Why does physics work?

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 prowl@hep.ph.liv.ac.uk and prowl@csc.liv.uk

Abstract. Physics works because it avoids directly characterizing nature. At the most fundamental level, a group symmetry between space, time, mass and charge ensures that every property assumed by a physical system is countered by its symmetrical opposite. In this way, we can apply the idea of measurement to nature without requiring that nature is characterized by measurability.

1 Introduction

The question I want to attempt to answer here is: why does physics work? Can a philosophical approach help us to find the answer, and can we use this to knowledge to tackle important physical questions? Can a theory of knowledge be used both to explain physics and to suggest possibilities for its future development? Will the much sought-after unified theory only be possible when we have located the structure which makes physics such a uniquely successful means of describing the processes of nature.

I believe that, to answer these questions, we must look at areas in which physics appears to be counter-intuitive. Physics could be described as the science of measurement, but its structure does not suggest that it is based on measurement convenience. There is, in fact, only one process of measurement known, the counting of spatial intervals, using a fixed scale of units defined arbitrarily. All measurements, even those supposedly representing other quantities, such as time, are really measurements of space. However, no attempt to structure the whole of physics on the basis of space alone has been so far successful. Physics seems to demand the existence of other fundamental quantities, which cannot be directly measured. The measurement process required by physics does not define the whole of nature.

The fundamental principles of physics have been developed on the basis of a long step-by-step process of conceptualizing and testing by experiment. There are theories such as classical mechanics, electromagnetic theory and quantum mechanics which have a vast range of applicability in describing the processes of nature. It is the requirement that the experimental justification must always accompany the theoretical developments that has led to their success, but this does not tell us why these particular theoretical structures are required by nature. Also, many of the principles used, such as the conservation laws and the irreversibility of time, appear to be intrinsically simple, and suggest that simpler theories are more likely to be inherently true than complex ones. Is there some more fundamental principle of knowledge which determines these structures and which privileges simplicity over complexity?

2 Historical background

It is interesting to look at the historical background of the development of physics as we know it today. This began in the late Middle Ages, with the attempt to explain processes, such as free fall and uniformly accelerated motion, through defining quantities based on the variations of space and time, the most abstract ideas then known, and still the only means of observing variation in nature. The theories then developed encompassed various states of uniform and nonuniform velocity, and uniform and nonuniform acceleration, which were expressed in both mathematical and graphical forms. The authors of these theories were not scientists in the modern sense, but theologians, more concerned with comprehending the mind of an unknowable God than with explaining physical observations. Hence, their explanations were pitched to the highest point of abstraction and to the greatest level of simplicity then attainable.

The success of these mathematical theologians is attributable to the extremism of their ideas, their relentless application of Ockham's Razor resulting in the creation of abstract concepts of a universal generality, rather than specific explanations of particular physical phenomena, as in the previous Aristotelian model. The method was adopted and extended by both Galileo and Newton, who were educated at universities which were still mediaeval in outlook, but Newton found it necessary to incorporate a third concept, mass, on the same level as space and time, to develop a more general system of dynamics. The significance of mass was that it established the principle that fundamental physical laws could be built around the fact that some concepts were conserved quantities while others were not. The Newtonian procedure also established the fact that the mathematical system used to describe fundamental physical laws, based as it was on differential equations and an infinite number of interacting particles, was not a direct description of nature, but an idealised abstraction which could never be observed in terms of direct physical measurements.

In principle, this separation between observables and the mathematical system carries over into quantum mechanics, and is extremely counter-intuitive. Clearly, we have not developed the present mathematical structure of physics for *convenience*, as is often maintained; it has rather evolved by natural selection because no other structure will work. None of the great abstract theories, such as classical mechanics, electromagnetic theory, or quantum mechanics, has been accepted wholeheartedly, even by physicists, as a self-evident fundamental truth, precisely because such theories run counter to our starting point of structuring the world of nature on direct observation and measurement. Yet no theory based on alternative principles has had the slightest success. Also, successful theories of this kind have led to the belief that theories explaining new phenomena must be similar in kind, and can be derived by analogy and symmetry.

This happened after the development of Newtonian mechanics and gravitational theory, when a two-hundred-year search for the laws of electricity, magnetism and optics was based on the starting assumption that analogies must exist with the Newtonian system. Through a long series of experimental and theoretical developments, the analogy was eventually successful through the creation of the mass-like parameter charge, and the additional component of the relativistic connection of space and time, which was then applied, by analogy, to the original Newtonian theory. In more recent times, it has been widely assumed that the two new physical forces discovered through radiaoctivity and particle physics, the strong and weak interactions, must be in some way analogous to the electric force, under ideal conditions; and that some process must exist which breaks an otherwise perfect symmetry between the three forces.

3 Symmetry

Symmetry (or analogy) has been the driving force of much of theoretical particle physics, as it was, previously, of classical physics, and physicists seem to *expect* to find symmetries in nature. There are many classic theorems which invoke symmetry, for example, Noether's theorem, which allows us to relate the conservation laws of energy, momentum and angular momentum to the translation and translation-rotation symmetries of time and space. It is this fact which gives us an important clue as to what really makes physics work. The other important clue is provided by the fact already stated that the work of mediaeval theologians aimed at expressing the unknowability of God has been successfully transported into physics and used with the totally different objective of describing the natural world in terms of abstract mathematics.

