

## THE FACTOR 2 AND PHYSICAL AND MATHEMATICAL DUALITY

**Peter Rowlands**

IQ Group and Science Communication Unit, Department of Physics, University of Liverpool, Oliver Lodge Laboratory, Oxford Street, P.O. Box 147, Liverpool, L69 7ZE, UK. e-mail: p.rowlands@liverpool.ac.uk

*Abstract.* Analysis of a numerical factor 2 or  $\frac{1}{2}$ , which occurs in many physics equations, suggests that, in all significant examples, classical, relativistic and quantum, it has a common origin, a principle of duality, in which both physics and mathematics structure themselves by trying to avoid creating ‘something’ from nothing.

### 1 Kinematics and the virial theorem

There is a purely geometric factor 2 in the formula for the area of a triangle,  $\frac{1}{2} \times \text{length of base} \times \text{perpendicular height}$ . This can be applied to kinematics if we represent a motion under uniform acceleration ( $a$ ) by a straight-line graph of velocity ( $v$ ) against time ( $t$ ). The area under the graph now becomes the distance travelled,  $\frac{1}{2} vt$ . If the motion had been under uniform velocity, the distance would have been represented as the area of a rectangle,  $vt$ . Here, the factor 2, distinguishes between steady conditions and steadily *changing* conditions. We can develop the idea further to produce the well-known kinematic equations for uniform acceleration, starting from initial velocity  $u$ , such as  $v^2 = u^2 + 2as$ . A body of mass  $m$ , acted on by a uniform force  $F = ma$ , then acquires kinetic energy  $\frac{1}{2} mv^2$  or  $p^2 / 2m$ , if we express it in terms of momentum,  $p = mv$ . It is easy, of course, to show that the last two formulae apply even when the acceleration is nonuniform.

The kinetic energy formula applies when a system is undergoing change. A classic example is that of a body of mass  $m$  escaping from a gravitational field, where

$$\frac{mv^2}{2} = \frac{GMm}{r}.$$

But if a system is in steady-state, as in a classical circular gravitational orbit, we use a *potential* energy relation, in which the potential energy has twice the value of kinetic:

$$mv^2 = \frac{GMm}{r}.$$

Strictly, this is true only for systems involving constant or inverse-square-law forces, for the more general *virial* theorem, relates the time-averaged kinetic and potential energies,  $\bar{T}$  and  $\bar{V}$ , in a conservative system, governed by forces proportional to  $r^{-n}$ , by the formula:

$$\bar{T} = \frac{(1-n)}{2} \bar{V}.$$

This brings in another aspect of the factor 2, for constant and inverse-square-law forces are characteristic of a universe structured within a three-dimensional Euclidean space, which brings us back to the geometric origin of the factor 2 in the formulae for the area of a triangle.

One way of looking at the kinetic and potential energy relations is to say that the first is concerned only with the action side of Newton's third law, while the second concerns both action and reaction. In fact, an old proof of Newton's of the  $mv^2 / r$  law for centripetal force, and thus of the formula  $mv^2$  for orbital potential energy, was based on a doubling of the momentum through action and reaction from an imagined infinite sided polygon defining the orbit of the motion. The same momentum doubling by reflection occurs in a more directly physical context in the elementary derivation of the pressure-density relation for an ideal gas in steady state,  $P = \rho \bar{c}^2 / 3$ , which has the same 'potential energy' format as  $mv^2$ , though divided randomly between three dimensions. To find the average kinetic energy of the gas molecules, we have to make an explicit use of the virial theorem for conditions equivalent to a constant force.

What is surprising, however, is that photons, which, unlike material particles, are relativistic objects, behave in exactly the same way in a 'photon gas', producing a radiation pressure of the form  $P = \rho c^2 / 3$ , with the relativistic energy  $E = mc^2$  behaving exactly like a classical potential energy term, and with no mysterious 'relativistic factor' at work. We can consider the photons as being reflected off the walls of the container in the exactly the same way as the molecules of materials although the real process obviously also involves absorption and re-emission.

## 2 Relativity

The doubling of the energy term in  $E = mc^2$ , by comparison with classical potential energy, is sometimes described as 'relativistic', but relativistic factors tend to be of the form  $\gamma = (1 - v^2 / c^2)^{-1/2}$ , suggesting some *gradual* change when  $v \rightarrow c$ , and abrupt transition involving a discrete integer.  $\Delta E = \Delta mc^2$ , used for material particles with rest mass  $m$ , is a relativistic equation because it incorporates the  $\gamma$  factor in the  $\Delta m$  term, but  $E = mc^2$  is simply a requirement needed by Einstein to reconcile special relativity with classical energy-conservation laws, and, since it is determined solely by an integration constant, it cannot be derived directly from relativity itself. In the case of photons, which have no rest mass and no kinematics, there is no distinction between a 'relativistic' approach and one based on classical potential energy, and it is perfectly possible to do classical calculations for photons, entirely independent of any concept of relativity.

In addition, even though free photons have no kinematics, it is also perfectly possible to treat photons acting under the constraint of certain forces as

though they have. One example occurs in a plasma, where they acquire an effective ‘rest mass’. Another occurs in a gravitational field, where the light ‘slows down’, and behaves exactly as if it had kinetic energy in the field. This is why it is possible to use the standard Newtonian escape velocity equation

$$\frac{mv^2}{2} = \frac{GMm}{r}$$

to derive the Schwarzschild limit for a black hole, as was done as early as the eighteenth century, by assuming  $v \rightarrow c$ , with no transition to a ‘relativistic’ value.

