Unity of fundamental forces, a paradigm in Physics

U. A. Yajnik*
Physics Department, Indian Institute of Technology Bombay,
Mumbai 400076

February 4, 2001

1 Introduction:

It is a previledge to be invited to this seminar, especially for someone whose specialty is not Philosophy of Science. One of the aims of the seminar is to create a better understanding of science by generating dialogue among people of different backgrounds. In the spirit of this goal, this contribution reports on some of the metaphysical inputs in a frontier area of science. Some of the justification for presenting such a report was provided by Prof. V. Singh's remarks in the keynote address. The remark being that Philosophy of Science must also be done in a scientific manner. The way science is done has itself been changing with the times. The scientist inherits the practice of science from his or her mentors and improvises upon it as demanded by circumstances. Thus the very subject matter of the Philosophy of Science has been evolving. This contribution presents an account some of this development. I shall talk about some technically distinct but philosophyically inter-related notions of symmetry that have been used in this century for guessing the laws governing the fundamental interactions. In this talk I am being neither rigourous nor exhaustive, but I hope I succeed in kindling some interest.

The notion of symmetry was used by Einstein to guess the General Theory of Relativity. This he did in the absence of any empirical evidence for the theory. In

*e-mail : yajnik@niharika.phy.iitb.ernet.in

apparently unrelated developments, Heisenberg proposed in the '30s a symmetry as the basis for understanding the nuclear forces among protons and neutrons. In 1956 Yang and Mills and independnently Shaw made a major breakthrough by extending Heisenberg's idea and making a theory closer in spirit to Einstein's postulate of General Relativity. At the time they did this, there was no empirical evidence to support the Yang-Mills theory as the theory of nuclear forces. A decade or so later, after more empirical evidence had accumulated and better understanding of collective quantum phenomena was obtained, the Yang-Mills theory could be suitably modified to account for all nuclear forces. The point is that the paradigm of symmetry principles was used to make fairly accurate guesses at theories for which scant or no empirical evidence existed. How is such a procedure justified? It is always successful? Not all theories guessed in this way are successful. One example is of Kaluza-Klein theory. Similarly, there is no empirical evidence for a supersymmetric theory. However, in the reductionist endeavour of theoretical physics, the symmetry principle has played a central role.

How useful has it been to understand the fundamental forces in this way? Do these theories allow us to calculate all the observed data from first principles? The answer is in the negative. While all the important features of the sub-nuclear world can be qualitatively explained, quantitative predictions are possible only for a limited class of phenomena. In the case of Gravity, Newtonian theory is sufficient for all but a few minute effects. The edifice of General Relativity may well appear to be a dispensable extravagance. With future technologies and techniques of computation, the theories will be surely exploited to their fuller potential. For the present, the overwhelming motivation is best expressed in Eintein's own words:

"I want to know how God created this world. I am not interested in this or that phenomenon, in the spectrum of this or that element. I want to know his thoughts, the rest are details".

This purist sentiment has been vindicated by the fact that the structure of the fundamental theories appear, with hindsight, to be remarkably elegant, in the sense that one or two fundamental ideas coupled with powerful mathematical machinery, the latter almost mechanical in nature, suffice to dedectively reconstruct the theory.

Thus it is not utilitarian motives that drive the effort in fundamental Physics, but rather a deep belief in the existence of a deductive structure based on simple axioms. The symmetry principles to be described in the following are a concrete realisation of this paradigm.

The remainder of this talk is divided as follows. In section 2 I discuss some of the mathematical and conceptual preliminaries. These include the ideas of a vector fields, symmetries, and the notion of coupling constant. In section 3 I describe the symmetry principles involved in General Theory of Relativity and in Electromagnetic and nuclear theories. Section 4 contains concluding remarks.

2 Prelininaries

2.1 Mathematical ingredients:

In the following I shall assume that the reader is familiar with the notion of fields as used in physics. For example temperature as a function of position, or the velocity field in a streamlined fluid flow. Another well known example is the electromagnetic field. All of these can be thought of as scalar or vector fields in Newtonian space and time. Vector nature of a field means that if the coordinate axes, (physically speaking the frame of reference) are rotated, the three components of the vector value at each point change in the well known manner. For example, if V_x, V_y, V_z are the components of vector V in the old frame and the corresponding components in the new frame are denoted by V_x' etc. then

$$\begin{aligned} V_x' &= a_{xx}V_x + a_{xy}V_y + a_{xz}V_z \\ V_y' &= a_{yx}V_x + a_{yy}V_y + a_{yz}V_z \\ V_z' &= a_{zx}V_x + a_{zy}V_y + a_{zz}V_z \end{aligned}$$

Here a_{xx}, a_{xy}, etc. are parameters relating the new and old reference axes. The main points to remember are that the relations are linear (i.e., do not contain terms quadratic or of other powers in the various components) and that it relates the components at the same physical point. The latter property is called locality.

