The most beautiful form of the Dirac equation, and some speculations re "123 mystery"

Warren D. Smith, June 2023

Abstract. Dirac 1928's relativistic electron wave equation had involved 4×4 matrices. We show how to reformulate it using 2×2 matrices only, and also that we can get rid even of them provided we make the wavefunction be biquaternion-valued. The Majorana equation then has quaternion-valued wavefunction. We then speculate that extensions of this idea might be able to solve one or two of the oft-cited "great mysteries" of physics.

In 1928, Dirac famously found his electron relativistic wave equation. To do so, he had to make the wavefunction Ψ be a vector with *four* complex components, and he needed to employ certain special 4×4 matrices with just the right algebraic properties. It was rather annoying that the matrices needed to be 4×4, but Dirac proved that no smaller matrix size was possible without disobeying at least some of his algebraic demands, so there the matter rested for the next 95 years.

However, I now point out that **you *can* meet Dirac's algebraic demands using 2×2 matrices only**, *provided* Ψ is not a 4-*vector*, but rather 2×2 *matrix*, with complex entries. I believe that it is *most natural* to regard "spinors" Ψ as 2×2 complex matrices, not 4-vectors, in which case Dirac and his followers were foolishly missing out on this opportunity. I find it quite surprising that Dirac, as probably the most noted connoisseur of mathematical beauty among 20th century physicists, failed to see this.

Let σ^1, σ^2, σ^3 denote the three Pauli 2×2 spin matrices

```
0 1        0 -i        1  0
1 0        i  0        0 -1
```

Each Pauli matrix is both Hermitian and unitary, has square equal to the identity, determinant=-1, trace=0, and they all anticommute. $\sigma^1\sigma^2=i\sigma^3$, $\sigma^2\sigma^3=i\sigma^1$, $\sigma^3\sigma^1=i\sigma^2$. Many people also define an additional "zeroth Pauli matrix" σ^0 to be the 2×2 identity matrix. It has det=+1, trace=2, and *commutes* with everything.

Let $D_u=(\partial_u-ieA_u)$ for u=0,1,2,3 denote the "Maxwell gauge-covariant derivative" operator, where A_u is the Maxwell 4-potential, (-e) is the electron charge, and ∂_u is partial differentiation with respect to the *u*th spatial coordinate (or if u=0 then with respect to ct where t is time and c the speed of light).

Dirac's form of his equation using four 4×4 "gamma matrices" and column-4-vector Ψ is

$$(\gamma^0 D_0 + \gamma^1 D_1 + \gamma^2 D_2 + \gamma^3 D_3 + i\, mc/\hbar)\, \Psi = 0$$

Here m is the electron mass and $\hbar$ the reduced Planck constant. This also is writeable as ($\not{D}$+imc/

$\hbar)\Psi=0$ using "Feynman slash" notation $\not{B}=\sum_u \gamma^u B_u$. My form of the gamma matrices γ^0, γ^1, γ^2, γ^3 (mine is one particular among many acceptable forms) is:

$$\begin{matrix} 0 & -i\sigma^0 \\ i\sigma^0 & 0 \end{matrix} \qquad \begin{matrix} 0 & i\sigma^1 \\ i\sigma^1 & 0 \end{matrix} \qquad \begin{matrix} 0 & i\sigma^2 \\ i\sigma^2 & 0 \end{matrix} \qquad \begin{matrix} 0 & i\sigma^3 \\ i\sigma^3 & 0 \end{matrix}$$

[My basis can be reached by starting with wikipedia's "alternate Weyl basis," then multiplying on the left by diag(1,-i) and on the right by its inverse diag(1,i), which plainly is a unitary conjugation.] The last three gammas are antiHermitian, but γ^0 is Hermitian.

My new form of the Dirac equation using 2×2 matrix-valued Ψ is

$$D_0 \Psi \sigma^3 + (\sigma^1 D_1 + \sigma^2 D_2 + \sigma^3 D_3) \Psi + \Psi \sigma^1 mc/\hbar = 0$$

My form is readily seen to be equivalent to the old form: multiply my equation on the right by $i\sigma^1$ to get the old form (the 4-vector entry-ordering 0123 corresponds to ordering the matrix entries Ψ_{ab} going down columns starting with the leftmost column).