Less obviously, but equally correctly, we can derive the full double gravitational bending of light, using the kinetic equation,

$$\frac{mc^2}{2} = \frac{GMm(e-1)}{r},$$

for an orbit which may be assumed to be hyperbolic with eccentricity  $e$ . We do *not* use the potential energy equation

$$mc^2 = \frac{GMm(e-1)}{r},$$

which requires steady-state conditions, which do not apply when an orbit is in the process of creation, as here. From the potential energy equation, with  $1 \ll e$ , the full angle deflection (in and out of the gravitational field) is easily derived as

$$\frac{2}{e} = \frac{2GM}{c^2 r}.$$

Contrary to popular opinion, Soldner, who attempted a calculation in 1801, did, in fact, use the correct kinetic energy, and not the incorrect potential energy equation, and only obtained an incorrect final result because he integrated over the single, rather than double, angle. Soldner rightly saw the procedure as being a kind of reverse analogy of Laplace’s black hole calculation, though using a hyperbolic rather than a circular orbit, the significant fact being that the photon’s speed outside the gravitational field is not determined by it.

Of course, the fact that we can do a calculation of the full deflection using a classical argument doesn’t mean that we *can’t* use a special or general relativistic argument to derive the effect. The work of many authors has shown that we can. What it does mean is that the cause of the effect itself is independent of the particular version of physics we use to calculate it. Something more profound is involved. This seems to be the fact that, in every case where a ‘relativistic’ correction (either special or general) seems to ‘cause’ the doubling of a physical effect, the relativistic aspect, like classical kinetic energy, is providing a way of incorporating the effect of *changing conditions*.

(And the same argument applies also to the calculation of planetary perihelion precession.)

It isn't necessarily important what particular physical phenomena we invoke to support the calculation, and the many disagreements over the 'cause' of the double deflection bear witness to this. Authors have generally agreed that there are two separate physical components involved, but not on what they are. In principle, it would seem that the potential energy equation is responsible for gravitational redshift, or time dilation, which gives half the effect, while relativity adds the corresponding length contraction. Now, this may be interpreted as redshift being 'Newtonian' while the length-contraction or 'space-warping' is relativistic, but it might be that the length contraction is Newtonian while the redshift is relativistic. It has also been argued that the 'Newtonian' effect has to be added to the Einstein equivalence principle calculation of 1911 (which again gives half the effect), but a counter-argument suggests that these two effects are the same, and need supplementing with a 'true' relativistic effect, like the Thomas precession. Amazingly, *all* of these arguments are correct! They are by no means mutually exclusive. In reality, it depends on the choice of classical energy equation. If we use the potential energy equation where the kinetic energy equation is appropriate, then we can find correct physical reasons for almost *any* additional term which doubles the effect predicted.

### 3 Spin and the anomalous magnetic moment

Almost exactly the same reasoning can be shown to apply to the anomalous magnetic moment or, equivalently, the gyromagnetic ratio, of a Bohr electron acquiring energy in a magnetic field. According to 'classical' reasoning, we are told, the energy acquired by an electron changing its angular frequency from  $\omega_0$  to  $\omega$  in a magnetic field  $\mathbf{B}$  will be of the form

$$m (\omega^2 - \omega_0^2) = e\omega_0 r B ,$$

with a corresponding angular frequency change  $\Delta\omega = eB / 2mr$ . But a relativistic effect (the Thomas precession, again!) replaces the classical  $e\omega_0 r B$  by  $2e\omega_0 r B$ , doubling the value of  $\Delta\omega$ . All we need to do, however, to obtain the correct value of  $\Delta\omega$  is to realise that we must use the *kinetic* energy equation when we have changing conditions, as, for example, *at the instant we 'switch on' the field*. Then, we automatically write

$$\frac{1}{2} m(\omega^2 - \omega_0^2) = e\omega_0 r B ,$$

which is nothing more than a version of the kinematic equation  $v^2 - u^2 = 2as$ . The Thomas precession is only needed as a 'relativistic' correction if we begin with the potential energy equation applicable to a steady state.

In showing that the gyromagnetic ratio of a Bohr electron is not truly ‘anomalous’ or relativistic in origin, but perfectly capable of a classical explanation, we are also showing that the origin of the factor 2 in the electron spin term is not of a fundamentally quantum origin either. Traditionally, of course, electron spin is derived from the relativistic Dirac equation by consideration of the commutator

$$[\hat{\mathbf{G}}, \mathcal{H}] = [\hat{\mathbf{G}}, i\gamma_0\boldsymbol{\gamma}\cdot\mathbf{p} + \gamma_0m] .$$

Purely formal reasoning reduces this to  $2\gamma_0 \boldsymbol{\gamma} \times \mathbf{p}$ , which, in the multivariate vector terminology used by the author (equivalent to Pauli matrices), becomes  $2ij \mathbf{1} \times \mathbf{p}$ . The factor 2 here emerges directly from the anticommuting properties of the vector operators  $\boldsymbol{\gamma}$  and  $\mathbf{p}$ , ultimately leading to

$$[\mathbf{L} + \hat{\mathbf{G}} / 2, \mathcal{H}] = 0 .$$

with the total angular momentum  $(\mathbf{L} + \hat{\mathbf{G}} / 2)$ , including the spin term  $\hat{\mathbf{G}} / 2$ .