These clues suggest that the important philosophical principle determining the structure of physics is that nature cannot be characterized. It is neither measurable nor unmeasurable; and in fact has no single defining characteristic. Any system which assumes a special characteristic of any kind limits us to the asymptotic discovery of our initial assumption. To create a universally applicable system, we have to avoid any such limiting assumption; and to make physics work, we have had to incorporate concepts which we would not have chosen on first principles, but which introduce a systematic contradiction of our starting assumptions. For example, though we started with space and time as variable quantities, we found that nature threw up quantities which we could not exclude which turned out to be *in*variable, such as mass and charge; and, though we started by assuming that we could measure everything directly (through space), we quickly found that we had to structure physics to incorporate quantities which could not be directly measured (time, mass and charge).

The in-built procedure which allows us to have these opposite characteristics within the same system is symmetry. If we apply a perfectly symmetrical structure to the explanation of nature, we ensure that, whatever characteristic we choose in one part of the system will be countered by its exactly symmetrical opposite in another. Symmetry thus allows us to do the apparently impossible, that it is to characterize nature using a measurement process (because the system as a whole is not measurable), while, at the same time allowing us to reduce our number of fundamental assumptions. By a process of choosing the only methodology which produces a match with experimental results, we have adopted a structure for physics which is perfectly symmetrical in its foundations, though we have not yet recognized this fact. However, if we do begin to recognize it, we will be able to design new physics according to the specifications which have, in this way, been forced upon us.

But what symmetry? There are many in physics, and we must decide which are the most important and fundamental. I will argue that there are just four basic parameters and that we can use such techniques as dimensional analysis to show that all other physical parameters arise from compounded versions of these elementary ones. We can also show that it is precisely these parameters which are assumed to be the elementary ones in the statement of the CPT theorem. It is perhaps slightly surprising that physicists have been so often prepared to tackle fundamental questions without taking proper account of these basic ideas. Although we can’t hope to *analyse* really basic ideas, we can learn a great deal by setting one off against the other. It would surely be profitable to examine the properties of these parameters as closely as possible and look for patterns, or symmetries of one sort or another, that would help to clarify their meaning and uses. Symmetry has been such a powerful tool in understanding particle physics and the fundamental interactions that we have every reason to expect to find it here.

The symmetry which I believe to underlie all the other symmetries in physics is that between the fundamental parameters space, time, mass and charge (where charge is a general term representing the sources of the three nongravitational interactions). Essentially, space is characterized by the properties required of a parameter of measurement: it is real, nonconserved, and countable. Nature, however, forces us to include, at the same level within our system, parameters which have directly opposing properties. These turn out to be those which we describe as time, mass and charge; and the symmetry between the four parameters appears to be absolute and representable in terms of a group structure. Classical mechanics, electrodynamics, relativity, quantum mechanics, and particle physics, can be accommodated within this structure. Many theorems can be established as consequences of the emerging symmetrical pattern; and new physical facts can be predicted, even in quantitative terms. The structure is at once more simple and powerful than any of the principles derived from it.

4 The meaning of the conservation laws

One of the most important aspects of Newtonian physics is its introduction of the parameter mass on the same fundamental level as space and time. Introducing mass also introduces the idea of conservation, which can be described in absolute terms, and leads to a whole series of conservation principles, which are invariably the bases of dynamical systems. Physical equations then become statements that one quantity is conserved while another is not. So, the conserved property of mass also shows up the contrasting, nonconserved natures of space and time, which are incorporated into physics in the differential forms with which these quantities are associated in fundamental equations.

In Newton's original formulation, the quantity force was defined as a product of the conserved mass and the differential forms of the nonconserved space and time, and the conservation property of mass was then established through the third law of motion, which effectively stated that force was zero within a conservative system. The method was powerful because it was extreme. It had to be true in all cases. Later versions of dynamics, associated with Euler, Maupertuis, Lagrange, and Hamilton, were structured on exactly the same pattern: a quantity was defined which incorporated mass and the differentials of space and time, and was then shown to be subject to extreme behaviour. The quantity was set to zero, or a constant value, or a maximum or minimum. And the extremum principles so derived, in particular the conservation laws of linear momentum, angular momentum and energy, were, in principle, nothing other than the fact that mass was conserved against all possible variations in space and time.

Subsequent developments in electrodynamics suggested that electric charge had precisely the same property. The conservation laws of mass and electric charge are among the most fundamental in physics (although we now, of course, interpret the meaning of 'mass' in terms of energy), and it is almost inconceivable to imagine any circumstance in which they would be violated. Again, it is almost certain that some kind of conservation law applies also to the sources of the weak and strong nuclear interactions, which have in the last two decades increasingly been referred to under the same generic label of 'charge'. Lepton and baryon conservation are obvious consequences. Of particular importance is the fact that the conservation laws of mass and charge are not merely global, applying to the total amount of each quantity in the universe, but also *local*, applying to the amount of each quantity at a given place in a given time. Each element of mass and charge has, it would seem, an *identity* which it retains throughout all interactions, subject only, in the case of charge, to its annihilation by an element with the opposite sign.

Strikingly, when we look at *nonconservation*, as manifested by space and time, we immediately see that it is the *exact opposite* of conservation and it is just as definite a property, though it manifests itself in many different ways. For example, if the elements of mass and charge have individual, specific and permanent identities, those of space and time have no identity whatsoever. One manifestation of this is the property of *translation symmetry* which applies to both space and time. This implies that every element of space and time is exactly like every other, and is not only indistinguishable in practice, but *must be stated to be indistinguishable* when we write down physical equations. The translation property has profound physical consequences, as Noether's theorem shows us that the translation symmetry of time is precisely identical to the conservation of energy, and that the translation symmetry of space is precisely identical to the conservation of linear momentum. The third conservation principle, that of angular momentum, is explained by Noether's theorem as a result of space's extra three-dimensional property of rotation symmetry, meaning that there is no more identity for spatial *directions* than there is for spatial locations. In addition to having no unique elements, space also lacks a unique set of dimensions. One direction in space is identical to any other.

