i\lambda}$ left by h for "the critical part of the theory": *the derivation of electromagnetism*.[3]

The standard "gauge principle" or "gauge argument" is sometimes attributed to Weyl.[4] Not only is his argument quite different, but it avoids the exact connection $A = d\lambda$ and vanishing field $F = d^2\lambda$ that vitiate the standard argument.

2. The standard gauge argument

One begins with a free field, of two-spinors $\psi \in \mathbb{C}^2$ for instance. The Lagrangian $\mathscr{L} = \bar{\psi}\,\partial\!\!\!/\,\psi$ is invariant under the *global* transformation $\psi \mapsto e^{i\xi}\psi$, where ξ is constant; $\partial\!\!\!/$ stands for the sum $\sigma^\mu \partial_\mu$, in which σ^0 is the identity and σ^k the three Pauli operators. It is then claimed[5] that $\mathscr{L}$ should also be invariant under the *local* transformation

$$\psi \mapsto \psi_\lambda = e^{i\lambda}\psi, \tag{2.1}$$

where $\lambda : M \to \mathbb{R}$ is a smooth function on the base manifold M which here is an appropriate spacetime. The Lagrangian $\mathscr{L}_\lambda = \bar{\psi}_\lambda\,\partial\!\!\!/\,\psi_\lambda = \bar{\psi}\sigma^\mu(\partial_\mu + i\partial_\mu\lambda)\psi$ is not invariant since the derivative ∂_μ has become $\partial_\mu + i\partial_\mu\lambda$. To offset (2.1) we therefore have to subtract the term $i\partial_\mu\lambda$ that alters $\mathscr{L}$, yielding the covariant differential $D = d - id\lambda$ with components $D_\mu = \partial_\mu - i\partial_\mu\lambda$. Writing $D\!\!\!\!/ = \sigma^\mu D_\mu$, the balanced Lagrangian $\mathscr{L}'_\lambda = \bar{\psi}_\lambda D\!\!\!\!/\psi_\lambda$ will be equal to $\mathscr{L}$ for all λ. It is then argued that an interaction $F = dA = d^2\lambda$ is thereby deduced, whose potential A is $d\lambda$. But since d^2 vanishes the interaction does too, as has often been pointed out.[6]

The gauge argument is fertile enough to produce another two Lagrangians,[7]

$$\mathscr{L}^A = -ij \wedge A = -ij^\mu A_\mu = -i\overline{\psi}\sigma^\mu A_\mu \psi \quad \text{and} \quad \mathscr{L}_F = F \wedge *F = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu},$$

[1] Weyl (1918)
[2] Weyl (1929a,b,c). See Straumann (1987), Brading (2002) and Scholz (2005) for more recent accounts.
[3] Weyl (1929b) p. 348.
[4] Brading (2002) pp. 3-4, Healey (2007) p. 160 for instance.
[5] See footnote 5 of Afriat (2013).
[6] See footnote 9 of Afriat (2013).
[7] *Cf.* Weyl (1929c) p. 283.

where the current density three-form $j = \varepsilon_{\mu\nu\sigma\tau} j^\mu dx^\nu \wedge dx^\sigma \wedge dx^\tau/3!$ corresponds to the vector with components $j^\mu = \overline{\psi}\sigma^\mu\psi$. One can either leave $A = d\lambda$ in $\mathscr{L}'_\lambda$ to offset (2.1), or balance $\mathscr{L}_\lambda$ with $\mathscr{L}^A_\lambda = -i\overline{\psi}_\lambda\sigma^\mu A_\mu\psi_\lambda$ in the sum $\mathscr{L}'_\lambda = \mathscr{L}_\lambda + \mathscr{L}^A_\lambda$. A Lagrangian $\mathscr{L}_F$ derived from the gauge argument will of course vanish. But once the argument has produced the exact potential $A = d\lambda$ and vanishing interaction $F = dA = d^2\lambda$ one can perhaps claim that A is no longer exact. The exact term $d\lambda$ would then be added to one that isn't[8] in the gauge transformation

$$A \mapsto A' = A + d\lambda, \tag{2.2}$$

which suggests a new differential $D = d - iA$ subject to

$$D \mapsto D' = D - id\lambda. \tag{2.3}$$

The Lagrangian $\mathscr{L}'_\lambda$ is sensitive to both (2.1) and (2.3) individually, but invariant under their joint action. Since $\mathscr{L}_F$ is indifferent to (2.3) and has nothing to do with (2.1), the total Lagrangian $\mathscr{L}'_\lambda + \mathscr{L}_F$ is also indifferent to (2.1) offset by (2.3).