Originally, with its automatic derivation from the Dirac equation, this term was thought to be related to the relativistic nature of the Dirac equation. However, the same result (or, more specifically, in its manifestation in the presence of a magnetic field) can be derived from the nonrelativistic Schrödinger equation, if we use a *multivariate* momentum operator, as we do automatically in the Dirac equation. Significantly for our purposes, the standard derivation of the Schrödinger equation proceeds by quantizing the classical expression for kinetic energy:

$$T = (E - V) = \frac{p^2}{2m} ,$$

using the operator substitutions  $E = i \partial / \partial t$  and  $\mathbf{p} = (-i\nabla + e\mathbf{A})$ , in the presence of a magnetic field determined by vector potential  $\mathbf{A}$ . Normally, the right-hand side of this equation is interpreted as  $-\nabla^2\psi / 2m$ , using the scalar product  $(-i\nabla + e\mathbf{A}).(-i\nabla + e\mathbf{A})$ . However, various authors (e.g. Gough<sup>1</sup>) have shown that, using a multivariate operator for  $\mathbf{p} = -i\nabla + e\mathbf{A}$ , we obtain:

$$2mE\psi = (-i\nabla + e\mathbf{A}) (-i\nabla + e\mathbf{A}) \psi$$

which leads ultimately to

$$2mE\psi = (-i\nabla + e\mathbf{A}).(-i\nabla + e\mathbf{A}) \psi + 2m \boldsymbol{\mu}\cdot\mathbf{B} ,$$

which we recognize as the form of the Schrödinger equation in a magnetic field, with a spin state supplied by the *ad hoc* addition of Pauli matrices. It becomes

an automatic component of our equation because we define a *full product* between multivariate vectors (or, equivalently, Pauli matrices or complex quaternions)  $\mathbf{a}$  and  $\mathbf{b}$  of the form  $\mathbf{a}\mathbf{b} = \mathbf{a}\cdot\mathbf{b} + i \mathbf{a} \times \mathbf{b}$ .

If real vectors, such as those representing space and momentum, are intrinsically multivariate, then a spin term will automatically result from taking the full product, and the  $\frac{1}{2}$ -integral value of fermionic spin (incorporated here in the  $2m \boldsymbol{\mu}\cdot\mathbf{B}$  term) becomes a consequence of the vectors' anticommuting properties. Relativity doesn't come into it; it is a vector, not a 4-vector effect, as we can also see from the well-known fact that the  $4\pi$  rotation involved in spin is purely a property of the rotation group. At the same time, the  $\frac{1}{2}$  in the Schrödinger equation itself clearly comes from the equation's initial derivation from the expression for classical kinetic energy.

A similar situation occurs when the Schrödinger equation is solved for the case of the quantum harmonic oscillator, with a varying *potential* energy term,  $\frac{1}{2} m\omega^2 x^2$ , taken directly from the classical kinetic energy term  $\frac{1}{2} mv^2$ , added to the Hamiltonian. The  $\frac{1}{2}$  in this expression then leads by direct derivation to the  $\frac{1}{2}$  in the expression for the ground state or 'zero-point' energy of the system,

$$E_0 = \frac{\hbar\omega}{2} .$$

This zero-point energy term can be related directly to the  $\hbar / 2$  in the Heisenberg uncertainty relation, but the formal derivation of this relation also shows that the factor  $\frac{1}{2}$  is generated by anticommutativity in the same way as it is for electron spin. We assume a state represented by a state vector  $\psi$  which is an eigenvector of the operator  $P$ . Then, if  $Q$  is an operator which anticommutes with  $P$ , we obtain

$$(\Delta p) (\Delta q) \geq (1/2) [P, Q] \geq \hbar / 2$$

where the noncommutation of the  $p$  operator introduces the factor 2.

#### 4 Fermions and bosons

The factor 2 in spin states establishes the distinction between bosons and fermions, with bosons occupying integral spin states and fermions half-integral ones. Ultimately, this factor can be shown to originate in the virial relation between kinetic and potential energies. To do this, it is most convenient to use the nilpotent formulation of the Dirac equation for fermions, derived in earlier papers.<sup>2</sup> In this procedure, we devise an algebra combining quaternions ( $i, j, k, 1$ ) and multivariate 4-vectors ( $i, \mathbf{i}, \mathbf{j}, \mathbf{k}$ ) to represent the five gamma matrices ( $\gamma^0 = i\mathbf{k}; \gamma^1 = i\mathbf{i}; \gamma^2 = \mathbf{j}\mathbf{i}; \gamma^3 = \mathbf{k}\mathbf{i}; \gamma^5 = ij$ ). We then take the classical relativistic energy-momentum conservation equation:

$$E^2 - p^2 c^2 - m_0^2 c^4 = 0 ,$$

where  $m_0$  is rest mass, and factorize using our quaternion-multivariate-4-vector algebra to give:

$$(\pm kE \pm i\mathbf{p} + ij m_0) (\pm kE \pm i\mathbf{p} + ij m_0) = 0 ,$$

before quantizing to obtain

$$\left( \pm ik \frac{\partial}{\partial t} \pm i \nabla + ij m_0 \right) \psi = 0 ,$$

where

$$\psi = (\pm kE \pm i\mathbf{p} + ij m_0) e^{-i(Et - \mathbf{p}\cdot\mathbf{r})}$$

for a free fermion.

It will be convenient here to refer to the expression  $(\pm kE \pm i\mathbf{p} + ij m_0)$  as the ‘quaternion state vector’ (QSV) for the fermion, and we may consider it formally as a row or column vector, with components  $(kE + i\mathbf{p} + ij m_0)$ ;  $(kE - i\mathbf{p} + ij m_0)$ ;  $(-kE + i\mathbf{p} + ij m_0)$ ; and  $(-kE - i\mathbf{p} + ij m_0)$ , which may be considered as the four creation / annihilation operators for fermion / antifermion, spin up / spin down. An antifermion QSV with the same state of spin would reverse the signs of  $E$  in all the components, while a spin reversal for either fermion or antifermion would reverse the signs of  $\mathbf{p}$ . A spin 1 boson QSV then becomes the sum of the terms