The meaning of the conservation laws now becomes more apparent. They are not only the absolute defining principles of all aspects of fundamental physics, but must also be accompanied by absolute *nonconservation* principles. We could, in fact, illustrate the exactly opposite nature of conservation and nonconservation by expressing the identity or uniqueness properties of mass and charge in terms of 'translation' *a*symmetries. Translation asymmetry then means that one element of mass or charge cannot be 'translated to' (or exchanged for) any other within a system, however similar. A similar concept of rotation asymmetry might be expected to apply to charge, but we will return to this later.

But there are also other manifestations of nonconservation in space and time. The absoluteness of the nonconservation properties is also apparent in the *gauge invariance* used in both classical and quantum physics. In classical or quantum electrodynamics, electric and magnetic fields terms remain invariant under arbitrary changes in the vector and scalar potentials, or phase changes in the quantum mechanical wavefunction, brought about, essentially, by translations (or rotations) in the space and time coordinates. Gauge invariance is simply a way of expressing the fact that a system will remain conservative under arbitrary changes in the coordinates which do not produce changes in the values of conserved quantities such as charge, energy, momentum and angular momentum. In other words, we cannot know the absolute phase or value of potential because we cannot choose to fix values of coordinates which are subject to absolute and arbitrary change. Even more significantly, in the Yang-Mills principle used in particle physics, the arbitrary phase changes are specifically *local*, rather than global. Nonconservation, therefore, must be local in exactly the same way as conservation.

5 The mathematical structure of physical quantities

Ancient knowledge tells us that space is three-dimensional. Three independent spatial axes can be drawn at right angles. In other words, the dimensions of space may be combined 'vectorially', or by the Pythagorean addition of their squared values. Special relativity, which emerged from electrodynamics in the early twentieth century, however, indicated that time, in many respects, could be treated as a fourth dimension of space, at the price of making the time component an imaginary number, with negative squared values, the whole combination being described as the Minkowski space-time 4-vector ($\mathbf{i}x$; $\mathbf{j}y$; $\mathbf{k}z$; it). Many people, of course, describe this as a mere mathematical convenience, or 'trick', but we still have to explain why such a convenient 'trick' actually works. It is instructive, therefore, to look at a representation of mass and charge.

Here, we may imagine, as is generally believed, that the electric, strong and weak source terms (say e, s, w) would, in some idealised regime, be of a similar nature, suggesting that we could describe them, by analogy with space, as 'dimensions' of a single quantity, generically known as 'charge'. This could be justified in terms of the fact the Newton and Coulomb force laws effectively square mass and charge terms, in the same way as space and time terms are squared in Pythagorean addition. Now, we have the intriguing fact, long known but never explained, that forces between like masses are attractive, whereas forces between like charges (of all kinds) are repulsive; that is, the forces between like masses and like charges have opposite signs. However, if we *choose* to represent charges by imaginary numbers and masses by real ones, we then have a symmetrical representation for the Newton and Coulomb force laws:

$$F = -\frac{Gm_1m_2}{r^2}$$

$$F = -\frac{iq_1iq_2}{4\pi\varepsilon_0 r^2}$$

The three types of source would, of course, have to be distinguished from each other in some way, and so would each require a different imaginary number, or square root of –1. However, the mathematics required for such a situation is already available and has been well-known for a hundred and fifty years. This is the *quaternion* system, discovered in 1843, in which $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$, the three square roots of –1, are related by the formulae:

$$\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = \boldsymbol{ijk} = -1 \ .$$

Effectively the quaternion components are the reverse of the 4-vectors used in Minkowski space-time: three imaginary parts and one real (ordinary real numbers), as opposed to three real parts and one imaginary. Their special significance is that they are unique; no other *associative* extension of ordinary complex algebra involving imaginary dimensions is possible. A system with two imaginary parts is impossible, and there are no systems with four, five or six imaginary parts. A system with seven imaginary parts is possible (octonions), but this requires breaking the rule of associativity; and there are no systems with more than seven.

It seems that, for dimensions determined by Pythagorean addition, the number three has a special significance. The three imaginary parts, in this representation, would be associated with the three components of charge (say, $\boldsymbol{i}e$, $\boldsymbol{j}s$, $\boldsymbol{k}w$), leaving the real fourth part to represent mass. Space and time would then become a three real- and one imaginary-part system by *symmetry*, and the necessary mathematical connection between space and time would be explained as a consequence of the necessary mathematical connection between charge and mass:

space-time	$\mathbf{i}x$	$\mathbf{j}y$	$\mathbf{k}z$	it
mass-charge	$\boldsymbol{i}e$	$\boldsymbol{j}s$	$\boldsymbol{k}w$	m

Quaternions, however, have different rules of multiplication, there being no such thing as a ‘full product’ (**ab**) between two vectors, **a** and **b**, such as exists between two quaternions. However, if we were to postulate that the symmetry between space-time and mass-charge should be an exact one, we could extend the vector property of space to incorporate a quaternionic-like ‘full’ product between two vectors, combining the scalar product with *i* times the vector product. This procedure has been fully justified by several decades of mathematical development, and it turns out that the extra vector terms in the full product are just those required to explain the otherwise ‘mysterious’ spin property in quantum mechanics.