But some people consider even 2×2 matrices to be too large – they want the ultimate in conciseness, 1×1 matrices, i.e. scalars! That desire also can be satisfied if we regard -i times the Pauli matrices as isomorphic to Hamilton's quaternions $\mathbb{i},\mathbb{j},\mathbb{k}$.

Review of reals, complexes, quaternions, and the "biquaternions." The reader hopefully is aware that the **complex numbers $\mathbb{C}$** arise by adjoining a symbol i obeying $i^2=-1$ to the reals $\mathbb{R}$. Then every complex number z may be written as z=x+iy with x,y real. Useful notation then includes the *complex conjugate* $z^*=x-iy$, the *absolute value* $|z|=[x^2+y^2]^{1/2}$, the nonnegative-real-valued *norm* $|z|^2=zz^*=z^*z=x^2+y^2$, the reciprocal $z^{-1}=z^*/|z|^2$ if $z\neq0$, the real part re(z)=x, and the imaginary part im(z)=y. If ab=c then ba=c and $a^*b^*=c^*$.

The **quaternions $\mathbb{H}$** (the "H" stands for their inventor W.R.*H*amilton) arise by adjoining three new symbols $\mathbb{i},\mathbb{j},\mathbb{k}$ (which I'll occasionally instead call $\mathbb{i}_1,\mathbb{i}_2,\mathbb{i}_3$), to the reals, obeying

$$\mathbb{i}^2 = \mathbb{j}^2 = \mathbb{k}^2 = -1, \quad \mathbb{i}\mathbb{j} = \mathbb{k} = -\mathbb{j}\mathbb{i}, \quad \mathbb{j}\mathbb{k} = \mathbb{i} = -\mathbb{k}\mathbb{j}, \quad \mathbb{k}\mathbb{i} = \mathbb{j} = -\mathbb{i}\mathbb{k}.$$

Note these three symbols **anticommute**. Then every quaternion Q may be uniquely written $Q=w+x\mathbb{i}+y\mathbb{j}+z\mathbb{k}$ with w,x,y,z real. If y=z=0 then the quaternions become the complexes (which form a subalgebra). Useful notation then includes the *quaternionic conjugate* $\overline{Q}=w-x\mathbb{i}-y\mathbb{j}-z\mathbb{k}$, where ab=c ⇔ $\overline{b}\overline{a}=\overline{c}$; the *absolute value* $|Q|=[w^2+x^2+y^2+z^2]^{1/2}$; the nonnegative-real-valued *norm* $|Q|^2=Q\overline{Q}=\overline{Q}Q=w^2+x^2+y^2+z^2$; the reciprocal $Q^{-1}=\overline{Q}/|Q|^2$ if $Q\neq0$; and the real part re(Q)=w. If Q=w is then Q is called "pure real," while if w=0 then Q is a "pure imaginary" quaternion.

If *both* i and $\mathbb{i},\mathbb{j},\mathbb{k}$ are adjoined to the reals, or equivalently if we adjoin $\mathbb{i},\mathbb{j},\mathbb{k}$ to the complex numbers, then we get the **biquaternions** $B=w+x\mathbb{i}+y\mathbb{j}+z\mathbb{k}$ with w,x,y,z complex. Our use of two different notations for complex and quaternionic conjugation then comes in handy: $\overline{B}^*=w^*-x^*\mathbb{i}-y^*\mathbb{j}-z^*\mathbb{k}$ and $B^*=w^*+x^*\mathbb{i}+y^*\mathbb{j}+z^*\mathbb{k}$ and $\overline{B}=w-x\mathbb{i}-y\mathbb{j}-z\mathbb{k}$ and re(B)=x (where here x in general is *complex*). But the biquaternions are not a "division algebra" because B has no reciprocal whenever $|B|^2=w^2+x^2+y^2+z^2=0$, (where note the left hand side can be a general complex number), which is a larger un-invertible set of complex 4-tuples than just (w,x,y,z)=(0,0,0,0)