$$\begin{aligned} & (kE + i\mathbf{p} + ij m_0) (-kE + i\mathbf{p} + ij m_0) \\ & (kE - i\mathbf{p} + ij m_0) (-kE - i\mathbf{p} + ij m_0) \\ & (-kE + i\mathbf{p} + ij m_0) (kE + i\mathbf{p} + ij m_0) \\ & (-kE - i\mathbf{p} + ij m_0) (kE - i\mathbf{p} + ij m_0) , \end{aligned}$$

with the components of the fermion state arranged in a row vector (represented, for convenience, as a column), and the components of the antifermion state in a column vector. A spin 0 boson state is obtained by reversing the signs of  $\mathbf{p}$  in the second column. From its original derivation, we can see that the fermion QSV is a nilpotent or square root of 0, while the boson wavefunction is a nonzero scalar, formed as a product of two nilpotents (each not nilpotent to the other). Significantly, the Dirac equation requires that a nilpotent QSV is matched exactly by the eigenvalue produced by the differential operator on the functional term ( $e^{-i(Et - \mathbf{p}\cdot\mathbf{r})}$  for a free particle).

While the Dirac and Schrödinger equations, which are ultimately concerned with kinetic energy states, produce fermions with half-integral spins, the Klein-Gordon equation, which applies to bosons, is a potential energy equation, based on  $E = mc^2$ , with  $m$  the ‘relativistic’, rather than rest mass  $m_0$ , and bosons derive their integral spin values from the fact that the energy term in this equation

incorporates unit values of the mass  $m$ . The Klein-Gordon equation is a direct quantization of the classical relativistic energy-momentum equation in the form:

$$\frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi = m_0^2 \psi,$$

and necessarily applies to fermions as well as bosons. This, we will see, is related to the fact that a fermion cannot be seen as isolated from its ‘environment’, and so effectively always acts as a member of a composite bosonic state.

Kinetic energy is always associated with rest mass, and cannot be defined without it; photons ‘slowing down’ in a gravitational field or condensed matter effectively acquire the equivalent of a rest mass. Potential energy, on the other hand, is associated with ‘relativistic’ mass because this term is actually *defined* through a potential energy-type expression ( $E = mc^2$ ). Light in free space provides the extreme case, with no kinetic energy or rest mass, and 100 % potential energy or relativistic mass. In both classical and quantum physics, we use the kinetic energy relation when we consider a particle as an object in itself, described by a rest mass  $m_0$ , undergoing a continuous change; and the potential energy relation when we consider a particle within its ‘environment’, with ‘relativistic mass’, in an equilibrium state requiring a discrete transition for any change.

The particle and its ‘environment’ can be considered as two ‘halves’ of a more complete whole. This is evident, in the case of a material particle, when we expand its relativistic mass-energy term ( $mc^2$ ) to find its kinetic energy ( $\frac{1}{2} m_0 v^2$ ). In effect, we either take the relativistic energy conservation equation

$$E - mc^2 = E^2 - p^2 c^2 - m_0^2 c^4 = 0 .$$

as a ‘relativistic’ mass or potential energy equation, incorporating the particle *and* its interaction with its environment, and then quantize to a Klein-Gordon equation, with integral spin; or, we separate out the kinetic energy term using the rest mass  $m_0$ , by taking the square root of

$$E^2 = m_0^2 c^4 \left( 1 - \frac{v^2}{c^2} \right)^{-1},$$

to obtain

$$E = m_0 c^2 + \frac{m_0 v^2}{2} + \dots$$

and, if we choose, quantize to the Schrödinger equation, and spin  $\frac{1}{2}$ . The  $\frac{1}{2}$  occurs in the act of square-rooting, or the splitting of 0 into two nilpotents in the Dirac equation; the  $\frac{1}{2}$  in the nonrelativistic Schrödinger approximation is a manifestation of this which we can trace through the  $\frac{1}{2}$  in the relativistic binomial approximation. If we go directly to the Dirac equation to obtain the spin  $\frac{1}{2}$  term, we see that the same result emerges from the behaviour of the



anticommuting terms, the anticommuting property is a direct result of taking the QSV as a nilpotent. So the anticommuting and binomial factors have precisely the same origin.

## 5 Radiation reaction

One way of looking at the factor 2 is that it links the continuous with the discontinuous. Expressions involving half units of  $\hbar$ , representing an average or integrated increase from 0 to  $\hbar$ , are characteristic of continuous aspects of physics, while those involving integral ones are characteristic of discontinuous aspects. The Schrödinger theory is an example of a continuous option, while the Heisenberg theory is discontinuous. Stochastic electrodynamics, which is based on the existence of zero-point energy of value  $\hbar\omega/2$ , is another completely continuous theory, which has developed as a rival to the purely discrete theory of the quantum with energy  $\hbar\omega$ .

It is important that we recognize that these alternative options do not represent different systems; they are different ways of interpreting the *same* system, and both are required for a complete explanation. Each has to incorporate the alternative option in some way. Thus, the Schrödinger approach is a continuous one, based on  $\frac{1}{2}\hbar$ , but incorporates discreteness (based on  $\hbar$ ) in the process of measurement – the so-called collapse of the wavefunction. The Heisenberg approach, by contrast, assumes a discrete system, based on  $\hbar$ , but incorporates continuity (and  $\frac{1}{2}\hbar$ ) in the process of measurement – via the uncertainty principle and zero-point energy. There seems always to be a route by which  $\frac{1}{2}\hbar\omega$  in one context can become  $\hbar\omega$  in another. A characteristic example is black-body radiation, where the spontaneous emission of energy of value  $\hbar\omega$  combines the effects of  $\frac{1}{2}\hbar\omega$  units of energy provided by both oscillators and zero-point field. In terms of fundamental particles, we see that a fermionic object on its own shows changing behaviour, requiring an integration which generates a factor  $\frac{1}{2}$  in the kinetic energy term, and a sign change when it rotates through  $2\pi$ , while a conservative ‘system’ of object plus environment shows unchanging behaviour, requiring a potential energy term, which is twice the kinetic energy.

