What Is the Meaning of Fractional Electric Charges?

Peter Rowlands¹

Dept. of Physics, Univ. of Liverpool, Oliver Lodge Laboratory, Oxford Street, Liverpool, L69 7ZE, UK

Abstract

The aim of this essay is not to challenge any accepted area of physics. The theory of quark phenomenology remains unaffected, as does the physical model of quarks based on fractional electric charges. However, it would seem that a different underlying or hidden symmetry may lead to a more successful resolution of at least two major problems in contemporary physics: Grand Unification and the application of the Higgs mechanism to fermion masses. In particular, it would seem that the fractional electric charges of quarks arise from the gauge invariance of the strong interaction and do not determine the deeper symmetries of the GU gauge group. The Han-Nambu representation has never been disproved, and fell out of favour more or less by accident. It would make sense to see if it can lead to insights in other problematic areas of physics, especially as it leads to testable predictions within our current experimental capacities.

Key Words: fractional electric charge, quark, gauge invariance, Grand Unification, physical assumption, FQXi, essay contest.

One of the most successful components of the Standard Model of particle physics is the theory of the strong interaction, which is explained as a force between ‘coloured’ quarks mediated by massless gluons and described by the non-Abelian gauge theory of quantum chromodynamics (QCD). There is no doubt that this theory has stood the test of experiment to a high level of precision and gives us the correct explanation of the structures of the composite particles known as baryons and mesons, which are respective three-quark and quark-antiquark ‘colourless’ combinations. Six quarks are now known. They are distributed between three generations, each composed of two weak isospin states – up / down, charm / strange, top / bottom, characterized in each generation by the weak isospin parameter, \( t_3 = \pm \frac{1}{2} \) – and there are a corresponding number of antiquarks. Each of the quarks comes in three colours, arbitrarily named red, green and blue. The three quarks in a baryon may be from any generation (or flavour) or from either isospin state, but must include one quark from each of the three colour options. So, protons are made up of two up and one down quark, and neutrons of two down quarks and one up, but each has one quark that is red, one that is green and one that is blue. Mesons are made up, similarly, of one quark and one antiquark of any flavour or isospin state, but they must be of the same colour / anticolour to preserve the colourless nature of the composite state.

In order to explain the structures of the baryons and mesons then known, the originators of the quark theory, Murray Gell-Mann and George Zweig, assumed that the up quark had an electric...
charge of $2e/3$, where $e$ is the fundamental electronic charge, while the down quark had a charge of $-e/3$. Subsequent discoveries showed that this pattern was repeated in the two further generations for the charm / strange and top / bottom quarks. Antiquarks were assumed to have the same electric charges with reversed sign. The phenomenology of quantum electrodynamics (QED) has shown over many experiments that quarks do behave as though constituted in exactly this way, with interactions between charges with fractional values of $e$. The three quarks could also be considered to contribute equally to the unit baryon number, $B = 1$, which indicated the presence of a source of the strong interaction and which is assumed to be identical for all baryons, however constituted.

<table>
<thead>
<tr>
<th></th>
<th>Blue</th>
<th>Green</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>up</td>
<td>$2e/3$</td>
<td>$2e/3$</td>
<td>$2e/3$</td>
</tr>
<tr>
<td>down</td>
<td>$-e/3$</td>
<td>$-e/3$</td>
<td>$-e/3$</td>
</tr>
<tr>
<td></td>
<td>$B/3$</td>
<td>$B/3$</td>
<td>$B/3$</td>
</tr>
</tbody>
</table>

There is no doubt that this is a successful and basically correct formulation. Quarks behave in QED in exactly the way that the theory predicts, and their fractional electrical charges are an established part of their structure, irrespective of the energy of the interaction. However, it is not necessarily an absolute guide to the possibly more fundamental symmetries that underlie the formulation. We have observed that baryons are constructed from three basic components that have charges of $2e/3$ or $-e/3$, but this does not mean that this is the structure demanded by the symmetry groups determining their behaviour at a more fundamental level. Physics has been shown to have a number of hidden symmetries which are not necessarily those that emerge directly from experiment. In working out the QED and QCD phenomenology of quarks we have necessarily developed a picture of three physical ‘particles’ interacting by a force known as the strong interaction, which has been observed to have certain identifiable characteristics. However, it may be that a more fundamental approach would privilege a more abstract basic principle which would lead to structures that could be physically interpretable as a force between a group of three particles, but that originated at a deeper level of explanation.