It is here that, now we understand the meaning of ‘dimensionality’ in the case of charge, we can return to the subject of its rotation asymmetry. If charge is absolutely conserved, then we can expect conservation in dimension as well as in quantity, which is what we mean by rotation *a*symmetry. That is, the sources of the electromagnetic, weak and strong interactions should be separately conserved, and incapable of interconversion. Immediately, this tells us that the proton, which has a strong charge measured by its baryon number, cannot decay to products like the positron and neutral pion, which have none. In fact, separate conservation laws should immediately lead to baryon and lepton conservation, baryons being the only particles with strong, as well as weak, components, and leptons being the only particles with weak, but no strong, components. We will see later how this produces a new manifestation of Noether’s theorem.

Dimensionality, however, is not the only advantage of an imaginary representation for charge, for imaginary numbers have yet another important property. This is the fact that equal representation must be given to positive and negative values of imaginary quantities. Neither positive nor negative imaginary values may be privileged in algebraic equations; every equation which has a positive solution also has an algebraically indistinguishable negative solution (the complex conjugate). Consequently, all our charges (but not necessarily real masses) must exist in both positive and negative states. This is exactly what we require to explain the existence of antiparticles, for even those particles, such as the neutron and neutrino, which have no electric charge still have antiparticles because they have strong and/or weak charges whose signs may be changed under the process of charge conjugation.

6 Where does dimensionality come from?

Dimensionality has two aspects – the fact of multidimensionality, and the Pythagorean addition of the multiple dimensions by squares. Assuming these two aspects, we can solve the problem of *three*-dimensionality, but dimensionality itself is too complex to be a basic property. It must be explicable on some more fundamental grounds. It is necessary, here, to compare the dimensional (or multidimensional) parameters, space and charge, with the nondimensional (or one-dimensional) parameters, time and mass.

Though space and time are associated mathematically in the Minkowski formalism, but there is good evidence that they are fundamentally different. Space, for example, is *always* used in direct measurement; it is, in fact, impossible to measure directly any other quantity. So-called 'time'-measuring devices, such as pendulums, mechanical clocks, and crystal and atomic oscillators, all use some concept of repetition of a spatial interval. Space also is reversible – and it is this reversibility which is used in the measurement of time – but time is not. There also seems to be a deep philosophical problem, as Zeno's paradoxes suggest, with the infinite divisibility of time. Whitrow, for example, writes: 'One can, therefore, conclude that the idea of the infinite divisibility of time must be rejected, or ... one must recognize that it is ... a logical fiction.'[1] And Coveney and Highfield conclude that: 'Either one can seek to deny the notion of 'becoming', in which case time assumes essentially space-like properties; or one must reject the assumption that time, like space, is infinitely divisible into ever smaller portions.'[2] The paradoxes seem to show, according to Whitrow, that motion is 'impossible if time (and, correlatively, space) is divisible ad infinitum'[1]

Whitehead thought that the pardoxes showed an 'instant of time' to be 'nonsense',[1] while Bergson, according to Whitrow, 'enthusiastically adopted the view' that time 'is wholly indivisible', 'as a means of escaping the difficulties raised by Zeno, concerning both temporal continuity and atomicity, without abandoning belief in the reality of time. ... Unfortunately, in attacking the geometrization (or spatialization) of time he went too far and argued that, because time is essentially different from space, therefore it is fundamentally irreducible to mathematical terms.'[1] According to our analysis, there is good evidence that one cannot simply assume that time can be indefinitely subdivided like space. There is every reason to believe, in fact, that time, unlike space, is an absolute continuum. There is no infinite succession of measurable instants in time, as supposed in the paradoxes, because there are no instants. Time cannot actually be divided. In more contemporary jargon, space is digital, time is analogue – and we have both concepts in nature because we have both parameters.

The space-time distinction has profound consequences for both mathematics and physics, and, if we believe (as I do) that physics is the inherent creator of mathematics, and not merely the employer of its techniques, then we will say that the mathematics which is possible is created because it is possible physically. We can, for example, say that time is the set of reals with the standard topology superimposed (and is nonalgorithmic); space is the set of reals without the topology (and is algorithmic). Abraham Robinson, in his *Non-Standard Analysis*,[3] has successfully treated infinitesimals as though they had the properties of real numbers, and has shown that proofs of many theorems become much simpler by this method, although all non-standard proofs may be duplicated by standard ones (and vice versa). Non-standard analysis is also closely related to Skolem's non-standard arithmetic of 1934, with its denumerable model of the reals, and the so-called non-Archimedean geometry, which relates this to space. These versions of non-standard mathematics are a reflection of the discrete nature of space while 'standard' results (based on limits) rely on the continuity of time.

Though we often describe real numbers in terms of a continuous line in space, the 'continuity' which we attributed to space because of its indefinite divisibility is not what is meant by the absolute continuity of time. Absolute continuity cannot be visualised and any process used to describe it would deny continuity. The property which space has that is often referred to as 'continuity' is indefinite elasticity, its 'continual' recountability or its unending divisibility. But it is this very divisibility of space which denies it *absolute* continuity; and the elastic nature of the divisibility comes from the entirely different property of nonconservation. A nonconserved quantity necessarily has nonfixed units, but they are units nonetheless. The whole process of measurement depends crucially on the divisibility of space, or creation of discontinuities within it. Thus the entire problem of Zeno's paradoxes disappears as soon as we accept that we can have discontinuities or divisibility in space, but not in time.

The discontinuities in space are in both quantity and direction; it can be reversed and changed in orientation; and, without both of these properties, measurement would be impossible. Time, however, cannot be reversed, precisely because it is absolutely continuous. Any reversal of time would require discontinuity. For the same reason, time cannot be multidimensional, or, in our terminology, 'dimensional'. It is also equally impossible for a discrete quantity, like space, to be nondimensional, for one cannot demonstrate discreteness in a one-dimensional system. Though we think of a line as one-dimensional, it is really a one-dimensional construction within a two-dimensional one. If our space was truly one-dimensional we would only have a point with no extension. We couldn't demonstrate discreteness, and certainly not discreteness with variability, as we demand of space.