The  $\frac{1}{2}h\nu$  or  $\frac{1}{2}\hbar\omega$  for black body radiation appears in both the theories of Planck, of 1911, and of Einstein and Stern, of 1913. In the Schrödinger version of quantum mechanics, as we have seen, the zero-point energy term is derived from the harmonic oscillator solution of the Schrödinger equation, showing the kinetic origins of the factor  $\frac{1}{2}$ , while, in the Heisenberg version, it comes from the  $\frac{1}{2}\hbar$  term involved in the uncertainty principle, suggesting an origin in continuum physics. The  $\frac{1}{2}\hbar\omega \rightarrow \hbar\omega$  transition for black body radiation can also be explained in terms of radiation reaction, which is connected again with the distinction between the relativistic and rest masses of an object. Rest mass effectively defines an isolated object, with *kinetic* energy. *Relativistic* mass, on the other hand, already incorporates the effects of the environment. For a

photon, which has no rest mass, and only a relativistic mass, the energy  $mc^2$  behaves exactly like a classical potential energy term, as when a photon gas produces the radiation pressure  $\rho c^2 / 3$ . We take into account both action and reaction because the doubling of the value of the energy term comes from doubling the momentum when the photons rebound from the walls of the container, or, alternatively, are absorbed and re-emitted. Exactly, the same thing happens with radiation reaction, thus explaining an otherwise ‘mysterious’ doubling of energy from  $\frac{1}{2} h\nu$  to  $h\nu$ . In a more classical context, Feynman and Wheeler require a doubling of the contribution of the retarded wave in electromagnetic theory, at the expense of the advanced wave, by assuming that the vacuum behaves as a perfect absorber and reradiator of radiation.

From our analysis, it would seem that by incorporating radiation reaction in the process we are also incorporating the effect of Newton’s third law. However, as in the parallel case of the anomalous magnetic moment of the electron, many of the same results are also explained by special relativity. Whitney<sup>3</sup> has argued that the correct magnetic moment for the electron is obtained, without relativity, by treating the transmission of light as a two-step process involving absorption and emission. In our terms, this is equivalent to incorporating action and reaction, and, as we have seen, the same result follows classically by taking the energy value at the moment when the field is switched on, which then becomes the new potential energy value when the system is in steady state. If, however, we use a one-step process, we also need relativity, because, once we introduce rest mass, we can no longer use classical equations. The two-step process is analogous to the use of radiation reaction, so it follows, in principle, that a radiation reaction is equivalent to adding a relativistic ‘correction’ (such as the Thomas precession).

Whitney also argues that the two-step process removes those special relativistic paradoxes which involve apparent reciprocity, and we could say that special relativity, by including only one side of the calculation, effectively removes reciprocity, and so leads to such things as asymmetric ageing in the twin paradox. Very similar arguments also apply to the idea that the problem lies in attempting to define a one-way speed of light that cannot be measured, because a two-way measurement of the speed of light also requires a two-step process. An argument by Morris<sup>4</sup> that the complete reciprocity involves a universal reference frame can be related to the notion here that reciprocity or reaction is the ‘environmental’ contribution as opposed to that of the particle.

## 6 Supersymmetry and the Berry phase

Taking ‘environment’ to apply to either a material or vacuum contribution, we can make sense, not only of the boson / fermion distinction and the spin 1 /  $\frac{1}{2}$  division in a fundamental way, but also such related concepts as supersymmetry, vacuum polarization, pair production, renormalization,

*zitterbewegung*, and so on, because the halving of energy in ‘isolating’ the fermion from its vacuum or material ‘environment’ is the same process as mathematically square-rooting the quantum operator via the Dirac equation. Taking this further, we can propose that energy principles determine that *all* fermions, in whatever circumstances, may be regarded either as isolated spin  $\frac{1}{2}$  objects or as spin 1 objects in conjunction with some particular material or vacuum environment, or, indeed, the ‘rest of the universe’. Characteristic examples of this occur when integral spins are produced automatically from half-integral spin electrons using the Berry phase, and, by generalizing this kind of result to all possible environments, we extend the principle in the direction of supersymmetry.

In the most general terms, we can consider that a relationship exists between any fermion and ‘the rest of the universe’, such that the *total* wavefunction representing fermion plus ‘rest of the universe’ is necessarily single-valued, automatically introducing the Berry phase. The Jahn-Teller effect and Aharonov-Bohm effect are examples of the action of this phase. Treated semi-classically, the Jahn-Teller effect couples the factors associated with the motions of the electronic and nuclear coordinates so that different parts of the total wave function change sign in a coordinated manner to preserve the single-valuedness of the total wave function. In the Aharonov-Bohm effect, electron interference fringes, produced by a Young’s slit arrangement, are shifted by half a wavelength in the presence of a solenoid whose magnetic field, being internal, does not interact with the electron but whose vector potential does. Effectively, the half-wavelength shift, or equivalent acquisition by the electron of a half-wavelength Berry phase, implies that an electron path between source and slit, round the solenoid, involves a *double-circuit* of the flux line (to achieve the same phase).

This duality between the fermion and its environment occurs with the actual creation of the fermion state. Splitting away a fermion from a ‘system’ (or ‘the universe’), we have to introduce a coupling as a mathematical description of the splitting. The converse effect must also exist, with bosons of spin 0 or 1 coupling to an ‘environment’ to produce fermion-like states. Both fermions and bosons, it would seem, always produce a ‘reaction’ within their environment, which couples them to the appropriate wavefunction-changing term, so that the potential / kinetic energy relation can be maintained at the same time as its opposite. It is possible to show that the whole process of renormalization depends on an infinite chain of such couplings through the vacuum. The coupling of the vacuum to fermions generates ‘boson-images’ and vice versa.