3. Weyl's argument

3.1 Gravity and electricity

Some history to begin with.

First, there was general relativity. Levi-Civita (1917) saw that the connection determined by Einstein's covariant derivative transported the *direction* of a vector anholonomically, but not its *length*, which was left unchanged. This was unfair, protested Weyl—length deserved the same treatment as direction.[9] To remedy he proposed a more general theory that propagated length anholonomically too. This *congruent* transport would also be governed by a connection, which Weyl defined as a bilinear mapping between neighbouring points: linear in the object propagated and in the direction of propagation. A connection transporting the (squared) length $l = \mathrm{g}(\dot\gamma, \dot\gamma)$ from $a = \gamma(a)$ to its neighbour[10] $b = \gamma(b)$ along the curve $\gamma : [a,b] \to M$ would therefore be a real-valued[11] one-form A applied to the direction $\dot\gamma \in T_aM$ and multiplied by the initial length l_a, yielding the small difference $\delta l = l_a - l_b = l_a\langle A, \dot\gamma\rangle$ subtracted from l_a. The final length l_b is $l_a(1 - \langle A, \dot\gamma\rangle)$—unless a and b are too far apart for γ to remain straight in between, in which case l_a has to be multiplied by $\exp\int_\gamma A$ instead.

To correct the geometrical injustice of Einstein's theory, the curvature $F = dA$ cannot vanish—unlike the three-form dF, which does. Seeing all this, Weyl couldn't help thinking of the electromagnetic four-potential A, the Faraday two-form $F = dA$ and Maxwell's two homogeneous equations $dF = 0$: he had unified gravity and electromagnetism, by mistake![12] And indeed Einstein

[8] One can wonder what the gauge argument is for if the curved potential A was already there to begin with. The exact term added in (2.2) has more to do with the invariance of $F = dA = dA'$ than with the gauge argument.

[9] See Afriat (2009).

[10] Which is so close to a it practically belongs to the tangent space T_aM; see Weyl (1926) p. 28, Weyl (1931b) p. 52.

[11] Here the structure group is the multiplicative group $\mathbb{R}^\times$ of dilations, generated by the Lie algebra $\langle\mathbb{R}, +, [\cdot,\cdot]\rangle$ or rather $\langle\mathbb{R}, +\rangle$; the Lie product $[\cdot,\cdot]$ vanishes since real numbers commute.

[12] See Ryckman (2005) pp. 149-54, 158.

would soon point out the mistake: the anholonomy on which Weyl based his theory is not observed in nature, as we'll see in §3.2.

Weyl sought to rectify general relativity using the curvature $F = dA$, which ensured geometrical justice. Differentiation is destructive, or rather irreversible; what (the nontrivial kernel of) d annihilates is the freedom (2.2) invisible to $F = dA = dA'$, in the sense that the inverse image $d^{-1}F$ of F under d is the whole equivalence class $[A] = [A + d\lambda]_\lambda$ given by the equivalence relation $A \sim (A + d\lambda)$. If A only served to produce the curvature F, (2.2) would be vacuous; but A appears elsewhere too, notably in the law of propagation

$$\nabla \mathrm{g} = A \otimes \mathrm{g}, \tag{3.1}$$

which is not indifferent to (2.2), where g is the metric. To make (3.1) invariant, (2.2) therefore has to be balanced by

$$\mathrm{g} \mapsto \mathrm{g}' = e^{\lambda}\mathrm{g}. \tag{3.2}$$

Such compensation is typical[13] of a gauge theory: an invariant expression (here (3.1)) is sensitive to a first transformation, and to a second as well—but indifferent to the two together, if their variations are appropriately constrained, and balance one another.

