The Cyclotron Note Books

Philip Gibbs

These notes reached their final form on 3 March 1996.

The cyclotron note books are a collection of essays about fundamental physics and metaphysics. The central essay is about the principle of event-symmetric space-time, a new theory about how to do quantum gravity.

The style is more technical in some parts than in others. Each essay can be read independently but if you find that one of them has terms which you don't understand you may find them explained in an earlier one. The later ones may be too difficult for the layman and even I don't understand the last one.

1. The Quantum Gravity Challenge

This first essay is about the problems facing physicists who are trying to combine the theories of general relativity and quantum mechanics into one theory of quantum gravity which might lead ultimately to a complete unified theory of the laws of physics.

2. Is Space-Time Discrete?

Here I discuss some of the motivation and evidence that have led many physicists to believe that, in some sense, space-time is discrete at very small length scales.

3. Metaphysics of Space-Time

Philosophers and physicists have been at odds about the nature of space-time for most of the duration of this century. Why? and who is right?

4. What About Causality?

To attack the principles of causality in physics is to attack our right to believe in free will, yet some physicists are prepared to cast aside causality altogether in order to progress in their understanding of physics.

5. Universal Symmetry

Physicists could never have progressed as far as they have if it were not for symmetry in the laws of physics. But the symmetry is often hidden from sight so just how much of it is there waiting to be revealed?

6. The Superstring Mystery

In 1982 the theory of superstrings suddenly rose to prominence as a possible unified theory of physics. 14 years on it is still not properly understood but the richness of the mathematics which it holds within is enough to keep many of the worlds most proficient theoretical physicists enthralled.

7. Event-Symmetric Space-Time

The principle of event-symmetric space-time is a simple but radical approach to quantum gravity. Proposed by a physicist working outside the academic establishment, it has received little support from within. What are the arguments in it's favour? is it correct? is it useful?

8. Is String Theory in Knots?

Twenty years ago the mathematical theory of knots might have been considered an esoteric discipline of interest to only pure mathematicians. In the last ten years it has become the mathematics of quantum gravity. This essay, more mathematical than others in this collection is an attempt to formulate the equations for an algebraic theory of gravity.

9. The Theory of Theories

If we succeed in finding a unified theory of the laws of physics what will it tell us? Can we hope to understand an answer to the ultimate question, why are we here?

1. The Quantum Gravity Challenge

Quantum Gravity is reputed to be one of the most difficult puzzles of science. In practical terms it is probably of no direct relevance and may even be impossible to verify by experiment. But for physicists it is the holy grail which may enable them to complete the unification of all fundamental laws of physics.

The problem is to put together general relativity and quantum mechanics into one self consistent theory. The difficulty is that the two parts seem to be incompatible, both in concept and in practice. Conceptually, it is the nature of space and time which present fundamental differences. A direct approach, attempting to combine general relativity and quantum mechanics, while ignoring conceptual differences, leads to a meaningless quantum field theory with unmanageable divergences.

There have, in fact, been many attempts to create a theory of quantum gravity. In this article I will first outline the nature of general relativity and quantum mechanics with emphasis on their similarities and differences. Then I will briefly review some of the main stream approaches to quantum gravity. Finally I will talk about some ways in which these ideas now seem to be converging.

General Relativity

General relativity is Einstein's monumental theory of gravity. It is based on two fundamental principles:
The principle of relativity which states that all basic laws of physics should take a form which is independent of any reference frame, and
The principle of equivalence which states that it is impossible to distinguish the effects of gravity from the effects of being in an accelerated frame of reference.

Einstein struggled with the consequences of these principles for several years, constructing many thought experiments to try to understand what they meant. Finally he learnt about Riemann's mathematics of curved geometry and realised that a new theory could be constructed in which the force of gravity was a consequence of the curvature of space-time.

In constructing that theory, Einstein was not significantly influenced by any experimental result which was at odds with the Newtonian theory of gravity. He knew, however, that Newtonian gravity was inconsistent with his theory of special relativity and he knew there must be a more complete self consistent theory. A similar inconsistency now exists between quantum mechanics and general relativity and, even though no experimental result is known to violate either theory, physicists now seek a more complete theory.

In the decades that have followed Einstein's discovery, a number of experimental confirmations of general relativity have been found but there still remains a possibility that it may not be accurate on very large scales, or under very strong gravitational forces. In any case, it is sure to break down under the conditions which are believed to have existed at the big bang where quantum gravity effects were important.

One of the most spectacular predictions of general relativity is that a dying star of sufficient mass will collapse under its gravitational weight into an object so compressed that not even light can escape its pull. These objects are known as black holes. Astronomers now have a growing list of celestial objects which they believe are black holes because of their apparent high density. The accuracy of Einstein's theory may be stringently tested in the near future when gravitational wave observatories such as LIGO come on-line to observe such catastrophic events as the collisions between black holes.

Quantum Mechanics

The Quantum theory was founded before Einstein began his theory of relativity and took much longer to be completed and understood. It was Planck's observations of quanta in the spectrum of black body radiation which first produced signs that the classical theories of mechanics were due for major revisions.

Unlike general relativity which was essentially the work of one man, the quantum theory required major contributions from Bohr, Einstein, Heisenberg, Schroedinger, Dirac and many others, before a complete theory of quantum electrodynamics was formulated. In practical terms, the consequences of the theory are more far reaching than those of general relativity. Applications such as transistors and lasers are now an integral part of our lives and, in addition, the quantum theory allowed us to understand chemical reactions and many other phenomena.

In the 1960's and 70's, further discoveries in quantum field theory have led to successful theories of the nuclear reactions and, in consequence, almost all ordinary physical phenomena can now be attributed to quantum interactions, even if the exact mechanisms are not always fully understood. The electromagnetic and weak nuclear interactions are unified into one force while the strong nuclear interaction is a force of a similar nature known as a gauge theory. Together these forces and all observed particles are combined into one self consistent theory known as the standard model of particle physics.

Despite such spectacular success, confirmed in ever more detail in high energy accelerator experiments, the quantum theory is still criticised by some physicists who feel that its indeterministic nature and its dependency on the role of observer suggest an incompleteness.

Unification

Since Newton set the foundations of physics, progress has come mostly in the form of unification. Maxwell unified electricity, magnetism and light into one theory of electromagnetism. Einstein unified space, time and gravity into one theory of general relativity. More recently, the nuclear forces have been (partially) unified with the electromagnetic force by Weinberg and others.

According to conventional wisdom among physicists, the process of unification will continue until all physics is unified into one neat and tidy theory. There is no *a priori* reason to be so sure of this. It is quite possible that physicists will always be discovering new forces, or finding new layers of structure in particles, without ever arriving at a final theory. It is quite simply the nature of the laws of physics as we currently know them that inspires the belief that we *are* getting closer to the end.

After physicists discovered the atom, they went on to discover that it was composed of electrons and a nucleus, then that the nucleus was composed of protons and neutrons, then that the protons and neutrons were composed of quarks. Should we expect to discover that quarks and electrons are made of smaller particles? This is possible but there are a couple of reasons to suppose not. Firstly there are far fewer particles at this level than there ever were at higher levels. Secondly, their interactions are described by a clean set of gauge bosons through renormalisable field theories. Composite interactions, such as pion exchange, do not take such a tidy form. These reasons in themselves are not quite enough to rule out the possibility that quarks, electrons and gauge bosons are composite but they reduce the number of ways such a theory could be constructed. In fact all viable theories of this type which have been proposed are now all but ruled out by experiment. There may be a further layer of structure but it is likely to be different. It is more common now for theorists to look for ways that different elementary particles can be seen as different states of the same type of object. The most popular candidate for the ultimate theory of this type is **superstring theory**, in which all particles are just different vibration modes of very small loops of string.

Note added: Just a few weeks after writing this, experimenters at Fermilab announced the discovery of evidence for structure within quarks!

Physicists construct particle accelerators which are sort of like giant microscopes. The higher the energy they can produce, the smaller the wavelength of the colliding particles and the smaller the distance scale they probe. In this way physicists can see the quarks inside protons, not through direct pictures but through scattering data. Other things that happen as the energy increases is that new heavy particles are formed and forces become unified. It is impossible to be sure about what will happen the next time a new, more powerful accelerator is built, but physicists can make theories about it.

In the next decade new accelerator experiments at CERN will probe beyond the electro-weak scale. There is some optimism that new physics will be found but nothing is certain.

Planck Scale

At first sight it might seem ridiculous to suppose that we can invent valid theories about physics at high energies before doing experiments. However, theorists have already demonstrated a remarkable facility for doing just that. The standard model of particle physics was devised in the 1960's and experimentalists have spent the last three decades verifying it. The reason for this success is that physicists recognised the importance of certain types of symmetry and self-consistency conditions in quantum field theory which led to an almost unique model for physics up to the electro-weak unification energy scale, with only a few parameters such as particle masses to be determined.

The situation now is a little different. Experimentalists are about to enter a new scale of energies and theorists do not have a single unique theory about what can be expected. They do have some ideas, in particular it is hoped that supersymmetry may be observed, but we will have to wait and see.

Despite these unknowns there are other more general arguments which tell us things about what to expect at higher energies. When Planck initiated the quantum theory he recognised the significance of fundamental constants in physics, especially the speed of light (known as c) and his newly discovered Planck constant (known as h). Scientists and engineers have invented a number of systems of units for measuring lengths, masses and time, but they are entirely arbitrary and must be agreed by international convention. Planck realised that there should be a natural set of units in which the laws of physics take a simpler form. The most fundamental constants, such as *c* and *h* would simply be one unit in that system.

If one other suitable fundamental constant could be selected, then the units for measuring mass, length and time would be determined. Planck decided that Newton's gravitational constant (known as *G*) would be a good choice. Actually there were not many other constants, such as particle masses known at that time, otherwise his choice might have been more difficult. By combining *c*, *h* and *G* Planck defined a system of units now known as the Planck scale. He calculated that the Planck unit of length is very small, about 10^{-35} (ten to the power of minus 35) metres. To build an accelerator which could see down to such lengths would require energies about 10^{15} times larger than those currently available. Note that units of speed and energy can be built from the three basic Planck units but to measure temperature and charge as well we have to also set Boltzmann's constant and the charge on the electron to one unit. In this way we can devise a fundamental system of measurement for all physical quantities.

Physicists have since sought to understand what the Planck scale of units signifies. Those who work with particles believe that at the Planck scale all the four forces of nature, including gravity, are unified. Physicists who specialise in general relativity have a different idea. In 1955 John Wheeler argued that when you combine general relativity and quantum mechanics you will have a theory in which the geometry of space-time is subject to quantum fluctuations, He computed that these fluctuations would become significant if you could look at space-time on length scales as small as the Planck length. Sometimes physicists talk about a space-time foam at this scale but we don't yet know what it really means. For that we will need the theory of quantum gravity.

Without really knowing too much for certain physicists guess that at the Planck scale all forces of nature are unified *and* quantum gravity is significant. It is at the Planck scale that they expect to find the final and completely unified theory of the fundamental laws of physics.

The Small Scale Structure of Space-Time

It seems clear that to understand quantum gravity we must understand the structure of space-time at the Planck length scale. In the theory of general relativity space-time is described as a smooth continuous manifold but we cannot be sure that this is correct for very small lengths and times. We could compare general relativity with the equations of fluid dynamics for water. They describe a continuous fluid with smooth flows in a way which agrees very well with experiment. Yet we know that at atomic scales water is something very different and must be understood in terms of forces between molecules whose nature is completely hidden in the ordinary world. If space-time also has a complicated structure at the tiny Planck length, way beyond the reach of any conceivable accelerator, can we possibly hope to discover what it is?

If you asked a bunch of mathematicians to look for theories which could explain the fluid dynamics of water, without them knowing anything about other physics and chemistry, then they would probably succeed in devising a whole host of mathematical models which work. All those models would probably be very different, limited only by the imagination of the mathematicians. None of them would correspond to the correct description of water molecules and their interactions. The same might be true of quantum gravity. Nevertheless, the task of putting together general relativity and quantum mechanics together into one self consistent theory has not produced a whole host of different and incompatible theories. The clever ideas which have been developed have enigmatic things in common. It is quite possible that all the ideas are partially correct and are aspects of one underlying theory which is within our grasp. It is time now to look at some of those ideas.

Attempts to do Quantum Gravity

The most direct way to try to quantise quantum gravity is to use **perturbative quantum field theory**. This is a procedure which has been applied with great success to electrodynamics. To do the same thing for gravity it is necessary to first construct a system of non-interacting *gravitons* which represent a zero order approximation to quantised gravitational waves in flat space-time. These hypothetical gravitons must be spin two massless particles because of the form of the metric field in general relativity.

The next step is to describe the interactions of these gravitons using the perturbation theory of quantum mechanics, which are defined by a set of Feynman diagrams derived from Einstein's gravitational field equations. For electrodynamics this can be made to work, but only after conveniently cancelling divergent anomalies which appear in the calculations. For gravity this simply cannot be done. The resulting quantum field theory is said to be unrenormalisable and is incapable of giving any useful result.

Because quantum gravity is an attempt to combine two different fields of physics, there are two distinct groups of physicists involved. These two groups form a different interpretation of the failure of the direct attack. The relativists say that it is because gravity cannot be treated perturbatively. To try to do so destroys the basic principles on which relativity was founded. It is, for them, no surprise that this should not work. Particle physicists say that if a field theory is non-renormalisable then it is because it is incomplete. The theory must be modified and new fields must be added to cancel divergences.

Supersymmetry

The first significant progress in the problem of quantum gravity was made by particle physicists. They discovered that a new kind of symmetry called supersymmetry was very important. Particles can be classed into two types; *fermions* such as quarks and electrons, and *bosons* such as photons and Higgs particles. Supersymmetry allows the two types to intermix. With supersymmetry we have some hope to unify the matter fields with radiation fields.

Particle physicists discovered that if the symmetry of space-time is extended to include supersymmetry, then it is necessary to supplement the metric field of gravity with other matter fields. Miraculously these fields led to cancellations of many of the divergences in perturbative quantum gravity. This has to be more than coincidence. At first it was thought that such theories of **Supergravity** might be completely renormalisable. After many long calculations this hope faded.

A funny thing about supergravity was that it works best in ten dimensional space-time. This inspired the revival of an old theory called Kaluza-Klein theory, which suggests that space-time has more dimensions than the four obvious ones. The extra dimensions are not apparent because they are curled up into a small sphere with a circumference as small as the Planck length. This theory provides a means to unify the gauge symmetry of general relativity with the internal gauge symmetries of particle physics.

The next big step taken by particle physicists came along shortly after. Green and Schwarz realised that a theory which had originally been studied as a theory of the strong nuclear force was actually more interesting as a theory of gravity. This was the beginning of string theory. Combining string theory and supergravity to form superstring theory quickly led to some remarkable discoveries. A small set of string theories in ten dimensions were perfectly renormalisable. This was exactly what they were looking for.

It seemed once again that the solution was near at hand, but nature does not give up its secrets so easily. The problem now was that there is a huge number of ways to apply Kaluza-Klein theory to the superstring theories. Hence there seem to be a huge number of possible unified theories of physics. The perturbative formulation of string theory makes it impossible to determine the correct way.

Recently there has been renewed hope for string theory from the discovery that different string theories are connected. They may all be parts of one unique theory after all.

Canonical Quantum Gravity

While particle physicists were making a lot of noise about superstring theory, relativists have been quietly trying to do things differently. Many of them take the view that to do quantum gravity properly you must respect its diffeomorphism symmetry. The *Wheeler-DeWitt equation* together with a *Hamiltonian constraint equation*, describe the way in which the quantum state vector should evolve according to this **canonical approach**.

For a long time there seemed little hope of finding any solutions to the Wheeler-DeWitt equation. Then in 1986 Ashtekar found a way to reformulate Einstein's equations of gravity in terms of new variables. Soon afterwards a way was discovered to find solutions to the equations. This is now known as the **loop representation** of quantum gravity. Mathematicians were surprised to learn that knot theory was an important part of the concept.

The results from the canonical approach seem very different from those of string theory. There is no need for higher dimensions or extra fields to cancel divergences. Relativists point to the fact that a number of field theories which appear to be unrenormalisable have now been quantised exactly. There is no need to insist on a renormalisable theory of quantum gravity. On the other hand, the canonical approach still has some technical problems to resolve. It could yet turn out that the theory can only be made fully consistent by including supersymmetry.

As well as their differences, the two approaches have some striking similarities. In both cases they are trying to be understood in terms of symmetries based on loop like structures. It seems quite plausible that they are both aspects of one underlying theory. Other mathematical fields are common features of both, such as knot theory and topology. Indeed there is now a successful formulation of quantum gravity in three dimensional space-time which can be regarded as either a loop representation or a string theory. A number of physicists such as Lee Smolin are looking for a more general common theory uniting the two approaches.

Black Hole Thermodynamics

Although there is no direct empirical input into quantum gravity, physicists hope to accomplish unification by working on the requirement that there must exist a mathematically self consistent theory which accounts for both general relativity and quantum mechanics as they are separately confirmed experimentally. It is important to stress the point that no complete theory satisfying this requirement has yet been found. If just one theory could be constructed then it would have a good chance of being correct.

Because of the stringent constraints that self consistency enforces, it is possible to construct thought experiments which provide strong hints about the properties a theory of quantum gravity has to have. There are two physical regimes in which quantum gravity is likely to have significant effects. In the conditions which existed during the first Planck unit of time in our universe, matter was so dense and hot that unification of gravity and other forces would have been realised. Likewise, a small black hole who's mass corresponds to the Planck unit of mass provides a thought laboratory for quantum gravity.