To understand this principle in more detail, we need to develop the nilpotent version of the Dirac wavefunction. In terms of the ‘environment’ principle, a fermion generates an infinite series of interacting terms of the form:

$$(kE + i\mathbf{p} + jm)$$

$$\begin{aligned}
& (\mathbf{k}E + i\mathbf{p} + ijm) (-\mathbf{k}E + i\mathbf{p} + ijm) \\
& (\mathbf{k}E + i\mathbf{p} + ijm) (-\mathbf{k}E + i\mathbf{p} + ijm) (\mathbf{k}E + i\mathbf{p} + ijm) \\
& (\mathbf{k}E + i\mathbf{p} + ijm) (-\mathbf{k}E + i\mathbf{p} + ijm) (\mathbf{k}E + i\mathbf{p} + ijm) (-\mathbf{k}E + i\mathbf{p} + ijm), \text{ etc.}
\end{aligned}$$

where  $(\mathbf{k}E + i\mathbf{p} + ijm)$  (abbreviated from the 4-component vector) represents a fermion state and  $(-\mathbf{k}E + i\mathbf{p} + ijm)$  an antifermion. The  $(\mathbf{k}E + i\mathbf{p} + ijm)$  and  $(-\mathbf{k}E + i\mathbf{p} + ijm)$  vectors are also an expression of the behaviour of the vacuum state, which acts like a ‘mirror image’ to the respective antifermion / fermion. An expression such as

$$(\mathbf{k}E + i\mathbf{p} + ijm) \mathbf{k} (\mathbf{k}E + i\mathbf{p} + ijm)$$

for a fermion creation operator is part of an infinite regression of images of the form

$$(\mathbf{k}E + i\mathbf{p} + ijm) \mathbf{k} (\mathbf{k}E + i\mathbf{p} + ijm) \mathbf{k} (\mathbf{k}E + i\mathbf{p} + ijm) \mathbf{k} (\mathbf{k}E + i\mathbf{p} + ijm) \dots$$

where the vacuum state depends on the operator that acts upon it, the vacuum state of  $(\mathbf{k}E + i\mathbf{p} + ijm)$ , for example, becoming  $\mathbf{k} (\mathbf{k}E + i\mathbf{p} + ijm)$ . In each case, the action simply reproduces the original state (after normalization). In addition,

$$(\mathbf{k}E + i\mathbf{p} + ijm) \mathbf{k} (\mathbf{k}E + i\mathbf{p} + ijm) \mathbf{k} (\mathbf{k}E + i\mathbf{p} + ijm) \mathbf{k} (\mathbf{k}E + i\mathbf{p} + ijm) \dots$$

is the same as

$$(\mathbf{k}E + i\mathbf{p} + ijm) (-\mathbf{k}E + i\mathbf{p} + ijm) (\mathbf{k}E + i\mathbf{p} + ijm) (-\mathbf{k}E + i\mathbf{p} + ijm) \dots$$

It thus appears that the infinite series of creation acts by a fermion / antifermion on vacuum is the mechanism for creating an infinite series of alternating boson and fermion / antifermion states as required for supersymmetry and renormalization. The nilpotent operators defined as QSVs for fermions and antifermions are also supersymmetry operators, which produce the supersymmetric partner in the particle itself. The  $Q$  generator for supersymmetry is simply the term  $(\mathbf{k}E + i\mathbf{p} + ijm)$ , and its Hermitian conjugate  $Q^\dagger$  is  $(-\mathbf{k}E + i\mathbf{p} + ijm)$ . Multiplying by  $(\mathbf{k}E + i\mathbf{p} + ijm)$  converts bosons to fermions, or antifermions to bosons. Multiplying by  $(-\mathbf{k}E + i\mathbf{p} + ijm)$  produces the reverse conversion of bosons to antifermions, or fermions to bosons. The supersymmetric partners, however, are not so much realisable particles, as the couplings of the fermions and bosons to vacuum states. The ‘mirror imaging’ process thus implies an infinite range of virtual  $E$  values in vacuum adding up to a single finite value, exactly as in renormalisation, with equal numbers of boson and fermion loops cancelling through their opposite signs. That is, if the supersymmetric virtual partners are merely vacuum images of the original particles, their mass values will be *identical*, so the infinite sum of boson masses

added to the vacuum will be identical to the infinite sum of fermion masses subtracted, so cancelling out exactly without requiring special assumptions about the nature of the vacuum or the masses of the supersymmetric states.

The existence of such ‘supersymmetric’ partner seemingly comes from the duality represented by the choice of fermion or fermion plus environment. The isolated fermion represents the action half of Newton’s third law, and characterized by kinetic energy, continuous variation, and spin in half-integral units, while in the case of the fermion interacting with its environment, it is the action and reaction pair, characterized by potential energy, a stable state, and spin in integral units. The combination then represents either a real boson with two nilpotents (which are not nilpotent to each other), or a bosonic-type state produced by a fermion interacting with its material environment or vacuum, and, as a consequence, manifesting Berry phase, the Aharonov-Bohm or Jahn-Teller effect, Thomas precession, relativistic correction, radiation reaction, the quantum Hall effect, Cooper pairing, *zitterbewegung*, or whatever else is needed to produce the ‘conjugate’ state in the fermion’s ‘environment’.