3.2 Einstein's objection

The tangent of a worldline's *image* $\bar{\gamma} \subset M$ only has a direction; the length l of the tangent $\dot{\gamma} = d\gamma/dt$ is given by the parameter rate $\partial\gamma/\partial t$. If the values of the parameter are identified with the readings of a clock describing γ, the length l giving the proper ticking rate should remain constant—the hands of a good clock don't accelerate. But far from remaining constant, lengths in Weyl's theory aren't even integrable: $l_b(\gamma) = l_a \exp \int_\gamma A$ depends on γ—an exact connection $A = d\mu$ would of course give $l_b = l_a \exp \int_a^b d\mu = l_a \exp \Delta\mu$ along any path joining a and b, $\Delta\mu$ being the difference $\mu(b) - \mu(a)$ between the final and initial values of μ. In addition to the *first* clock effect (Langevin's twins) already present in Einstein's theory, Weyl's theory therefore involves a *second* clock effect expressed in the anholonomy of ticking rates.

Einstein objected that *nature provides integrable clocks*.[14] Two clocks trace out a loop $\bar{\gamma} = \partial\omega$ enclosing a region ω (without holes): starting from the same point a they describe worldlines γ_1, γ_2 that meet at b. They tick at the same rate if A is exact, for then $\oint_{\partial\omega} d\mu = \iint_\omega d^2\mu$ vanishes—in fact (without holes) it is enough for A to be closed, $\oint_{\partial\omega} A = \iint_\omega dA$ vanishes too provided dA does. But if the loop encloses an electromagnetic field $F = dA$, one of the clocks will tick faster than the other once they're compared at b. In any case the theory didn't work: from the beginning it rested on an anholonomy not seen in nature.

3.3 Gravity, electricity and matter

But then wavefunctions appeared in the twenties, and Weyl accordingly developed a *quantum* gauge theory of gravity, electricity *and matter*. As long as there was only gravity and electricity, the gauge relation—worth preserving in some form or other—could only hold between *them*;

[13] Typical but mysterious, even for Weyl (1931b) p. 54: "insbesondere [. . .] Eichfaktor e^{λ}."
[14] missives to Weyl dated 15, 19 April 1918

but now, with a third element, as many compensations were in principle possible, of which only two were plausible: the old relation (2.2)-(3.2) between gravity and electricity, and a new one between electricity and the matter wave. With (2.2)-(3.2) the theory would have remained subject to Einstein's objection—all the more convincing in the new quantum-mechanical context which provided an absolute (integrable!) standard of length or ticking rate allowing the distant comparisons Weyl wanted to prevent in 1918.[15] The other possibility was left: (2.2) with a quantum version of (3.2),[16] of which the simplest and most obvious[17] was (2.1), where $\mathbb{U}(1)$ replaced the multiplicative group $\mathbb{R}^\times$ of (3.2).[18] As the wave function was now part of a four-dimensional space-time theory, it could no longer obey the Schrödinger equation, which violates relativity by treating space and time very differently.[19] Weyl adopted what amounted to a Dirac equation, but cut in half, without mass or the associated crisscrossing of component pairs ...

3.4 Dirac-Weyl theory

We can first take $H = p_1^2$ as the simplified Hamiltonian of a particle whose mass is one-half (and whose motion is one-dimensional). Momentum p in quantum mechanics is represented by differentiation, in the sense that[20] $p \mapsto id$, in components $p_\mu \mapsto i\partial_\mu$. Our quantum Hamiltonian will therefore be $-\partial_1^2 = -(\partial/\partial x^1)^2$, which means that Schrödinger's equation $i\partial_t \psi = \partial_1^2 \psi$ differentiates space twice as much as time. But by what should it be replaced? The d'Alembertian $\Box = \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2$ and (massless) Klein-Gordon equation $\Box\psi = 0$ treat space about the same way as time, they have the right transformation properties; but $\Box$ is 'squared' and there are reasons to prefer a wave operator and especially a time derivative[21] that aren't. In seeking a square root $\sqrt{\Box}$ Dirac found $\partial\!\!\!/ = \gamma^\mu \partial_\mu$, where the γ^μ's have the algebraic properties needed to get rid of the cross terms that appear when squaring. He therefore proposed the *Dirac equation* $(m - i\partial\!\!\!/)\psi = 0$ which not only treats the three spatial derivatives $\gamma^k \partial_k$ the same way as the time derivative $\gamma^0 \partial_0$, but differentiates with respect to time only once.[22] The γ^μ's, which do not commute, cannot be numbers; they admit for instance the canonical representations