Then **my new biquaternion-valued-Ψ form of the Dirac equation** involving the quaternions $\mathbb{i},\mathbb{j},\mathbb{k}$ becomes

$$D_0\,\Psi\,\mathbb{k} + (\mathbb{i}\,D_1 + \mathbb{j}\,D_2 + \mathbb{k}\,D_3)\,\Psi + \Psi\,\mathbb{i}\,mc/\hbar = 0$$

I personally consider the biquaternions more annoying than the 2×2 complex matrices (although the two are isomorphic algebras) hence would not normally be excited about this biquaternion reformulation. However, what is interesting here is that we also can use this equation with ordinary (not bi) quaternion Ψ, because all the constants in the equation are plain quaternions. What we then have is **Majorana's wave equation** instead of Dirac's. E.Majorana in 1937 proposed his equation as suitable for uncharged spin-1/2 positive-mass fermions, such as neutrinos. Since they are uncharged, set $A_u=0$ in the formula for the covariant derivative, replacing D by ∂:

$$\partial_0\,\Psi\,\mathbb{k} + (\mathbb{i}\,\partial_1 + \mathbb{j}\,\partial_2 + \mathbb{k}\,\partial_3)\,\Psi + \Psi\,\mathbb{i}\,mc/\hbar = 0$$

We then see Majorana's equation has a very natural formulation using quaternion Ψ.

Also, H.Weyl in 1929 had proposed two equations suitable for *massless* fixed-helicity-chirality neutrinos. This was before neutrinos were discovered to have mass. If you do not believe that any massless fermions exist, then Weyl's equations are physically irrelevant. But in any case, our form of **Weyl's equation** is the same as our form of Dirac's equation with A=0 and m=0

$$\partial_0\,\Psi + (\sigma^1\,\partial_1 + \sigma^2\,\partial_2 + \sigma^3\,\partial_3)\,\Psi = 0$$

except that Ψ now is a *column 2-vector* with complex entries. We also have the **second Weyl equation**

$$\partial_0\,\Psi - (\sigma^1\,\partial_1 - \sigma^2\,\partial_2 - \sigma^3\,\partial_3)\,\Psi = 0;$$

the first Weyl equation governs right-handed massless neutrinos, and the second left-handed.

Now let us **translate** various common expressions between the old and new formats. Here is a **dictionary** (A^H where A is a matrix, denotes the complex-conjugated transpose of that matrix; A^T the plain transpose, so $A^T=A^{H*}$, $(AB)^H=B^HA^H$. For Majorana plain quaternions, simply omit all * in the biquaternion column):

name	column 4-vector Ψ	2×2 matrix Ψ	biquaternion Ψ
real-valued inner product	$A^* \cdot B = A^H B$	$\mathrm{trace}(A^H B) = \mathrm{trace}(A B^H)$	$2\,\mathrm{re}(\overline{A}^* B) = 2\,\mathrm{re}(A \overline{B}^*)$
Charge conjugation=complex conjugation	Ψ^*	Ψ^*	Ψ^*
Left-multiply by γ^0	$\gamma^0\,\Psi$	$-\Psi\,\sigma^2$	$i\,\Psi\,\mathbb{j}$
Left-multiply by γ^u, $u \in \{1,2,3\}$	$\gamma^u\,\Psi$	$\sigma^u\,\Psi\,\sigma^1\,i$	$-\mathbb{i}_u\,\Psi\,\mathbb{i}\,i$
Feynman slash $\not{B}$	$(\gamma^0,\gamma^1,\gamma^2,\gamma^3)\cdot B$	$(\overleftarrow{\sigma}^3, \vec{\sigma}^1, \vec{\sigma}^2, \vec{\sigma}^3)\cdot B\,\sigma^1 i$	$-i\,(\overleftarrow{\mathbb{k}}, \vec{\mathbb{i}}, \vec{\mathbb{j}}, \vec{\mathbb{k}})\cdot B\,\mathbb{i}$
Left-multiply by " γ^5 "	$i\gamma^0\gamma^1\gamma^2\gamma^3\,\Psi$	$\Psi\,\sigma^3$	$-i\,\Psi\,\mathbb{k}$
Matrix Hermitian adjoint		Ψ^H	$\overline{\Psi}^*$
Matrix transpose		Ψ^T	$\overline{\Psi}$
Matrix trace (complex valued)		$\mathrm{trace}(\Psi)$	$2\,\mathrm{re}(\Psi)$
Matrix determinant (complex valued)		$\det(\Psi)$	$\lvert\Psi\rvert^2$
(Nonneg. real) Probability density	$\Psi^* \cdot \Psi = \Psi^H \Psi$	$\lVert\Psi\rVert_F^2 = \mathrm{trace}(\Psi^H \Psi) = \mathrm{trace}(\Psi \Psi^H)$	$2\,\mathrm{re}(\overline{\Psi}^* \Psi) = 2\,\mathrm{re}(\Psi \overline{\Psi}^*)$
Dirac "spinor adjoint"	$(\gamma^0 \Psi)^H$	$-\sigma^2\,\Psi^H$	$-\mathbb{j}\,\overline{\Psi}^*\,i$
Pauli spin matrices		σ^u for $u \in \{1,2,3\}$	$\mathbb{i}_u\,i = i\,\mathbb{i}_u$ (σ^0 becomes 1)