Physics and mathematics, however, are not isolated from each other at the fundamental level, and the process may also be seen in mathematical contexts. The half-wavelength shift in the Aharonov-Bohm effect, for example, is also well known to be a feature of the topology of the space surrounding the discrete flux-lines of the solenoid, which is not *simply-connected*, and cannot be deformed continuously down to a point. A path that goes round a circuit twice cannot be continuously deformed into a path which goes round once (as would be the case in a space without flux-lines). The presence of the flux line is equivalent, as in the quantum Hall effect and fractional quantum Hall effect (which also involve fermions and flux lines), to the extra fermionic  $\frac{1}{2}$ -spin which is provided by the electron acting in step with the nucleus in the Jahn-Teller effect and makes the potential function single-valued, and the circuit for the complete system a single loop.

It is of particular significant that the  $U(1)$  (electromagnetic) group responsible for the fact that the vacuum space is not simply connected is isomorphic to the integers under addition. The spin- $\frac{1}{2}$ ,  $\frac{1}{2}$ -wavelength-inducing nature of the fermionic state (in the case of either the electron or the flux line) is, in effect, a product of discreteness in both the fermion (and its charge) and the space in which it acts. In principle, the very act of creating a discrete particle requires a splitting of the continuum vacuum into *two* discrete halves, or (relating the concept of discreteness to that of dimensionality) two square roots of 0.

## 7 Physics and duality

From the foregoing discussion, it would appear that the factor 2 may be seen as a result of action and reaction (A); commutation relations (C);

absorption and emission (E); object and environment (O); relativity (R); the virial relation (V); or continuity and discontinuity (X). Many of these explanations, however, overlap in the case of individual phenomena, suggesting that they are really all part of some more general overall process:

Kinematics				V	X
Gases	A			V	
Orbits	A			V	X
Radiation pressure	A	E		V	
Gravitational light deflection				R	V
Fermion / boson spin		C	O	R	V
Zero-point energy	A	C		V	X
Radiation reaction	A	E		R	V
SR paradoxes	A	E			

In addition, the complete description of the system tends to lead to the overall elimination of the factor. The use of the factor 2 is a two-way process: halving in one direction and doubling in another, and the system can only be described in complete terms by taking both into account. Physical phenomena involving the factor tend to incorporate, in some form, the opposing sets of characteristics. Kinetic energy variation, for example, is continuous, but it starts from a discrete state; potential energy variation, on the other hand, is discrete, but starts from a continuous state. Neither kinetic nor potential exists independently: each creates the other, just like action and reaction.

In the most general terms, the factor 2 is an expression of fundamental dualities which result from the attempt at creating something from nothing. (In principle, the attempt is only possible if 0 is also the end result. *Nihil ex nihil fit!*) Essentially, they are the result of three distinct mathematical processes, which may be described as *conjugation*, *complexification* and *dimensionalization*. In an abbreviated form of the argument, we start, with the most fundamental dual group ( $C_2$ ), which we describe in mathematical terms, using the elements 1 and  $-1$ , but which, physically, is just anything and its conjugate, the totality being 0. (In the most general case, we begin without assuming units, or even numbers at all, by simply assuming an undefined category  $\mathcal{M}$ ; discreteness and numbering then arise as a *result* of anticommutativity or dimensionalization.<sup>4</sup>) Then we proceed by a further series of ‘duallings’ to create higher order groups. Assuming conjugation at all subsequent levels, we extend to four elements, to find an equivalent to  $C_2 \times C_2$ , by introducing complexification. The group of  $1, -1, i, -i$  is not, of course,  $C_2 \times C_2$ , but  $C_4$ , but it contains the same *information* as  $C_2 \times C_2$ , if we write this in the form of the complex ordered pairs:  $1, i; 1, -i; -1, i; -1, -i$ . Dualling further, we can complexify indefinitely, as in:

order 2	$(1, -1)$
order 4	$(1, -1) \times (1, i_1)$
order 8	$(1, -1) \times (1, i_1) \times (1, j_1)$
order 16	$(1, -1) \times (1, i_1) \times (1, j_1) \times (1, i_2)$
order 32	$(1, -1) \times (1, i_1) \times (1, j_1) \times (1, i_2) \times (1, j_2)$
order 64	$(1, -1) \times (1, i_1) \times (1, j_1) \times (1, i_2) \times (1, j_2) \times (1, i_3)$ , etc.

or, multiplying out:

order 2	$\pm 1$
order 4	$\pm 1, \pm i_1$
order 8	$\pm 1, \pm i_1, \pm j_1, \pm ij_1$
order 16	$\pm 1, \pm i_1, \pm j_1, \pm ij_1, \pm i_2, \pm i_2i_1, \pm i_2j_1, \pm i_2ij_1$
order 32	$\pm 1, \pm i_1, \pm j_1, \pm ij_1, \pm i_2, \pm i_2i_1, \pm i_2j_1, \pm i_2ij_1,$ $\pm j_2, \pm j_2i_1, \pm j_2j_1, \pm j_2ij_1, \pm j_2i_2, \pm j_2i_2i_1, \pm j_2i_2j_1, \pm j_2i_2ij_1$
order 64	$\pm 1, \pm i_1, \pm j_1, \pm ij_1, \pm i_2i_1, \pm i_2i_1, \pm i_2j_1, \pm i_2ij_1,$ $\pm j_2, \pm j_2i_1, \pm j_2j_1, \pm j_2ij_1, \pm j_2i_2, \pm j_2i_2i_1, \pm j_2i_2j_1, \pm j_2i_2ij_1$ $\pm i_3, \pm i_3i_1, \pm i_3j_1, \pm i_3ij_1, \pm i_3i_2, \pm i_3i_2i_1, \pm i_3i_2j_1, \pm i_3i_2ij_1,$ $\pm i_3j_2, \pm i_3j_2i_1, \pm i_3j_2j_1, \pm i_3j_2ij_1, \pm i_3j_2i_2, \pm i_3j_2i_2i_1, \pm i_3j_2i_2j_1,$ $\pm i_3j_2i_2ij_1$