Black holes have the property that the surface area of their event horizons must always increase. This is suggestively similar to the law that entropy must increase, and it led Bekenstein to conjecture that the area of the event horizon of a black hole is in fact proportional to its entropy. If this is the case then a black hole would have to have a temperature and obey the laws of thermodynamics. In the 1970's Stephen Hawking investigated the effects of quantum mechanics near a black hole using **semi-classical approximations to quantum gravity**. He discovered the unexpected result that black holes do emit thermal radiation in a way consistent with the entropy law of Bekenstein.

This forces us to conclude that black holes can emit particles and eventually evaporate. For astronomical sized black holes the temperature of the radiation is minuscule and certainly beyond detection, but for small black holes the temperature increases until they explode in one final blast. Hawking realised that this creates a difficult paradox which would surely tell us a great deal about the nature of quantum gravity if we could understand it.

The entropy of a system can be related to the amount of information required to describe it. When objects are thrown into a black hole the information they contain is hidden from outside view because no message can return from inside. Now if the black hole evaporates, this information will be lost in contradiction to the laws of thermodynamics. This is known as the *black hole information loss paradox*.

A number of ways on which this paradox could be resolved have been proposed. The main ones are,

- The lost information escapes to another universe
- The final stage of black hole evaporation halts leaving a remnant particle which holds the information.
- There are strict limits on the amount of information held within any region of space to ensure that the information which enters a black hole cannot exceed the amount represented by its entropy.
- Something else happens which is so strange we can't bring ourselves to think of it.

The first solution would imply a breakdown of quantum coherence. We would have to completely change the laws of quantum mechanics to cope with this situation. The second case is not quite so bad but it does seem to imply that small black holes must have an infinite number of quantum numbers which would mean their rate of production during the big bang would have been divergent. It might be possible to find a way round this but anyway, it is an ugly solution!

Assuming that I have not missed something out, which *is* a big assumption, we must conclude that the amount of entropy which can be held within a region of space is limited by the area of a surface surrounding it. This is certainly counterintuitive because you would imagine that you could write information on bits of paper and the amount you could cram in would be limited by the volume only. This is false because any attempt to do that would eventually cause a black hole to form. Note that this rule does not force us to conclude that the universe must be finite because there is a hidden assumption that the region of space is static which I did not mention.

If the amount of information is limited then the number of physical degrees of freedom in a field theory of quantum gravity must also be limited. Inspired by this observation, Gerard 't Hooft, Leonard Susskind and others have proposed that the laws of physics should be described in terms of a discrete field theory defined on a space-time surface rather than throughout space-time. They liken the way this might work to that of a hologram which holds a three dimensional image within its two dimensional surface.

Rather than being rejected as a crazy idea, this theory has been recognised by many other physicists as being consistent with other ideas in quantum gravity.

Quantised Space-time

Although there has been considerable progress on the problem of quantising gravity, it seems likely that it will not be possible to complete the solution without some fundamental change in the way we think about space-time. All the approaches I have described suggest that the Planck units of length and time define a minimum scale of measurement. Indeed the same conclusion can be reached using fairly general arguments based on the Heisenberg uncertainty principle applied to the metric field of gravity.

One possibility would be that space-time is some kind of lattice structure at small scales. A regular cubic lattice structure is generally regarded as an unacceptable alternative because it destroys space-time symmetry. A random lattice is more plausible. Numerical studies of statistical randomly triangulated surfaces are quite encouraging. The **Regge calculus** describes such a discretisation of gravity and is akin to topological lattice quantum field theories as models of quantum gravity in three dimensions.

As far back as 1947, Synder attempted to quantise space-time by treating space-time co-ordinates as non-commutating operators. The original formulation was unsuccessful but recent work on quantum groups have initiated a revival of this approach. This approach also leads to a discrete interpretation of space-time. Another related topic is **non-commutative geometry** in which space-time itself is regarded as secondary to the algebra of fields which can be generalised to have non-commuting products.

Still this seems to be not quite radical enough to account for quantum gravity. Some physicists believe that we must modify our views sufficiently to allow for dynamical changes in the number of space-time dimensions.

To face the quantum gravity challenge we need new insights and new principles like those which guided Einstein to the correct theory of gravity.

Stephen Hawking, DAMTP Cambridge

Related Reading

Prima Facie Questions in Quantum Gravity, Chris Isham

Quantum Gravity: A Mathematical Physics Perspective, Abhay Ashtekar

The World as a Hologram, Leonard Susskind

Knots and Quantum Gravity, John Baez

Experimental Signatures of Quantum Gravity, Lee Smolin

The Quest for Quantum Gravity, Gary Au

Virtual Black Holes, Stephen Hawking

Structural Issues in Quantum Gravity, Chris Isham

2. *Is Space-Time Discrete?*

Discrete matter

At a seaport in the Aegean around the year 500BC two philosophers, Leucippus and his student Democritus, pondered the idea that matter was made of indivisible units separated by void. Was it a remarkable piece of insight or just a lucky guess? At the time there was certainly no compelling evidence for such a hypothesis. Their belief in the atom was a response to questions posed earlier by Parmenides and Zeno. Perhaps they were also inspired by the coarseness of natural materials like sand and stone. Democritus extended the concept as far as it could go claiming that not just matter, but everything else from colour to the human soul must also consist of atoms.

The idea was subsequently supplanted by the very different philosophies of Plato and Aristotle who believed that matter was infinitely divisible and that nature was constructed from perfect symmetry and geometry. Matter was composed of four elements, Earth, Air, Fire and Water.

In the 1660's Robert Boyle, a careful chemist and philosopher was one of the first to seek a revision. He proposed a corpuscular theory of matter to explain behaviour of gases such as diffusion. According to Boyle there was only one element, all corpuscles would be identical. Different substances would be constructed by combining the corpuscules in different ways. The theory was based as much on the alchemist's belief in the existence of a philosopher's stone which could turn lead into gold, as is was on empirical evidence. Newton built on the corpuscular theory with the mechanistic philosophy of Descartes. He saw the corpuscles as units of mass and introduced the laws of mechanics to explain their motion.

In 1811 the atomic theory was again resurrected by John Dalton to explain chemical composition. Avogadro developed the molecular theory and his law that all gases at the same temperature, pressure and volume contain the same number of molecules even though their weights are different. By the mid nineteenth century the number of molecules could be measured. Maxwell and Boltzmann went on to explain the laws of thermodynamics through the statistical physics of molecular motion. Despite this indirect evidence many scientists were sceptical of the kinetic theory until Einstein supported it. In the early eighteenth century, A biologist Robert Brown had observed random motion of particles suspended in gases. Einstein explained that this Brownian motion could be seen as direct experimental evidence of molecules which were jostling the particles with their own movements.

How far has modern physics gone towards the ideal of Democritus that everything should be discrete?

The story of light parallels that of matter. Newton extended Boyle's corpuscular theory to light but without empirical foundation. Everything he had observed and much more was later explained by Maxwell's theory of Electromagnetism in terms of waves in continuous fields. It was Planck's Law and the photoelectric effect which later upset the continuous theory. These phenomena could best be explained in terms of light quanta. Today we can detect the impact of individual photons on CCD cameras even after they have travelled across most of the observable universe from the earliest moments of galaxy formation.

Those who resisted the particle concepts had, nevertheless, some good sense. Light and matter, it turns out, are both particle and wave at the same time. The paradox is at least partially resolved within the framework of Quantum Field Theories where the duality arises from different choices of basis in the Hilbert space of the wave function.

Discrete Space-Time

After matter and light, history is repeating itself for a third time and now it is space-time which is threatened to be reduced to discrete events. The idea that space-time could be discrete has been a recurring one in the scientific literature of the twentieth century. A survey of just a few examples reveals that discrete space-time can actually mean many things and is motivated by a variety of philosophical or theoretical influences.

It has been apparent since early times that there is something different between the mathematical properties of the real numbers and the quantities of measurement in physics at small scales. Riemann himself remarked on this disparity even as he constructed the formalism which would be used to describe the space-time continuum for the next century of physics. When you measure a distance or time interval you can not declare the result to be rational or irrational no matter how accurate you manage to be. Furthermore it appears that there is a limit to the amount of detail contained in a volume of space. If we look under a powerful microscope at a grain of dust we do not expect to see minuscule universes supporting the complexity of life seen at larger scales. Structure becomes simpler at smaller distances. Surely there must be some minimum length at which the simplest elements of natural structure are found and surely this must mean space-time is discrete.

This style of argument tends to be convincing only to those who already believe the hypothesis. It will not make many conversions. After all, the modern formalism of axiomatic mathematics leaves no room for Zeno's paradox of Archiles and the tortoise. However, such observations and the discovery of quantum theory with its discrete energy levels and the Heisenberg uncertainty principle led physicists to speculate that space-time itself may be discrete as early as the 1930's. In 1936 Einstein expressed the general feeling that ... *perhaps the success of the Heisenberg method points to a purely algebraic method of description of nature, that is, to the elimination of continuous functions from physics. Then, however, we must give up, by principle, the space-time continuum* .... Heisenberg himself noted that physics must have a fundamental length scale which together with Planck's constant and the speed of light permit the derivation of particle masses. Others also argued that it would represent a limit on the measurement of space-time distance. At the time it was thought that this length scale would be around 10^{-15}m corresponding to the masses of the heaviest particles known at the time but searches for non-local effects in high energy particle collisions have given negative results for scales down to about 10^{-19}m and today the consensus is that it must correspond to the much smaller Planck length at 10^{-35}m.

The belief in some new space-time structure at small length scales was reinforced after the discovery of ultraviolet divergences in Quantum Field Theory. Even though it was possible to perform accurate calculations by a process of renormalisation many physicists felt that the method was incomplete and would break down at smaller length scales unless a natural cut-off was introduced.

A technique which introduces such a minimum length into physics by quantising space-time was attempted by Snyder in 1947. Snyder introduced non-commutative operators for space-time co-ordinates. These operators have a discrete spectrum and so lead to a discrete interpretation of space-time. The model was Lorentz invariant but failed to preserve translation invariance. Similar methods have been tried by others since and although no complete theory has come of these ideas there has been a recent upsurge of renewed interest in quantised space-time, now re-examined in the light of quantum groups.

Another way to provide a small distance cut-off in field theory is to formulate it on a discrete lattice. This approach was introduced in 1940 by Wentzel but only later studied in any depth. If the continuum limit is not to be restored by taking the limit where the lattice spacing goes to zero then the issue of the loss of Lorentz invariance must be addressed.

None of these ideas were really very inventive in the way they saw space-time. Only a rare few such as Finkelstein with his space-time code or Penrose with twistor theory and spin networks could come up with any concrete suggestions for a more radical pregeometry before the 1980's.

Another aspect of the quantum theory which caused disquiet was its inherent indeterminacy and the essential role of the observer in measurements. The Copenhagen interpretation seemed inadequate and alternative hidden variable theories were sought. It was felt that quantum mechanics would be a statistical consequence of a more profound discrete deterministic theory in the same sense that thermodynamics is a consequence of the kinetic gas theory.

Over the years many of the problems which surrounded the development of the quantum theory have diminished. Renormalisation itself has become acceptable and is proven to be a consistent procedure in perturbation theory of gauge field physics. The perturbation series itself may not be convergent but gauge theories can be regularised non-perturbatively on a discrete lattice and there is good reason to believe that consistent Quantum Field Theory can be defined on continuous space-time at least for non-abelian gauge theories which are asymptotically free. In Lattice QCD the lattice spacing can be taken to zero while the coupling constant is rescaled according to the renormalisation group. In the continuum limit there are an infinite number of degrees of freedom in any volume no matter how small.

Quantum indeterminacy has also become an acceptable aspect of physics. In 1964 John Bell showed that most ideas for hidden variable theories would violate an important inequality of quantum mechanics. This inequality was directly verified in a careful experiment by Alain Aspect in 1982. There are still those who try to get round this with new forms of quantum mechanics such as that of David Bohm, but now they are a minority pushed to the fringe of established physics. Everett's thesis which leads us to interpret quantum mechanics as a realisation of many worlds has been seen as a resolution of the measurement problem for much of the physics community.

Without the physical motivation discrete space-time is disfavoured by many. Hawking says *Although there have been suggestions that space-time may have a discrete structure I see no reason to abandon the continuum theories that have been so successful*. Hawking makes a valid point but it may be possible to satisfy everyone by invoking a discrete structure of space-time without abandoning the continuum theories if the discrete-continuum duality can be resolved as it was for light and matter.

Discreteness in Quantum Gravity

It is only when we try to include gravity in Quantum Field Theory that we find solid reason to believe in discrete space-time. With quantisation of gravity all the old renormalisation issues return and many new problems arise. Whichever approach to quantum gravity is taken the conclusion seems to be that the Planck length is a minimum size beyond which the Heisenberg Uncertainty Principle prevents measurement if applied to the metric field of Einstein Gravity.

Does this mean that space-time is discrete at such scales with only a finite number of degrees of freedom per unit volume? Recent theoretical results from String Theories and the Loop-representation of Gravity do suggest that space-time has some discrete aspects at the Planck scale.

The far reaching work of Bekenstein and Hawking on black hole thermodynamics has led to some of the most compelling evidence for discreteness at the Planck scale. The *black hole information loss paradox* which arises from semi-classical treatments of quantum gravity is the nearest thing physicists have to an experimental result in quantum gravity. Its resolution is likely to say something useful about a more complete quantum gravity theory. There are several proposed ways in which the paradox may be resolved most of which imply some problematical breakdown of quantum mechanics while others lead to seemingly bizarre conclusions.

One approach is to suppose that no more information goes in than can be displayed on the event horizon and that it comes back out as the black hole evaporates by Hawking radiation. Bekenstein has shown that if this is the case then very strict and counter-intuitive limits must be placed on the maximum amount of information held in a region of space. It has been argued by 't Hooft that this finiteness of entropy and information in a black-hole is also evidence for the discreteness of space-time. In fact the number of degrees of freedom must be given by the area in Planck units of a surface surrounding the region of space. This has led to some speculative ideas about how quantum gravity theories might work through a holographic mechanism, i.e. it is suggested that physics must be formulated with degrees of freedom distributed on a two dimensional surface with the third spatial dimension being dynamically generated.

At this point it may be appropriate to discuss the prospects for experimental results in quantum gravity and small scale space-time structure. Over the past twenty years or more, experimental high energy physics has mostly served to verify the correctness of the Standard Model as proposed theoretically between 1967 and 1973. We now have theories extending to energies way beyond current accelerator technology but it should not be forgotten that limits set by experiment have helped to narrow down the possibilities and will presumably continue to do so.

It may seem that there is very little hope of any experimental input into quantum gravity research because the Planck energy is so far beyond reach. However, a theory of quantum gravity would almost certainly have low energy consequences which may be in reach of future experiments. The discovery of supersymmetry, for example, would have significant consequences for theoretical research on space-time structure.

It from Bit

In the late 1970's the increasing power of computers seems to have been the inspiration behind some new discrete thinking in physics. Monte Carlo simulations of lattice field theories were found to give useful numerical results with surprisingly few degrees of freedom where analytic methods have made only limited progress.

Cellular automata became popular at the same time with Conway's invention of the Game of Life. Despite its simple rules defined on a discrete lattice of cells the game has some features in common with the laws of physics. There is a maximum speed for causal propagation which plays a role similar to the speed of light in special relativity. Even more intriguing is the accidental appearance of various species of **glider** which move through the lattice at fixed speeds. These could be compared with elementary particles.

For those seeking to reduce physics to simple deterministic laws this was an inspiration to look for cellular automata as toy models of particle physics despite the obvious flaw that they broke space-time symmetries. The quest is not completely hopeless. With some reflection it is realised that a simulation of an Ising model with a metropolis algorithm is a cellular automaton if the definition is relaxed to allow probabilistic transitions. The Ising model has a continuum limit in which rotational symmetry is restored. It is important to our understanding of integrable quantum field theories in two dimensions. 't Hooft has also looked to cellular automata as a model of discrete space-time physics. His motivation is somewhat different since indeterminacy in quantum mechanics is, for him, quite acceptable. He suggests that the states of a cellular automata could be seen as the basis of a Hilbert space on which quantum mechanics is formed.

The influence of computers in physics runs to deeper theories than cellular automata. There is a school of thought which believes that the laws of physics will ultimately be understood as being a result from information theory. The basic unit of information is the binary digit or bit and the number of bits of information in a physical system is related to its entropy.

J.A Wheeler has sought to extend this idea, .. *every physical quantity, every it, derives its ultimate significance from bits, a conclusion which we epitomise in the phrase, It from Bit* . For Wheeler and his followers the continuum is a myth, but he goes further than just making space-time discrete. Space-time itself, he argues, must be understood in terms of a more fundamental **pregeometry** . In the pregeometry there would be no direct concepts of dimension or causality. Such things would only appear as emergent properties in the space-time idealisation.

So is it or isn't it?

There do seem to be good reasons to suppose that space-time is discrete in some sense at the Planck Scale. Theories of quantum gravity suggest that there is a minimum length beyond which measurement can not go, and also a finite number of significant degrees of freedom. In canonical quantisation of gravity, volume and area operators are found to have discrete spectra, while topological quantum field theories in 2+1 dimensions have exact lattice formulations.

At the same time, the mathematics of continuous manifolds seem to be increasingly important. Topological structures such as instantons and magnetic monopoles appear to play their part in field theory and string theory. Can such things be formulated on a discrete space?

The riddle will most probably be resolved through a dual theory of space-time which has both discrete and continuous aspects.

