

THE ROLE OF MATHEMATICS IN PHYSICAL SCIENCES

Jayant V. Narlikar

Abstract

This article highlights the important role mathematics has been playing in the development of physical sciences. Although its practitioners look upon mathematics as an art to be enjoyed for its aesthetic values, the subject has found applications in other disciplines. The physicists in particular have found mathematics indispensable to their search for the basic laws governing natural phenomena.

The inputs provided by mathematics to physics will be illustrated by examples drawn from gravity and electrodynamics, from classical as well as quantum physics.

PART I

Mathematics vs the Sciences

The scientist aims at finding the basic laws of nature, the rules that govern the various phenomena we see around us. The underlying assumption behind this aim is that there do exist such basic laws and that they are unique. Experience of several centuries tells us that the laws can be described by mathematical models. That is how maths enters the discussion. And, since the scientist judges the success or failure of his model from the extent of its ability to account for what is observed and to predict what should be observed, the survival and acceptability of a theory are decided by these pragmatic criteria.

Not so in mathematics. Each branch of mathematics has a basic set of postulates and logical rules of making deductions from them. The question of 'factuality' does not enter. Nor can one say that one branch of mathematics is better or more realistic than another. As in the world of sports different rules govern different games, so it is for maths. It is ludicrous to say that algebra is better than topology as it is to say that cricket is better than football. Of course a person may like cricket better than football just as a mathematician may prefer to work in algebra rather than topology. But these value judgements are aesthetic and subjective. You cannot give a logical and objective explanation of why you like Beethoven better than Mozart nor can you say why you enjoy one painting better than the other. There is no universality in these comparisons. It is in this sense that mathematics is an art. Like a painting or a piece of music it is created for aesthetic appreciation. Considerations of its relevance to reality let alone usefulness in life do not matter.

The scientist on the other hand can say purely objectively why Theory A is better than Theory B. The criterion is simply that the set of phenomena explicable by Theory B is a proper subset of the set of phenomena explicable by Theory A. Likewise a theory is no longer viable if there is a single observation decisively against it. With all its successes, Newton's law of gravitation was supplanted by another 'better' theory, the theory of relativity, when a few crucial tests conducted in the solar system went in favour of the latter.

Having outlined the basic difference between mathematics and science I will now illustrate how mathematics nevertheless remains an important component in the evolution of scientific thought. I will illustrate my ideas with an example from astronomy.

The mathematical inputs come at various levels :

1. *Pattern searching* : the 'raw' data may conceal a pattern that holds the key to some natural law. How do we extract the 'signal' from the 'noise'?
2. *Empirical law* : The signal so extracted will exhibit the pattern. But what is the quantitative description of that pattern?
3. *Basic physical theory* : Is there a theoretical model that predicts the empirical law observed in (2)?
4. *Global principle* : Does the theoretical framework of (3) fit into a global picture that also includes other theoretical frameworks describing different sets of phenomena?
5. *Predictive power* : To be viable as a scientific theory, it must make new testable and quantitative predictions.
6. *Design of tests* : Suitable 'hardware' needs to be designed to test the predictions in (5)

From Planets to the Law of Gravitation

The above inputs were present in the story of how and why planets move, a story that took two millennia to unfold, beginning with Aristotle (384-322 BC).

Aristotle was a pupil of the great Greek philosopher Plato and the teacher of another famous Greek personality, Alexander the Great. But Aristotle was a distinguished scholar in his own right, and his ideas on the nature of the universe were to dominate not only the Greek civilization, but also other civilizations in Europe and Asia. Indeed, in medieval Europe, Aristotle's ideas were so well established as to become part of religious dogma. The key to Aristotle's ideas lies in his classification of different types of motion.

Aristotle distinguished two types of motion seen in the universe : *natural motion*, which he supposed always to be in circles, and *violent motion*, which was a departure from circular motion, and implied the existence of a disturbing agency. Why circles? Because Aristotle was fascinated by a beautiful property of circles which no other curve seemed to possess. Take any portion of a circle (what we usually call a 'circular arc') and move it anywhere along the circumference : that portion will coincide exactly with the part of the circle underneath it. (The straight line also has this property, but it can be considered a circle of infinite radius.)

In the jargon of modern theoretical physics, the above property is one of *symmetry*. A one-dimensional creature moving along the circle will find all locations on it exactly similar, there being no privileged position. We find that present-day cosmologists employ similar symmetry arguments about the large-scale structure of the universe.

If you expose a photographic film for a long time to the light of stars in the night sky, you will develop pictures of circular trajectories, as stars rise in the east and set in the west. The sun and the moon also appear to follow circular paths in the sky. We thus see why a prima-facie case existed for Aristotle singling out circular trajectories as nature's favourites.

There was a fly in the ointment, however! A handful of heavenly bodies did not appear to follow circular tracks. Known as 'planets' (meaning 'wanderers' in Greek), these bodies appeared to defy Aristotle's edict of natural motion. Did the planets possess some special power which enabled them to wander at will?

Those who answered this question in the affirmative went the way of the astrologer, who argues that planets not only possess special powers but wield them on human destinies! It is ironical that even though the mystery of planetary motion was fully resolved by Kepler and Newton, and these celestial bodies were shown to be moving involuntarily under the force of gravitation, belief in astrology is still widespread today.

Greek astronomers who did not go the way of astrology nevertheless missed the chance of possibly discovering the law of gravitation. For, had they considered planetary motion as of the violent kind, they might have been led to wonder about the disturbing agency. Instead, they stuck to natural motion in circles, and to reconcile the manifest lack of circular motion, invented 'epicycles'.