A particularly interesting fact, which is seldom now mentioned in textbooks, is that the original coloured quark theory of M. Y. Han and Yoichiro Nambu which followed on from the first quark theories of Gell-Mann and Zweig proposed a different assignment of electric charges to the quarks. $^5$ Han and Nambu saw that exactly the same results could be obtained using integral and zero charges and assigning an integral baryon number to a single quark.

<table>
<thead>
<tr>
<th></th>
<th>Blue</th>
<th>Green</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>up</td>
<td>$e$</td>
<td>$e$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$B$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>down</td>
<td>0</td>
<td>0</td>
<td>$-e$</td>
</tr>
<tr>
<td></td>
<td>$B$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
In this representation, ‘colour’ came about because the quarks were actually different in structure. In the Gell-Mann-Zweig version (GMZ), this had to be added subsequently as an extra property. It is important to realise, however, that the two models are, in fact, merely different representations of the same physical theory, and, from the point of view of quark phenomenology, represent exactly the same physics. Though it has been sometimes claimed that the Han-Nambu formulation (HN) implies that there must be a high energy regime where the integral nature of the charges will be revealed and the colours become distinguishable, this is looking at the problem from a phenomenological, rather than an abstract fundamental, point of view. If the fractional nature results from an exact symmetry of nature, such as the perfect gauge invariance of the strong interaction, then no such transition will occur.

This used to be recognised by authors of textbooks on particle physics. Frank Close, for example, wrote in 1979: ‘Imagine what would happen if the colour nonsinglets were pushed up to infinite masses. Clearly only colour 1 [singlets] would exist as physically observable states and quarks would in consequence be permanently confined. At any finite energy we would only see the ‘average’ quark changes and phenomenologically we could not distinguish this from the Gell-Mann model where the quarks form three identical triplets.’ Of course, if the strong interaction SU(3) is an exact symmetry, as experimental evidence so far seems to suggest, then the fractional electric charges we observe will not be ‘averages’, but exact values, reflecting a perfect equivalence between the different coloured states or phases of the interaction. They will be QED or electroweak eigenstates.

When the quark theory was first proposed, no fractional charges had been observed in any form of physics. Subsequently, however, the fractional quantum Hall effect was discovered, and electrons appeared with effective charges of $e/3$, $e/5$ and other fractional values. In principle, this was because an electron or other fermion could form a pseudobosonic combination with an odd number of magnetic flux lines and so effectively share itself out between them, with the consequent appearance of nonintegral charge values. It is interesting to imagine what would have happened if the effect had been discovered before the quark theory was postulated. It might never have occurred to anyone that true fractional charges could actually exist.

In fact, the discoverer of the fractional quantum Hall effect, Robert Laughlin went so far as to suggest the connection with particle physics in his Nobel Lecture: ‘The fractional quantum Hall effect is fascinating for a long list of reasons, but it is important in my view primarily for one: It establishes experimentally that both particles carrying an exact fraction of the electron charge $e$ and powerful gauge forces between these particles, two central postulates of the standard model of elementary particles, can arise spontaneously as emergent phenomena. Other important aspects of the standard model, such as free fermions, relativity, renormalizability, spontaneous symmetry breaking, and the Higgs mechanism, already have apt solid-state analogues and in some cases were even modeled after them, but fractional quantum numbers and gauge fields were thought to be fundamental, meaning that one had to postulate them. This is evidently not true.’