$$\gamma^0 \leftrightarrow \begin{pmatrix} 0 & \sigma^0 \\ -\sigma^0 & 0 \end{pmatrix} \qquad \gamma^k \leftrightarrow \begin{pmatrix} 0 & \sigma^k \\ \sigma^k & 0 \end{pmatrix}, \tag{3.3}$$

where all four quaternions $\sigma^\mu : \mathbb{C}^2 \to \mathbb{C}^2$ are hermitian and unitary; σ^0 is the identity, and the three traceless operators σ^k satisfy $2i\sigma^j = \varepsilon_{jkl}[\sigma^k, \sigma^l]$.

The wave ψ on which the γ^μ's act therefore has four (complex) components—*embarras de richesses* which Weyl found most troubling: "doppelt zu viel Energieniveaus"! The anti-diagonality of the γ^μ's governs the embarrassing excess by swapping the two two-spinors making

[15] Weyl (1929c) pp. 284, 290; and Weyl (1931b) p. 55: "Die Atomistik [...] Überzeugungskraft."

[16] Weyl (1929c) p. 284: "this principle [...] position and time."

[17] The conservation requirement $\|\psi_\lambda\|^2 = \|\psi\|^2$ being very natural. And (2.1) isn't even observable (with respect to position at any rate); *cf.* Weyl (1931a) p. 87.

[18] Weyl (1931a) p. 89, Weyl (1931b) pp. 55, 57

[19] Weyl (1931a) pp. 187-8

[20] Weyl (1931a) p. 89

[21] Weyl (1931a) pp. 188, 193

[22] Weyl (1931a) p. 190

up ψ. As the embarrassment is due to the *sign* that distinguishes between the different interweavings[23] produced by the γ^μ's, Weyl deals with it by choosing the only mass—none at all—that doesn't distinguish between plus and minus.[24] Without mass and cut in half, the Dirac equation becomes $\sigma^\mu \partial_\mu \psi = 0$. The reduced wave has two complex components but four (null) real ones: the three Hermitian quadratic forms $j^k = \overline{\psi}\sigma^k\psi$ and the squared length $j^0 = \|\psi\|^2 = \overline{\psi}\sigma^0\psi = \sqrt{(j^1)^2 + (j^2)^2 + (j^3)^2}$.

3.5 Electricity out of gravity and matter

Weyl's first gauge theory generalized Einstein's general relativity by adding electricity to gravity—to which Weyl then adds a further ingredient, *matter*, in his second theory. But to understand Weyl's quantum-mechanical gauge argument one has to begin not with gravity and electricity but with gravity and matter, for it is out of their group-theoretical relationship that electricity emerges. We have just seen that the matter Weyl introduces in 1929 is represented by a two-spinor-valued wavefunction, presumably subject to something like $\mathbb{SL}(2,\mathbb{C})$; gravity by an *Achsenkreuz* (tetrad) field subject to $\mathbb{SO}^+(1,3)$. The relationship between gravity and matter might be given by the 2-1 homomorphism between $\mathbb{SL}(2,\mathbb{C})$ and $\mathbb{SO}^+(1,3)$; but what about electricity and the angular freedom needed to produce it? $\mathbb{SL}(2,\mathbb{C})$ is *too small*.