The primary hypothesis behind the epicyclic theory was that the earth is at rest in the universe, and that all heavenly bodies go round it. This theory came to be known as the *geocentric* theory, since everything was viewed with the earth as the centre. The actual motion of bodies round the earth may be of two or more circles. Thus, in the simplest version a planet was considered to be moving in a circular path whose centre moved on another circular path around the fixed earth. If this description was found inadequate to represent planetary motion and to forecast accurately the position of a planet in the sky, more circles were added to the picture. Ptolemy carried the epicyclic theory to a high degree of sophistication in his classic book *Syntaxis*.

The epicyclic theory was thus no different from the kind of parameter-fitting exercise that goes on in modern times, when the resolution of apparent conflicts between observations and a favoured theory is sought by introducing adjustable parameters into the theoretical framework. Such an exercise tells us more about the freshly introduced parameters than it does about the basic hypothesis of the original theory. In fact, as happened with the Greek epicyclic theory, a theory requiring too much patchwork of this sort eventually has to be abandoned.

The first person to take issue with Greek astronomy was the Indian mathematician and astronomer Aryabhata. In his treatise entitled *Aryabhatiya* there is a verse which translates as follows: 'Just as a man moving in a boat downstream sees the trees on the river bank go in the opposite direction, so it is with stars which appear to move in the westerly direction when in fact they are fixed in space.' Aryabhata had thus discovered the fact of the earth's revolution about its polar axis as the cause of the rising and setting of fixed stars. So the mystery of Aristotle's natural motion in circles was resolved as far as stellar motions were concerned.

But so deep-rooted were Aristotelian ideas in India in the fifth to eighth centuries that distinguished scholars who venerated Aryabhata in other respects either ignored this argument or tried to reinterpret it in a way which did not conflict with the 'fixed earth' hypothesis.

A more decisive challenge to the Greek dogma came in the fifteenth and sixteenth centuries, first from Nicolaus Copernicus and then from Galileo Galilei. Copernicus challenged the geocentric theory *in toto*. Not only was the earth rotating about an axis, but it also travelled around the sun, which was fixed in space; or so argued Copernicus in his volume *De Revolutionibus Orbium Celestium*. In this *heliocentric theory* (in which, as its name implies, everything is viewed with the sun as the centre), epicycles are still needed, but the constructions are simpler than those given by Greek astronomers like Hipparchus and Ptolemy. Moreover, by arguing that the earth and the planets go round the sun, Copernicus paved the way for future theoreticians, who could look to the Sun for a possible cause of planetary motion.

Although Copernicus encountered considerable opposition to his revolutionary hypothesis, this was nothing compared to what Galileo had to suffer. Galileo not only supported the heliocentric theory; he also attacked the very basis of the geocentric theory - namely, the ideas of Aristotle.

When the Establishment could take no more of this, it reacted in the only way it knew. Galileo was subjected to the Inquisition, for arguing against the established tenets of religion. In the end, he recanted, but privately he retained his convictions about the correctness of the heliocentric theory and his objections to Aristotelian ideas to the end of his life.

It is interesting to note in this context that Tycho Brahe, one of the greatest observational astronomers of his time, shared the geocentric view of the Establishment. He had an excellent observatory at Uraniborg in Denmark, where he conducted observations to establish that the earth is at rest after all! Tycho later moved to Bohemia, where in 1601 he engaged a young assistant by the name of Johannes Kepler to work towards his goal of disproving the Copernican view through detailed observations. Tycho died in 1601, however, leaving behind all the wealth of his observations for Kepler to analyse. Kepler undertook the task in his characteristic meticulous manner, and after some twenty-five years of further observations and data analysis, came to the opposite conclusion, that Copernicus was right after all.

In fact, Kepler took the heliocentric theory considerably further than Copernicus and Galileo. He found, for example, that the planets moved round the Sun not in circles (or in circles on circles, and so on), but in ellipses. The elliptical orbits of all the planets had the Sun as a common focus. Kepler also discovered empirically the manner in which a planet moved along its orbit. These discoveries were summarized by Kepler in three laws of planetary motion.

Mathematically it is possible to describe the planetary motion in elliptical orbit as made up of a series of epicycles. In principle, an infinite number of epicycles is needed; but in practice, a small finite number gives a good approximation, because the orbits are nearly circular. Had the planetary orbits been considerably more elongated than they actually are, both Ptolemy and Copernicus would have found it hard to sustain the epicyclic construction.

Although Kepler had his own somewhat bizarre interpretation of why planets move according to the three laws, the correct answer came much later, from Isaac Newton. In 1687 Newton published his masterpiece *Philosophiae Naturalis Principia Mathematica* ('*Mathematical Principles of Natural Philosophy*') The *Principia*, published in three parts, may be regarded as the first book in theoretical physics as we know it today. Whereas Galileo's famous *Dialogue* had demolished the Aristotelian framework, the *Principia* filled the vacuum thereby created by presenting the laws of motion and gravitation. Newton was able to give a mathematical derivation of Kepler's three laws of planetary motion on the basis of his own laws of motion and gravitation.

Predictions by Newton's Law

We thus see how a mathematical description of planetary motion became possible. But as a scientific theory Newtonian laws needed to do more : they would need to make predictions of new phenomena and test them. It was fitting therefore that Edmund Halley, who had been

instrumental in getting the *Principia* published, should have applied Newton's laws to comets. Unlike planets, which are almost always seen, comets were noted as occasional visitors to the neighbourhood of the earth and the sun. It was Halley who pointed out that a comet describes a highly eccentric elliptical orbit around the Sun. Thus the same comet is seen at regular intervals when it is near the Sun, only to disappear as it moves away. Halley regarded sightings of a comet in 1456, 1531, 1607 and 1682 as visits of the same comet which, he predicted, would return in 1758. The comet did come as forecast (although Halley was not alive to see it). Known since then as Halley's comet, it has been coming near the Sun every 75-6 years, its most recent arrival having been in 1985-6.