If we are in the domain of Special Relativity, we have to allow for rotations as well as velocity "boosts", i.e., a change to a frame moving with constant speed relative to the old one. The vectors will now be four-vectors, and hence will also possess a time-like component. In this case too, the relationship between old and new components is linear and local.

2.2 Symmetries:

Consider an equilateral triangle cut out of a cardboard. Let a particular way of placing it, as in part (a) of fig 1, correspond to some physical situation. If the traingle is now rotated by 120° as in part (b), there is no discernible change in

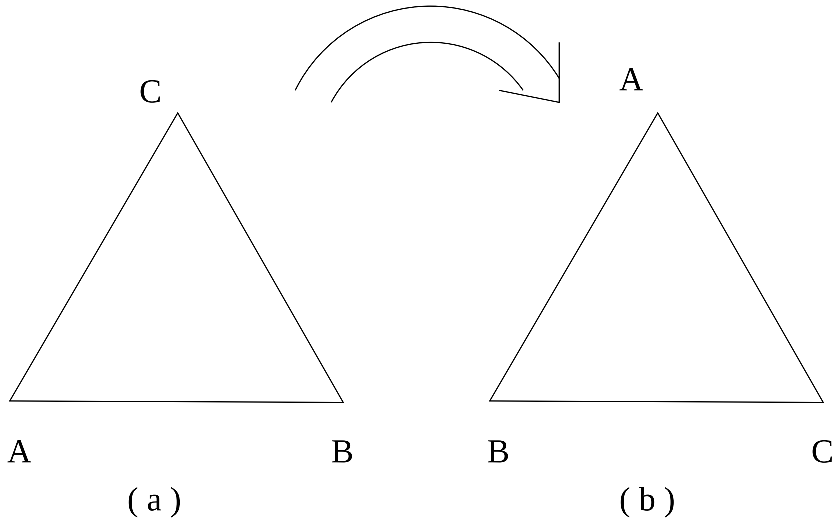

Figure 1: Example of a ymmetry operation on a triangle

the situation. If the triangle is a good representation of the physics, the physical situation will be unchanged under 120° rotations of the triangle.

The equivalence to real physics is that instead of a triangle, we have vector fields. Physics is supposed to remain unchanged under rotations or Lorenz transformation of frames of reference. Fig. 2 shows two choices of co-ordinates, with the $x' - y'$ inclined at θ with respect to the orginal $x - y$ system. The components of a given vector field change under such transformations but not the physical answer calculated from them. Such operation are called symmetry operations. In 1930's the notion of such symmetry was extended by Heisenberg in a very profound way. It is known that the strong nuclear force does not distinguish between the proton and the neutron, or separately, among the three pions π^+, π^0 and π^-. He proposed that the strong nuclear force does not depend separately on the physical state of the proton or that of the neutron. In Quantum Mechanics, the physical state of a system is described by a complex wavefunction denoted ψ. Heisenberg's proposal was that instead of using ψ_p (for proton) and ψ_n (for neutrons), we could define new wavefunctions ψ'_p and ψ'_n given by

$$\begin{aligned} \psi'_p &= a\psi_p + b\psi_n \\ \psi'_n &= c\psi_p + d\psi_n \end{aligned}$$

and the physics would remain unchanged. Here the a, b, c, d are complex numbers satisfying some constraints. The relations above can be thought of as a com-

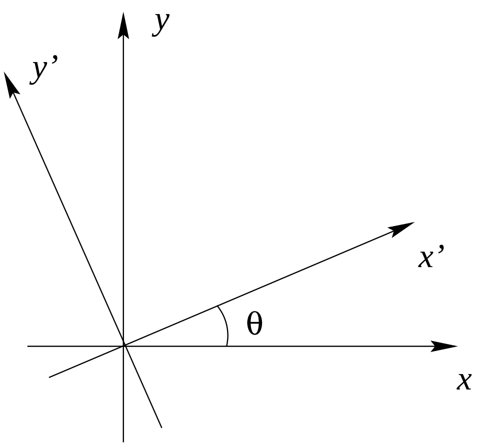

Figure 2: Choosing new co-odinates axes rotated with respect to a given set. Such operations are symmetries of Physics.

plex rotation, an abstract generalisation from the case of real vectors discussed in section 2.1. This rotation, called an isospin rotation is a symmetry (although approximate) of the strong nuclear force.

2.3 Coupling Constant:

In shall complete the preliminary discussion by explaining the essential aspect of the notion of unification. A good example is electricity and magnetism. It was at first though that two independent constants ϵ_0 and μ_0 are required to give the force laws between charges and magnets respectively. But later it was realised that magnetism and electricity are interlinked through relative motions of charges and magnets. It was found that the product $1/(\epsilon_0\mu_0)$ in free space gives c^2, the square of the speed of light, a universal constant. Thus there was only one coupling constant needed to describe electricity and magnetism, and the latter two were "unified".