In the "Feynman slash" line, the **arrows** indicate on which side (left or right) of the matrix or quaternion is the object (component of indexed-quantity B) it operates upon.

Charge conjugation is an outer automorphism of the gamma group. For my choice of gamma matrices, complex conjugation of Ψ, is, physically, charge conjugation: $\Psi^{(\mathrm{chargeconj})} = \Psi^*$. That is, Ψ^* obeys the same Dirac equation as Ψ except with negated charge: $e \leftrightarrow -e$. The existence and properties of the charge conjugation operation were first realized by H.Weyl, and not by P.A.M.Dirac, who indeed initially regarded his equation as a "theory of electrons and protons" and thought that there was some other factor which somehow disturbed the symmetry, thus explaining why protons were 1836 times heavier than electrons. However, Weyl then pointed out the *exact* charge conjugation symmetry, causing Dirac to predict the existence of the "positron," which then was observed in cosmic ray photographs by C.D.Anderson in 1932. I suggest that if Dirac had employed my "beautiful" forms of his equation instead of his uglier form, then this symmetry would have been too obvious to escape his notice.

Dirac's old non-negative-real conserved "**probability density**" $|\Psi|^2$ where $|X|^2$ denotes sum of absolute squared vector entries, now is replaced by the squared Frobenius matrix norm

$$\text{probdens} = ||\Psi||_F^2 = \text{trace}(\Psi^H\Psi) = \text{trace}(\Psi\Psi^H)$$

where the H denotes Hermitian matrix conjugation (i.e. complex conjugate transpose of 2×2 matrix).

Here note that $\Psi^H\Psi$ automatically is a nonnegative-definite hermitian matrix, so both its eigenvalues are nonnegative reals. Thus $\det(\Psi^H\Psi)$ is also always a nonnegative real, although in general $\det(\Psi)$ is complex-valued.

In the biquaternion or quaternion world, this is

$$\text{probdens} = 2\ \text{re}(\overline{\Psi}^*\Psi) = 2\ \text{re}(\Psi\overline{\Psi}^*).$$

Confusingly, many authors often write the "Dirac spinor adjoint of Ψ" as "$\overline{\Psi}$" but we refuse to do that because we reserve the overline symbol for quaternion conjugation.