It is tempting to imagine that Weyl chooses the slightly larger group[25] $\mathbb{W}(2,\mathbb{C})$ *in order to produce electricity*; but the choice seems dictated rather by the object $j^\mu = \overline{\psi}\sigma^\mu\psi$ which (aside from representing the four-current density) expresses the homomorphism between matter and gravity—and is indifferent to phase.[26] The 2-1 homomorphism can be understood as the correction of $g \in \mathbb{SL}(2,\mathbb{C})$ by $h(g) \in \mathbb{SO}^+(1,3)$ in $j^\mu = h(g)^\mu_\nu(\overline{\psi}\overline{g}\sigma^\nu g\psi)$. The further $\mathbb{U}(1)$ seems to come from the disappearance of $e^{i\lambda}$ in $\overline{\psi}e^{-i\lambda}\overline{g}\sigma^\mu g e^{i\lambda}\psi$, which Weyl takes to indicate the existence of *an angular degree of freedom caught between gravity and matter*. One can also see the loose angle by writing $h(g) = h(e^{i\lambda}g) \in \mathbb{SO}^+(1,3)$ or $h^{-1}(h(g)) = [e^{i\lambda}g]_\lambda \subset \mathbb{W}(2,\mathbb{C})$. Infinitesimally we have the Lie algebras $\mathfrak{o}(1,3) = \mathfrak{sl}(2,\mathbb{C})$ and $\mathfrak{w}(2,\mathbb{C}) = \mathfrak{sl}(2,\mathbb{C}) \oplus i\mathbb{R}\mathbb{1}_2$ (where $i\mathbb{R} = \mathfrak{u}(1) = \mathrm{Lie}\,\mathbb{U}(1)$), and the homomorphism $\mathfrak{h} : \mathfrak{w}(2,\mathbb{C}) \to \mathfrak{o}(1,3)$; the free phase $e^{i\lambda} \in \mathbb{U}(1)$ becomes the additive freedom $i\lambda\mathbb{1}_2 \in i\mathbb{R}\mathbb{1}_2$ in $\mathfrak{h}(\mathfrak{g}) = \mathfrak{h}(\mathfrak{g} \oplus i\lambda\mathbb{1}_2) \in \mathfrak{o}(1,3)$ and $\mathfrak{h}^{-1}(\mathfrak{h}(\mathfrak{g})) = [\mathfrak{g} \oplus i\lambda\mathbb{1}_2]_\lambda \subset \mathfrak{w}(2,\mathbb{C})$.[27] Once the freedom is there, surely it deserves propagation![28] So a connection A is needed; and why should it be flat?

[23] Symplectic for time but simply '`NOT`' for space. The interweaving produced by a purely `NOT` γ^0 would be gratuitous; the symplecticity given by the sign difference is essential—with respect to the three γ^k's with merely `NOT` anti-diagonality.

[24] Weyl (1929c) pp. 292, 294

[25] Weyl (1929b) p. 333: "man beschränke sich auf solche lineare Transformationen U von ψ_1, ψ_2, deren Determinante den absoluten Betrag 1 hat."

[26] Weyl (1929c) p. 291, Weyl (1931a) p. 195: "Aus der Natur, dem Transformationsgesetz der Größe ψ ergibt sich, daß die vier Komponenten ψ_ρ relativ zum lokalen Achsenkreuz nur bis auf einen gemeinsamen Proportionalitätsfaktor $e^{i\lambda}$ durch den physikalischen Zustand bestimmt sind, dessen Exponent λ willkürlich vom Orte in Raum und Zeit abhängt, und daß infolgedessen zur eindeutigen Festlegung des kovarianten Differentials von ψ eine Linearform $\sum_\alpha f_\alpha dx_\alpha$ erforderlich ist, die so mit dem Eichfaktor in ψ gekoppelt ist, wie es das Prinzip der Eichinvarianz verlangt."

[27] Weyl (1929b) p. 348: "Dann ist aber auch die infinitesimale lineare Transformation dE der ψ, welche der infinitesimalen Drehung $d\gamma$ entspricht, nicht vollständig festgelegt, sondern dE kann um ein beliebiges rein imaginäres Multiplum $i \cdot df$ der Einheitsmatrix vermehrt werden." See also Weyl (1929c) p. 291.

[28] Again, Weyl (1931a) p. 195: "infolgedessen [. . .] verlangt."