Minimizing the number of phenomenological parameters is at the heart of the predictive power of theories. To describe properties of real substances we need a large number of phenomenological parameters. But most of these can be ascribed to the particulars of the state in which we find the substance. For instance the elastic moduli of a crystal can be ascribed to the lattice arrangement of the ions, th shape of the electron clouds in individual ion which determines its interaction with its neighbors and so on. But the fundamental interaction involved is Electromagnetism, determined by one coupling consonant. As is well known, at the present state of knowledge there are four interactions we may call fundamental. The strong nuclear, weak nuclear, electromagnetism and Gravitation. The hypothesis of unification is that the need for four different couling constants is a consequence of the particular "state" in which we find ourselves. If only we could see the specific conditions characterizing this state, we could progress towards identifyintg the universal principle.

Another example which seems simpler due to hindsight is Newton's Law of Universal Gravitation. No one suspects at first that the fall of an apple on the earth is governed by the same law that governs the orbital motion of the moon around the earth, and indeed the motion of all "heavenly" bodies. Given the masses of the objects concerned, only one universal constant G is needed to give the Gravitational force law.

3 The Symmetry Principles

3.1 General Theory of Relativity:

Newtonian theory of Gravity envisages an instantaneous force acting at any arbitrarily long distance. This conflicts with Special Relativity, where physical influences cannot travel at infinitely large speeds. Einstein took up to modify the Newtonian law.

It was well known that the mass of a particle required in Newton's second law, the inertial mass, was the same as that which entered the law of Gravitation. The consequence of this is that given a large gravitating body, all the other particles when released from a given point with given velocity would trace out the same trajectory, regardless of the particular values of their masses. This fact about Gravitation had been demonstrated by Galelio in his famous experiments at the Tower of Pisa. It had been checked to a high degree of accuracy by the Hungarian experimenter Etovos in the nineteenth century.

Einstein promoted this observation, the equivalence of Gravitational and inertial masses to a physical principle, the Equivalence Principle. As we have seen, the Principle means that in the vicinity of a large mass, the trajectories of particles are independent of their inertia and are determined purely by their initial velocities. The trajectories can then be thought of as a property of the space-time itself rather than that of the particular test particle passing by. The space-time around the Sun can be thought of as a rubbur sheet stretched by the weight of

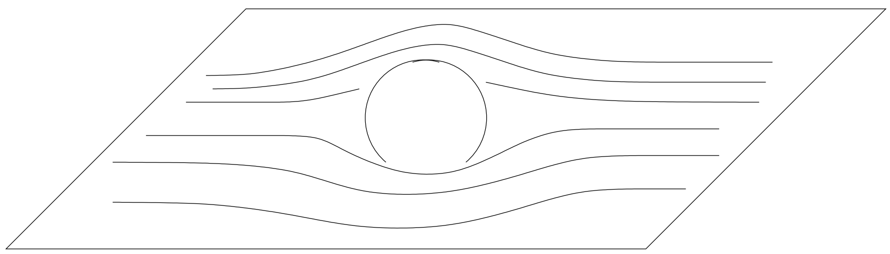

Figure 3: A sheet of rubber stretched by a ball placed on it serves as a model for the gravitation field of the ball.

the sun. This is schematically shown in fig. 3

Elinstein's observation upto this point amounted to elegant geometric reformulation of known facts. But too guess the full form of the Gravitational law, he used the following device. Suppose we know a law in the absence of Gravity. This should be Lorentz covariant. That is to say the equations will involve vector fields which will transform linearly if we make a change from one inertial frame too another. In a curved s pace-time an inertial frame can be defined only locally, in a small region of space and time. Suppose we designate a closed cabin freely falling in the local Gravitational field as an inertial frame. Since all material bodies inside this cabin fall with the same acceleration, if this cabin small enough, the observer inside will be able to do short experiments in which the curvature of space and time (i.e. effects of Gravity) will not shown up. If the cabin is too large or the experiment lasts too long, the curvature effects will shown up. Thus, a small cabin will indeed work like an inertial frame provided it is falling freely in the local Gravitational field. At another point we will need another frame, again falling freely there. Thus, in a curved space one does not choose rigid, Cartesian system of coordinates, but any convenient, curvilinear coordinates. An example is shown in fig. 4. The complete freedom in the choice of a coordinate system has the implication that the local inertial observers at different points, may use completely different Lorentz frames to describe their physics. This is indicated by different co-ordinate systems in tangent planes at points P and Q in the fig. 4. It is easy to see that the local co-ordinate systems at P and Q are not the same.