It was the French mathematician Pierre Simon de Laplace who really pushed the application of Newtonian theory to the limits then possible. During the years 1799-1825 Laplace solved the complex problem involving the motion of the many planets and their satellites moving in the gravitational influence of one another and of the Sun. A sum total of *eighteen* bodies was involved in this calculation! Laplace's aim was to demonstrate that this complex system moved not at the dictates of a divine agency but according to the laws of dynamics and gravity. Even Newton had felt that the clockwork precision of planetary motion breaks down once in a while and needs to be set right by the Almighty. Laplace's calculations could, however, account for all the observed features of the planetary motion within the mathematical framework of the inverse square law.

Laplace's work gave stimulus to the subject now known as *celestial mechanics*. In general this subject deals with the motion of heavenly bodies moving under the law of gravity. Its main applications have been to the solar system where considerable astronomical data are available to compare with the theoretical predictions. Laplace and his successors evolved elaborate procedures to handle the complicated set of equations whose solutions often involved long series of mathematical terms. Nowadays these equations are readily solved on fast electronic computers.

A Success and a Failure of Newtonian Gravity

Let me now come to two astronomical observations, one a success for Newtonian celestial mechanics, the other a failure.

In 1781 the English astronomer William Herschel discovered the planet Uranus. At that time Uranus was the outermost known planet of our Solar System. However, by 1840 it was becoming noticeable that the orbit of Uranus was not strictly what would be expected on the basis of Newtonian celestial mechanics. The orbit showed departures from the theoretically calculated elliptical tracks.

Whenever there is a disagreement between theory and observation one of two possible deductions can be drawn : (1) the theory is wrong or (2) the observations are incorrect or incomplete. In the case of Uranus two

astronomers, J. C. Adams in England and U. J. J. Le Verrier in France placed their confidence in Newtonian theory and opted for the second deduction. Both independently looked for the cause of the discrepancy in terms of the gravitational pull of a new object in the vicinity of Uranus. Where should this object be located and what should be its estimated mass?

Adams got to the answer first in 1843 and he communicated his conclusions to two leading observers in England, Challis at the observatory in Cambridge and Airy the Astronomer Royal at Greenwich. Neither Airy, nor Challis took much notice of Adams's calculations and did nothing so far as instituting a telescopic search for the object was concerned. Meanwhile, in September 1846 Le Verrier also arrived at similar results which he communicated to J. G. Galle at the Berlin Observatory. Galle promptly looked for the object and found it. It is the planet Neptune, located even further away from the Sun than Uranus.

This episode, apart from illustrating the varying responses of observers to the predictions of theoreticians, is a good example of the second of the two alternatives described above, viz : a discrepancy between theory and observations may imply a new breakthrough on the observational front. We now consider an example of the first alternative. This relates not to a remote planet of the Solar System but to the planet situated closest to the Sun. The discrepancy in this case was not known during Newton's lifetime. Had it been known then it would have caused him a great deal of anxiety!

The planet nearest to the Sun is Mercury which takes nearly 88 days to make a complete round of the Sun. However, the planet does not seem to be following an ellipse exactly. Its track is nearly elliptical but at the same time it rotates round the Sun. This is most clearly seen if we look at the point on the orbit situated closest to the Sun. Such a point is called the *perihelion*. The line pointing from the Sun to the perihelion of Mercury seems to rotate slowly in the sense of the motion, from one orbit to the next. The actual rate of change is small, about 575" (arc second) per century. (An arc second is a measure of angle and it corresponds to 3600th of a degree.) The shift in the perihelion of Mercury is therefore hardly noticeable. Yet it was large enough to worry the theoreticians.

Like the disturbances of the orbit of Uranus, could this progressive shift of Mercury's orbit arise from the gravitational pull of other planets? Certainly the other planets of the Solar System do exert a pull on Mercury significant enough to cause a steady shift of its orbit. Calculations of celestial mechanics which took into account the gravitational influences of all other planets showed that more than 90 percent of the observed perihelion shift of Mercury could indeed be explained that way.

So out of the original 575" per century there still remained a small part, 43" per century, unexplained. This discrepancy was known at the time of Le Verrier who tried to account for it by a similar technique which led to the discovery of Neptune. Le Verrier postulated an innermost planet *Vulcan*, whose gravitational force could have caused the unexplained discrepancy in the motion of Mercury. This planet was looked for but not found, thus marking a failure for Newton's theory.

Enter Einstein

Conceptual difficulties rather than observational discrepancies led Albert Einstein (1879-1955) to cast a critical look at the Newtonian law of gravitation. The most bothersome aspect of the law of gravitation was its concept of *instantaneous action at a distance*. The Sun and the Earth attract each other, according to this law, by a force that acts across the vast distance separating them - a distance of some 90 million miles. Not only is this force acting across a great distance, it supposedly also acts *instantaneously*.

Such a situation is contrary to the *special theory of relativity* developed by Einstein in 1905. This theory places an upper limit on the speed with which an observable effect can propagate from one point in space to another. *This upper limit is represented by the speed of light*. Thus the motion of instantaneous propagation of gravitational pull across vast distances is inconsistent with the basic tenet of special relativity. This is why, having developed the theory of special relativity, Einstein was compelled to revise the law of gravitation.