Related Reading

Quantum Gravity and Minimum Length, Luis Garay

The Nature of Space and Time, Stephen Hawking

Partons and Black Holes, Leonard Susskind

Small Scale Structure of Space-Time, Phil Gibbs

The Bekenstein Bound, Lee Smolin

Specimen of Theory Construction from Quantum Gravity, Rafael Sorkin

Quantisation and Space-Time Discreteness, Gerard 't Hooft

3. The Metaphysics of Space-Time

Space and time have been favourite subjects for philosophers since at least the ancient Greeks. The paradoxes of the infinite and the infinitesimal are reinvented each day by children with inquisitive minds. How can space be infinite? If it is not infinite what would lie beyond the end? Can the universe have a beginning and an end? How have modern physicists and philosophers learnt to deal with these questions?

The simplest answer is that they use mathematics to construct models of the universe from basic axioms. Mathematicians can define the system of real numbers from set theory and prove all the necessary theorems of calculus that physicists need. With the system of real numbers they can go on to define many different types of geometry. In this way it was possible to discover non-Euclidean geometries in the nineteenth century which were used to build the theory of general relativity in the twentieth.

The self consistency of general relativity can be proven mathematically from the fundamental axioms. This does not make it correct, but it does make it a viable model who's accuracy can be tested against observation. In this way there are no paradoxes of the infinite or infinitesimal. The universe could be infinite or finite, with or without a boundary. There is no need to answer questions about what happened before the beginning of the universe because we can construct a self-consistent mathematical model of space-time in which time has a beginning with no before. Notice that I say *no before* not *nothing before*. There is not even a time before when there was nothing.

So long as we have a consistent mathematical model we know there is no paradox, but nobody yet has an exact model of the whole universe. Newton used a very simple model of space and time described by Euclidean geometry. In that model space and time are separate, continuous, infinite and absolute.

This is consistent with what we observe in ordinary experience. Clocks measure time and normally they can be made to keep the same time within the accuracy of their working mechanisms. It as if there was some universal absolute standard of time which flows constantly. It can be measured approximately with clocks but never directly.

So long as there is no complete theory of physics we know that any model of space-time is likely to be only an approximation to reality which applies in a certain restricted domain. A more accurate model may be found later and although the difference in predicted measurement may be small, the new and old model may be very different in nature. This means that our current models of space and time may be very unrealistic descriptions of what they really are even though they give very accurate predictions in any experiment we can perform.

Philosophers try to go beyond what physicists can do. Using thought alone they consider what space and time might be beyond what can be observed. Even at the time of Newton there was opposition to the notion of absolute space and time from his German rival Leibnitz. He, and many other philosophers who came after, have argued that space and time do not exist in an absolute form as described by Newton. Newton himself appreciated that he was making a big working assumption.

If we start from the point of view of our experiences, we must recognise that our intuitive notions of space and time are just models in our minds which correspond to what our senses find. This is a model which exists like a computer program in our head. It is one which has been created by evolution because it works. In that case there is no assurance that space and time really exist in any absolute sense.

The philosophical point of view developed by Leibnitz, the Bishop Berkeley and Mach is that space and time should be seen as formed from the relationships between objects. Objects themselves are formed from relationships between our experiences. Only our experiences are absolute. The mathematical models used by physicists turn this upside-down. They start with space and time, then they place objects in it, then they predict our experiences as a result of how the objects interact.

Mach believed that space and time do not exist in the absence of matter. The inertia of objects should be seen as being a result of their relation with other objects rather than their relation with space and time. Einstein was greatly influenced by Mach's principle and hoped that it would follow from his own principles of relativity.

In the theory of special relativity he found that space and time do not exist as independent absolute entities but space-time exists as a combination of the two. In General Relativity he found, ironically, that the correct description of his theory must use the mathematics of Riemannian geometry. Instead of confirming Mach's principle he found that space-time can have a dynamic structure in it's own right. Not only could space-time exists independent of matter but it even had interesting behaviour on it's own. His most startling prediction that there should exist gravitational waves, ripples in the fabric of space-time itself, may soon be directly confirmed by detection in gravitational wave observatories.

Einstein's use of geometry was so elegant and compelling that physicists thereafter have always sought to extend the theory to a unified description of matter through geometry. Examples include the Kaluza-Klein models in which space-time is supposed to have more than four dimensions with all but four compacted into an undetectably small geometry. Thus physicists and philosophers have become alienated during the twentieth century.

Recent theories of particle physics have been so successful that it is now very difficult to find an experimental result which can help physicists go beyond their present theories. As a result they have themselves started to sound more philosophical and are slowly reviewing old ideas. The fundamental problem which faces them is the combination of general relativity and quantum theory into a consistent model.

According to quantum theory a vacuum is not empty. It is a sea of virtual particles. This is very different from the way that space and time were envisioned in the days of Mach. In a theory of quantum gravity there would be gravitons, particles of pure geometry. Surely such an idea would have been a complete anathema to Mach. But suppose gravitons could be placed on a par with other matter. Perhaps then Mach would be happy with gravitons after all. The theory could be turned on its head with space-time being a result of the interactions between gravitons.

In string theory, the most promising hope for a complete unified theory of physics, we find that gravitons are indeed on an equal footing with other particles. All particles are believed to be different modes of vibration in loops of string. Even black holes, one of the ultimate manifestations of the geometry of space-time are thought to be examples of single loops of string in a very highly energised mode. There is no qualitative distinction between black holes and particles, or between matter and space-time.

The problem is that there is as yet no mathematical model which makes this identity evident. The equations we do have for strings are somewhat conventional. They describe strings moving in a background space-time. And yet, the mathematics hold strange symmetries which suggest that string theories in different background space-times and even different dimensions are really equivalent. To complete our understanding of string theory we must formulate it independently of space-time. The situation seems to be analogous to the status of electrodynamics at the end of the 19th century. Maxwell's equations were described as vibrations in some ether pervading space. The Michelson-Morley experiments failed to detect the hypothetical ether and signalled the start of a scientific revolution.

Just as Einstein banished the ether as a medium for electromagnetism we must now complete his work by banishing space-time as a medium for string theory. The result will be a model in which space-time is recovered as a result of the relationship between interacting strings. It will be the first step towards a reconciliation of physics and philosophy. Perhaps it will be quickly followed by a change of view, to a point from where all of our universe can be seen as a consequence of our possible experiences just as the old philosophers wanted us to see it. What other ways will we have to modify our understanding to accommodate such a theory? Not all can be foreseen.

4. Causality

Causality on the Radio

I heard a news article on the radio recently. It reported a survey which had shown that people who smoke are more likely to suffer from tooth decay than those who don't. The news presenter concluded that smoking contributed to tooth decay, another good reason to give up cigarettes.

But how right was he? The survey would have looked at the rates of tooth decay among people who smoke and among those who don't. It must have found that there was a statistically significant positive correlation between how many fillings people have in their teeth and how many cigarettes they smoke each day. Given such data, can we conclude that smoking *causes* tooth decay? Would it not have been equally valid to conclude that having fillings causes people to smoke?

The news presenter applied his prejudices and drew a conclusion which sounds reasonable without realising that the converse was also a possible explanation of the survey results. In fact the real reason behind the correlation was probably a third factor such as general variations in health awareness among the people surveyed. Those who care about their good health are more likely to brush their teeth *and* not to smoke.

The difference between the possible conclusions from the survey is not just one of semantics. People listening to the radio might have thought that if they gave up smoking then they would have less tooth decay. They would be wrong. The correct way for them to prevent tooth decay is to brush their teeth more often. In this case the false conclusion is not very dangerous, but such false conclusions about causal relationships drawn from surveys are common. If it is a survey which shows a correlation between race and crime rates then the wrong conclusion could lead to increased racism. In truth it is probably the consequences of racism, not race, which are the real cause of the crime problem identified by the survey in the first place.

Causality in Physics

If your mind is opened a little by my story of the survey in the news article, then now would be a good time to ask yourself if you are drawing the wrong conclusion about causality in physics. Suppose you saw your child bump into a table and an expensive vase fell off smashing into pieces on the floor. Would you conclude that her carelessness *caused* the vase to be broken? Probably you would. Why would you not conclude that the vase falling off the table caused her to bump, quite innocently, into the table?

Your response might be that, for one thing, the vase was broken after her collision with the table so the direction of the causal link is incontestable. Do the laws of physics support such a stance?

To keep things simple, let's start by considering just classical Newtonian mechanics. The form which the laws of physics take is crucial to our understanding of causality. Newton's laws take the form of a set of differential equations describing the motion of particles under forces that act between them. If we know the initial positions and velocities of all the particles at an initial time then their positions are determined at any future time. So does this form for the laws of physics allow us to justify our concept of causality. It would seem so because the initial conditions seem to be causing all that happens in the future.

There is a catch. The laws of physics in this form can be made to work identically in reverse. If we know the final state of a system we can just as easily determine its past. Newton's laws do not explain why past events are the cause of future events.

How about the laws of thermodynamics? If we have a system of many particles then we can not determine all their positions and velocities exactly. When we know only some statistical information about them they obey laws which seem not to be reversible. The second laws of thermodynamics says that entropy must always increase. Could this be linked to causality?

Indeed, the continual increase of entropy is intimately linked to our perception of causality. Entropy is a measure of disorder in a system and defines an arrow of time which can be linked to the psychological arrow. There is, however, a catch. The second law of thermodynamics is inexplicable in terms of the underlying laws of physics which, as far as we know, are reversible. This is enshrined in a theorem of relativistic quantum field theory which proves the necessity of CPT conservation.

The increase of entropy can be understood in certain idealised experiments. For example, if we take two closed containers filled with gases which are each in thermal and chemical equilibrium, and allow them to mix by connecting the two systems without allowing any energy to escape or enter, then when the system comes back into equilibrium the entropy of the final state can be shown theoretically to be higher than the combined entropies in the two original systems. This seems to be theoretical evidence for increasing entropy and it is confirmed by experiment, but we must not be missled. The assumption that systems tend towards equilibrium has been justified. We are victims of our prejudices about causality again and have devised an argument with circular reasoning to support it.

Physicists have devised many other arguments for why entropy always increases, trying to get round the problem of CPT symmetry. Here are a few possibilities:

- CPT symmetry exchanges matter for antimatter so perhaps entropy would decrease for antimatter.
  - Fault: Electromagnetic radiation can not be distinguished from its antimatter image, and yet it obeys the second law of thermodynamics.
- CPT symmetry does not apply to the collapse of the wavefunction in quantum mechanics which is a time asymmetric process.
  - Query: Does this mean that the third law of thermodynamics is not valid for classical statistical mechanics?
- CPT conservation is violated by quantum gravity
  - This could be true but can the laws of thermodynamics be a result of quantum gravity who's effects are normally thought to be irrelevant except in the most extreme physical regimes.
- Entropy increases on account of the fact that it started very low at the beginning of time. Thus it is due to the initial conditions being set in a special way. and from then on it could only increase.
  - But then why were initial conditions set rather than final or mixed boundary conditions.

When I was an undergraduate I naively thought that physicists understood entropy. Some have produced arguments based on any or all of the above possibilities. In retrospect I think now that I should be no more convinced by any of those arguments than I should if I heard someone arguing that smoking causes tooth decay based on the correlation reported in the survey.

One of the difficulties is that we don't really have an ideal definition of entropy. We can understand it as a measure of disorder in a closed equilibrium system. More generally we have to resort to some kind of coarse graining process in which we imagine that a non-equilibrium system can be seen as made of small sub-systems, or grains, which are in equilibrium themselves but not in equilibrium with each other.

Entropy might be better understood in terms of information. It can be linked to the number of bits which are needed to describe a system accurately. In a hot disordered system you need to specify the individual state of each particle, while a cold lattice can be described in terms of its lattice shape, size and orientation. Far less information is needed for the low entropy system.

The claim that entropy increases because it started low in the big bang is perhaps the one which has fallen into conventional wisdom, even if it is admitted that we don't understand why it started low. Perhaps it is because of some huge unknown symmetry which was valid at the high temperatures of the big bang and broken later. It is not really clear why it should increase all the time either. Why can't it just go up and down?

In a completely deterministic system the evolution of the system is equally well determined by its final state as by its initial so we could argue that the amount of information in the system must be constant. The difficulty there is that we are assuming an exact knowledge of state which is impossible. In any case, quantum mechanics is not deterministic. If we make a perfect crystal with an unstable isotope, as time passes some of the atoms will decay. The amount of information needed to track the decayed atoms increases. Perhaps, then, it really is quantum mechanics and the collapse of the wave function which is responsible.

If physicists used to think they understood entropy then their faith was deeply shaken when Hawking and Bekenstein discovered that the laws of thermodynamics could be extended to the quantum mechanics of black holes. The entropy is given by the area of the black hole but its temperature can only be understood through quantum mechanical effects. This shows that classical understanding of thermodynamics is indeed incomplete and perhaps only a complete theory of quantum gravity can explain the laws fully.

We might try to understand the quantum state of the entire universe by using Feynman's path integral formulation of quantum mechanics. We must form a sum over all possible space-time manifolds allowed in general relativity. Hawking has argued that we can understand entropy in this way if the universe is an entirely closed system, bounded in both time and space. He claims that there are two possible ways a universe could start or end. One has low entropy the other high. The only consistent picture is one in which it is low at one end and high at the other hence temporal symmetry is broken.

If this argument could be made solid then it would be a powerful one. The path integral formulation avoids problems of time since it is a sum over all possible universes rather than an evolutions with separate boundary conditions. The argument can only be made complete when we understand quantum gravity better.

At this point our belief in causality seems to rest in our faith that the universe has a simple topology as described by standard big bang models. This rests on little more than some limited observations and an application of Occam's razor. Our measurements of the cosmic microwave backgrounds show a high degree of isotropy and the universe seems to be homogenous on large scales in so far as we can tell. Our observation is limited by a horizon defined by the age of the universe and the speed of light. Thus we cannot observe anything beyond about 15 billion light years distance. Why should we imagine that the size of the universe is a similar order of magnitude to its current age? We have been unable to measure the extent to which space is curved and can not place limits on its size, or even be sure it is finite.

It seems to be only an application of Occam's razor which justifies the assumption that space is homogenous on scales hundreds of orders of magnitude larger than the observable horizon. It is quite possible, as far as we can tell, that the big bang is actually just a huge white hole which formed in a larger universe. Perhaps on some huge scale there are a population of black and white holes of vastly different sizes. What does that say about the laws of thermodynamics?

Apart from entropy there are other aspects of causality. We know that in general relativity causal effects are limited by the light cones which are part of the geometry of space-time. But the geometry is itself dynamic. In general relativity it is possible to construct space-time models which have closed time-like paths. If such things really exist in the universe we would be able to travel back to our past.

Traditionally physicists have simply said that such universes must be ruled out because if we could travel back to our past we could change our history, which seems to raise contradictions. Recently some physicists have started to question this assumption. It seems possible that quantum mechanics may allow closed time like curves through space-time wormholes to be constructed, at least in principle. The contradictions which were thought to be a consequence of time travel do not stand up to close examination.

Perhaps it would be possible to travel back to the past and see our parents but some chance event would prevent us from being able to change their lives in ways which we know never happened. If that is a correct interpretation then it attacks our faith in our own free will.

There is perhaps little that we can conclude reliably about causality from our current understanding of physics. Only when we have found and understood the correct theory for quantum gravity will we be able to know the truth. We may be prevented from finding that theory if we hold fast to our prejudices.

Occam's razor does not have a very good track record in cosmology. Usually space turns out to contain more complexity than we imagined before we looked. It will be billions of years before we are able to see beyond the current horizon defined by the speed of light . In the nearer term theory is our only hope to know what the structure of space-time is like on very large scales.

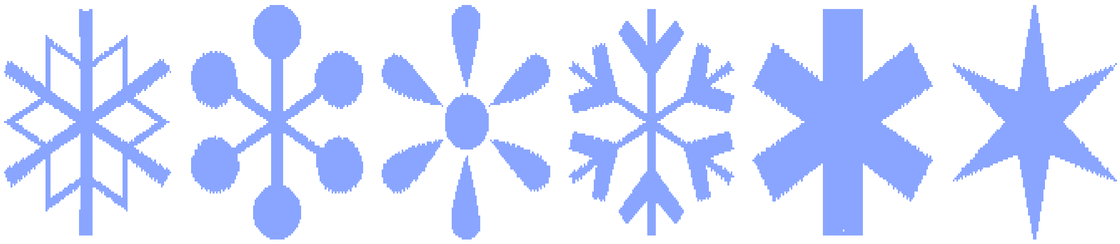
Related Reading

Cosmology, Time's Arrow, and that old double standard, Huw Price

Nonlocality as an Explanation for Finetuning and Field Replication in Nature., Bennet, Froggatt and Nielsen

Time Machines and the principle of least action, Carlini, Frolov, Mensky, Novikov, Soleng

5. Universal Symmetry



A Snowflake

We don't have to examine nature very closely to see its beauty. A bird, a forest or a galaxy has a form of beauty which is typical of complex organised systems. A snowflake has another element to its beauty which is also very common in nature but which is often only evident on close inspection. We call it *symmetry*.

The snowflake begins its life as a minute hexagonal crystal forming in a cloud. During its passage from there to the ground, it experiences a sequence of changes in temperature and humidity which cause it to grow at varying rates. Its history is recorded in the variations of thickness in its six petals as it grows. This process ensures that each petal is virtually identical and accounts for the snowflakes symmetry.