Despite all this, the underlying representation based on integral charges has seemingly dropped out of view entirely in the last thirty years. It was never disproved, but just faded away, despite the repeated citation of the HN paper as the origin of colour theory. If a reason needed to be
given for this, it would be that the phenomenology supports fractional charges, which is, of course, true for both representations. It would seem that the building of large accelerators and the opportunities they provided for experimental investigation of models tended to concentrate effort into providing detailed prediction and interpretation of the phenomenology and that, if an established model was able to do this, then there was no point in looking at alternatives which provided identical results.

So, why does it matter which representation we use if they both lead to the same model for QED? The answer here is that some physical theories depend on deeper and sometimes hidden symmetries which may not be obvious but have a significant effect on the predictions that can be made. There are two areas where the deeper symmetries might produce significantly different results for different representations of a model which gives the same QED phenomenology. One is in Grand Unified Theory (GUT) and the other is in the Higgs mechanism as applied to fermion masses. Now, Grand Unification might be considered a successful idea but not yet a successful theory. That is, the principle seems sound, but has yet to achieve a successful resolution. The theory begins with the (successful) Glashow-Weinberg-Salam SU(2) × U(1) unification of the electric and weak interactions. This is governed by the weak mixing angle parameter sin²θw, which is effectively the ratio between the weak and electric couplings (α/αe).

Georgi and Glashow were able to extend this towards including the strong interaction in a Grand Unified (GU) scheme, based on the SU(5) group, by showing that, in any such scheme determined by a single GU gauge group, sin²θw would be given by the ratio of the sum of all the squared units of weak isospin for the fermions of the Standard Model to the sum of all their squared units of electric charge (Q). \[ \sin^2 \theta_w = \frac{Tr(t^2_3)}{Tr(Q^2)}. \]

If we take the weak components with only left-handed contributions to weak isospin, for the first generation of quarks and leptons, that is, for 3 colours of u, 3 colours of d, and the leptons e and ν, we obtain:

\[ Tr(t^2_3) = \frac{1}{4} \times 8 = 2. \]

Quarks and leptons have identical units of weak isospin, and so this summation will be the same for GMZ and HN, and will also be the result expected for phenomenology. But for the electric charge structure, the summations of GMZ and HN will diverge. For, GMZ and phenomenology, with both left- and right-handed contributions in the first generation, we obtain

\[ Tr(Q^2) = 2 \times \left( \frac{4}{9} \times 3 \times \frac{1}{9} \times 3 + 1 + 1 + 1 + 0 \right) = \frac{16}{3} \]

from which

\[ \sin^2 \theta_w = 0.375. \]

For HN, however, we have

\[ Tr(Q^2) = 2 \times (1 + 1 + 0 + 0 + 1 + 1 + 0) = 8, \]
Steven Weinberg is one of a large number of authors who have observed that the value 0.375 for \( \sin^2 \theta_W \) is in ‘gross disagreement’ with the experimental value of 0.231 at around the mass-energy of the Z particle \( (M_Z = 91 \text{ GeV}) \). On the other hand, 0.25 is relatively close to this value and would be even closer (with some small second order corrections) if the effect of the direct production of W and Z bosons at their mass-energies \( M_W \) and \( M_Z \) is taken into account (or if the 0.25 occurs at the vacuum expectation energy \( (246 \text{ GeV}) \) rather than at \( M_W \) or \( M_Z \)). In addition, 0.25 is the value that would be obtained purely from the leptonic contribution, and it is rather curious that the value for a purely electroweak parameter should be different in the quark and lepton sectors.