Another difficulty related to the definition of inertial observers that are so crucial to special relativity. An inertial observer is one on whom no force acts. Can we actually pinpoint any observer or any physical object that is free from all forces? In our discussion of the first law of motion, we encountered friction as a force that retards motion. However, under idealized conditions, we can achieve situations where the friction is made very, very small. A man attempting to walk on an ice rink knows how difficult it is to proceed with very little friction. A projectile fired in a vacuum chamber encounters no resistance from air. In these cases, however, there is another force which has been ignored, the force of the Earth's gravity. Gravity acts on all material objects and can never be eliminated as a force. Even if we go far, far away from the Earth, we still have other objects in the universe to contend with.

In short, there is nowhere in the universe that we can go to eliminate gravity as a force, and so our definition of the inertial observer seems unrealizable in practice. Since the inertial observers form the starting point of special relativity, it looks as if the theory is based on unrealistic foundations. Thus, not only does special relativity make the Newtonian law of gravitation inconsistent, but its own foundations appear threatened by the existence of gravity.

Einstein discovered an ingenious way out of these difficulties by proposing an entirely new approach to the phenomenon of gravity. His theory of gravitation, called the *General Theory of Relativity*, is not a patchwork repair job on Newtonian gravitation and special relativity; rather, it is a radically new attempt at understanding the nature of gravity.

The general theory of relativity goes far beyond the special theory. Special relativity brought into physics the important notion that space and time together form a joint entity. The measurements of spatial distances and time intervals in the special theory are performed according to flat

space geometry. The notion of curvature of space-time and its relation to gravity is the remarkable new feature of the general theory.

In 1916, a few months after Einstein's equations were published, Karl Schwarzschild solved them in order to calculate the nature of curved space - time outside the Sun. Knowing the geometry of this space - time it is then a relatively simple matter to calculate the geodesics in it. *These are the world lines of planets.* In the space-time diagram the world line spirals up as time passes. Its projection in space gives an almost *elliptical* shape, showing the similarity between the predictions of Newtonian gravity and Einstein's general relativity. The adjective 'almost' needs further elucidation. The ellipse along which the planet moves according to Einstein's theory, slowly rotates so that the perihelion advances steadily. This effect is negligible for most planets, being noticeable only in the case of planet Mercury because its orbit is closest to the Sun and is the most eccentric. General relativity predicts that the perihelion of Mercury advances at an angular rate of 43" per century.

We now recall that this is precisely the effect which had remained unexplained in Newton's gravity. The planetary orbits calculated according to general relativity using the Schwarzschild solution differ from the orbits calculated by the Newtonian theory by small amounts. The biggest estimated difference between the two theories is for the planet Mercury and the observations favour general relativity. Indeed, the circumstance general relativity accounted almost exactly for this outstanding discrepancy went a long way towards establishing the credibility of this theory.

To conclude, the search for the right solution is a long and patient exercise to which *mathematics is indispensable*. In this connection, I may mention that frequently I get letters from people claiming to have disproved (or improved upon) Newton or Einstein. These people have vague qualitative ideas with no mathematical training. My standard (polite) reply to such aspirants for scientific glory is :

"Dear Mr. X,

Because of my shortcomings I am not able to appreciate ideas that are not founded on a solid mathematical framework. I therefore cannot comment on your work.

Your etc. etc."

PART II

New Maths for New Physical Ideas

As our observing techniques and experimental ability improve, the underlying physical concepts become more sophisticated, demanding new mathematical techniques to express them. I wish to illustrate this development with an example from electrodynamics. This also enables me to demonstrate how a global principle operates in physics.

Let me start with the Lagrange's equations in the usual format

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_r} \right) - \frac{\partial L}{\partial q_r} = 0 \quad (1)$$

for a set of dynamical variables $q_1 \dots q_N$, say. When first learning Lagrange's equations one feels that it is an unnecessary complication for mechanics. Why not use Newton's laws of motion

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i \quad (2)$$

instead, for masses m_1, \dots, m_n forming the system?

The Lagrange equations are in fact the first step in a process of globalization of physics. Through these equations we see how a unifying thread binds such diverse disciplines as mechanics and the electromagnetic theory. For, the q_r 's pinpoint the essential degrees of freedom which m_i, \mathbf{r}_i cannot. Moreover, we can generalize q_r 's to an uncountably infinite degrees of freedom. In particular, if we *define* q as a function of the three space coordinates x_μ ($\mu = 1, 2, 3$) as well as time, then we arrive at the notion of a field. Note that the Lagrange equations remain the same. We can make them covariant by writing $q_{,i}$ for $\partial q / \partial x_i$ and writing

$$\eta_{kl} \frac{\partial}{\partial x_k} \left(\frac{\partial L}{\partial q_{,l}} \right) - \frac{\partial L}{\partial q} = 0 \quad (3)$$

where η_{kl} is the Lorentz metric. (We use Latin indices to indicate values 0, 1, 2, 3, with 0 standing for timelike components. We also use the summation convention.)

What about L ? For the simplest free particle case in Newtonian mechanics we have

$$L = \frac{1}{2} m \dot{x}^2 \Rightarrow m \ddot{x} = 0 \quad (4)$$

The simplest field version of this is for a scalar field ϕ :

$$L = \frac{1}{2} \eta^{kl} \phi_{,k} \phi_{,l} \quad (5)$$

Using Lagrange's equations, we easily get

$$\eta^{kl} \phi_{,kl} = 0, \quad (6)$$

the scalar wave equation ! So the underlying similarity of structure tells us that the wave equation satisfied by ϕ is analogous in field theory to the unaccelerated motion of a free particle.

If we go from a scalar field to a vector field A^i ($i = 0, 1, 2, 3$) we similarly write

$$L = \frac{1}{2} \eta^{il} \eta^{km} A_{i,m} A_{k,l} \quad (7)$$

with slight rearrangements of terms this Lagrangian can be seen to lead to the Maxwell equations for sourcefree fields! Thus we have a similarity between two important branches of physics : mechanics and electromagnetic theory.