When a snowflake is rotated through an angle of 60 degrees about its centre, it returns to a position where it looks the same as before. It is said to be *invariant* under such a transformation and it is invariance which characterises symmetry. The shape of the snowflake is also invariant if it is rotated through 120 degrees. It is invariant again if it is turned over. By combining rotations and turning over it is possible to find 12 different transformations (including the identity transformation which does nothing). We say that the *order* of the snowflakes symmetry is 12.

Consider now the symmetry of a regular tetrahedron. That is a solid shape in the form of a pyramid with a triangular base for which all four faces are equilateral triangles. The shape of a regular tetrahedron is invariant when it is rotated 120 degrees about an axis passing through a vertex. It is also invariant when rotated 180 degrees about an axis passing through the midpoints of opposite edges. If you make a tetrahedron and experiment with it you will find that it has a symmetry of order 12. But the symmetry of the tetrahedron is not quite the same as that of a snowflake because the snowflake has a transformation which must be repeated six times to restore it to its original position and the tetrahedron does not.

Mathematicians have provided precise definitions of what I meant by *not quite the same*. The invariance transformations of any shape form an algebraic structure called a *group* when you consider composition of transformations as multiplication. Two groups are isomorphic if there is a one-to-one mapping between them which respects the multiplication. Groups can be considered to be a mathematical abstraction of symmetry and mathematicians have spent a great deal of effort in classifying them but with only partial success. One spectacular achievement is the complete classification of finite simple groups which culminated in the discovery of the monster group which has

808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000 elements. It is a subject of great beauty in itself.

There are also infinite order symmetries described by infinite groups. The simplest example is the group of rotations in a plane which describes the symmetry of a circle. Mathematicians have also succeeded in classifying an important class of infinite dimensional groups known as semi-simple Lie groups.

Symmetry in physics

Symmetry is important in physics because there are all kinds of transformations which leave the laws of physics invariant. For example, we know that the laws of physics are the same everywhere. I.e. we can detect no difference in the results of any self contained experiment which depends on where we do it. Another way to say the same thing is that the laws of physics are invariant under a translation transformation. The infinite dimensional group of translation transformations is a symmetry of the laws of physics.

The next important example is rotation symmetry. The laws of physics are invariant under rotations in space about any axis through some origin. An important difference between the translation symmetry and the rotation symmetry is that the former is *abelian* while the latter is *non-abelian*. An Abelian group is one in which the order of multiplication does not matter, they *commute*. This is true of translations but is not true of rotations about different axis.

If the laws of physics are invariant under both rotations and translations then they must also be invariant under any combination of a rotation and a translation. In this way we can always combine any two symmetries to form a larger one. The smaller symmetries are contained within the larger one. Note that the symmetry of a snowflake is already contained within rotation symmetry. Mathematicians say that the invariance group of the snowflake is a *subgroup* of the rotation group.

Hidden Symmetry

Symmetry in physics is not always evident at first sight. When we are comfortably seated on the ground we notice a distinct difference between up and down, and between the horizontal and the vertical. If we describe the motion of falling objects in terms of physical laws which have the concept of vertical and horizontal built in then we do not find the full rotational symmetry in those laws. Many ancient philosophers thought that the Earth marked a special place at the centre of the universe. In such a case we could not say that the laws of physics were invariant under translations.

It was the Copernican revolution that changed all that. Newton discovered a law of gravity which could at the same time account for falling objects on Earth and the motion of the planets in the Solar system. From that point on it could be seen that the laws of physics are invariant under rotations and translations. It was a profound revelation. Whenever new symmetries of physics are discovered the laws of physics become more unified. Newton's discovery meant that it was no longer necessary to have different theories about what was happening on Earth and what was happening in space.

Once the unifying power of symmetry is realised and combined with the observation that symmetry is not always recognised at first sight, the great importance of symmetry is revealed. Physicists have discovered that as well as the symmetries of space transformations, there are also more subtle internal symmetries which exist as part of the forces of nature. These symmetries are important in particle physics. In recent times it has been discovered that symmetry can be hidden through mechanisms of spontaneous symmetry breaking. Such mechanisms are thought to account for the apparent differences between the known forces of nature. This increases the hope that there are other symmetries not yet found. Ultimately we may discover the universal symmetry which combines all other symmetries of physics.

Conservation Laws

During the centuries which followed Newton's work physicists and mathematicians came to realise that there is a deep relationship between symmetry and conservation laws in physics. The law of conservation of momentum is related to translation invariance, while angular momentum is related to rotation invariance. Conservation of energy is due to the invariance of the laws of physics with time.

The relationship was finally established in a very general mathematical form known as Noether's theorem. Mathematicians had discovered that classical laws of physics could be derived from a philosophically pleasing *principle of least action*. Noether showed that any laws of this type which have a continuous symmetry would have a conserved quantity which could be derived from the action principle.

Although Noether's work was based on classical Newtonian notions of physics. The principle has survived the quantum revolution of the twentieth century. In quantum mechanics we find that the relationship between symmetry and conservation is even stronger. There are even conservation principles related to discrete symmetries.

An important example of this is *parity*. Parity is a quantum number which is related to symmetry of the laws of physics when reflected in a mirror. Mirror symmetry is the simplest symmetry of all since it has order two. If the laws of physics were indistinguishable from their mirror inverse then according to the rules of quantum mechanics parity would be conserved. It was quite a surprise to physicists when they discovered that parity is not conserved in weak nuclear interactions. Because these interactions are not significant in our ordinary day-to day life, we do not normally notice this asymmetry of space.

Simple laws of mechanics as well as those of gravity and electrodynamics are symmetric under mirror inversion. They are also invariant under time reversal. This is a little surprising because our everyday world does not appear to be symmetric in this way, there is a clear distinction between future and past. Time reversal is also broken by the weak interaction but not enough to account for the perceived difference. There is a combined operation of mirror inversion and time reversal and a third operation which exchanges a particle with its antiparticle image. This is known as CPT. Again the universe does not appear to realise particle-antiparticle symmetry macroscopically because there seems to be more matter than anti-matter in the universe. However, CPT is an exact symmetry of all interactions, as far as we know.

Relativity

There is another symmetry which is found in ordinary mechanics. If you are travelling in a modern high speed train like the French TGV, on a long straight segment of track, it is difficult to tell that you are moving without looking out of the window. If you could play a game of billiards on the train, you would not notice any effects due to the speed of the train until it turned a corner or slowed down.

This can be accounted for in terms of an invariance of the laws of mechanics under a *Galilean transformation* which maps a stationary frame of reference onto one which is moving at constant speed.

When you examine the laws of electrodynamics discovered by Maxwell you find that they are not invariant under a Galilean transformation. Light is an electrodynamic wave which moves at a fixed speed c. Because c is so fast compared with the speed of the TGV, you could not notice this on a train. However, towards the end of the nineteenth century, a famous experiment was performed by Michelson and Morley. They hoped to detect changes in the speed of light due to the changing direction of the motion of the Earth. To everyone's surprise they could not detect the difference.

Maxwell believed that light must propagate through some medium which he called ether. The Michelson-Morley experiment failed to detect the ether. The discrepancy was finally resolved by Einstein when he discovered special relativity. The Galilean transformation, he realised, is just an approximation to a Lorentz transformation which is a perfect symmetry of electrodynamics. The correct symmetry was there in Maxwell's equations all along but symmetry is not always easy to see. In this case the symmetry involved an unexpected mixing of space and time co-ordinates. Minkowski later explained that Einstein had unified space and time into one geometric structure which was thereafter known as space-time.

It seems that Einstein was more strongly influenced by symmetry principles than he was by the Michelson-Morley experiment. According to the scientific principle as spelt out by Francis Bacon, theoretical physicists should spend their time fitting mathematical equations to empirical data. Then the results can be extrapolated to regions not yet tested by experiment in order to make predictions. In reality physicists have had more success constructing theories from principles of mathematical beauty and consistency alone. Symmetry is an important part of this method of attack.

Einstein demonstrated the power of symmetry again with his dramatic discovery of general relativity. This time there was no experimental result which could help him. Actually there was an observed discrepancy in the orbit of Mercury but this could just as easily have been corrected by some small modification to Newtonian gravity. Einstein knew that Newton's description of gravity was inconsistent with special relativity, and even if there were no observation which showed it up, there had to be a more complete theory of gravity which complied with the principle of relativity.

Since Galileo's experiments on the leaning tower of Pisa, it was known that inertial mass is equal to gravitational mass. Einstein realised that this would imply that an experiment performed in an accelerating frame of reference could not separate the apparent forces due to acceleration from those due to gravity. This suggested to him that a larger symmetry which included acceleration might be present in the laws of physics.

It took several years and many thought experiments before Einstein completed the work. He realised that the equivalence principle implied that space-time must be curved, and the force of gravity is a direct consequence of this curvature. In modern terms the symmetry he discovered is known as diffeomorphism invariance. It means that the laws of physics take the same form when written in any 4d co-ordinate system on space-time.

I would like to stress that the symmetry of general relativity is a much larger symmetry than any which had been observed in physics before. We can combine rotation invariance, translation invariance and Lorentz invariance to form the complete symmetry group of special relativity which is known as the Poincare group. The Poincare group can be parameterised by ten real numbers. We say it has *dimension* 10. Diffeomorphism invariance, on the other hand, cannot be parameterised by a finite number of parameters. It is an infinite dimensional symmetry.

Diffeomorphism invariance is a hidden symmetry. If the laws of physics were invariant under any change of co-ordinates in a way which could be clearly observed, then we would expect the world around us to behave as if everything could be deformed like rubber. The symmetry is hidden by the local form of gravity just as the constant vertical gravity seems to hide rotational symmetry in the laws of physics. On cosmological scales the laws of physics do have a more versatile form allowing space-time to deform, but on smaller scales only the Poincare invariance is readily observed.

Einstein's field equations of general relativity which describe the evolution of gravitational fields, can be derived from a principle of least action. It follows from Noether's theorems that there are conservation laws which correspond to energy, momentum and angular momentum but it is not possible to distinguish between them. A special property of conservation equations derived from the field equations is that the total value of a conserved quantity integrated over the volume of the whole universe is zero, provided the universe is closed. This fact is useful when sceptics ask you where all the energy in the universe came from if there was nothing before the big bang! However, the universe might not be finite.

A final remark about relativity is that the big bang breaks diffeomorphism invariance in quite a dramatic way. It singles out one moment of the universe as different from all the others. It is even possible to define absolute time as the proper time of the longest curve stretching back to the big bang. This fact does not destroy relativity provided the big bang can be regarded as part of the solution rather than being built into the laws of physics. In fact we cannot be sure that the big bang is a unique event in our universe. Although the entire observable universe seems to have emerged from this event it is likely that the universe is much larger than what is observable. In that case we can say little about its structure on bigger scales.

Gauge Symmetry

What about electric charge? It is a conserved quantity so is there a symmetry which corresponds to charge according to Noether's theorem? The answer comes from a simple observation about electric voltage. It is possible to define an electrostatic potential at any point in space. The voltage of a battery is the difference in this potential between its terminals. In fact there is no way to measure the absolute value of the electrostatic potential. It is only possible to measure its difference between two different points. In the language of symmetry we would say that the laws of electrostatics are invariant under the addition of a value to the potential which is the same everywhere. This describes a symmetry which through Noether's theorem can be related to conservation of electric charge.

In fact the electric potential is just one component of the electromagnetic vector potential which can be taken as the dynamical variables of Maxwell's theory allowing it to be derived from an action principle. In this form the symmetry is much larger than the simple one parameter invariance I just described. It corresponds to a change in a scalar field of values defined throughout space-time. Like the diffeomorphism invariance of general relativity this symmetry is infinite dimensional. Symmetries of this type are known as gauge symmetries.

Both diffeomorphism invariance and the electromagnetic symmetry are local gauge symmetries because they correspond to transformation which can be parameterised as fields throughout space-time. In fact there are marked similarities between the forms of the equations which describe gravity and those which describe electrodynamics, but there is an essential difference too. Diffeomorphism invariance describes a symmetry of space-time while the symmetry of electromagnetism acts on some abstract internal space of the components of the field.

The gauge transformation of electrodynamics acts on the matter fields of charged particles as well as on the electromagnetic fields. The phase of the matter fields is multiplied by a phase factor. Through this action the transformation is related to the symmetry group of the circle which is known as U(1).

In the 1960s physicists were looking for quantum field theories which could explain the weak and strong nuclear interactions as they had already done for the electromagnetic. They realised that the U(1) gauge symmetry could be generalised to gauge symmetries based on other Lie groups. As I have already said, an important class of such theories has been classified by mathematicians. They can be described as matrix groups which fall into three families parameterised by an integer *N* and five exceptional groups:

The orthogonal groups SO(N)

The unitary groups SU(N)

The symplectic groups Sp(N)

Exceptional Groups G2 F4 E6 E7 E8

The best thing about gauge symmetry is that once you have selected the right group the possible forms for the action of the field theory are extremely limited. Einstein found that for general relativity there is an almost unique most simple form with a curvature term and an optional cosmological term. For internal gauge symmetries the corresponding result is Yang-Mills field theory. From tables of particles physicists were able to conjecture that the strong nuclear interactions used the gauge group SU(3). The weak interaction was a little more difficult. It turned out that the symmetry was SU(2)xU(1) but that it was broken by a Higgs mechanism. By this use of symmetry theoretical physicists were able to construct the complete standard model of particle physics which has kept the experimentalists busy for 30 years.

Super symmetry

Symmetry is proving to be a powerful unifying tool in particle physics because through symmetry and symmetry breaking, particles which appear to be different in mass, charge etc.

can be understood as different states of a single unified field theory. Ideally we would like to have a completely unified theory in which all particles and forces of nature are related through a broken symmetry.

A possible catch is that fermions and bosons cannot be related by the action of a classical group based symmetry. One way out of this problem would be if all bosons were revealed to be bound states of fermions. but the gauge bosons appear to be fundamental.

A more favourable possibility is that fermions and bosons are related by supersymmetry. Supersymmetry is an algebraic construction which is a generalisation of the Lie-group symmetries already observed in particle physics. It is a type of symmetry which can not be described by a classical group, but it has most of the essential algebraic properties.

If supersymmetry existed in nature we would expect to find that fermions and bosons came in pairs of equal mass. In other words there would be bosonic squarks and selectrons with the same masses as the quarks and electrons, as well as fermionic photinos and higgsinos with the same masses as photons and Higgs. The fact that no such partners have been observed implies that supersymmetry must be broken if it exists.

It is probably worth adding that there may be other ways in which supersymmetry is hidden. For example, If quarks are composite then the quark constituents could be supersymmetric partners of gauge particles. Also the creation and annihilation operators of fermions define a supersymmetry. Finally, Witten has found a mechanism which allows particles to have different masses even though they are supersymmetric partners and the symmetry is not broken.

There is now some indirect experimental evidence in favour of supersymmetry, but the main reasons for believing in its existence are purely theoretical. During the 1970s it was discovered that a form of space-time supersymmetry applied to general relativity provides a perturbative quantum field theory for gravity which has better renormalisation behaviour. This was one of the first breakthroughs of quantum gravity.

The big catch with supergravity theories is that they work best in ten or eleven dimensional space-time. To explain this discrepancy with nature theorists revived an old idea called Kaluza-Klein theory which was originally proposed as a way to interpret internal gauge theories geometrically. According to this idea space-time has more dimensions than are apparent. All but four of them are compacted into a ball so small that we do not notice it. Particles are then supposed to be modes of vibration in the geometry of these extra dimensions. If we believe in supergravity then even fermions fall into this scheme.

At one point supergravity looked very promising as a theory which might unify all physics. At the time I was a student at Cambridge University where Stephen Hawking was taking up his position as new Lucasian professor. There was great anticipation of his inaugural lecture and even though I made a point of turning up early I found only standing room in the auditorium. It was an exciting talk at which Hawking made some of his most famous comments. He confidently predicted that the end of physics was in sight. But early hopes faded as the perturbative calculations in supergravity became difficult and it seemed less likely that it defined a renormalisable field theory. Hawking maintains his controversial claim.

Supergravity was quickly superseded by superstring theory. String theories had earlier been considered as a model for strong nuclear forces but, with the addition of supersymmetry it became possible to consider them as a unified theory including gravity. In fact, supergravity is present in superstring theories.

Enthusiasm for superstring theories became widespread after Schwartz and Green discovered that a particular form of string theory was not only renormalisable, it was even finite to all orders in perturbation theory. That event started many research projects which are a story for another article. All I will say now is that string theory is believed to have much more symmetry than is understood, but its nature and full form is still a mystery.

String theory has also excited some mathematicians. They have found that in certain background space-times it has a symmetry described by the *fake monster super Lie algebra* which is related to the finite monster group.

Universal Symmetry

We have seen how symmetry in nature has helped physicists uncover the laws of physics. Symmetry is a unifying concept. It has helped combine the forces of nature as well as joining space and time. There are other symmetries in nature which I have not yet mentioned. These include the symmetry between identical particles and the symmetry between electric and magnetic fields in Maxwell's equations of electrodynamics. Symmetry is often broken or hidden so it is quite possible that there is more.

There is plenty of good reason to think that there is more. Both general relativity and quantum mechanics are full of symmetry so it would be natural to imagine that a unified theory of quantum gravity would combine those symmetries into a larger one. String theory certainly seems to have many forms of symmetry: conformal symmetries, W-symmetries dualities etc. There is evidence within string theory that it contains a huge symmetry which has not yet been revealed.