Even if the (electromagnetic) ‘structure’ group of Weyl’s second theory is no longer $\mathbb{R}^\times$ but $\mathbb{U}(1)$, the two Lie algebras are the same since the groups are locally equivalent; characterized infinitesimally (and not integrally), the connection A propagating λ is the same as the length connection we saw in §3.1. Here the infinitesimal variation $\delta\lambda$ will be linear in λ and in the direction $\dot{\gamma}$. Applied to the direction $\dot{\gamma}$, the one-form A yields the infinitesimal generator $\langle A, \dot{\gamma}\rangle$, which then multiplies λ to produce the increment $\delta\lambda = \lambda\langle A, \dot{\gamma}\rangle$.[29] So there’s a connection for tetrads, another for spinors, *and a third one—A—for the residual* $\mathbb{U}(1)$ *freedom between tetrads and spinors*. The values $\langle A, \dot{\gamma}\rangle$ are in the Lie algebra $\mathfrak{u}(1)$ of the group $\mathbb{U}(1)$ caught between gravity and matter. The values $\langle \mathscr{A}, \dot{\gamma}\rangle = \mathscr{A}^r_\mu \dot{\gamma}^\mu \mathbf{T}_r$ of the gravitational connection $\mathscr{A} = \mathscr{A}^r_\mu dx^\mu \otimes \mathbf{T}_r$ are in $\mathfrak{o}(1,3)$, the values $\langle \mathfrak{A}, \dot{\gamma}\rangle = \mathfrak{A}^r_\mu \dot{\gamma}^\mu \mathbf{U}_r$ of the material connection $\mathfrak{A} = \mathfrak{A}^r_\mu dx^\mu \otimes \mathbf{U}_r$ in $\mathfrak{w}(2,\mathbb{C})$.

The whole point of allowing the propagation of λ to depend on direction is to admit anholonomies. So the curvature $F = dA$ of A will not necessarily vanish; and since F is exact, it is also closed: $dF = d^2A = 0$. In F, A and $dF = 0$ Weyl (again) saw[30] the electromagnetic field, its potential and Maxwell’s two homogeneous equations (which are the same—up to Hodge duality—as the other two, away from sources).

3.6 The curved electromagnetic connection

Nothing in Weyl’s argument indicates a flat connection. The electromagnetic connection of Weyl’s 1918 gauge theory not only isn’t necessarily flat; *it has to be curved* (as a matter of geometrical justice)—and we have seen that the electromagnetic connection of Weyl’s 1929 theory *is its direct descendant*.

We have a theory of gravity, matter and electricity, with connections for all three. Gravity and matter are clearly governed by curved connections; why not electricity too? The three connections are related by their Lie algebras $\mathrm{Lie}\,\mathbb{W}(2,\mathbb{C}) = \mathrm{Lie}\,\mathbb{SL}(2,\mathbb{C}) \oplus \mathrm{Lie}\,\mathbb{U}(1)$; the equation $\mathbb{SL}(2,\mathbb{C}) \times \mathbb{U}(1) = \mathbb{W}(2,\mathbb{C})$ makes more sense locally than globally, where it becomes a 2-1 homomorphism.

$\mathbb{SO}^+(1,3) = G$ and $\mathbb{W}(2,\mathbb{C}) = G'$ are just ‘structure’ groups, acting at a generic spacetime point. What about the corresponding gauge groups $\mathscr{G}, \mathscr{G}'$ acting on all of spacetime M? In special relativity “there’s just a single tetrad”; so there’s just one $\mathbb{SO}^+(1,3) = G = \mathscr{G}$, one $\mathbb{W}(2,\mathbb{C}) = G' = \mathscr{G}'$, and above all one $e^{i\lambda}$.[31] But with spacetime curvature the tetrad varies,[32] and so does λ. This could mean the following: Only a *flat* gravitational connection allows the assignment of the *same* tetrad to distant points—only with flatness can there be *global* constancy or ‘sameness.’ With

[29] Weyl (1929b) p. 348: “Zur eindeutigen Festlegung des kovarianten Differentials $\delta\psi$ von ψ hat man also außer der Metrik in der Umgebung des Punktes P auch ein solches df für jedes von P ausgehende Linienelement $\overrightarrow{PP'} = (dx)$ nötig. Damit $\delta\psi$ nach wie vor linear von dx abhängt, muß $df = f_p(dx)^p$ eine Linearform in den Komponenten des Linienelements sein. Ersetzt man ψ durch $e^{i\lambda}$, so muß man sogleich, wie aus der Formel für das kovariante Differential hervorgeht, df ersetzen durch $df - d\lambda$.” See also Weyl (1929c) p. 291. Weyl’s notation is confusing: whereas the one-form $d\lambda$ (which *is* a differential) is necessarily exact, df (my A) isn’t.