The next step lies in *deducing* Lagrange's equations from a Lagrangian. Here too, the Hamilton's principle

$$\delta \int L dt = 0, \quad (8)$$

which gives the Lagrange's equations in classical mechanics can be generalized to

$$\delta \int L d^4x = 0, \quad (9)$$

in classical field theory. This is commonly known as the principle of stationary action and has proved useful as a starting point for numerous developments in theoretical physics.

Electrodynamics of Action-at-a-Distance

I now come to another development in electrodynamics where mathematical techniques have evolved to satisfy the physicists' demand to describe a certain concept, namely the concept of action at a distance. It started as a set of two inverse square laws - one for the attraction or repulsion of electric charges and the other for the attraction or repulsion of magnetic poles. Both laws, known as Coulomb's laws, are similar in structure to Newton's law of gravitation, and they seemed to give a fair description of the observed phenomena.

However, in the nineteenth century, discrepancies began to appear, especially when the motion of rapidly moving electric charges was involved. There were several phenomena which these laws could not explain. In 1845 the famous theoretician, Gauss, in a letter to Weber (another well-known physicist of that time) gave his conjecture as to the cause of these failures. He thought that the trouble lay in the fact that, according to these laws, electric and magnetic interactions propagated instantaneously. He felt that the interactions should really propagate with some finite velocity - he even made the prophetic suggestion that this could be the velocity of light. However, this idea was not followed up either by Gauss or by others.

In the 1860s, James Clerk Maxwell resolved the problem in an entirely different way. He abandoned the concept of 'action at a distance' between electric charges (or magnetic poles) and instead proposed the existence of a new entity - the 'electromagnetic field'. According to him two charges a and b interact in the following way. Both are surrounded by the electromagnetic field. When a moves, it causes a disturbance in the field which propagates from a to b with the speed of light. This disturbance acts on b , and b begins to move. Thus the action of a on b is divided into three steps : the influence of a on the electromagnetic field, the propagation of the disturbance in the field, and the action of the electromagnetic field on b . These steps can be described mathematically (by a set of equations known as the Maxwell-Lorentz equations).

The Maxwellian concept was very successful in describing diverse electromagnetic phenomena. Since success is the sole criterion by which scientific theories are judged, it is not surprising that Maxwell's approach completely replaced the earlier action-at-a-distance description of electromagnetism given by Coulomb.

Somewhat belatedly (in the present century), action at a distance was revived along the lines suggested by Gauss. Important work by Schwarzschild, Tetrode, and Fokker led to the formulation of *delayed* action at a distance.

$$\sum m_a \int da - \sum_{a < b} \sum e_a e_b \int \int \delta (s^2_{AB}) da_i db_k \eta^{ik}. \quad (10)$$

The first term consists of the usual action for classical relativistic mechanics of free particles a, b, \dots of masses m_a, m_b, \dots . The second is the new input describing action at a distance between pairs of charges e_a, e_b . The delta function vanishes unless its argument s^2_{AB} vanishes, i. e., unless the invariant square of distance between points A and B vanishes. In special relativity this means that A and B interact only if light can go from A to B or vice versa. This is the mathematical formulation conjectured by Gauss. Let us visualize it further.

Here we have tracks of two charged particles a and b in the space-time diagram. Let A be a typical point on the track of the charge a . From A draw a dotted line towards b to denote the track of a light ray. This intersects b 's track at a point B , say. What this indicates is that the disturbance generated by the electric charge a at the space-time point A on its track is communicated to b not at once but at a later time denoted by B . Thus if a and b are at rest a light-hour apart and the disturbance from a leaves at 4 p. m., it will reach b at 5 p. m.

So far so good. But now a difficulty arises. Newton's third law of motion says that for every action there is an equal and opposite reaction. In the present picture it means that the action from A to B must be accompanied by a reaction from B to A , that is, *backward in time*. Hence if

a acts on *b* via a *retarded effect*, *b* must react back on *a* via an *advanced effect* - and vice versa. Thus advanced and retarded effects are on an equal footing in this theory. This seemed to be so much against the experience in everyday life that the theory was again abandoned for a considerable period, until it was resurrected by John Wheeler and Richard Feynman, towards the end of the Second World War.

The Wheeler-Feynman theory

To understand the Wheeler-Feynman approach, let us begin with a typical paradox that seems to arise when both advanced and retarded interactions are present in the universe. Suppose two observers *A* and *B*, a light - hour apart, set up a system of communication with the following rules. The observer *A* sends out a signal to *B* at 4 p. m. if and only if he does not hear from *B* at 4 p.m. The observer *B* sends out a signal to *A* at 5 p.m. if and only if he hears from *A* at 5 p.m. Does this system work? It would work in a world of retarded interactions only - but not in a world of both advanced and retarded interactions. For if *A* sends a signal at 4 p.m. *B* will receive it both at 3 p.m. (via advanced waves) and at 5 p.m. (via retarded waves). Thereupon *B* will send a signal at 5 p.m. which will reach *A* at 4 p.m. and 6 p.m. But this violates the condition for *A* to send a signal at 4 p.m. If, on the other hand, *A* does not send a signal at 4 p.m., *B* does not send a signal at 5 p.m., and so *A* received no signal from *B* at 4 p.m. - again violating the condition for *A* not sending a signal at 4 p.m.! Thus neither possibility is consistent.

Paradoxes like these are bound to arise if free communication between past and future is possible in the world. So a theory which permits situations like this must be abandoned.