The information paradox in the thermodynamics of black holes might be solved by a hologram mechanism. This would require that the number of effective degrees of freedom in quantum gravity can be reduced by a large gauge transformation; more evidence for a peculiar hidden symmetry in quantum gravity.

It seems that there is some universal symmetry in nature that has yet to be found. It will be a symmetry which includes the gauge symmetries and perhaps also the symmetry of identical particles. The existence of this symmetry is a big clue to the nature of the laws of physics and may provide the best hope of discovering them with little further empirical data.

The importance of the symmetry in a system of identical particles is often overlooked. The symmetry group is the permutation group acting to exchange particles of the same type. The reason why this symmetry is not considered to be as important as gauge symmetry lies in the relationship between classical and quantum physics. There is an automatic scheme which allows a classical system of field equations derived from a principle of least action to be quantised. This can be done either through Dirac's canonical quantisation or Feynman's path integral. The two are formally equivalent. In modern quantum field theory a classical field is quantised and particles arise as a result. Gauge symmetry is a symmetry of the classical field which is preserved in the process of quantisation. The symmetry between identical particles, however, does not exist in the classical theory. It appears along with the particles during the process of quantisation. Hence it is a different sort of symmetry.

But the matter can not simply be left there. In a non-relativistic approximation of atomic physics it is possible to understand the quantum mechanics of atoms by treating them first of all as a system of classical particles. The system is quantised in the usual way and the result is the Schroedinger wave equation for the atom. This time we have gone from a classical particle picture to a field theory and the symmetry between particles existed as a classical symmetry.

This observation suggests that the relationship between classical and quantum systems is not so clear as it is often portrayed and that the permutation group could also be a part of the same universal symmetry as gauge invariance. This claim is now supported by string theory which appears to have a mysterious duality between classical and quantum formulations. A further clue may be that the algebra of fermionic creation and annihilation operators generate a supersymmetry which includes the permutation of identical particles.

What will the universal symmetry look like? The mathematical classification of groups is incomplete. Finite simple groups have been classified and so have semi-simple Lie groups. But infinite dimensional groups appear in string theory and these are so far beyond classification. Further more, there are new types of symmetry such as supersymmetry and quantum groups which are generalisations of classical symmetries. These symmetries are algebraic constructions which preserve an abstract form of invariance. They turn up in several different approaches to quantum gravity including string theory so they are undoubtedly important. This may be because of their importance in understanding topology. At the moment we don't even know what should be regarded as the most general definition of symmetry let alone having a classification scheme.

There seems little doubt that there is much to be learnt in both mathematics and physics from the hunt for better symmetry. The intriguing idea is that there is some special algebraic structure which will unify a whole host of subjects through symmetry, as well as being at the root of the fundamental laws of physics.

Related Reading

Asymptotic Freedom and the Emergence of QCD, David Gross

Kaluza-Klein Theory in Perspective, Michael Duff

The Life and Times of Emmy Noether, Nina Byers

Exact Electromagnetic Duality, David Olive

The Status of Supersymmetry, Jonathon Bagger

The Unreasonable Effectiveness of Quantum Field Theory, R. Jackiw

6. The Superstring Mystery

Theory of Everything?

In 1982 Michael Green and John Schwarz made a discovery which might turn out be the greatest scientific advance of all time, if it is right. What they found was that a particular quantum field theory of supersymmetric strings in 10 dimensions gives finite answers at all orders in perturbation theory.

This was a breakthrough because the superstring theory had the potential to include all the particles and forces in nature. It could be a completely unified theory of physics. By 1985 the press had got hold of it. Articles appeared in Science and New Scientist. They called strings a *Theory Of Everything*.

The term *Theory of Everything* is a desperately misleading one. Physicists usually try to avoid it but the media can't help themselves. If physicists find a complete unified set of equations for the laws of physics, then that would be a fantastic discovery. The implications would be enormous, but to call it a theory of everything would be nonsense.

For one thing, it would be necessary to solve the equations to understand anything. No doubt many problems in particle physics could be solved from first principles, perhaps it would be possible to derive the complete spectrum of elementary particles. However, there would certainly be limits to the solvability of the equations. We already find that it is almost impossible to derive the spectrum of hadrons composed of quarks, even though we believe we have an accurate theory of strong interactions. *In principle* any set of well defined equations can be solved numerically given enough computer power. The whole of nuclear physics and chemistry ought to be possible to calculate from the laws we now have. *In practice* computers are limited and experiments will always be needed.

Furthermore, it is not even possible to derive everything *in principle* from the basic laws of physics. Many things in science are determined by historical accident. The foundations of biology fall into this category. The final theory of physics will not help us to understand how life on Earth originated. The most ardent reductionist would retort that, in principle, it would be possible to derive a list of all possible forms of life from the basic laws of physics.

Finally it must be said that even given a convincing unified theory of physics, it is likely that it would still have the indeterminacy of quantum mechanics. This would mean that no argument could finally lay to rest questions about paranormal, religion, destiny or other such things, and beyond that there are many matters of philosophy and metaphysics which might not be resolved, not to mention an infinite number of mathematical problems. Clearly the term *Theory of Everything* is misleading.

Theory of Nothing?

Following the media reports about string theory there was an immediate backlash. People naturally asked what this *Theory Of Everything* had to tell us. The answer was that it could not yet tell us anything, even about physics. On closer examination it was revealed that the theory is not even complete. It exists only as a perturbation series with an infinite number of terms. Although each term is well defined and finite, the sum of the series will diverge. To understand string theory properly it is necessary to define the action principle for a non-perturbative quantum field theory. In the physics of point particles it is possible to do this at least formally, but in string theory success has evaded all attempts. To get any useful predictions out of string theory it will be necessary to find non-perturbative results. The perturbation theory simply breaks down at the Planck scale where stringy effects should be interesting.

More bad news was to come. Systematic analysis showed that there were really several different 10 dimensional superstring theories which are well defined in perturbation theory. If you count the various open and closed string theories with all possible chirality modes and gauge groups which have no anomalies, there are four in all. This is not bad when compared to the infinite number of renormalisable theories of point particles, but one of the original selling points of string theory was its uniqueness. Worse still, to produce a four dimensional string theory it is necessary to compactify six dimensions into a small curled up space. There are estimated to be many thousands of ways to do this. Each one predicts different particle physics. With the *Heterotic string* it is possible to get tantalisingly closed to the right number of particles and gauge groups, At the moment there are just too many possibilities and the problem is made more difficult because we do not know how the supersymmetry is broken.

All this makes string theory look less promising. Some physicists called it a *theory of nothing* and advocated a more conservative approach to particle physics tied more closely to experimental results. But a large number of physicists have persisted. There is something about superstring theory which is very persuasive.

Why String Theory?

The most common question from lay-people about string theory is *Why?*. To understand why physicists study string theory rather than theories of surfaces or other objects we have to go back to its origins. In 1968 physicists were trying to understand the nature of the strong nuclear interactions which held the quarks together in nucleons. There was an idea about duality between scattering interactions which led Veneziano and Virasoro to suggest exact forms for the dual resonance amplitude. These amplitudes turned out to have interesting properties in 26 dimensions and various independent lines of research by Nambu, Nielson and Susskind led to the revelation that the amplitudes were derivable from a theory of strings.

String theory was considered as a theory of strong interactions for some time. Physicists thought that the explanation for confinement of quarks was that they were somehow bound together by strings. Eventually this theory gave way to another theory called Quantum Chromo Dynamics which explained the strong nuclear interaction in terms of colour charge on gluons.

String theory suffered from certain inconsistencies apart from its dependence on 26 dimensions of space-time. It also had *Tachyonic modes* which destabilised the vacuum. But string theory had already cast its spell on a small group of physicists who felt there must be something more to it. Ramond, Neveu and Schwarz looked for other forms of string theory and found one with fermions in place of bosons. The new theory in 10 dimensions was supersymmetric and, magically, the tachyon modes were removed.

But what was the interpretation of this new model? Scherk and Schwarz found that at low energies the strings would appear as particles. Only at very high energies would these particles be revealed as loops of string. The strings could vibrate in an infinite tower of quantised modes in an ever increasing range of mass, spin and charge. The lowest modes could correspond to all the known particles. Better still, the spin two modes would behave like gravitons. The theory was necessarily a unified theory of all interactions including quantum gravity. Still only a small group pursued this idea until the historic paper of Green and Schwarz with the discovery of almost miraculous anomaly cancellations in one particular theory.

To come back to the original question, *why string theory?*. The answer is simply that it has the right mathematical properties to be able to reduce to theories of point particles at low energies, while being a perturbatively finite theory which includes gravity. The simple fact is that there are no other known theories which accomplish so much. Of course physicists have studied the mathematics of vibrating membranes in any number of dimensions. The fact is that there are only a certain number of possibilities to try and only the known string theories work out right in perturbation theory.

Ed Witten

Of course it is possible that there are other completely different self-consistent theories but they would lack the important perturbative form of string theories. The fact is that string theorists are now turning to membrane theories, or *p-brane theories* as they are known, where *p* is the number of dimensions of the membrane. Harvey, Duff and others have found equations for certain p-branes which suggest that self-consistent field theories of this type might exist, even if they do not have a perturbative form.

Dualities

In the past couple of years there have been some new developments which have inspired a revival of interest in string theory. The first of these concerns duality between electric and magnetic monopoles.

Maxwell's equations for electromagnetic waves in free space are symmetric between electric and magnetic fields. A changing magnetic field generates an electric field and a changing electric field generates an magnetic one. The equations are the same in each case, apart from a sign change which is irrelevant here. However, it is an experimental fact that there are no magnetic monopole charges in nature which mirror the electric charge of electrons and other particles. Despite some quite careful experiments only dipole magnetic fields which are generated by circulating electric charges have ever been observed.

In classical electrodynamics there is no inconsistency in a theory which places both magnetic and electric monopoles together. In quantum electrodynamics this is not so easy. To quantise Maxwell's equations it is necessary to introduce a vector potential field from which the electric and magnetic fields are derived by differentiation. This procedure can not be done in a way which is symmetric between the electric and magnetic fields.

40 years ago Paul Dirac was not convinced that this ruled out the existence of magnetic monopoles. He always professed that he was motivated by mathematical beauty in physics. He tried to formulate a theory in which the gauge potential could be singular along a string joining two magnetic charges in such a way that the singularity could be displaced through gauge transformations and must therefore be considered physically inconsequential. The theory was not quite complete but it did have one saving grace. It provided a tidy explanation for why electric charges must be quantised as multiples of a unit of electric charge.

In the 1970's it was realised by 't Hooft and Polyakov that grand unified theories which might unify the strong and electro-weak forces would get around the problem of the singular gauge potential because they had a more general gauge structure. In fact these theories would predict the existence of magnetic monopoles. Even their classical formulation could contain these particles which would form out of the matter fields as topological solitons.

There is a simple model which gives an intuitive idea of what a topological soliton is. Imagine first a straight wire pulled tight like a washing line with many clothes pegs strung along it. Imagine that the clothes pegs are free to rotate about the axis of the line but that each one is attached to its neighbours by elastic bands on the free ends. If you turn up one peg it will pull those nearby up with it. When it is let go it will swing back like a pendulum but the energy will be carried away by waves which travel down the line. The angle of the pegs approximate a field along the one dimensional line. The equation for the dynamics of this field is known as the sine-Gordon equation. It is a pun on the Klien-Gordon equation which is the correct linear equation for a scalar field and which is the first order approximation to the sine-Gordon equation for small amplitude waves. If the sine-Gordon equation is quantised it will be found to be a description of interacting scalar fields in one dimension.

The interesting behaviour of this system appears when some of the pegs are swung through a large angle of 360 degrees over the top of the line. If you grab one peg and swing it over in this way you would create two twists in the opposite sense around the line. These twists are quite stable and can be made to travel up and down the line. A twist can only be made to disappear in a collision with a twist in the opposite direction.

These twists are examples of topological solitons. They can be regarded as being like particles and antiparticles but they exist in the classical physics system and are apparently quite different from the scalar particles of the quantum theory. In fact the solitons also exist in the quantum theory but they can only be understood non-perturbatively. So the quantised sine-Gordon equation has two types of particle which are quite different.

What makes this equation so remarkable is that there is a non-local transformation of the field which turns it into another one dimensional equation known as the Thirring model. The transformation maps the soliton particles of the sine-Gordon equation onto the ordinary quantum excitations of the Thirring model, so the two types of particle are not so different after all. We say that there is a duality between the two models, the sine-Gordon and the Thirring. They have different equations but they are really the same.

The relevance of this is that the magnetic monopoles predicted in GUT's are also topological solitons, though the configuration in three dimensional space is more difficult to visualise than for the one dimensional of the clothes line. Wouldn't it be nice if there was a similar duality between electric and magnetic charges as the one discovered for the sine-Gordon equation? If there was then a duality between electric and magnetic fields would be demonstrated. It would not quite be a perfect symmetry because we know that magnetic monopoles must be very heavy if they exist.

In 1977 Olive and Montenen conjectured that this kind of duality could exists, but the mathematics of field theories in 3 space dimensions is much more difficult than that of one dimension and it seems beyond hope that such a duality transformation can be constructed. But they made one step further forward. They showed that the duality could only exist in a supersymmetric version of a GUT. This is quite tantalising given the increasing interest in supersymmetric GUT's which are now considered more promising than the ordinary variety of GUT's for a whole host of reasons.

Until 1994 most physicists thought that there was no good reason to believe that there was anything to the Olive-Montenen conjecture. Then Seiberg and Witten made a fantastic breakthrough. By means of a special set of equations they demonstrated that a certain supersymmetric field theory did indeed exhibit electro-magnetic duality. As a bonus their method can be used to solve many unsolved problems in topology and physics.

Now at last we turn to string theory with the realisation that duality in string theory is very natural. In the last year physicists have discovered how to apply tests of duality to different string and p-brane theories in various dimensions. A series of conjectures have been made and tested. This does not prove that the duality is correct but each time a test has had the potential to show an inconsistency it has failed to destroy the conjectures. What makes this discovery so useful is that the dualities are a non-perturbative feature of string theory. Now many physicists see that p-brane theories can be as interesting as string theories in a non-perturbative setting. The latest result in this effort is the discovery that all four string theories which are known to be perturbatively finite are now thought to be derivable from a single theory in 11 dimensions known as M-theory. M-theory is a hypothetical quantum field theory which describes 2-branes and 5-branes related through a duality.

It would be wrong to say that very much of this is understood yet. There is still nothing like a correct formulation of M-theory or p-brane theories in their full quantum form, but there is new hope because now it is seen that all the different theories can be seen as part of one unique theory.

Black Strings

As if one major conceptual breakthrough was not enough, string theorists have been coming to terms with another which turned up last year. Just as physicists have been quietly speculating about electro-magnetic duality for decades, a few have also speculated that somehow elementary particles could be the same things as black holes so that matter could be regarded as a feature of the geometry of space-time. The idea actually goes back at least as far as Riemann.

The theory started to look a little less ridiculous when Hawking postulated that black holes actually emit particles. The process could be likened to a very massive particle decaying. If a black hole were to radiate long enough it would eventually lose so much energy that its mass would reduce to the Planck scale. This is still much heavier than any elementary particle we know but quantum effects would be so overwhelming on such a black hole that it would be difficult to see how it might be distinguished from an extremely unstable and massive particle in its final explosion.

To make such an idea concrete requires a full theory of quantum gravity and since string theory claims to be just that it seems a natural step to compare string states and black holes. We know that strings can have an infinite number of states of ever increasing spin, mass and charge. Likewise a black hole, according to the *no hair conjecture* is also characterised only by its spin, mass and charge. It is therefore quite plausible that there is a complementarity between string states and black hole states, and in fact this hypothesis is quite consistent with all tests which have been applied. It is not something which can be established with certainty simply because there is not a suitable definition of string theory to prove the identity. Nevertheless, many physicists now consider it reasonable to regard black holes as being single string states which are continually decaying to lower states through Hawking radiation.

The recent breakthrough due to Strominger, Greene and Morrison is the discovery that if you consider Planck mass black holes in the context of string theory then it is possible for space-time to undergo a smooth transition from one topology to another. This means that many of the possible topologies of the curled up dimensions are connected and may pave a way to a solution of the selection of vacuum states in string theory.

String Symmetry

Superstring theory is full of symmetries. There are gauge symmetries, supersymmetries, covariance, dualities, conformal symmetries and many more. But superstring theory is supposed to be a unified theory which should mean that its symmetries are unified. In the perturbative formulation of string theory that we have, the symmetries are not unified.

One thing about string theory which was discovered very early on was that at high temperatures it would undergo a phase transition. The temperature at which this happens is known as the Hagedorn temperature after a paper written by Hagedorn back in 1968, but it was in the 1980's that physicists such as Witten and Gross explored the significance of this for string theory.

The Hagedorn temperature of superstring theory is almost very high, such temperatures would only have existed during the first 10^{-43} seconds of the universe existence, if indeed it is meaningful to talk about time in such situations at all.

Calculations suggest that certain features of string theory simplify above this temperature. The implication seems to be that a huge symmetry is restored. This symmetry would be broken or hidden at lower temperatures, presumably leaving the known symmetries as residuals.

The problem then is to understand what this symmetry is. If it was known then it might be possible to figure out what string theory is really all about and answer all the puzzling questions it poses. This is the superstring mystery.

A favourite theory is that superstring theory is described by a *topological quantum field theory* above the Hagedorn temperature. TQFT is a special quantum field theory which has the same number of degrees of gauge symmetry as it has fields, consequently it is possible to transform away all field variables except those which depend on the topology of space-time. Quantum gravity in 2+1 dimensional space-time is a TQFT and is sufficiently simple to solve, but in the real world of 3+1 dimensional Einstein Gravity this is not the case, or so it would seem.