The distinction in status between space and time is even responsible for the fundamental fact that time, in the definition of velocity and acceleration, the basic quantities used in dynamics, is the independent variable, whereas space is the dependent variable. This situation arises because time, as a continuous quantity, unlike space, is not susceptible to measurement. We have no control over the variation of time, and so its variation is necessarily independent.

The existence of both 'standard' and 'nonstandard' versions of analysis and arithmetic are consequences of the prior existence of space and time. The mathematical options that are available, here and elsewhere, are almost certainly a reflection of the availability of physical options. Continuity and discontinuity, finiteness and infinity, and so on, probably exist as mathematical categories because they are also physical categories. For example, the discrete process of differentiation (using infinitesimals) is essentially modelled on variation in space; while the continuous process (using limits) is modelled on variation in time. Each is a valid option, as differentiation is a property linked to nonconservation, and not concerned, in principle, with the difference between absolute continuity and indefinite divisibility. (It is significant that the solutions of Zeno's paradoxes which invoke the concept of limit tacitly assume the time-based definition of differentiation.) In arithmetical terms, the Cantorian definition of an absolutely continuous set of real numbers has equal validity with the idea of an infinitely constructible, though not absolutely continuous, set of real numbers based on algorithmic processes. Of significance here is the Löwenheim-Skolem theorem, that any consistent finite, formal theory has a denumerable model, with the elements of its domain in a one-to-one correspondence with the positive integers.

The continuous-discrete distinction also occurs, as we would expect, between mass and charge. Mass (in the sense that it incorporates fields and energy) is an absolute continuum present in all systems and at every point in space; this is why there it is unipolar, unlike charge, for negative mass would necessarily require a break in the continuum, as would any multidimensionality. Charge, on the other hand, is divisible and observed in units. Naturally because charge is also a conserved quantity, unlike space, these units must be fixed, unlike those of space. Again, charge as a noncontinuous quantity is also dimensional.

Our analysis will also allow us to deal with two well-known physical paradoxes. One is the 'reversibility paradox', where time, according to the laws of physics is reversible in mathematical sign, when it is clearly not reversible in physical consequences. Time, however, is characterised by imaginary numbers, and imaginary numbers are not privileged according to sign. Thus, it is quite possible to have a time which has equal positive and negative mathematical solutions because it is imaginary, but which has only one physical direction because it is continuous. The corresponding unipolarity, or single sign, of mass is the reason why we have a CPT, rather than an MCPT, theorem, C standing for charge conjugation, P for space reflection and T for time reversal, all of which have two mathematical sign options.

The other apparent paradox is wave-particle duality. This arises from the fact that, when we mathematically combine space and time in Minkowski's 4-vector formalism, as symmetry apparently requires us to do, we have two options: we can either make time space-like (or discrete) or space time-like (or continuous). Using the discrete options, we obtain particles, special relativity and Heisenberg's quantum mechanics.[4,5] Using the continuous options, we obtain waves, Lorentzian relativity and Schrödinger's wave mechanics. Heisenberg makes everything discrete, so mass becomes charge-like quanta in quantum mechanics; Schrödinger, on the other hand, makes everything continuous, so charge becomes mass-like wavefunctions in wave mechanics. In measurement, the true situations are restored, for Heisenberg reintroduces continuous mass via the uncertainty principle and the virtual vacuum, while Schrödinger reintroduces discrete charge via the collapse of the wavefunction.

7 A group of order 4

We have shown that the four basic parameters may be distributed between three sets of opposing paired categories: real / imaginary, conserved / nonconserved, divisible / indivisible, with each parameter paired off with a different partner in each of the categories, according to the following scheme:

space	real	nonconserved	divisible
time	imaginary	nonconserved	indivisible
mass	real	conserved	indivisible
charge	imaginary	conserved	divisible

The properties where they match, seem to be exactly identical, and where they oppose, to be in exact opposition. Mathematically, the scheme incorporates a group of order 4, in which any parameter can be the identity element and each is its own inverse. We can easily generate an algebraic representation by representing the properties of space (real, nonconserved, divisible) by, say, a, b, c, with the opposing properties (imaginary, conserved, indivisible) represented by $-a$, $-b$, $-c$ (though this representation is not, of course, unique). The group now becomes:

space	a	b	c
time	$-a$	b	$-c$
mass	a	$-b$	$-c$
charge	$-a$	$-b$	c

With group multiplication rules of the form:

$$a * a = -a * -a = a$$
$$a * -a = -a * a = -a$$
$$a * b = a * -b = 0$$

and similarly for b and c, we can establish a group multiplication table of the form:

*	space	time	mass	charge
space	space	time	mass	charge
time	time	space	charge	mass
mass	mass	charge	space	time
charge	charge	mass	time	space

This is the characteristic multiplication table of the Klein-4 group, with space as the identity element and each element its own inverse. However, there is no reason to privilege space with respect to the other parameters, since the symbols a and $-a$, b and $-b$, c and $-c$ are arbitrarily selected, and any of the other three parameters may be made the identity by defining its properties as a, b, c.

8 The Dirac state

We can proceed, from this mathematical structure, to derive relations between the parameters which represent the binary operations between the group members.[6-8] In particular, using the numerical relations between the units of space and time and mass and charge necessitated by the 4-vector and quaternion structures, we can generalise to group relations between each parameter and every other, and to the equivalent inverse relations, suggesting the existence of fundamental constants with the dimensions of G, h and c. Here, for the first time, we see the significance of the squaring operation involved in dimensionality, or the multiplication of a unit of any parameter by an identically-valued unit of the same parameter.