Another curious aspect of the original ‘minimal SU(5)’ GUT proposed by Georgi and Glashow is that it doesn’t actually unify the pure interactions, for, though the theory begins with the equations for the running weak and strong coupling constants, derived from their respective \( SU(2) \) and \( SU(3) \) structures:

\[
\frac{1}{\alpha_2(\mu)} = \frac{1}{\alpha_G} - \frac{5}{6\pi} \ln \frac{M_X^2}{\mu^2},
\]

and

\[
\frac{1}{\alpha_3(\mu)} = \frac{1}{\alpha_G} - \frac{7}{4\pi} \ln \frac{M_X^2}{\mu^2},
\]

(where \( M_X \) is the GU energy scale, \( \alpha_G \) is the fine structure constant at this energy and \( \mu \) is the energy scale of measurement) it assumes that the grand unified gauge group structure will modify the equivalent \( U(1) \) equation for the electromagnetic coupling \( (1/\alpha) \), assumed (in this theory) to be

\[
\frac{1}{\alpha(\mu)} = \frac{1}{\alpha_G} + \frac{5}{3\pi} \ln \frac{M_X^2}{\mu^2},
\]

to one in which it is mixed with the weak value, based on \( SU(2) \times U(1) \). So, now we have

\[
\frac{1}{\alpha_1(\mu)} = \frac{1}{\alpha_G} + \frac{1}{6\pi} \ln \frac{M_X^2}{\mu^2},
\]

where

\[
\frac{5}{3\alpha_1(\mu)} + \frac{1}{\alpha_2} = \frac{1}{\alpha}.
\]

From these equations, we derive a grand unified mass scale \( (M_X) \) of order \( 10^{15} \text{ GeV} \), and from

\[
\sin^2 \theta_W = \frac{\alpha(\mu)}{\alpha_2(\mu)},
\]

we find ‘renormalized’ values of \( \sin^2 \theta_W \) at the measurement scale of order 0.19 to 0.21.
The idea was a breakthrough when first proposed, but the problem, as is well known, is that it doesn’t work. The curves representing the variations of the parameters $\alpha_1$, $\alpha_2$ and $\alpha_3$ at different energy scales ($\mu$) don’t actually cross at a point or anything very close to one, leading to the somewhat ad hoc proposal that a supersymmetric model may be the only solution. In addition we are forced to use a combined electroweak parameter which makes assumptions about group structure, and relies on a particular value for the squared ‘Clebsch-Gordan coefficient’ of the group, $C^2 = 1 / \sin^2 \theta_W - 1 = 5 / 3$, that has, as yet, no experimental or theoretical justification – though it is clear that using the equation for the purely electric fine structure constant would clearly not have produced anything closely resembling unification. Unifying electroweak, weak and strong parameters seems to be rather less convincing than if we had used the original electric, weak and strong, while the assumed value of $\sin^2 \theta_W = 0.375$ at GU suggests that the electroweak unification is not even then complete, as the two forces are not an equal footing. In addition, the convergence, such as it is, is three or four orders of magnitude below the scale of the Planck energy at which quantum gravity is assumed to operate, suggesting that another principle will be needed to include gravitation. But, even worse than all these is the fact that, compensating errors in the combination tend to disguise the massive inconsistencies between the separate equations for the coupling constants. In particular, recalculating the value of $\sin^2 \theta_W$ at $\mu = 10^{15}$ GeV gives 0.6 rather than the 0.375 which was initially assumed in setting up the equations.

It has long been recognised that the symmetries of the $SU(5)$ group, when applied at a fundamental level to quarks using the GMZ representation, do not provide the correct answers, though $SU(5)$ has many of the aspects that we would require from a GU group. What happens, then, if we switch to the HN representation? Here, we have an independent value for $\sin^2 \theta_W = \alpha / \alpha_2$ of the right order, and we can perform a much simpler calculation for $M_X$ without making assumptions about the group structure, by avoiding the problematic running coupling constant equation for $1 / \alpha$, using only the more secure equations for $1 / \alpha_2$ and $1 / \alpha_3$. In addition, the hypercharge numbers for the $U(1)$ electromagnetic running coupling equation will now be no longer identical to those for a quark model based purely on QED phenomenology. The fermionic contribution to QED vacuum polarization is, for GMZ,$^{14}$