But TQFT in itself is not enough to solve the superstring mystery. If space-time topology change is a reality then there must be more to it than that.

Most physicists working in string theory believe that a radical change of viewpoint is needed to understand it. At the moment we seem to be faced with the same kind of strange contradictions that physicists faced exactly 100 years ago over electromagnetism. That mystery was finally resolved by Einstein when he dissolved the ether. To solve string theory it may be necessary to dissolve space-time altogether.

In string theory as we understand it now, space-time curls up and changes dimension. A fundamental minimum length scale is introduced, below which all measurement is possible. It will probably be necessary to revise our understanding of space-time to appreciate what this means.

Even the relation between quantum mechanics and classical theory seems to need revision. String theory may explain why quantum mechanics works according to some string theorists.

All together there seem to be rather a lot of radical steps to be made and they may need to be put together into one leap in the dark.

Those who work at quantum gravity coming from the side of relativity rather than particle physics see things differently. They believe that it is essential to stay faithful to the principles of diffeomorphism invariance from general relativity rather than working relative to a fixed background metric as string theorists do. They do not regard renormalisability as an essential feature of quantum gravity.

Working from this direction they have developed a canonical theory of quantum gravity which is also incomplete. It is a theory of loops, tantalisingly similar in certain ways to string theory, yet different. Relativists such as Lee Smolin hope that there is a way to bridge the gap and develop a unified method.

Related Reading

Space-Time Topology Change in String Theory, Aspinwall, Greene, Morrison

Unity of Superstring Dualities, C. Hull

Classical Quantum Duality, Michael Duff

What is String Theory?, Joseph Polchinski

String Theory Dynamics in Various Dimensions, Edward Witten

The Power of M Theory, John Schwarz

Black Hole Condensation and the Duality in String Theory, Andrew Strominger

World Volumes an String Target Spaces, Michael Green

7. *Event-Symmetric Space-Time*

Quantum Gravity

The scientific revolution of the twentieth century has constructed two major pillars; General Relativity and Quantum Theory, on which all known physics is built. Through general relativity we have come to understand the force of gravity, while through quantum field theory we understand the other three known forces in particle physics. Physicists are now trying to complete the final step, to put together these two theories into one theory of quantum gravity.

To do so is proving to be a difficult task because General Relativity and Quantum Theory seem to be utterly incompatible if we try to put them together in the most obvious ways. It is now generally believed that a full theory of quantum gravity will not be possible without a radical revision in the way we think of space-time.

Symmetry in String Theories

In 1981 a new theory of quantum gravity known as superstring theory was born. According to this idea, all fundamental particles are different resonance modes of tiny loops of string. Many theorists believe that superstring theory is going to turn out to be a *Theory Of Everything*, by which they mean a complete unified theory of the fundamental laws of physics to which all other physical laws can be reduced. This hope is rather frustrated by the fact that they don't have a completely self consistent and elegant set of equations to define it. All they *do* have are some perturbative calculation schemes and classical limits.

One of the discoveries which have enabled physicists to get so far towards a complete understanding, is the importance of the role played by symmetry in physics. In general relativity it is a form of symmetry known as *diffeomorphism invariance* which is important, while in particle physics it is *internal gauge symmetry*.

If symmetry is so important in both general relativity and quantum theory then we should surely expect it to be of at least as much importance in a combined theory of quantum gravity. In superstring theory we do indeed find all kinds of symmetries. There are clues in the mathematics of string theory which seem to suggest that at a very high temperature, known as the Hagedorn Temperature, a huge order symmetry manifests itself. This could be the universal symmetry which includes all known symmetries unified together. The problem is that we don't know what that symmetry is. If we did then we might understand superstring theory much better.

Topology Change

Back in 1958 John Wheeler suggested that when general relativity and quantum theory were put together there would be astonishing things going on at the very small length scale known as the Planck length (about 10^{-35} metres). If we could look down to such distances we would see the topology of space-time changing through quantum fluctuations. The structure of this space-time foam has been a mysterious area of research ever since.

Witten's Puzzle

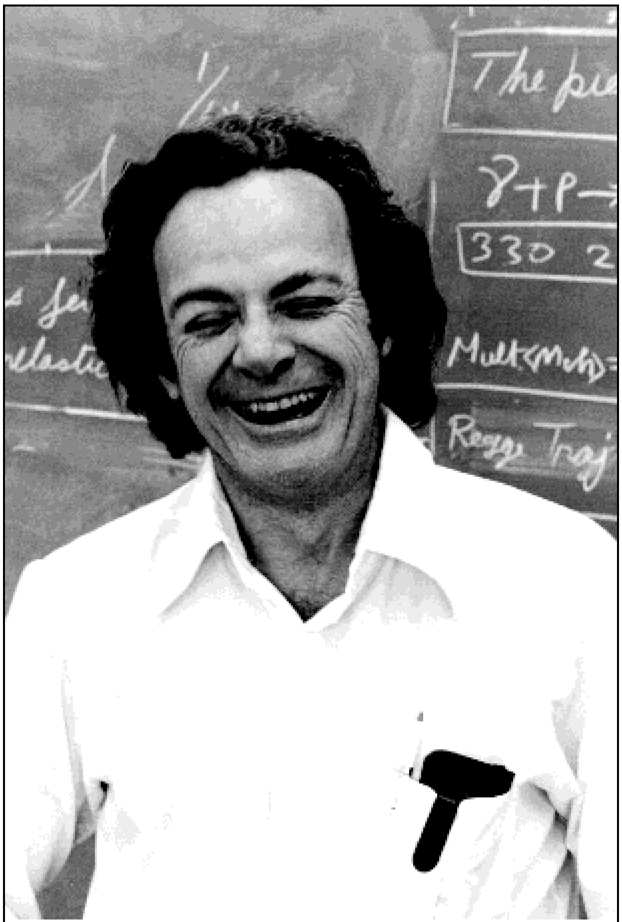
Topology change is found to be an important part of superstring theory, so again string theorists seem to be on the right track. But, when they try to understand together the concepts of topology change and universal symmetry they come up against a strange enigma known as Witten's Puzzle after the much cited string theorist, Ed Witten, who first ran across it.

The difficulty is that both diffeomorphism invariance and internal gauge symmetry are strictly dependent on the topology of the space-time on which they act. Different topologies lead to non-equivalent symmetries. If topology change is permitted then it follows that the universal symmetry must, in some fashion, contain the symmetry structures for all allowable topologies at the same time. Witten admitted he could think of no reasonable solution to this problem.

An old maxim of theoretical physics says that once you have ruled out reasonable solutions you must resort to unreasonable ones. As it happens there is one unreasonable but simple solution to Witten's puzzle.

Consider diffeomorphisms to begin with. A diffeomorphism is a suitably smooth one to one mapping of a space-time onto itself. The set of all such mappings form a group under composition which is the diffeomorphism group of space-time. A group is an algebraic realisation of symmetry. One group which contains all possible diffeomorphism groups as a subgroup is the group of all one-to-one mappings irrespective of how smooth or continuous they are. This group is known as the symmetric group on the manifold. Unlike the diffeomorphism groups, the symmetric groups on two topologically different space-times are algebraically identical. A solution of Witten's puzzle would therefore be for the universal group to contain the symmetric group acting on space-time events.

This is called *The Principle of Event Symmetric Space-Time* which states that **The universal symmetry of the laws of physics includes a mapping onto the symmetric group acting on space-time events**.



Richard Feynman

Hidden Symmetry

There are a number of reasons why this principle may seem unreasonable. For one thing it suggests that we must treat space-time as a discrete set of events. In fact there are plenty of reasons to believe in discrete space-time. Theorists working on quantum gravity in various forms agree that the Planck scale defines a minimum length beyond which the Heisenberg uncertainty principle makes measurement impossible. In addition, arguments based on black hole thermodynamics suggest that there must be a finite number of physical degrees of freedom in a region of space.

A more direct reason to doubt the principle would be that there is no visible or experimental evidence of such a symmetry. The principle suggests that the world should look the same after permutations of space-time events. It should even be possible to swap events from the past with those of the future without consequence. This does not seem to accord with experience.

To counter this it is only necessary to point out that symmetries in physics can be spontaneously broken. The most famous example of this is the Higgs Mechanism which is believed to be responsible for breaking the symmetry between the electromagnetic and weak interactions. Could space-time symmetry be hidden through an analogous mechanism of spontaneous symmetry breaking?

There are, in fact, other ways in which symmetries can be hidden without being broken. Since the symmetric group acting on space-time can be regarded as a discrete extension of the diffeomorphism group in general relativity, it is worth noting that the diffeomorphism invariance is not all that evident either. If it were then we would expect to be able to distort space-time in ways reminiscent of the most bizarre hall of mirrors without consequence. Everything around us would behave like it is made of liquid rubber. Instead we find that only a small part of the symmetry which includes rigid translations and rotations is directly observed on human scales. The rubbery nature of space-time is more noticeable on cosmological scales where space-time can be distorted in quite counterintuitive ways.

If space-time is event-symmetric then we must account for space-time topology as it is observed. Topology is becoming more and more important in fundamental physics. Theories of magnetic monopoles, for example, are heavily dependent on the topological structure of space-time. To solve this problem is the greatest challenge for the event-symmetric theory.

The Soap Film Analogy

To get a more intuitive idea of how the event-symmetry of space-time can be hidden we use an analogy. Anyone who has read popular articles on the Big Bang and the expanding universe will be familiar with the analogy in which space-time is compared to the surface of an expanding balloon. The analogy is not perfect since it suggests that curved space-time is embedded in some higher dimensional flat space, when in fact, the mathematical formulation of curvature avoids the need for such a thing. Nevertheless, the analogy is useful so long as you are aware of its limitations.

We can extend the balloon analogy by imagining that space-time events are like a discrete set of particles populating some higher dimensional space. The particles might float around like a gas of molecules interacting through some kind of forces. In any gas model with just one type of molecule the forces between any two molecules will take the same form dependent on the distance between them and their relative orientations. Such a system is therefore invariant under permutations of molecules. In other words, it has the same symmetric group invariance as that postulated in the principle of event-symmetric space-time.

Given this analogy we can use what we know about the behaviour of gases and liquids to gain a heuristic understanding of event symmetric space-time. For one thing we know that gases can condense into liquids and liquids can freeze into solids. Once frozen, the molecules stay fixed relative to their neighbours and form rigid objects. In a solid the symmetry among the forces still exists but because the molecules are held within a small place the symmetry is hidden.

Another less common form of matter gives an even better picture. If the forces between molecules are just right then a liquid can form thin films or bubbles. This is familiar to us whenever we see soap suds. A soap film takes a form very similar to the balloon which served as our analogy of space-time for the expanding universe.

More Symmetry

So far we have seen how the principle of event-symmetric space-time allows us to retain space-time symmetry in the face of topology change. Beyond that we would like to find a way to incorporate internal gauge symmetry into the picture too. It turns out that there is an easy way to embed the symmetric group into matrix groups. This is interesting because, as it happens, matrix models are already studied as simple models of string theory. String theorists do not normally interpret them as models on event-symmetric space-time but it would be reasonable to do so in the light of what has been said here.

In these models the total symmetry of the system is a group of rotation matrices in some high dimensional space. The number of dimensions corresponds to the total number of space-time events in the universe, which may be infinite. Permutations of events now corresponds to rotations in this space which swap over the axes.

So does this mean that the universal symmetry of physics is an infinite dimensional orthogonal matrix? The answer is probably *no* since an orthogonal matrix is too simple to account for the structure of the laws of physics. For example, orthogonal groups do not include supersymmetry which is important in superstring theories. The true universal symmetry may well be some much more elaborate structure which is not yet known to mathematicians.

Identical Particles

Theorists often talk about unifying the gauge symmetries which are important to our understanding of the four natural forces. There are, however, other symmetries in nature which are rarely mentioned in the context of unification. These symmetries take the form of an invariance under exchange of identical particles. For example, every electron in the universe is the same, they all have the same charge, mass etc. If we swap one electron in the universe with another the universe will carry on as before.

The symmetry involved here is described by the symmetric groups, just like event-symmetric space-time. Obviously we should ask ourselves whether or not there is any connection between the two. Could the symmetric group acting to exchange identical particles be part of the symmetric group acting on space-time events? If it were, then that would suggest a deep relation between space-time and matter. It would take the process of unification beyond the forces of nature towards a more complete unification of matter and space-time.

Mach's Space and Time

At this point it is worth taking a little philosophical diversion. Philosophers and scientists have often discussed what each believe to be the true nature of space and time. It seems that the way we perceive space and time is rather indirect. You can mark notches which measure distance on a ruler but you can't mark notches on space itself. Likewise, time is measured by the ticking of clocks which are assumed to be related to some passage of absolute time of the universe. Our minds have evolved a process of placing objects into a model of space and time in our heads but to what extent is this model an accurate model of the real world?

In the 17th century Leibnitz was of the opinion that space-time was some sort of illusion. Newton took a more pragmatic approach and decided to take it for granted that space and time exist as part of the arena within which the laws of physics play out their roles. This dichotomy between the preferred philosophy and the progress of science has existed ever since with philosophers such as Berkeley, Kant and Mach insisting that space and time should not be absolute despite the success of physical theories based on the contrary assumption. Ultimately, they argued, space and time would be understood in terms of relationships among physical objects which we observe more directly. Some philosophers go further and claim that physical objects themselves are illusions which can only be understood as artefacts of our own perceptions.

Most ironic of all was the part played by Einstein in the development of this debate. He was much impressed by the work of Mach and hoped that Mach's principle would follow from his theory of general relativity. Minkowski had already shown that Einstein's theory of special relativity could be interpreted as a unification of space and time into a single geometry of space-time. Space and time were not absolute if treated independently. It turned out that the correct formulation of Einstein's ideas resulting from his equivalence principle needed the mathematics of curved space-time as developed by Gauss, Riemann and others in the previous century. The result was that space-time seemed to take on a more absolute form than before. It was no longer a static background, but could itself move dynamically and have a complicated existence even in the absence of any matter at all.

The mathematical elegance of general relativity, as much as its empirical success, has led to its strong influence over the way physicists have formed their theories. They seem to be more and more contrary to the philosophical view. These days we are used to trying to find ways to describe matter as some kind of manifestation of geometry such as in Kaluza-Klein theories where matter is seen as a result of higher dimensions of space-time.

It could simply be that the philosophers are wrong. Many physicists of the mid-twentieth century, including Dirac, Feynman and Weinberg, have been openly disdainful of the ways of philosophers. More recently a slight change in the wind might be felt. As theoretical physicists themselves have moved further away from the influence of experimental data, they have themselves become a lot more philosophical. If the absoluteness of space-time is to breakdown it will probably be at the scales of quantum gravity that it is found to happen. Already we see the influence of non-commutative geometry in which space-time takes a secondary role to the matter fields defined on it as functions. Many physicists now express a wish to find a more algebraic theory for quantum gravity. In former years Einstein had been alone in expressing such wishes.

Clifford's Legacy

On its own, the principle of event symmetric space-time is not very fruitful. What is needed is a mathematical model which incorporates the principle and which gives body to some of the speculative ideas outlined above.

It turns out that such a model can be constructed using Clifford algebras. These algebras are very simple in principle but have a remarkable number of applications in theoretical physics, They first appeared to physicists in Dirac's relativistic equation of the electron. They also turn out to be a useful way to represent the algebra of fermionic annihilation and creation operators.

If we regard a Clifford algebra as an algebra which can create and annihilate fermions at space-time events then we find we have defined a system which is event-symmetric. It can be regarded as an algebraic description of a quantum gas of fermions.

This is too simple to provide a good model of space-time but there is more. Clifford algebras also turn out to be important in construction of supersymmetries and if we take advantage of this observation we might be able to find a more interesting supersymmetric model.

Back to Superstrings

Since superstring theory was an important part of the motivation for proposing the principle of event-symmetric space-time in the first place. String theorists seem to believe that the subject they are studying is already the correct theory of physics, but they are probably missing the key to understanding its most natural formulation.

The situation seems to parallel Maxwell's theory of electromagnetism as it was seen at the end of the 19th century. Many physicists did not accept the validity of the theory at that time. This was largely because of the apparent need for a medium of propagation for light known as the ether, but experiment had failed to detect it. Einstein's theory of special relativity showed why the ether was not needed. He did not have to change the equations to correct the theory. Instead he introduced a radical change in the way space and time were viewed.

It is likely that the equations we have for string theory are also correct, although they are not as well formed as Maxwell's were. To complete the theory it is again necessary to revise our concept of space-time and remove some of its unnecessary structure just as Einstein removed the ether.

It would be natural to search for an event-symmetric string model. We might try to generalise the fermion model described by Clifford algebras to something which was like a gas of strings. A string could be just a sequence of space-time events connected in a loop. The most significant outcome of the event-symmetric program so far is the discovery of an algebra which does just that. It is an algebraic model which can be interpreted as an algebra of strings made of closed loops of fermionic partons.

The result is not sophisticated enough to explain all the rich mathematical structures in string theory but it may be a step towards that goal. Physicists have found that new ideas about knot theory and deformed algebras are important in string theory and also in the canonical approach to quantisation of gravity. This has inspired some physicists to seek deeper connections between them. Through a turn of fate it appears that certain knot relation have a clear resemblance to the relations which define the discrete event-symmetric string algebras. This suggest that there is a generalisation of those algebras which represents strings of anyonic partons, that is to say, particles with fractional statistics.