Dirac's 3-vector "**probability current**"

$$J^u = (\gamma^0\ \Psi)^* \cdot \gamma^u\ \Psi$$

(also valid when u=0, in which case we get a 4-current; J^0 is the probability density), in the 2×2 matrix world becomes (albeit the following is only valid for $u \in \{1,2,3\}$ and *not* when u=0):

$$J^u = \text{trace}(\ \sigma^3\ \Psi^H\ \sigma^u\ \Psi\ ) = \text{trace}(\ \Psi^H\ \sigma^u\ \Psi\ \sigma^3\ )$$

while in the biquaternion world it becomes

$$J^u = -2\ \text{re}(\ \mathbb{k}\ \overline{\Psi}^*\ \mathbb{i}_u\ \Psi\ ) = -2\ \text{re}(\ \overline{\Psi}^*\ \mathbb{i}_u\ \Psi\ \mathbb{k}\ )$$

Parity (changing the signs of x,y,z): $\Psi^{(par)} = \gamma^0\ \Psi$ obeys the negated-(x,y,z) Dirac equation, including if also multiply by any complex scalar. However, only the particular values ±1 for that scalar cause "par" to be a self-inverse transformation. All other unit-norm-complex values would yield transformations which are self-inverse *except* for introducing a constant complex phase angle factor, which may or may not bother you. In the 2×2 matrix world this is $\Psi^{(par)} = -\Psi\ \sigma^2$. In the biquaternion world this is $\Psi^{(par)} = -\Psi\ \mathbb{j}\ i$. For Majorana (plain quaternion Ψ) this fails to preserve realness, but if we ignored the phase factor i then it would be $\Psi^{(par)} = \Psi\ \mathbb{j}$, which would cause $\Psi^{(par)(par)} = -\Psi$. Also in the Weyl case, one is not allowed to multiply a column-2-vector on the *right* by a 2×2 matrix, therefore again this map fails. There are reasons for these failures: the Majorana and Weyl equations are *not* symmetric under either parity or time reversal!

Time reversal (changing sign of t): By the CPT theorem time reversal is the same thing as charge-conjugation *combined* with parity, $\Psi^{(timerev)} = \gamma^0 \Psi^*$ with our gamma matrices. In the 2×2 matrix world this becomes $\Psi^{(timerev)} = -\Psi^* \sigma^2$. In the biquaternion world it is $\Psi^{(timerev)} = -\Psi^* \mathbb{j}\, i$. For Majorana, if we were to ignore the phase factor i which destroys realness, we would have $\Psi^{(timerev)} = \Psi^* \mathbb{j}$. In any case (i.e. without any phase factor cheating) with either biquaternions or Majorana we have $\Psi^{(timerev)(par)} = \Psi^*$. Similarly with Weyl, multiplying a column-2-vector on the *right* by a 2×2 matrix, although forbidden in general, is unobjectionable if the matrix happens to be the identity matrix, which, here, it is because $(-\sigma^2)^2$ equals the identity matrix. So again $\Psi^{(timerev)(par)} = \Psi^*$.

The **Dirac Lagrangian** is $L=(\gamma^0\Psi)^H(i\hbar c \not{D}-mc^2)\Psi$. To save space switch to units in which $\hbar$=c=1, so that

$$L = (\gamma^0\Psi)^H (i\not{D}-m) \Psi = (\gamma^0\Psi)^H (\gamma^0 D_0+\gamma^1 D_1+\gamma^2 D_2+\gamma^3 D_3+im) \Psi\, i.$$

In the 2×2 matrix world this becomes

$$L = \sigma^2\Psi^H [iD_0\Psi\sigma^2 +\sigma^1 D_1\Psi\sigma^1 +\sigma^2 D_2\Psi\sigma^1 +\sigma^3 D_3\Psi\sigma^1 +m\Psi].$$

In the biquaternion world it is

$$L = 2\, \mathrm{re}(\mathbb{j}\bar{\Psi}^* i\, [-D_0\Psi\mathbb{j} -\mathbb{i}D_1\Psi\mathbb{i} -\mathbb{j}D_2\Psi\mathbb{i} -\mathbb{k}D_3\Psi\mathbb{i} +m\Psi]).$$

A speculation

Somebody might complain "this all was merely a reformulation which does not yield any new physics" – therefore its value is merely aesthetic/cosmetic. But I speculate that complaint might not be correct... which will lead to a wider interesting idea.