[30] Weyl (1929b) p. 349, Weyl (1929c) pp. 291-2

[31] Weyl (1929b) p. 348: “In der speziellen Relativitätstheorie muß man diesen Eichfaktor als eine Konstante ansehen, weil wir hier ein einziges, nicht an einen Punkt gebundes Achsenkreuz haben.”

[32] Weyl (1929b) p. 348: “Anders in der allgemeinen Relativitätstheorie: jeder Punkt hat sein eigenes Achsenkreuz und darum auch seinen eigenen willkürlichen Eichfaktor; dadurch, daß man die starre Bindung der Achsenkreuze in verschiedenen Punkten aufhebt, wird der Eichfaktor notwendig zu einer willkürlichen Ortsfunktion.” See also Weyl (1929c) p. 291.

curvature it becomes meaningless to say that tetrads at distant points are the same. Where tetrads cannot remain constant, one has to suppose they *vary*. A flat electromagnetic connection alongside a curved $\mathscr{A}$ can of course be countenanced, but it is in the spirit of Weyl's argument for both to be flat or both curved. So if the tetrad varies (anholonomically), λ might as well too.[33]

4. Final remark

Whatever its idiosyncrasies, Weyl's gauge argument at least avoids the exact connection $A = d\lambda$ and vanishing curvature $F = d^2\lambda = 0$ produced by the standard argument.

I thank Ermenegildo Caccese, Johannes Huisman, Marc Lachièze-Rey, Thierry Levasseur and Jean-Philippe Nicolas for many valuable conversations and clarifications.

References

Afriat, A. (2009) "How Weyl stumbled across electricity while pursuing mathematical justice" *Studies in history and philosophy of modern physics* **40**, 20-5

Afriat, A. (2013) "Weyl's gauge argument" *Foundations of physics* **43**, 699-705

Brading, K. (2002) "Which symmetry? Noether, Weyl, and the conservation of electric charge" *Studies in History and Philosophy of Modern Physics* **33**, 3-22

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

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

Levi-Civita, T. (1917) "Nozione di parallelismo in una varietà qualunque e conseguente specificazione geometrica della curvatura riemanniana" *Rendiconti del Circolo matematico di Palermo*" **42**, 173-205

Ryckman, T. (2005) *The reign of relativity: philosophy in physics 1915-1925*, Oxford University Press

Scholz, E. (2005) "Local spinor structures in V. Fock's and H. Weyl's work on the Dirac equation (1929)" pp. 284-301 in D. Flament *et al.* (editors) *Géométrie au vingtième siècle, 1930-2000*, Hermann, Paris

Straumann, N. (1987) "Zum Ursprung der Eichtheorien bei Hermann Weyl" *Physikalische Blätter* **43**, 414-21

Weyl, H. (1918) "Gravitation und Elektrizität" pp. 147-59 in *Das Relativitätsprinzip*, Teubner, Stuttgart, 1990

Weyl, H. (1926) *Philosophie der Mathematik und Naturwissenschaft*, Oldenbourg, Munich

[33]Here I am indebted to Johannes Huisman.

Weyl, H. (1929a) “Gravitation and the electron” *Proceedings of the National academy of sciences, USA* **15**, 323-34

Weyl, H. (1929b) “Elektron und Gravitation” *Zeitschrift für Physik* **56**, 330-52

Weyl, H. (1929c) “Gravitation and the electron” *The Rice Institute Pamphlet* **16**, 280-95

Weyl, H. (1931a) *Gruppentheorie und Quantenmechanik*, Hirzel, Leipzig

Weyl, H. (1931b) “Geometrie und Physik” *Die Naturwissenschaften* **19**, 49-58