Event-Symmetric Physics

What can this theory tell us about the universe? Since it is incomplete it is limited. The one place where a theory of quantum gravity would have most significance would be at the big bang. In the first jiffy of existence the temperature was so high that the structure of space-time would have been disrupted. It is known that in string theory there is a high temperature phase transition in which the full symmetry is realised. If the principle of event-symmetric space-time is correct then that is a much larger symmetry than people have previously imagined. At such high temperature space-time would cease to exist in the form we would know it, and only a gas of interacting strings would be left. A reasonable interpretation of this state of affairs would be to say that space-time has evaporated. The universe started from such a state, then space-time condensed and the rest is history.

Related Reading

The Space-Time Manifold as a Critical Solid, Peter Orland

Space-Time Transitions (Witten's Puzzle), Edward Witten

Event-Symmetric Physics, Phil Gibbs

8. *Is String Theory in Knots?*

Surely Not Knots

When I was a mathematics student at Cambridge back in 1980, I remember going to one of John Conway's popular lectures which he gave to the mathematics clubs. This one was about knot theory. Conway performed a series of tricks with bits of rope to demonstrate various properties of knots. A fundamental unsolved problem in knot theory, he told us, is to discover an algorithm which can tell when a loop of string is a knot or not.

It is possible to tie up closed loops of string into complicated tangles which can nevertheless be untied without cutting the string. But suppose I gave you a tangled loop of string. How could you determine if it could be untied?

Conway showed us a clever trick with groups which enabled him to determine that some knotted loops could not be untied, but there were others which were not classified in this way. Conway had generalised a polynomial invariant of knots first discovered by Alexandria many years ago. The Conway Polynomial was quite a powerful tool to distinguish some knots from others, but it could not separate all.

I remember thinking at the time that this was a piece of pure maths which would never have any useful applications apart from providing a way of proving that your boat can't slip its moorings, perhaps. Mathematicians delight in these kind of problems.

Ten years later a dramatic change had taken place. Knot theory now looked like it was going to have applications to solving quantum gravity and probably other problems in condensed matter theory. Louis Kauffman had even written a substantial book called *Knots and Physics* (World Scientific). Conway's Knot Polynomial had been generalised and the problem of classifying knots seemed all but solved.

I am not going to try to review all those things which have been found out recently about knots. Instead I refer you to Kauffman's book and John Baez: *This weeks finds in Mathematical Physics* which is where I learnt just about all of what little I know on the subject.

To summarise, I will list just a few points of interest here

- Knot theory is important in understanding the physics of particles with fractional statistics: anyons or parafermions. These particles, which can exist in one or two dimensions have properties between fermions and bosons. The symmetric group is the symmetry of fermions and bosons, while the braid group from knot theory plays the same role for anyons.
- Knot theory is important in canonically quantised quantum gravity. Where knotted loop states provide a basis of solutions to the quantum gravity equations. This is described in the important loop representation of quantum gravity.
- Knot theory is closely related to quantum groups. These are a generalisation (or deformation) of classical Lie groups and are important in condensed matter theory, string theory and other physics. Knot theory seems to be very closely related to symmetry.
- Quantum groups are also used to construct Topological Quantum Field theories which can be used to find invariants of manifolds.

From this point on things are going to get more technical and I am going to assume that the reader knows some maths.

From the Symmetric Group to the Braid Group

The principle of event symmetric space-time states that the universal symmetry of physics must have a homomorphism onto the symmetric group acting on space-time events. Now the symmetric group can be defined by the following relations among the transposition generators a_1, a_3, a_3,...

$$a_i a_j a_i = a_j a_i a_j$$
$$a_i^2 = 1$$

The braid group is defined in the same way but with only the former relation. Put into words, this means that the braid group describes a symmetry where it does not matter in which order you exchange things but if you exchange two things then exchange them again you don't necessarily get back to what you had before.

There is a homomorphism from the braid group onto the symmetric group generated by the second relation. This means that the braid group is also a candidate for part of the universal symmetry according to the principle of event-symmetric space-time. In that case space-time events would behave like particles with fractional statistics.

Discrete String Theory

Now I will turn to another question. Are strings discrete? In string theory as we currently know it there is not much indication that string theory is discrete. Strings are described as continuous loops in space. However, there has been some interesting work by Susskind and others which does seem to suggest that string theory could be discrete. It may be possible to describe strings as objects made of small partons strung together. These partons would not exist as hard objects but can be conceptually subdivided and rejoined. They are points on the string which describe the topology of its interactions.

If the partons can be subdivided then they must be permitted to have fractional statistics. They must live on the string world sheet. The statistics of a whole loop of string would be the sum of the fractional statistics of its partons and would be an integer or half integer so that the string can live in three dimensional space.

If space-time is event-symmetric and we wish to consider event-symmetric string field theory, then a discrete string approach is essential. The partons of the string can be tied to the events through which the string passes. It will be permitted to pass through space-time events in any order it likes. In this way strings can tie together the events of space-time and provide an origin of topology in an otherwise unstructured event-symmetric universe.

If strings are formed from loops of partons with fractional statistics then it seems natural to allow them to be knotted. We should look for ways of describing this algebraically in an event-symmetric string theory.

String theorists are now also turning to higher dimensional membrane theories. If strings can be made of partons then surfaces, or 2-branes, can be made from strings. The process could continue ad infinitum. Space-time itself might be viewed as a membrane built in this way. There may be structures of all dimensions in physics. The 2 dimensional string world sheets and 3 dimensional space-time are more visible only because they stand out as a consequence of some as yet unknown quirk in the maths.

Event-Symmetric Open String Theory

As a first step I have constructed new types of super-symmetry inspired by the concept of discrete event-symmetric strings. The first of these is for open strings.

Imagine space-time as a large number N of discrete events which are arbitrarily numbered $1, 2, \dots,$ N. An open ended string will be defined simply by the sequence of events it passes through. An example would be

$$A = 15213$$

A general string of length 4 might be written

$$B = abcd$$

a, b, c, d are variables for the events the string passes through.

Strings can be any finite length starting from the null string of length zero which will be denoted by a pair of empty parentheses, *()*.

These strings are taken as the defining basis of a vector space.

I define multiplication of strings by joining them together and summing over all possibilities where identical events are cancelled. e.g., using a dot for the product

$$543.2 = 5432$$
$$1234.4351 = 12344351 + 123351 + 1251$$

The null string acts as an identity and it can be checked that the multiplication is associative. What I have defined then, is an infinite dimensional unital associative algebra.

The associativity is not entirely trivial. To check that,

$$(A.B).C = A.(B.C)$$

there are two main cases to consider which can be represented like this

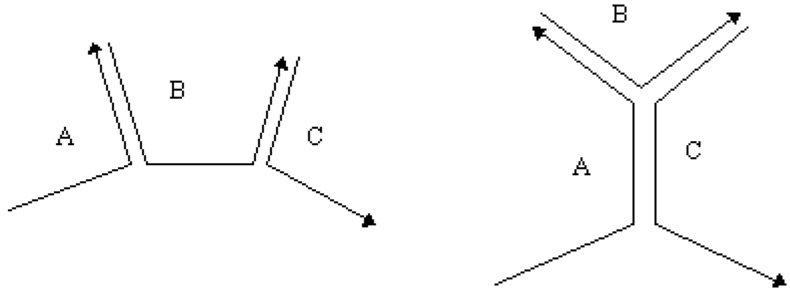


fig 8.1

Here *V*, *W*, *X*, *Y*, *Z* are arbitrary bits of strings which are concatenated together to form the three complete strings.

Given any associative algebra it is possible to define the Lie algebra by defining the Lie product as the anticommutator of the algebra.

$$[A,B] = A.B - B.A$$

The Jacobi identity follows directly from the associativity.

The Lie algebra can be regarded as the generators of the symmetry of the discrete open string theory. The way I have defined it is inspired by a description of symmetry algebras for topological strings due to Michio Kaku for continuous strings.

A benefit of the discrete string version is that it is easy to go from the bosonic discrete open string to the supersymmetric version. The associative algebra is graded by parity of the length of strings. I.e. the product of two even length strings or two odd length strings is a sum of even length strings, while the product of an odd and an even is odd. It follows that a Lie superalgebra can be defined using the graded commutator.

$$[A,B]_{+-} = A.B +- B.A$$

where the + sign is chosen if both A and B have odd parity, and the - sign is chosen otherwise.

This describes a rather simple sort of string theory which does not do very much except have super-symmetry. The interpretation is that these are open strings made of discrete *fernionic* or *bosonic* partons at space-time events. The model is event-symmetric in sense that the order in which the events are numbered is irrelevant.

Event-Symmetric Closed String Theory

Can we do the same thing with discrete closed strings? Kaku had attempted this with his topological formulation of universal string theory so why not?

What is needed is a Lie superalgebra defined on a basis of closed discrete cycles. It actually took me quite a lot of investigation before I discovered the correct way to do this. I will spare you the details of how I found it and go directly to the result even though this will make it look like I have just pulled the answer out of a hat. As we shall see, the properties of the closed discrete string superalgebra are much more promising than for the open version.

Start with a set *E* of *N* events. Write sequences of events in the same way as for the open strings.

A = abcdef, a, b, .. are from E

To introduce closed loops we define permutations on these sequences. The permutation can be shown as arrows going from each event to another (or itself). An example would look like this,

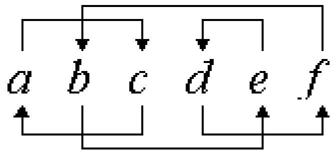


fig 8.2

The permutation is composed of cycles. In the example there are two cycles, one of length 2 and one of length 4. But the order of the events across the page is also important.

As before these objects form the basis of a vector space. An associative algebra is defined on these objects by simply taking multiplication to be concatenation of two of these objects together. The empty sequence is a unit for this algebra.

A more interesting algebra is formed by factoring out a set of relationships among these elements. The relations are defined in the following diagram.

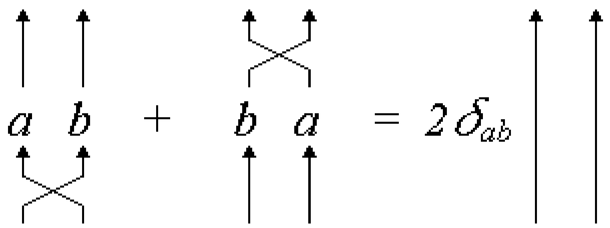


fig 8.3

This says that the order of two events can be interchanged keeping the loop connections intact. The sign is reversed and if the two events are the same an extra reduced term must be included. To get a complete relation the ends of the string in these diagrams must be connected to something. If they are just joined together the following two equations can be formed,

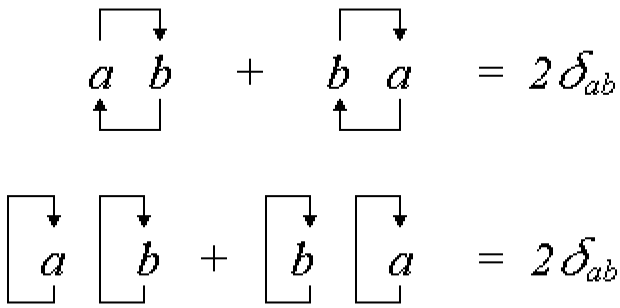


fig 8.4

The first shows the cyclic relationships for a loop of two events. The second is the anti-commutation relation for two loops of single events.

Since the relationship can be used to order the events as we wish, it is possible to reduce every thing to a canonical basis which is a product of ordered loops. A more convenient notation without the connections shown is then introduced.

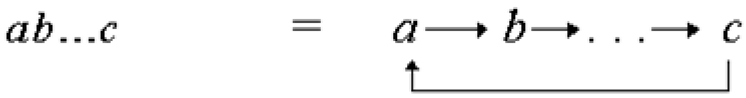


fig 8.5

This notation allows the relations to be written in a way similar to those of the open strings, but now the cyclic relations mean that they must be interpreted as closed loops.

The algebra is again graded by parity and so generates an interesting supersymmetry. As far as I know this infinite dimensional supersymmetry has never been studied by mathematicians. It is possible that it can be reduced to something well known but until this is demonstrated I will assume that it is original and interesting.

Here are a few important properties of the discrete closed string algebra which did not apply to the open string algebra.

- closed strings which do not have any events in common commute or anticommute. This is important because it can be interpreted to mean that strings only interact when they touch.
- the algebra contains a subalgebra isomorphic to a Clifford algebra. It also has a homomorphism onto a Clifford algebra which is defined by stripping out the string connections. This is important because the algebra of creation and annihilation operators for fermions is also isomorphic to a Clifford algebra. This justifies the interpretation of this algebra as a model of discrete closed strings made from fermionic partons.
- The length two strings generate an orthogonal group acting on the vector space spanning events. The symmetric group permuting events exists as a subgroup of this. It follows that the symmetry of event-symmetric space-time is included in this supersymmetry.

A String made of anyons?

It is almost certainly incorrect to model strings as loops of fermions. They must have some continuous form. To achieve this in an event-symmetric framework it will be necessary to replace the fermions with partons having fractional statistics which can be divided, i.e. anyons.

Defining creation and annihilation operators for anyons is not a simple matter. Various schemes have been proposed but none seem ideal. However, here we have the advantage that our anyons are strung together. The statistics and symmetries of anyons must be described by knot theory.

The commutation relations used to generate the closed string algebra will remind anyone who knows about knot polynomials of Skein relations. This suggests a generalisation may be possible if the string connections are replaced by knotted cords which can be tied. These could be subject to the familiar Skein relations which define the HOMFLY polynomial.

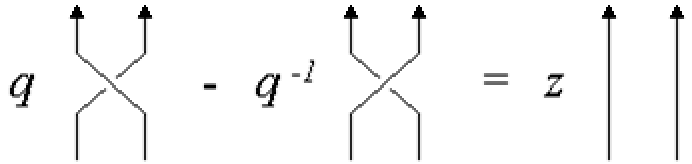


fig 8.6

In the special case where $q=1$ and $z=0$ this relation says that string can pass through itself. This is what we have for the strings which join the fermions. The crucial question is, are there generalisations of the parton commutation relations which are consistent with the general Skein relation? If there are then I have not found them yet.

The ladder of dimensions

Louis Crane, working towards algebraic quantum gravity with category theory, has found a way to construct a ladder of algebras of increasing dimension. In string theory there is evidence that membranes and space-times of various different dimensions play important roles. It is possible to go down the scale of dimensions by compactifying space-times. From M-theory in 11 dimensions or F-theory in 12 dimensions it is possible to construct the important critical string theories in 10 dimensions.

I shall now demonstrate a simple dimensional ladder construction which generalises the discrete fermion string symmetry. This construction may explain why structures of so many different dimensions are important in string theory.

The fermionic operators which are strung together in the discrete string model form a Heisenberg Lie superalgebra when the strings are removed. The universal enveloping algebra of this is then a Clifford algebra. I would like to repeat the string construction starting from a general Lie superalgebra. To keep things simple I will begin with just an ordinary Lie algebra ***A***.

As before, the elements of the Lie-algebra can be strung together on strings but this time the commutation relations will look like this,

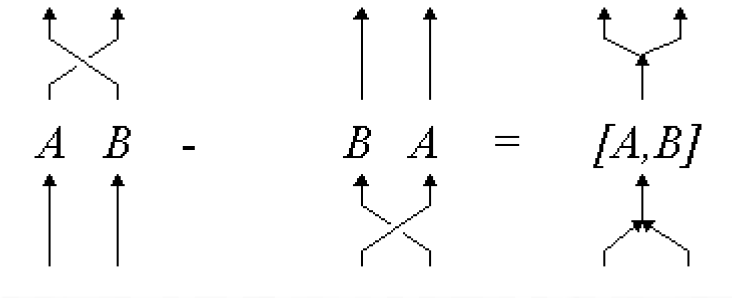


fig 8.7

I have introduced the possibility that the strings can join and separate. This kind of thing is best understood in the context of category theory. The Lie algebra elements are connected to themselves by a *functor* which can be shown as a network of strings which join and separate like this,

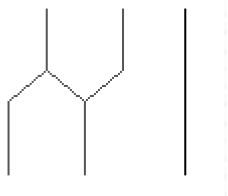


fig 8.8

The commutation relations of fig 8.7 can be shown to be consistent with the Jacobi relations provided the functors satisfy the following associativity relationship,

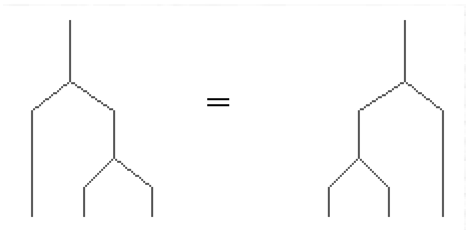


fig 8.9

and also the similar coassociativity relationship upside down.

In this way we can take out Lie algebra ***A*** and generate a new Lie algebra *L(**A**)*. The process can be generalised to a Lie superalgebra. In the case where ***A*** is a Heisenberg superalgebra there is a homomorphism from *L(**A**)* onto the discrete string algebra which I defined previously. So this process can be regarded as a generalisation.

The interesting thing to do now is look at what happens if we apply the *L* ladder operator to the string algebra. This can be visualised by circling the discrete strings around the network so that they are replaced with tubes. The interpretation is that we generate a supersymmetry algebra as string world sheets. The ladder operator can be applied as many times as desired to generate higher dimensional symmetry algebras. Furthermore. There is always a homomorphism from *L(**A**)* back onto ***A***. This makes it possible to apply the ladder operator an infinite number of times to generate a single algebra which contains all the previous ones.