-Wheeler and Feynman, however, pointed out that so far as electromagnetism is concerned the above paradox does not represent the correct state of affairs. Suppose *A* and *B* in the above example are electric charges. In the actual universe there are more than two charges. So when *A* and *B* set up their communication system they do not confine the electromagnetic signals to themselves alone. Take another charge *C* located *n* light - hours away from *A*. He gets *A*'s signal at $(4 + n)$ p.m. Because of this charge *C* moves and generates its own influence which travels backward as well as forward in time. The former will arrive at *A* *n* hours *earlier* than when it was emitted, that is, $4 + n - n = 4$ p.m. Note that this conclusion does not depend on *n*, that is *C*'s distance from *A*, no matter how far away *C* is located. We have already seen the vastness of the universe. Unless we take into account the responses of *all* electric charges in it, we cannot determine the net outcome of the signal sent out by *A*. This important point was made by Wheeler and Feynman, and they then proceeded to calculate the net response of the universe when an electric charge accelerates.

Wheeler and Feynman's work involves intricate mathematical details, but here we will try to get the gist of it in as simple terms as possible.

When an electric charge a is set in motion it generates, according to the action-at-a-distance theory, an equal mixture of retarded and advanced waves. Suppose we write this as the following combination :

$$F = \frac{1}{2}(\text{retarded}) + \frac{1}{2}(\text{advanced}). \quad (11)$$

Here (retarded) stands for the full disturbance generated if we accept the full retarded (outgoing wave) solution of Maxwell's equation. Thus we begin with complete symmetry between advanced and retarded waves. There is no prior commitment to the latter on the grounds of, say, the principle of causality.

The next step consists of setting a in motion and calculating the reaction from all other charges in the universe. This may be called the *response* of the rest of the universe to the motion of a . Assuming that the net result is the full (retarded) and using the same half-retarded-plus-half-advanced rule for all charges in the universe, Wheeler and Feynman calculated the response of the rest of the universe to the motion of a to be

$$R = \frac{1}{2}(\text{retarded}) - \frac{1}{2}(\text{advanced}). \quad (12)$$

When this response is added to the disturbance generated by a , the answer comes out to be

$$F + R = (\text{retarded}). \quad (13)$$

In other words the response of the universe neatly cancels the awkward part: the one which goes backwards in time. We shall refer to this as the '*correct response*' from the universe. The entire calculation is self-consistent with (13) *deduced* being consistent with (13) *assumed* in calculating it.

Wheeler and Feynman performed their calculation in a static universe. A static universe has no cosmological arrow of time. If we change the sign of the time coordinate, the static universe looks just the same as before. And this raises a paradoxical situation.

Suppose we do change the sign of the time coordinate in the Wheeler-Feynman argument. This leaves the universe unchanged, but it interchanges the 'advanced' and 'retarded' waves. This means that we have a new response from the universe which has exactly the opposite sign to that of the old response. This in turn results in the cancellation of the retarded half rather than the advanced half. The new solution obtained also being self-consistent, must have the same validity as the old one obtained before. How do we resolve this ambiguity? Moreover, we are still far from achieving our objective, that is, answering why the universe contains only retarded waves.

The Expanding Universe

In 1962 Jack Hogarth pointed out that the ambiguity which led Wheeler and Feynman to think in terms of thermodynamics need not have arisen if they had worked in an expanding model of the universe. In an expanding universe there is a cosmological time arrow. If we change the sign of the time coordinate we also change the universe from an expanding one to a contracting one - that is, to a *new* universe. Suppose we find that the response in an expanding universe cancels all advanced waves. It does not follow automatically from this that it will cancel all retarded waves in the other solution - simply because the transformation of time reversal (that is, change of sign of the time coordinate) is no longer permitted in an expanding universe.

Hogarth himself, and later other authors including Hoyle, Narlikar and Roe, examined specific cosmological models to see which of them yield the correct response. The general rule for obtaining a correct response from a given model of the universe is as follows. We divide the rest of the universe with respect to the present position of the electric charge by two light cones drawn with the charge at the apex in the future and the past. These are the tracks of light rays from a into the future and the past. The interior of the cones represent the parts of the universe accessible to physical disturbances arising from the motion of a . The cone and its interior in the future is the 'future absorber', while the cone and its interior in the past is the 'past absorber'. The rule for the correct response from a universe model is that the future absorber is *perfect* and the past absorber is *imperfect*. If the reverse is true, the response is of opposite sign, that is, it leads to the cancellation of retarded solutions. If both absorbers are perfect, as in the static universe, the result is ambiguous, and an appeal to another arrow of time, like the thermodynamic one, is necessary.

I will summarize the situation in the better - known cosmological models. It is clear that only the steady - state model gives the correct response. None of the big - bang models give the retarded solution in an unambiguous manner. To understand qualitatively the reason for this we have to take account of two effects present in an expanding cosmological model.

Take the case of the future absorber. All radiation emitted by a which travels in the future is red-shifted, that is, it gets diminished in frequency. So it is comparatively easy to absorb. However, in the ever - expanding big - bang models the density of matter in the future absorber decreases to zero so rapidly that there is not enough matter present in it to absorb even the low - frequency radiation. This results in the future absorber being imperfect. In the steady - state model, on the other hand, the density of matter is always the same and so there is enough matter in the future to absorb this radiation. Thus the future absorber in the steady - state model is perfect. In the case of the past absorber the situation is different. The radiation going into the past is blue - shifted, that is, it has

enormous frequency and energy. It cannot therefore be absorbed in the steady - state model, with its constant density of matter. In the big - bang models the density of matter in the past was very high - being infinite at the big - bang. Their past absorbers are therefore able to absorb even the highly energetic radiation coming from the future.

An ambiguous situation exists in the big - bang models which subsequently contract. Here infinite density in the past and the future ensures that both past and future absorbers are perfect.