$$\frac{4}{3} \times \frac{1}{2} \left( \frac{1}{36} \times 3 + \frac{1}{36} \times 3 + \frac{1}{3} \times 3 + \frac{4}{9} \times 3 + \frac{1}{4} \times 1 + \frac{1}{4} \times 1 + 1 \right) \frac{n_g}{4\pi} = \frac{5}{3\pi},$$

where $n_g = 3$ is the number of fermion generations, and the terms in the bracket represent, respectively, the squared average charge in the isospin quark doublet, the squared charges of the quarks, the squared average charge of the isospin lepton doublet, and the squared charges of the leptons, all for both left- and right-handed states,$^{14}$ but, modifying this for HN, we obtain:

$$\frac{4}{3} \times \frac{1}{2} \left( \frac{1}{4} \times 3 + \frac{1}{4} \times 3 + \frac{1}{4} \times 1 + 1 + 0 + 0 + 0 + 1 + \frac{1}{4} \times 1 + \frac{1}{4} \times 1 + 1 \right) \frac{n_g}{4\pi} = \frac{3}{\pi}.$$
This result corresponds to a change in the squared Clebsch-Gordan coefficient from $C^2 = \frac{5}{3}$ to $C^2 = \frac{3}{5}$, when $\sin^2 \theta_W = 1 / (1 + C^2)$ changes from 0.375 to 0.25. With the new values we have obtained for the hypercharge numbers, the running coupling of the pure electromagnetic interaction, will be:

$$
\frac{1}{\alpha(\mu)} = \frac{1}{\alpha_G} + \frac{3}{\pi} \ln \frac{M_X^2}{\mu^2}.
$$

Leaving out the speculative equation for $1/\alpha_1$, and, for the moment, this new one for $1/\alpha$, but using the well-established ones for $1/\alpha_2$ and $1/\alpha_3$, and $\sin^2 \theta_W = \alpha / \alpha_2$, we obtain

$$
\sin^2 \theta_W(\mu) = \alpha(\mu) \left( \frac{1}{\alpha_3(\mu)} + \frac{11}{6\pi} \ln \frac{M_X}{\mu} \right).
$$

Taking typical values for $\mu = M_Z = 91.2$ GeV, $\alpha_3(M_Z^2) = 0.118$ (or 0.12), $\alpha(M_Z^2) = 1 / 128$, and $\sin^2 \theta_W = 0.25$, we obtain a value for the GU energy scale ($M_X = 2.8 \times 10^{19}$ GeV) which is extraordinarily close to the Planck value ($1.22 \times 10^{19}$ GeV), and may well be exactly so, as purely first-order calculations overestimate the value of $M_X$. Assuming that $M_X$ is the Planck mass, we obtain $\alpha_G$ (the GU value for all interactions) = 1 / 52.4, and $\alpha_3(M_Z^2) = 1 / 31.5$, which is exactly the kind of value we would expect for the weak coupling with $\sin^2 \theta_W = 0.25$ close to $M_Z$. To provide an independent check on the validity of the procedure, we can now make direct use of the equation we have derived for $1/\alpha$, with the new hypercharge numbers and GU at the Planck mass, to obtain $1/\alpha(M_Z^2) = 128$, which is, of course, exactly the value obtained experimentally at energies corresponding to $\mu = M_Z$. This appears to be a striking confirmation of the assumptions made in the first calculation, leading to $M_X$, as coincidental agreements are most unlikely for equations involving logarithmic terms, and it is also potentially very significant, for it would now appear that the unification which occurs at $M_X$ might well involve a direct numerical equalization of the strengths of the three, or even four, physical force manifestations, without reference to the exact unification structure.