This last observation raises some interesting mathematical puzzles. The algebra formed by applying the ladder operator an infinite number of times will have the property that it is isomorphic to the algebra formed by applying the ladder operator to itself. It is certainly of interest to ask whether this situation actually arises after just a finite number of steps of the ladder. Would it be too daring to conjecture that the algebra becomes complete after only 26 steps in the ordinary Lie algebra case and 10 steps in the supersymmetric case?

To progress further it will be necessary to study more general categories like those defined by Skein relations. Mathematical physicists are looking at ways to construct n-categories by stepping up a ladder of dimensions. The symmetries I have described here could be a related to such structures. The hope is that a full theory of quantum gravity and string theory can be constructed algebraically in such a fashion.

Related Reading

Is quantum gravity algebraic?, Louis Crane

Symmetry in the Topological Phase of String Theory, Phil Gibbs

What can we learn from the study of non-perturbative quantum general relativity?, Lee Smolin

Strings. Loops, Knots and Gauge Fields, John Baez

Is String Theory in Knots?, Phil Gibbs

9. *Theory of Theories*

The Theory That Flies

As everybody knows, the job of a theoretical physicist is to invent theories of the universe. A layman might ask a physicist *What is charge?* or *What is Time?*. He will be disappointed when the physicist replies that his theories do not even try to explain what these things are. Theories are just mathematical models which make predictions about how they will behave in experiments.

When pressed the physicist will probably admit that he does physics because he too seeks deeper explanations for why things are the way they are in the universe. One day he hopes to understand the most basic laws of physics and he hopes that they will provide an answer to the most difficult question of all, *Why do we exist?*. Physicists can be justly proud of the fact that almost everything in physics can be accounted for with just a small number of basic equations embracing General Relativity and the Standard Model of particle physics. There remain many puzzles but those will probably be solved once a unified theory of quantum gravity and the other forces is found. Such a theory would be the final theory. It may be cast in other forms but they would always be exactly equivalent. There is no *a priory* reason why such a theory should exist but, as Steven Weinberg argues in *Dreams of a Final Theory*, the convergence of principles in modern physics seems to suggest that it does.

How many physicists have not wondered what principle of simplicity and beauty underlies that final theory? Could we not take an intellectual leap and figure it out from what we already know? John Wheeler wrote his thoughts thus, ... *Paper in white the floor of the room, and rule it off in one-foot squares. Down on one's hands and knees, write in the first square a set of equations conceived as able to govern the physics of the universe. Think more overnight. Next day put a better set of equations into square two. Invite one's most respected colleagues to contribute to other squares. At the end of these labours, one has worked oneself out into the door way. Stand up, look back on all those equations, some perhaps more hopeful than others, raise one's finger commandingly, and give the order* ***Fly!*** *Not one of those equations will put on wings, take off, or fly. Yet the universe* ***flies****.*

The Nature of Nature

If there is really a unique principle on which the laws of physics are founded then to understand it we should look for clues in the nature of nature, or as Feynman puts it *the character of physical law*. One thing is clear: Nature uses mathematics. If this were not the case, if nature was governed instead by a committee of demons who made nature follow their whims, then there would be little hope for us to understand physics and predict the outcome of experiments or invent new technology. Scientists would be replaced by sorcerers.

But the relationship between physics and mathematics seems to be much deeper than we yet understand. In Early history there was little distinction between a mathematician and a physicist but in modern times pure mathematicians have explored their subject independently of any potential application. Mathematics has an existence of its own. Those mathematicians have constructed a huge web of logical structures which have a remarkable inner beauty only visible to those who take the time to learn and explore it. They would usually say that they discovered new mathematics rather than invented it. It is almost certain that another intelligent race on another planet, or even in a different universe, would have mathematicians who discover the same theorems with just different notation.

What becomes so surprising is the extent to which mathematical structures are applicable to physics. Sometimes a physicist will discover a useful mathematical concept only to be told by mathematicians that they have been studying it for some time and can help out with a long list of useful theorems. Such was the case when Heisenberg formulated a theory of quantum mechanics which used matrix operations previously unfamiliar to physicists. Other examples abound, Einstein's application of non-Euclidean geometry to gravitation and, in particle physics, the extensive use of the classification of Lie groups. Recently the mathematical theory of knots has found a place in theories of quantum gravity. Before that, mathematicians had considered it an area of pure mathematics without application (except to tying up boats of course). Now the role played by knots in fundamental physics seems so important that we might even guess that the reason space has three dimensions is that it is the only number of dimensions within which you can tie knots in strings. Such is the extent to which mathematics is used in physics that physicists find new theories by looking for beautiful mathematics rather than by trying to fit functions to empirical data as you might expect. Dirac explained that it was this way that he found his famous equation for the electron. The laws of physics seem to share the mathematician's taste for what is beautiful. It is a deep mystery as to why this should be the case. It is what Wigner called *The unreasonable effectiveness of mathematics in the natural sciences*.

It has also been noted by Feynman that physical law seems to take on just such a form that it can be reformulated in several different ways. Quantum mechanics can be formulated in terms of Heisenberg's matrix mechanics, Schroedinger's wave mechanics or Feynman's path integrals. All three are mathematically equivalent but very different. It is impossible to say that one is more correct than the others.

Perhaps there is a unique principle which determines the laws of physics and which explains why there is such a tight relationship between mathematics and physics? If the laws of physics were merely some isolated piece of mathematics chosen for its simple beauty then there would be no explanation why so much of mathematics is incorporated into physics. There is no reason why one set of equations should **fly**. The fundamental principle of physics must be something more general. Something which embraces all of mathematics. It is the principle which explains the nature of nature. So what is it?

Many Anthropic Principles

There are other aspects of the universe which provide clues as to what principle determines the laws of physics. The universe is populated by an impressive menagerie of objects which exhibit organised complexity. They exist on all length scales from the atomic to the cosmological. Most impressive of all (that we know of) are living beings like ourselves.

Examination of the way that chemistry, nuclear physics, astrophysics, cosmology and other sciences are dependent on the details of the laws of physics suggests that the existence of so much complexity is no accident. The precise values of various constants of nature, such as the fine structure constant, seem to be just right to allow organised complexity to develop. Perhaps we might even say, to allow life to develop.

This observation has inspired much faith among physicists and philosophers in the *Anthropic Principle*. The anthropic principle supposes that the laws of physics are indeed selected so that intelligent life has a maximum chance of developing in the universe. Believers ask us to consider first why our planet Earth is so well suited to the evolution of life while other planets in the solar system seem to be more hostile. The answer is that we would not be on this planet to consider the question if it were not suitable for life to evolve here. The same principle can then be extended to the whole universe.

One way to understand the Anthropic Principle is to imagine that all possible universes exist with a validity which is equal to our own. When we say all possible universes we might mean any system which can be described by mathematics. Each such system has a set of physical laws which allow its structure to be determined in principle. Sometimes they will be simple and beautiful and often they will be complex and ugly. Sometimes the phenomenology of such a system will be dull or easily determined and nothing interesting will happen. Sometimes it will be so complicated that nothing can be determined, even a hypothetical computer simulation would reveal little of interest in the turmoil of those universes. Somewhere in between would exist our universe which has just the right balance of equations in its physical laws for intelligent life to exist and explore the nature of its environment.

Another interpretation of the anthropic principle, developed by Lee Smolin, is that there is one universe with a set of physical laws much as we know them. Those laws may have a number of variables which determine the physical *constants* but which can vary in certain extreme situations such as the collapse of massive stars into black holes. Universes governed by such laws might give birth to baby universes with different physical constants. Through a process of natural selection universes might evolve over many generations to have constants which are conducive to further procreation. This might mean they are optimised for the production of black holes and, from them, more baby universes. Within this population of worlds there will be some with laws conducive to life, indeed, the production of black holes may be linked to the existence of advanced life-forms which could have an interest in fabricating black holes as energy sources. This scenario makes a number of demands on the nature of physical laws. In particular it is essential that some physical parameters such as the fine structure constants should be able to vary rather than being determined by some equation. Future theories of quantum gravity may tell us if this is so.

Is the Anthropic Principle Enough?

The Anthropic Principle is compelling enough for us to wonder if it can determine the laws of physics on its own. I know of no convincing argument that it can. There is nothing in the anthropic principle which explains why so many of the most elegant discoveries of mathematics are so important in physics. There is nothing to explain why there is so much symmetry in physics, or why the elegant principle of least action is important or even why the laws of physics should be the same in one place as they are in another.

You might try to argue that the laws of physics have to take a certain form because otherwise they would be impossible to understand. I don't buy it! I am convinced that a suitable mathematical system, perhaps even something as simple as a cellular automata, can include sufficient complexity that intelligent life would evolve within it. There must be a huge variety of possible forms the laws of physics could have taken and there must be many in which life evolves. In the case of cellular automata, the cellular physicists living in it would probably be able to work out the rules of the automata because its discrete nature, and simple symmetry would be clear and easily uncovered. They would not need to know so much sophisticated mathematics as we do.

Whether or not the principle is valid as an explanation for some of the characteristics of nature and the values of its parameters I believe that there must be some other principle which explains those other things.

Universality

For centuries mathematicians confined themselves to looking at specific structures with simple definitions and interesting behaviour. With the advent of powerful computers they are now looking at general behaviour of complex systems. It was Feigenbaum who made the discovery that a large class of complex systems of chaotic non-linear equations exhibit a universal behaviour characterised by the Feigenbaum constants. This type of universality has an independent existence which transcends details of the specific equations which generate it. Other examples of universality can be identified in physics and mathematics. Statistical physics looks at the behaviour of systems with many degrees of freedom. Such systems exhibit a universal behaviour which can be described by the laws of thermodynamics. The microscopic details of the forces between particles are reduced to just a few macroscopic parameters which describe the thermodynamic characteristics.

A more mathematical example is the notion of computability. Computability of a sequence of integers can be defined in terms of a hypothetical programming language such as a Turing machine or a Minsky machine. Those languages and a large number of other possibilities turn out to give an equivalent definition of computability despite the fact that they look very different. There is no most natural or most simple way to define computability but computability itself is a natural and unambiguously defined concept. If we made contact with an alien intelligence we would probably find that they had an equivalent concept of computability but probably not quite the same definitions. Computability, then, can be seen as a universal characteristic of computing languages.

The message I wish to draw from this is that the laws of physics may themselves be a universal behaviour of some general class of systems. If this is the case then we should not expect the laws of physics to be given by one most natural formulation. Like computability there may be many ways to describe them. The universal behaviour of a class of complex systems would be likely to display organised complexity itself. Furthermore, there is evidence that thermodynamics runs deeper than just a behaviour of particle systems. It is also found to be a useful description of black hole dynamics. We can also remark that quantum mechanics and statistical physics are closely related through an exchange of real and imaginary time. All these things are intimately related and hint at the importance of universality in nature at its most fundamental level.

The Theory of Theories

At last we come to the main hypothesis of this essay. If the laws of physics are to be seen as a universal behaviour of some class of systems then it is necessary to ask what class to choose. We can regard any possible mathematical system as a theory of physics. I suggest that the laws of physics are a universal behaviour to be found in the class of all possible mathematical systems. This is known as *The Theory of Theories*

To understand the Theory of Theories we start from the same premise as we do with the anthropic principle, i.e. that all mathematically consistent models exist just as our own universe exists. We can simply take this to be our definition of existence.

We know from Feynman's Path Integral formulation of quantum mechanics that the evolution of the universe can be understood as a supposition of all possible histories that it can follow classically. The expectation values of observables are dominated by a small subset of possibilities whose contributions are reinforced by constructive interference. The same principle is at work in statistical physics where a vast state space is dominated by contributions at maximum entropy leading to thermodynamic behaviour. We might well ask if the same can be applied to mathematical systems in general to reveal the laws of physics as a universal behaviour which dominates the space of all possible theories and which transcends details of the construction of individual theories. If this was the case then we would expect the most fundamental laws of physics to have many independent formulations with no one of them standing out as the simplest. This might be able to explain why such a large subset of mathematics is so important in physics.

Can we use the Theory of all Theories to explain why symmetry is so important in physics? There is a partial answer to this question which derives from an understanding of critical behaviour in statistical physics. Consider a lattice approximation to a Yang-Mills quantum field theory in the Euclidean sector. The Wilson discretisation preserves a discrete form of the gauge symmetry but destroys the space-time rotational symmetry. If we had more carelessly picked a discretisation scheme we would expect to break all the symmetry. We can imagine a space of discrete theories around the Yang-Mills theory for which symmetry is lost at almost all points. The symmetric continuum theory exists at a critical point in this space. As the critical point is approached correlation lengths grow and details of the discretisation are lost. Symmetries are perfectly restored in the limit, and details of all the different discretisations are washed out.

If this is the case then it seems that the critical point is surrounded by a very high density of points in the space of theories. This is exactly what we would expect if universal behaviour dominating in theory space was to exhibit high symmetry. It also suggests that a dominant theory could be reformulated in many equivalent ways without any one particular formulation being evidently more fundamentally correct than another. Perhaps ultimately there is an explanation for the unreasonable effectiveness of mathematics in physics contained in this philosophy.

If physics springs in such a fashion from all of mathematics then it seems likely that discovery of these laws will answer many old mathematical puzzles. There is no *a priori* reason to believe that mathematical theories should have some universal behaviour, but if they did it might explain why there is so much cross-reference in mathematics. Perhaps mathematicians sense intuitively when they are near the hot spots in the space of theories. They notice the hightened beauty, the multitude of unexpected connections. Eventually, left to their own devices mathematicians might be capable of finding the central source of the heat, if physicists don't get there first.

I think therefore I am...

So, is it really possible to derive the laws of physics from pure mathematics without any reference to empirical observations? If the Theory of Theories is correct then the answer should be *yes*. At first it seems that it is rather hard to make progress with the theory of theories beyond the philosophical conception, since it is necessary to define an appropriate topology and measure in the space of all mathematical theories. Mathematics is just too large for this, or is it?

Perhaps we could search for a universal behaviour in the set of all possible computer programs. The set is sufficiently diverse to cover all mathematics because, in principle, we can write a computer program to explore any mathematical problem. John Wheeler proposed this as a place to start and called it *It From Bit*. Simple computer programs can be very complex to understand, but we are not interested in understanding the details of any one. We are concerned about the universal behaviour of very big programs randomly written in some (any) computer language.

The variables of a large program would evolve in some kind of statistical manner. Perhaps the details would fade into the background and the whole could be understood using the methods of statistical physics. So let's look at general systems of statistical physics. Suppose one system (one theory, one universe) had a number *N* of variables, its degrees of freedom.

$$a_1, a_2, \dots a_N$$

In addition there must be an energy function,

$$E(a_1, a_2, \dots a_N)$$

In the system, a possible set of values for these variables would appear with a weight given by

$$Z = e^{-E(a_1, a_2, \dots a_N)}$$

I have not said much about the values of these variables. They could be discrete variables or real numbers, or points on a higher dimensional manifold. Somewhere in this complete set of systems you could find something close to any mathematical universe you thought of. For example, cellular automata would exist as limiting cases where the energy function forced discrete variables to follow rules.

But what did I mean when I said *close*? Two different systems would be *isomorphic* if there was a one to one mapping between them which mapped the weight function of one onto the weight function of the other. We could define a distance between two systems by finding the function mapping one to the other which minimised the correlations between them. This defines a metric space with the minimum correlation as metric.

A powerful property of metric spaces is that they can be *completed* by forming Cauchy sequences. Hence we can define a larger set of theories as the completed metric space of statistical systems. By means of this technique we include even renormalisable lattice gauge theories into the theory space. The renormalisation process can be defined as a Cauchy sequence of finite statistical systems. It remains to define a natural measure on this space and determine if it has a universal point where the total measure within any small radius of this point is larger than the measure on the rest of the space.

Needless to say, this is quite a difficult mathematical problem and I am not going to solve it. Perhaps I didn't really get much further than Descartes!

The Story Tellers Philosophy

On the off chance that someone has read this far, or skipped to the end, I will conclude with a lighter description of the theory of theories.

Suppose I were to tell you a story. It might be a story about wars between elves and goblins or a story about men exploring space or a story about creatures living in a two dimensional world or perhaps just the story of my life. A story defines a universe just as a mathematical theory does, but a story is incomplete. The same story might be played out in many different universes, or many times in one very large universe. When you listen to the story you add more details in your imagination. You picture the characters. Perhaps you see the hero wearing a blue coat which was not in the original tale.

Our universe is very large and will exist for a long time. It may even be infinite. There are probably many stories being told. Perhaps some are so close that they can be seen as accidentally telling different parts of the same story. If there are indeed many possible universes then stories will overlap sufficiently that complete universes can be built up from them. Details can be made consistent. The story of every possible universe will be told in the others. We could just as well regard ourselves as living in a story. There is no way to distinguish a real universe from a fictional one if all the details are told somewhere.

If a story teller ends his tale, there will always be another who continues from that point on. Indeed, there will be many who will all continue in different ways so we should not expect the future to be determined. But overall there would be some order to what was told. Some stories might get told more often than others. For each time a story is told where something too ridiculous and unexpected happens, there will be many others for which the story follows a more reasonable thread. Your story may never be told exactly but it can be pieced together many times over from those that do get told.

Related Reading

Theory of Theories Approach to String Theory, Hiroyuki Hata

Possible Implications of the Quantum Theory of Gravity, Louis Crane

The Fate of Black Hole Singularities, Lee Smolin

Cosmology as a Problem in Critical Phenomena, Lee Smolin

Quantum Cosmology and the Constants of Nature, Alexander Vilenkin