This resolution of the problem of action at a distance could not, however, be looked at in isolation. If the idea can work in one part of physics can it not work in another? Or, to put it more strongly, a concept like this must hold for *all* physical interaction, if it has to acquire any credibility. So, for example, we must try to express gravity also in an action-at-a-distance format, thus reversing the historical development which made Coulomb base his inverse square laws for electricity and magnetism on the success of Newton's law for gravity.

Inertia and Gravity

In 1964 Sir Fred Hoyle and I put forward an action at a distance theory of gravitation. In this we assumed, as a starting point, that the inertia of any piece of matter owes its origin to the rest of the matter in the universe. This is the idea which had been suggested by Ernst Mach; but he did not give it a quantitative form. It is commonly known as *Mach's principle*. In our formulation such a form was given, and its consequences examined.

As in general relativity, we assumed that the geometry of space-time is non-Euclidean and that its character can be determined only with a full knowledge of all physical interactions in it. The equations of the Hoyle-Narlikar theory are obtained with the help of the principle of stationary action, which I referred to earlier. The equations are much more complicated than Einstein's and have the following consequences.

1. When the number of particles in the universe is large the equations can be approximated to look like those of general relativity - except near isolated particles. In these equations the gravitational constant G appears as a quantity determined by all the masses in the universe. Thus, while in general, the gravitational constant can vary with space and time, in this approximation it remains constant and *positive*. The *sign* of the gravitational constant is important. In the theories of Newton and Einstein, G is tacitly *assumed* to be positive. In the Hoyle - Narlikar theory it is *deduced* to be positive.
2. The exceptional case mentioned above - that of an isolated particle-is interesting in the following way. By an isolated particle we mean a particle well separated from all others in the universe. If we examine the motion of a test - particle in a region in the neighbourhood of such a particle, we find that it is first attracted by the isolated particle. But when the test particle gets very close to the isolated particle *it is repelled*. Does this happen for two ordinary particles? This is an

intriguing question, not yet answered. If gravitation changes sign and attraction changes to repulsion at very close range, this may explain why for example, the massive galactic nuclei seem to be exploding rather than imploding. The problem is at present under investigation.

3. If we imagine the universe to be empty, or to have just one particle in it, then the Hoyle-Narlikar theory predicts no interaction at all. Thus a single particle will have no inertia. This is fully in accord with Mach's conjecture. In general relativity, on the other hand, even empty universes have non-trivial gravitational properties, which do not conform with Machian concepts.

Eventually we may have to extend this work further to include strong and weak interactions, the two remaining basic forces known to physicists. The work is likely to be more difficult in view of the complexity of high energy physics at the experimental level.

Intuition vs Rigour

Fred Hoyle once said that no mathematician, howsoever great a genius he may be, would be able to produce high quality work if he works entirely in isolation, say being locked up in a cell! Of course this statement has not been (and probably cannot be) tested; but it is worth elaborating on. What Hoyle meant was that the human brain is constantly being influenced by the environment and even the so-called abstract ideas of mathematics have had intuitive beginnings. The process of abstraction - of making the result as 'general' as possible - follows later.

Mathematicians may deny this, arguing that these results have no usefulness whatsoever and that it would be futile to link their abstract conclusions with any real experience. This may be true of the final product; but Hoyle was talking about the beginning - about the genesis of a great idea.

Leaving mathematicians aside, let us look at physicists. Very often we find a conflict between what they intuitively perceive and how they express it mathematically. Because, their mathematical expressions often lack rigour. Had they waited for someone to bring in rigour, they would certainly have lost out in their race against nature. So they have accepted compromises in terms of empirical laws, dubious mathematical operations or ad-hoc approaches to their mathematical problems. The development of quantum theory provides numerous examples.

To begin at the beginning, the thermodynamics of trapped radiation - the so called *black body radiation* led Planck in 1900 to postulate the concept of a packet of radiant energy, called the *quantum*. The Rayleigh - Jeans law of classical thermodynamics gave the radiation-spectrum in equilibrium as $I(\nu) d\nu$ where

$$I(\nu) = \frac{2\pi kT}{c^2} \nu^2 \quad (14)$$

The total intensity of radiation as per (14) however, diverges!

On the other hand the black-body spectrum showed a maximum of $I(\nu)$ at a specific frequency ν_m that is proportional to temperature (Wien's law). How to explain this turn around of $I(\nu)$ at high frequencies that also fulfilled the requirement of finite total intensity? Planck's approach met both these criteria. Yet, it appeared ad-hoc and certainly not elegant let alone rigorous by mathematical standards.

There were several steps in this century which led to a clarification of the situation, e.g.

- (a) The notion of a photon introduced by Einstein in the first decade
- (b) The Bose-Einstein statistics of the 1920s
- (c) Quantum field theory of radiation attempted by Dirac in the 1930s
- (d) A working theory of quantum electrodynamics developed independently by Feynman, Schwinger and Tomonaga in the late 1940s

Although a physicist today may confidently assert that he has a working theory of quantum electrodynamics (QED), the mathematician may legitimately point out that a working theory does not necessarily mean a rigorous theory. Before I proceed further let me look at the evolution of quantum theory of charged particles, since they form the other component of QED.