It is significant here that units of mass and charge have individual identities, unlike those of space and time, and so the 'squaring' of their units becomes the multiplication of individual units, such as m_1m_2 and q_1q_2, and such 'squaring' must be a universal operation between any units of mass and charge, no individual unit being privileged. It will be convenient to give this process the name of 'interaction'. (It will be recognised that 'interaction' in this sense is nonlocal.) The numerical relations established between the parameters through the group, can be combined with conservation and nonconservation conditions to provide mathematical derivations of the laws of classical mechanics, electromagnetic theory and quantum mechanics, in terms of these 'interactions'. Another development suggests that the relationship between the quaternion representation and the requirement of separate conservation for charges might affect the fundamental particle (or 'charge') structures that are possible.

Making direct use of the 3-dimensionality of charge and space, we can devise an overall structure requiring a quaternion and a 4-vector within another overall quaternion-type arrangement. This can be accomplished using an octonion, with sixteen members ($\pm 1m$, $\pm \boldsymbol{i}s$, $\pm \boldsymbol{j}e$, $\pm \boldsymbol{k}w$, $\pm \boldsymbol{e}t$, $\pm \boldsymbol{f}x$, $\pm \boldsymbol{g}y$, $\pm \boldsymbol{h}z$), though this is no longer a group. The nonassociativity of the dimensional terms in the octonion extension seems to be lost within terms which effectively cancel each other out, and are of no physical significance.

A related development occurs when we combine the units of 4-vector space-time with those of quaternion mass-charge. The combination of these two constructs, putting the four parameters onto an equal overall footing in a single mathematical representation, creates a 32-part algebra which is identical in all respects to the 32-part algebra used in the Dirac equation for the electron. The 32 parts are then conveniently derived from an anticommuting pentad with just five primitive components ($\boldsymbol{ik}$, $\boldsymbol{i}\mathbf{i}$, $\boldsymbol{i}\mathbf{j}$, $\boldsymbol{i}\mathbf{k}$, $\boldsymbol{j}$). It will be seen that these are the coefficients in the Dirac equation for the respective energy term (E), momentum term ($\mathbf{p}$) in three vector components, and rest mass term (m). Analysis shows that these terms are created by the superposition of the charge quaternion labels ($\boldsymbol{k}$, $\boldsymbol{i}$, $\boldsymbol{j}$) onto the respective parameters time, space and mass. The same process creates the fundamental variation seen in the behaviour of the weak, strong and electromagnetic charges, and suggests their idealised unification in an $SU(5)$ / $U(5)$ group structure.

The Dirac algebra can also be generated by taking charge as the identity element of the parameter group, and representing it by a scalar; the remaining structure for time, space and mass (and, implicitly, the energy, momentum and mass operators) becomes that of the Dirac algebra, and $SU(5)$ or $U(5)$. Such representations do not determine the properties of space, time, mass and charge. They exist because the group has four components, and can, therefore, be represented by a 4-component structure like quaternions, in which the link between elements is made by a binary operation (squaring); but the link between a group with four components and a 4-dimensional space-time or mass-charge may be in itself significant.

9 Noether's theorem revisited

The symmetries we have discussed are prescriptive as well as descriptive. New results may be derived on symmetry grounds even before we have the means of working out their mathematical or physical consequences. One example is associated with an extension of Noether's theorem. This theorem, as we have seen, requires the translation symmetry of time to be linked to the conservation of energy. Of course, since energy is related to mass by the equation $E = mc^2$, then the translation symmetry of time is also linked to the conservation of mass (that is, mass in the general sense, not rest mass). To put it another way, the nonconservation of time is responsible for the conservation of mass. This result could have been derived from symmetry alone, as it is inherent in the Klein-4 group structure; and so, extending the analogy, we can link the conservation of the quantity of charge with the nonconservation, or translation symmetry of space; and since the latter is already linked with the conservation of linear momentum, we can propose a theorem in which the conservation of linear momentum is responsible for the conservation of the quantity of charge (of any type). By the same kind of reasoning, we can make the conservation of *type* of charge linked to the rotation symmetry of space, and so to the conservation of angular momentum, as in the following scheme:

symmetry	**conserved quantity**	**linked conservation**
space translation	linear momentum	value of charge
time translation	energy	value of mass
space rotation	angular momentum	type of charge

When I first proposed these theorems more than a decade ago,[7] I could only give some special cases of their applications. Thus, the conservation of electric charge within a system has been known, since 1927, to be identical to invariance under transformations of the electrostatic potential by a constant representing changes of phase, and the phase changes are of the kind involved in the conservation of linear momentum. Since, in a conservative system, electrostatic potential varies only with the spatial coordinates, this is, in effect, a statement of the principle that the quantity of electric charge is conserved because the spatial coordinates are not, which is a special case of the first predicted relation. In the second case, the relation between spin and statistics observed in fundamental particles could be explained by saying that fermions and bosons have different values of spin angular momentum, and they also differ in that fermions carry weak units of charge, where bosons do not. In some way, then, the presence of a particular *type* of charge determines the angular momentum state of the particle, so conservation of this type of charge is linked with the value of angular momentum. As a result of recent developments, I have now been able to show that the conservation of angular momentum does indeed require the separate conservation of weak, strong and electric charges, through the conservation of the separate properties of orientation (with respect to the linear momentum), direction, and magnitude.