Note Ψ now is a member of the non-commutative *ring* of 2×2 complex matrices – or in the Majorana case, the division ring of quaternions. Therefore we now are allowed to *multiply* Ψs, and in the Majorana case also *divide* them, whereas with Dirac's original formalism we only could add and subtract them. And if for some reason we did want to multiply two Ψs, then their complex-valued determinants, and nonnegative-real-valued |determinants|, would also multiply – or in the Majorana case their nonnegative-real-valued quaternion norms would multiply. (The 2×2 matrix determinant equals the biquaternion norm in the biquaternion view.)

These new abilities to multiply and/or divide Ψs open interesting new possibilities that simply could never happen with the old way of doing things.

Consider – what many physicists claim are among the top mysteries:

1. Why did Nature base the standard model (SM) on the gauge group U(1)×SU(2)×SU(3), from among the infinite number of gauge groups Nature could have chosen? Here U(1)×SU(2), loosely speaking, arises from the Glashow-Weinberg-Salam electroweak lepton model, while SU(3) comes from quark-gluon chromodynamics. SU(3) has 8 real degrees of freedom (the

same count as the Cayley-Graves octonions 𝕆), SU(2) has 3, and U(1) has 1.
2. Why does SM contain *three* "generations" of quarks and leptons? (As opposed to some other number?) Where the hell does the number "3" come from? Threes tend not to arise.

Well, two obvious possible sources for the number "3" are (a) the number of space dimensions, and (b) the fact that in our biquaternion form of the Dirac equation, the symbols (𝕚,𝕛,𝕜) could be replaced by any of its three cyclic shifts, to get an equivalent equation. Actually, we also could swap two while negating the third symbol e.g. (𝕛,𝕚,-𝕜), or negate all three while reversing their order (-𝕜,-𝕛,-𝕚), ... the full group of allowable permutations and symbol-negations has cardinality 4!=24 and consists of the even permutations with an even number of sign changes, and the odd permutations with an odd number of sign changes. (This group is isomorphic to S_4.) But anyhow 24 includes a factor of 3, and this "3" is indeed coming from the number of space dimensions.

But I'm not sure that all makes sense when we consider that, more generally, (𝕚,𝕛,𝕜) could be replaced by (Q^{-1}𝕚Q, Q^{-1}𝕛Q, Q^{-1}𝕜Q) for *any* quaternion Q, wlog with |Q|=1, which is a continuum-infinite number of symmetries. The new Dirac equation is equivalent to the old one in the sense that: If Ψ solved the original equation, then $Q^{-1}\Psi Q$ would solve the new one. These symmetries Q form the Lie group SU(2), which is isomorphic to the multiplicative group of unit-norm quaternions, and also to the automorphism group of that. The 3-element cyclic group discussed above is a tiny subgroup of that, but perhaps nevertheless could explain the "3 generations" mystery given suitable "symmetry breaking."

Which brings us to my possible solution to the U(1)×SU(2)×SU(3) question.

> **"Nature likes normed division rings" speculation:** Nature wants wavefunctions Ψ *not* merely to be objects with nonnegative-real-valued norms that can be added and subtracted, but *also* demands the ability to multiply and divide them (and when you do, their norms multiply or divide).

This speculation would tremendously reduce the number of possibilities, because **famous theorems** state that:

1. The *only* linear division algebras over the reals with a multiplicative identity element 1 (or we can even permit 1 not to exist if we demand power-associativity for powers≤4: AA·A=A·AA, AA·AA=A·AAA=AAA·A) and a nonnegative-real-valued multiplicative norm are ℝ, ℂ, ℍ, and the non-associative 8-dimensional Cayley-Graves "octonions" 𝕆.
2. The *only* linear division algebras over the reals with multiplication obeying or AA·B=A·AB are ℝ, ℂ, ℍ, 𝕆. (and this remains true even if the word "division" is weakened to the demand that there be no nilpotent ideals besides {0}).
3. Any continuous map from $S^{n-1}\times S^{n-1}\to S^{n-1}$ with a 2-sided identity, is 1, 2, 4 or 8 dimensional.