As in the case of radiation, the classical Maxwell theory of charged particles appeared to fail at the microscopic level in explaining the simplest system, the hydrogen atom. The Newton and Coulomb laws gave the following relation for the motion of an electron of mass m and charge $-|e|$, moving in a circular orbit of radius r around a proton of charge $|e|$:

$$mr\omega^2 = \frac{e^2}{r^2} \quad (15)$$

This relation tells us the angular velocity ω of the orbit in terms of its radius. Secure though this relation appears, it ignores a consequence of the Maxwell theory, viz the radiation by an accelerated electron to which I referred earlier. The loss of energy by radiation and the consequent radiation damping would reduce the radius of the orbit steadily so much so that $r \rightarrow 0$ in a time scale of the order

$$r = e^2 / mc^3 \approx 10^{-23} \text{ s} . \quad (16)$$

Thus we run into the problem of extremely short life-time of such a common atom as hydrogen! Clearly something is missing in the picture. That the classical theory is incomplete was also becoming apparent from spectroscopy which revealed discrete wavelength lines or bands (both dark and light) in the spectra of radiation from terrestrial and astronomical objects. The first breakthrough came from Niels Bohr in 1912 when he added a further relation to the dynamics of the H - atom :

$$mr\omega^2 = nh, \quad n \text{ an integer } \geq 1, \quad (17)$$

where \hbar is none other than Planck's constant \hbar divided by 2π . By requiring that the angular momentum of the electron must be an integral multiple of \hbar , Bohr brought in the discrete element that seemed to be needed. By further demanding that $n \geq 1$ he permitted only discrete orbits of *finite* size for the electron. Thus spiralling inwards to $r = 0$ is avoided. The spectral features could also be related to this picture.

Nevertheless Bohr's prescription appeared ad-hoc. That it worked indicated that the solution was along the right lines but further steps were needed to understand *why* it worked. Progressively these steps came with the following landmarks :

- a) Einstein's *A* and *B* coefficients in 1917 for atomic transitions
- b) The matrix mechanics of Heisenberg in 1925
- c) The wave equation of Schrödinger in 1926
- d) The noncommutative operator for the electron in 1928
- f) The path integrals approach developed by Feynman in the 1940s
- g) QED (Feynman, Schwinger, Tomonaga)

Thus Bohr's prescription could be explained in terms of the mathematical problem of eigen values of a second order differential operator implied by steps (b) - (d). The step (f) in the particle stream above is the same as the step (d) in the radiation stream. So in QED came the final working model of how electrically charged particles interact with radiation. Yet, as I mentioned earlier, even today we are far from a rigorous mathematical formulation. To illustrate the kind of problem encountered consider the following result of calculating an intergral

$$K(\lambda) \equiv \int_0^\lambda f(x) dx = A \ln \lambda + g(\lambda) \quad (18)$$

where, as $\lambda \rightarrow \infty$, $g(\lambda)$ is well defined and finite. Clearly, $K(\lambda)$ in the limit $\lambda \rightarrow \infty$ is not definable. Yet a result of this kind emerges when the physicist uses QED to evaluate such basic entities as the effective charge and mass of an electron! Integrals of this kind appear in other measurable quantities also. Yet the saving grace is that if we neglect the logarithmically divergent term, the balance [$g(\lambda) \rightarrow$ finite limit] does agree remarkably well with the experimentally measured quantities as in the so called Lamb Shift of the *H* - atom and the anomalous magnetic moment of the electron! Further, the rule for subtraction of infinities, once defined, can be applied unambiguously.

So this working rule (commonly dignified by the title '*renormalization programme*') has some germ of truth! Our present QED is along the right lines but still not complete. Some believe that very esoteric techniques like the *string theory* may provide the answer. Others suspect that quantum gravity may hold the key. There is also the possibility that I outlined earlier, of invoking the response of the universe in what (misleadingly) appears to be a local microscopic problem.

Need-based Mathematics

At this stage I would like to return to the inputs provided by mathematics. The steps (a) - (d) or (a) - (f) just described use increasingly sophisticated mathematical techniques. Yet, it is worth recording here the circumstances under which they were evolved. In *none* of the cases did the physicist approach a mathematician with his conceptual problem and the mathematician provided the necessary inputs! Instead the physicist more or less created his own techniques to deal with his needs. How have mathematicians reacted to these need-based creations? Their response has been twofold.

First they have been critical of the standard of rigour. Thus the Dirac delta function, the Feynman path integrals, the use of spinors by Dirac, the renormalization programme etc. are not considered respectable enough by the mathematicians.

Second, in some cases the mathematicians have pointed out earlier work which did give a rigorous treatment of the concepts, albeit in forms unfamiliar to the physicist. Thus the Dirac spinors turned out to be part of theory of group representations discussed by algebraists. In other cases the mathematicians took hold of a physicist's formulation (like the delta function) and recast it in a 'respectable' rigorous form.

The counter response of the physicists to these reactions of mathematicians has ranged from appreciation to 'couldn't care less.' I once attended a course of lectures by Feynman on path integrals in which he would take frequent jibes at mathematicians for their obsession with rigour and perfection, saying that if he waited to make his concepts good enough for them, he would make no progress towards his understanding of physics. Perhaps I can do no better than end with the following anecdote :

A traveller riding a horse on the Texan plane during the frontier days arrived - at the end of a long day - at a small town. At the outskirts he saw a shop with a big sign :

"Laundry! Your clothes cleaned while-U-wait."

"Just what I need" thought the tired horseman as he approached the shop and dumped his shirt and trousers on the counter. "Will you please clean them right away? This is the only set I have," he urged.

"What do you mean, Sir?" exclaimed the horrified man behind the counter. "We don't wash clothes here." The irate customer pointed to the sign outside. The shopkeeper replied calmly :

"Oh that? Well, we are sign painters."

This has the obvious moral for the physicist that he should wash his dirty linen himself rather than expect help from the mathematician. Yet, in the present company I would like to add a postscript that it would help the physicist to once in a while (and not always!) explore what laundries the mathematicians do have : may be he will find one that will wash his particular linen shining white!