10 Analytic versus synthetic

Our examination of the structure of fundamental physics has revealed that symmetry is an essential ingredient in a successful theory. It has also suggested that symmetries at the fundamental level must be absolute. Such absoluteness cannot be achieved at any but the most abstract or analytic level. The major fundamental ideas in physics have always been analytic rather than synthetic. Strangely, such analytic ideas have always been resisted, even by other physicists, especially when their bases are abstract rather than concretely realisable. The Newtonian method is a classic instance. Like quantum mechanics in our own time, it was opposed *in principle* by nearly all leading scientists of the time, because it postulated an abstract concept of gravitational force independent of any known mechanism which could produce it. It was used only because it worked. That is, it succeeded ultimately, not on account of its fundamental analytical validity, but because it could be applied *synthetically* to a wide range of physical phenomena.

Newton's method separated the abstract *system* from physical *measurement*. The system had a perfection that could never be physically realised. The principle underlying all his work was that there were a few certain types of information which were more fundamental than others and that these were abstract and could be defined precisely in an abstract way without regard to any model of nature based on concrete terms. For Newton, though not for his mechanistically-inclined contemporaries, the ultimate causes of things were abstract rather than mechanical. The laws describing the system did not depend on any physical hypotheses. As he said in Query 28 of the *Opticks* of 1717: 'the main Business of natural Philosophy is to argue from Phaenomena without feigning Hypotheses, and to deduce Causes from Effects, till we come to the very first Cause, which certainly is not mechanical ...'

According to this way of doing physics, universal laws are abstract definitions and do not primarily describe nature. Scientific knowledge is not organised around hypotheses, mechanistic or otherwise, but around fundamental abstract laws which are not derived directly from experiment, but which are the basic parameters in terms of which experimental information is organised. Since they only concern details and since assumptions of any kind can be made in hypotheses, experimental tests are not tests of the validity of fundamental laws.

All other physical laws than the universal ones are solutions of general equations which are ultimately approximate or local. Again, the universal law cannot describe any particular physical system, but is rather a totally abstract statement of a relationship between fundamental parameters of measurement such as mass, space and time. Mathematically, universal laws are expressed by differential equations of which there is no exact solution. The solution always involves an approximation, which does not directly relate to the equation. *Differential equations* are expressed, not in terms of algebraic relations between the quantities themselves, but in terms of their rates of change, or rates of rates of change. To convert from relations involving rates of change to simple and direct relations between the original quantities (which is described as 'solving' the equations), one has to reduce the general equation to a particular case by imposing 'boundary conditions' and this is essentially a process of approximation. In effect, general and exact laws cannot give us direct knowledge; to obtain the latter, we have to reduce the infinite number of possible solutions to a particular and individual case using some kind of approximation.

In Newton's own system, the universal law of gravitation, everything attracting everything else, means that the system was fundamentally indeterminate. Perfection certainly existed in the inverse-square law of attraction between all particles, but the very universality of this law made it impossible to have perfection in a system of such particles in motion. The motion of every particle depended on an infinite number of interactions; strictly speaking, it was not even possible to specify the motion of a particle unless the effect of all these interactions was known. There was no perfection in observed nature, only in the abstract system.

It is significant that mass, in Newton's theory, is something other than the mere quantity of matter. He even describes it as a kind of 'force', the 'impressed force', and gives it force-like properties. The relation between matter and force in the theory allows something outside of matter and opposing it, which can be treated abstractly. Newton, with his theological inclinations, called it 'spirit', which brought him into conflict with the more strictly materialist followers of Descartes. However, in later physics, there is always something other than matter, which has this characteristic. In the nineteenth century, it might be the field concept or aether; in the twentieth century it might be energy or vacuum. A version of it comes into particle physics with the distinction between fermions and bosons. The ultimate origin, in my view, is the distinction between mass and charge (which roughly correspond to the nineteenth-century aether and matter), but it is important that physicists have always found the need for something in opposition to the matter concept.

11 The power of analogy

The idea of 'charge' developed as the eighteenth century struggled to establish the electrostatic force as inverse-square, like gravity, while Kant shows that this was a natural result of 3-dimensional space. The whole thrust of this work was to produce an analogy between the different physical forces. Though current electricity and electromagnetism complicated the picture, Maxwell, eventually set down, in mathematical form, all the laws that were known to be valid for electric and magnetic fields – in particular, those of Coulomb, Ampère and Faraday – and in doing so noted that there was an asymmetry which could be corrected by the addition of another term to the law of Ampère. This term was the so-called displacement current – a current which he supposed must exist when charge was supplied to a parallel plate capacitor. Even though there was no physical justification for such a current except by analogy, the assumption had a remarkable effect, for Maxwell was immediately able to generate wave equations whose velocity is exactly that of light.

The objection to Maxwell's theory of the electromagnetic field for many years was its intrinsically abstract nature. William Thomson (Lord Kelvin) famously declared: 'I never satisfy myself until I can make a mechanical model of a thing. If I can make a mechanical model I can understand it. As long as I cannot make a mechanical model all the way through I cannot understand; and that is why I cannot understand the electromagnetic theory.' (*Baltimore Lectures*, 1884) Following the special theory of relativity, the whole of electromagnetic theory became explicable as an extension of Coulomb's inverse-square law by the addition of a fourth-dimension onto that of space in Minkowski's space-time (which did away with the need even of Einstein's simplified kinematics). The simple parallel between electromagnetism and gravity (or between mass and charge) had at last been established, and it was natural to assume (as in general relativity) that the 4-dimensional space-time connection applied to the gravitational, as well as the electromagnetic.