The analysis suggests that, at grand unification, $C^2 = 0$ and $\sin^2 \theta_W = 1$, creating an exact symmetry in every respect between weak and electric interactions, as well as between weak and strong, which is completely different from the only partial unification achieved using GMZ, and linking this with the scale associated with quantum gravity. The mixing parameter, $\sin^2 \theta_W$, as normally understood, may then be interpretable as the electroweak constant for a specifically broken symmetry, taking the value of 0.25 at the energy range where the symmetry breaking occurs (presumably at $M_Z$), or, alternatively, the expectation value of the Higgs field, 246 GeV, and gradually decreasing from the maximum ($\sin^2 \theta_W = 1$) to this value at intermediate energies. At GU, we may suppose, all four forces are reduced to scalar phases, with $U(1)$ symmetry and purely Coulombic interaction, all distinguishing aspects of the weak and strong interactions having diminished to zero. One of the most significant aspects of the calculation is that it leads to completely testable predictions, as the values of the three coupling constants can be calculated for any energy with relative precision from the known values of $\alpha_G$ and $M_X$. In particular, the value of $\alpha$ changes rapidly in a way that can be determined at energies now
available to us experimentally. At 14 TeV, for instance, it would have the value of \(1/118\), compared to \(1/125\) from the minimal \(SU(5)\) theory of Georgi and Glashow.\(^{13,15}\)

Relatively simple considerations based on results from the extensive quantum field theories of the electric, weak and strong interactions, which require only a small amount of arithmetical and algebraic manipulation, thus suggest that, if the HN representation is valid, then it has major consequences for GU, which are accessible by experiment. Of course, it would also have other advantages in making quark-lepton unification (in symmetry terms) much more likely, as both sectors would now be characterized by integral charges. In addition, we would not expect the deeper symmetry of the GU gauge group to be determined only by the one component in which a division by 3 appears, but by principles which are clearly more general. In fact, it is not unreasonable to suppose that the fundamentally integral nature of charges is a strong argument in favour of unification in that the QED phenomenology of quarks is determined by the behaviour of a nonelectric (strong) force, just as the QED phenomenology of electrons in the fractional quantum Hall effect is seemingly determined by the nonelectric (presumably weak) force involved in creating a pseudobosonic state. The problem faced by having 3 separate ‘units’ of charge \(e/3, 2e/3\) and \(e\), with its implication that the electron, with \(e\), may not actually be ‘elementary’, would also disappear. Yet, there is an even more significant consequence in relation to the application of the Higgs mechanism for fermion masses.

This has always been recognized as one of the most problematic and \textit{ad hoc} aspects of the Higgs theory. Apart from the fact that there is no known scale for the coupling which would generate the observed masses, there is also the fact that to generate separate masses for the two isospin states in each generation, we require two different hypercharge (or \(2 \times \) average charge) units of 1 and \(-1\), yet in GMZ there is only one hypercharge value for all quarks, and that is the fractional value, \(2/3\). The only expedient then is to ‘invent’ two hypercharges not justified by the assumed charge structure.\(^{17}\) In the HN representation, however, the different colours of quark automatically produce the two hypercharge values, 1 and \(-1\), which we require for both isospin states and which would be repeated in each generation.\(^{13}\)

Of course, the leptons are not fractionally charged, but there is a separate area of difficulty, here, for both GMZ and HN. In the past, the lepton mass mechanism could be accommodated by assigning the single hypercharge value in the first generation to electrons, but the discovery of neutrino masses means that the opposite hypercharge value is now required for neutrinos. It is possible that this difficulty can be resolved in both representations if the neutrino is a Majorana particle, with a low mass resulting from the low probability of the neutrino transforming to its antistate with the opposite hypercharge.

The aim of this essay has not been to challenge any accepted area of physics. The theory of quark phenomenology remains unaffected, as does the physical model of quarks based on fractional electric charges. However, it would seem that a different underlying or hidden symmetry may lead to a more successful resolution of at least two major problems in contemporary physics: Grand Unification and the application of the Higgs mechanism to fermion masses. In particular, it would seem that the fractional electric charges of quarks arise from the gauge invariance of the strong interaction and do not determine the deeper symmetries of the GU gauge group. The HN representation, suggested originally by one Nobel Prize winner and
indirectly supported by another, has never been disproved, and fell out of favour more or less by accident. It would make sense to see if it can lead to insights in other problematic areas of physics, especially as it leads to testable predictions within our current experimental capacities.

References