If Nature includes scalar particles with *complex-valued* wavefunctions (and it does: the Higgs boson), then physics should be expected to be symmetric under the ℂ-norm-preserving group, that is, U(1). Also if Nature includes photons obeying the Maxwell equations (and it does) we also get U(1). If Nature includes particles with *quaternion-valued* wavefunctions (Majorana fermions) then physics should be expected to be symmetric under the Aut(ℍ) group, that is, SU(2). If Nature includes particles with *octonion-valued* wavefunctions then physics should be expected to be symmetric under the Aut(𝕆) group, that is, the simple exceptional Lie group G_2, which has 14 real degrees of freedom.

Oops – G_2 is not what Standard Modelers wanted to hear, which would have been SU(3). And another apparent "oops" is that we've argued Dirac fermions may be regarded as *bi*quaternions, and the biquaternions are algebraically isomorphic to $GL_2(\mathbb{C})$, and the $\mathrm{Aut}(GL_2(\mathbb{C}))$ group is $SL_2(\mathbb{C})$ at least (which has 6 real degrees of freedom), anyhow something considerably larger than SU(2).

However, I riposte that SU(2) is the stabilizer of ℍ inside the biquaternions. And SU(3) is the elementwise stabilizer of ℂ inside G_2=Aut(𝕆). And $SO(4)\subset G_2$ is the stabilizer of ℍ inside 𝕆; and the intersection of that SO(4) and that SU(3) is U(2). And G_2 does not contain any simple Lie groups besides these and their subgroups. And a theorem by Gell-mann, Glashow, and Weinberg (see appendix A of Weinberg's ch.15) indicates that the only physically permissible gauge groups are direct products of simple Lie groups, and U(1)'s.

Given these facts, it is plausible to hope that the "Nature likes division" speculation plus some structural compatibility demands or partial-gauge-fixing terms in the Lagrangian might actually *force* U(1)×SU(2)×SU(3), or at most that plus a small finite set of rival possibilities. Certainly it would suffice to forbid F_4, E_6, E_7, E_8, SU(n) for all n≥4, SO(n) for all n≥5, and USp(2n) for all n≥2. In other words, this speculation would *explain* the "great U(1)×SU(2)×SU(3) mystery" up to at most a small finite number of rival options.

Suggested Future Project: Explore reformulations of the Standard Model based on quaternions, octonions, and possibly bi-octonions.

References

J.F.Adams: On the non-existence of elements of Hopf invariant one, Annals of Math. 72 (1960) 20-104.

J.F.Adams: Vector fields on spheres, Annals of Math. 75 (1962) 603-632.

A.A.Albert: On the right-alternative algebras, Annals of Math. (ser. 2) 50 (1949) 381-328; The structure of right-alternative algebras, 51 (1954) 408-417.

G.M.Benkart, D.J.Britten, J.M.Osborn: Real flexible division algebras, Canadian J. Math. 34,3 (1982) 550-588.

R.H.Bruck & Erwin Kleinfeld: The structure of alternative division rings, Proc. Amer. Math. Soc. 2 (1951) 878-896.

H-D.Ebbinghaus & 7 others: Numbers, Springer-Verlag (GTM#123) 1991. QA241.Z3413

A.Elduque: Third power associative composition algebras, Manuscripta Math. 84,1 (1994) 73-87.

Adolf Hurwitz: Über die Komposition der quadratischen Formen von beliebig vielen Variabeln, Nachrichten von den Königlichen Gesellschaft der Wissenschaften in Göttingen, Math.-Phys. Klasse (1898) 309-316 [Math. Werke II, 565-571].

I.L.Kantor & A.S.Solodovnikov: Hypercomplex numbers, An elementary introduction to algebras, Springer-Verlag 1989.

David I. Stewart: The reductive subgroups of G_2, J. Group Theory 13 (2010) 117-130.

Steven Weinberg: The Quantum theory of fields, vol. 1 foundations, 2 Modern applications, QC174.45 W45. (errata.)

Max Zorn: Alternativkörper und quadratische Systeme, Abh. Math. Seminar Hamburg 9 (1933) 395-402.