The method of analogy presupposes the more fundamental concept of symmetry, and this would seem to be the magic ingredient which makes physics work. Symmetry allows us to do what Newton and other analytic physicists have wished to do: to define an abstract, unknowable reality, combined with a process of observation or measurement of its parts. Symmetry between two concepts means absolute identity in most respects, combined with absolute opposition in one. So symmetry allows us to characterize a part of reality without characterizing the whole. Only through symmetry can unity result in diversity. And physics works in such a way that when you characterize a part of reality in a certain way, you are necessarily characterizing the rest as different (i.e. opposite). This is what explains Newton's success in introducing mass as a conserved quantity opposed to the variables space and time, and also the success of his opposition of matter and what he called 'spirit' (ultimately resolving itself as charge and mass). But he didn't consciously set out to do this; he developed by the ruthless application of analytical techniques the only procedures that would work.

The same applies to the creators of quantum mechanics. Though the abstract aspects of the Newtonian and Maxwellian theories were long resisted on account of their intrinsically abstract nature, quantum mechanics has forced modern physicists into the same abstract positions. When Werner Heisenberg introduced his new mechanics, strongly influenced by the formalized dynamical tradition dating back to Lagrange, in which relations were expressed only between observable quantities, he abandoned the reality of Bohr's physical electron orbits and the concept of orbital radius, in order to retain the measurable quantity of frequency as a fundamental quantity. This led to Bohr's Copenhagen interpretation, in which the abstract system was effectively separated from the physical measuring apparatus. The subsequent development of the ideas of nonlocality and entangled states, backed up by strong experimental evidence, has led physics back to the indeterminate infinity of interacting states required even in the classical Newtonian theory, but ignored by his successors.

Quantum mechanics has left many people puzzled. It is clearly a highly successful theory, which can make predictions to eleven places of decimals (in the case of the magnetic moment of the electron), but why does it imply that there is no fundamental 'naïve' reality in which real particles with real positions and real momentum states interact with each other with real forces? The answer ought to be simple, and I believe that it is. We have no right to believe that nature can be described according to the principles of measurement, or according to those of 'naïve' realism. Quantum mechanics, in fact, merely takes to an extreme the principles of conservation and nonconservation which underlie the more classical areas of physics. We simply construct conservation equations for charge and mass which allow for the *complete* variation of the nonconserved parameters space and time which is part of their original specification. It describes the ultimate abstract system based on symmetry.

12 The nature of reality

There is no such thing as 'reality'. Physics has been constructed in such a way that it avoids creating any such concept. The power and the generality of the subject originates entirely in this. However, circumscribed beings like ourselves cannot avoid thinking in 'realistic' terms, and so we have created a system in which apparent reality in one aspect is countered by total nonreality in another. Hence, at the most fundamental level, physics is described by an abstract system, whose relationship to the original concept of measurement is only ever indirect, though it must always be present. Measurement is a component of the system, but it cannot describe it completely. In addition, the mathematical structures which we employ are not a separate system which we *apply* to physics but an integral component of it. Valid mathematical structures have an ultimately physical origin.

It is symmetry which makes it possible to avoid characterizing reality. If to every concept there is an exactly symmetrical opposite, then we never have to specify reality if we use both at the same time. Symmetry can also be exact, in a way that no specific idea can, and it helps us to reduce our starting assumptions without reducing our range of options. Symmetry, however, does *not* mean identity. Space and time are not identical, but *typically symmetrical* concepts: they have some points of identity and others which are different, and indeed symmetrically opposite – facts which ultimately explain wave-particle duality and Zeno's paradoxes. It follows also that the programme to reduce everything to an aspect of space in multiple dimensions is fundamentally misconceived because physics needs its symmetrical opposites for its success.

It is clear that the search for a *unified* theory, however secular it has now become, is essentially at one with the originally theologically-inspired project of the fourteenth century, subsequently continued into the seventeenth century by Galileo and Newton, and we now have a better understanding of what such a theory would actually look like. It would certainly be characterized by abstraction, simplicity and symmetry. It would also be, in principle, extreme, no compromise being allowed for an ultimate theory. There would be no mathematics, other than that derived through symmetry principles, no model-dependent structures of any kind, and no arbitrary cosmology. It would certainly look different from any theory yet devised for a particular aspect of physics, yet these would all be ultimately deducible from it. Our analysis of such structures as we consider to be at the heart of physics suggests that they are not there either for mathematical or for measurement convenience. Physics does not work because it provides a simple or convenient description of 'reality'. Physics works because it has successfully, and uniquely, avoided characterizing nature.

References

1 G. J. Whitrow, *The Natural Philosophy of Time* (Nelson, London, 1961), pp. 135-57

2 P. Coveney and R. Highfield, *The Arrow of Time* (London, 1990), pp. 28, 143-4, 157.

3 A. Robinson, *Non-Standard Analysis* (Princeton University Press, 1996, original publication, 1966).

4 P. Rowlands, *Waves Versus Corpuscles: The Revolution That Never Was* (PD Publications, Liverpool, 1992)

5 P. Rowlands, 'Quantum uncertainty, wave-particle duality and fundamental symmetries', in S. Jeffers, S. Roy, J-P. Vigier and G. Hunter (eds.), *The Present Status of the Quantum Theory of Light: A Symposium in Honour of Jean-Pierre Vigier* (*Fundamental Theories of Physics*, vol. 80, Kluwer, 1997), 361-372.

6 P. Rowlands, 'The Fundamental Parameters of Physics', *Speculat. Sci. Tech*., 6, 69-80, 1983.

7 P. Rowlands, *The Fundamental Parameters of Physics* (PD Publications, Liverpool 1991).

8 P. Rowlands, 'Physics; Let's Get Down to Basics', in *Aspects II*, ed. K. G. Bowden, (ANPA, 1999).