Here we have an option. We can choose either  $(ij_1)^2 = 1$  or  $(ij_1)^2 = -1$ , implying respective commutation or anticommutation. The first allows an infinite number of possibilities; the second produces a closed system, with no further options available (quaternions). This we may describe as dimensionalization, and the infinite series of groups may be considered as an infinite number of independent quaternion systems. We then obtain the sequence:

order 2	real scalar
order 4	complex scalar (pseudoscalar)
order 8	quaternions
order 16	complex quaternions or multivariate vectors
order 32	double quaternions
order 64	complex double quaternions or multivariate vector quaternions

with the processes defined by:

$C_2$	$C_2$	$\pm 1$	conjugate
$C_4$	$C_2 \times C_2$	$\pm 1, \pm i$	complexify

$Q_8$	$C_2 \times C_2 \times C_2$	$\pm 1, \pm i, \pm j, \pm k$	
	dimensionalize		
$V_{16}$	$C_2 \times C_2 \times C_2 \times C_2$	$\pm 1, \pm i, \pm j, \pm k$	complexify
$QQ_{32}$	$C_2 \times C_2 \times C_2 \times C_2 \times C_2$	$\pm 1, \pm I, \pm J, \pm K, \pm i, \pm j, \pm k$	
	dimensionalize		
$VQ_{64}$	$C_2 \times C_2 \times C_2 \times C_2 \times C_2 \times C_2$	$\pm 1, \pm i, \pm I, \pm J, \pm K, \pm i, \pm j, \pm k$	complexify

Products of all terms are here assumed, and the complex quaternion terms ( $\pm ii, \pm ij, \pm ik$ ) implied by  $V_{16}$  can be more conveniently written as multivariate vectors ( $\mathbf{i}, \mathbf{j}, \mathbf{k}$ ), like the ( $\pm iI, \pm iJ, \pm iK$ ) terms of  $VQ_{64}$ . The structure repeats at order 16 but to incorporate all those of orders 2, 4, 8 and 16 as independent units, we need order 64. We can identify the four new algebraic units introduced as being those of mass (-energy) (real scalar, 1), time (pseudoscalar  $i$ ), charge (quaternions  $i, j, k$ ) and space (multivariate vectors,  $\mathbf{i}, \mathbf{j}, \mathbf{k}$ ); and the combined algebra of order 64 as the Dirac state vector, which is, significantly, a nilpotent (which takes us immediately back to the required 0), while the Hilbert space incorporating all the Dirac state vectors then takes the order of the structure to infinity. We can also express the properties of the fundamental parameters in terms of the three processes which create the entire system in the table:

<b>space</b>	nonconjugated	real	dimensional
<b>time</b>	nonconjugated	complex	nondimensional
<b>mass</b>	conjugated	real	nondimensional
<b>charge</b>	conjugated	complex	dimensional

Conjugated here is equivalent to conserved, so a positive charge (or positive source of mass-energy) cannot be created without also creating a negative one. It is also equivalent to the unchanging state (as implied in potential energy equations) as opposed to the changing or unconjugated (kinetic) state. As recognized in earlier work, only the (3-)dimensional quantities, space and charge, are countable. As in conventional mathematics, two versions of the 'real' numbers emerge from this structure: the uncountable ones of the Cantor continuum and standard analysis (which apply to mass), and the countable ones of the Löwenheim-Skolem arithmetic and Robinson's non-standard analysis (which apply to space).

In terms of the mathematical structure here proposed, it would be possible to classify the physical processes involving the factor 2 as resulting from the following processes:



action and reaction (A)	conjugation	
commutation relations (C)	dimensionalization	
absorption and emission (E)	conjugation	
object and environment (O)	conjugation	
relativity (R)	complexification	
the virial relation (V)	conjugation	
continuity and discontinuity (X)	conjugation	/
dimensionalization		

However, overlap is possible in most, or even all, of these cases, and physical systems which apply a doubling through one route will involve a halving through another. For example, gravitational light deflection, if treated as relativistic, doubles its value because of complexification, both space and time being considered. However, the double deflection can also be derived from the use of a kinetic energy term being half the total (potential) energy, because it represents the unconjugated rather than the conjugated case. This is why there are so many physical phenomena involving the factor 2 with alternative explanations. The factor appears when we look at a process from a one-sided point of view. Though a single duality separates alternative theories, such as Heisenberg and Schrödinger, or quantum mechanics and stochastic electrodynamics, it is invariably open to more than one interpretation because each pair of parameters is always separated by two distinct dualities, and the separate interpretations ultimately act together when we consider a phenomenon in relation to its place in the overall ‘environment’ of the physical universe.

## REFERENCES

1. W. Gough, *Eur. J. Phys.*, **11**, 326-33 (1990).
2. P. Rowlands, ‘A New Algebra for Relativistic Quantum Mechanics’, *Proceedings of Conference on Physical Interpretations of Relativity Theory V*, British Society for Philosophy of Science, London, September 1996, 381-7. See also *ibid.*, 1998, 308-15; *Speculat. Sci. Tech.*, 17, 279-282, 1994; 243-51, 1996; *Nuclear Physics A* **684**, 713-5, 2001; arXiv:quant-ph/0103036; 0109069; 0106111.
2. T. Morris, ‘The unique reference frame – is it superfluous or essential in special relativity?’, *Proceedings of Conference on Physical Interpretations of Relativity Theory VII*, British Society for Philosophy of Science, London, September 2000, 267 (abstract, plus oral presentation).

3. C. K. Whitney, 'How can paradox happen?', *Proceedings of Conference on Physical Interpretations of Relativity Theory VII*, British Society for Philosophy of Science, London, September 2000, 338-51.

4. P. Rowlands. and B. Diaz, arXiv:cs.OH/0209026.