Einstein insisted that in the presence of Gravity, the laws should be such as to be covariant under arbitrary changes of the curvilinear coordinates. In fig. 5 we have shown a new coordinate system on the space of fig. 4. Furthermore, under a change of co-ordinates, the two frames suffer different transformations at the two different points, as indicates by the new frames at the same points P and Q in fig.

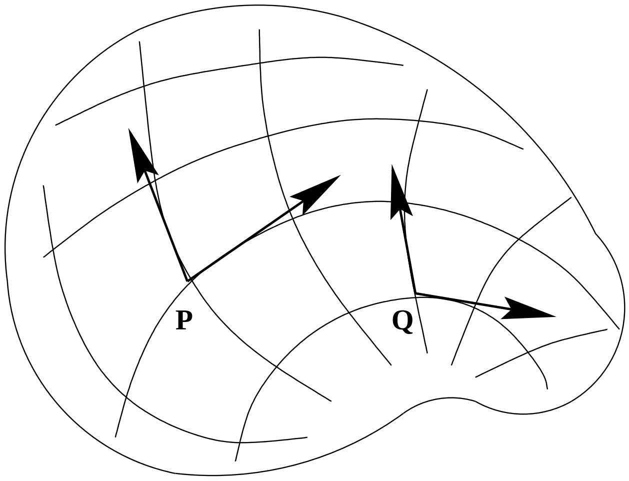

Figure 4: A curved surface with a coordinate system with local "inertial frames" shown for "observers" P and Q.

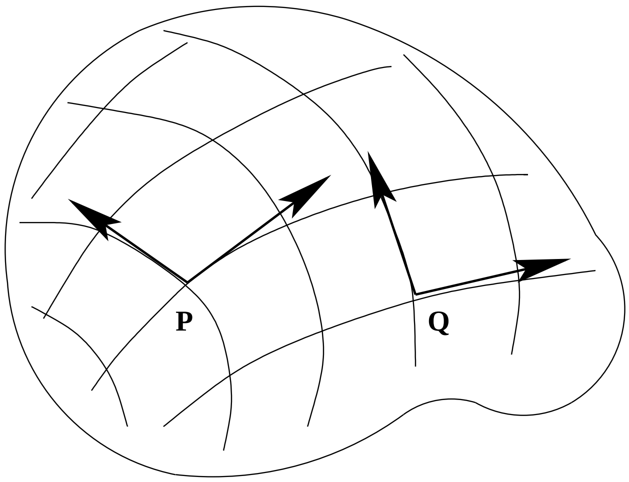

Figure 5: The surface of fig. 4 with a different coordinate system.

5.

Thus the freedom in the choice of the Cartesian Lorentz frames is greatly expanded to a freedom in choice of Lorentz frames independently at each point. This principle was called by Einstein the Principle of General Covariance. This symmetry principle served as a tool for guessing correct laws of motion for particles and the laws of Electromagnetism in the presence of Gravity. Further, the same principle was exploited by Einstein to guess the correct law of motion for the Graviational field itself.

In Einstein's own words he had generalised the Special Relativity of inertial frames under Lorentz transformations to General Relativity of arbitrary curvilinear coordinate systems. In the decades since, a different view is taken of the Principle of General Covariance. Stated in the form of a symmetry (operations that do not affect Physics) it really is a prescriptiion for guessisng the form of a physical interaction (very much affecting Physics), viz., Gravity. But the details of this discussion cannot be included here.

In the 1930's T. Kaluza and independently Oskar Klein, proposed a theory in which Electromagnetism as well as nuclear interactions appear as a consequence of General Covariance in space-time of larger dimension. The extra dimensions of space would be too tiny for us to directly experience. The program however did not work, and under closer investigation in the 1980's it was shown to have fundamental disagreement with some known facts of the nuclear world.

3.2 Gauge invariance of Electromagnetism:

In the 1920's Hermann Weyl pointed out a symmetry of Electromagnetism that was different in technical detail but very similar in spirit to the Principle of General Covariance. This symmetry he called the Gauge Principle.

Suppose we write the Maxwell Equations, not in terms **E** and **B** fields, but in terms of the four-vector potential field $(A_t,\ A_x,\ A_y,\ A_z)$. It can then be shown that the Physics is unchanged if we choose a different vector field $(A_t',\ A_x',\ A_y',\ A_z')$ such that

$$\begin{aligned} A_t' &= A_t + \frac{df}{dt} \\ A_x' &= A_x + \frac{df}{dx} \end{aligned}$$

etc., where f is a function of space as well as time, t. This is like the traingle we had in sec. 2.2. We need an equilateral traingle, but somehow the Physics does not depend on which corner is which. Here we need a four vector field but if the components are all changed in the specific form given above, it does not affect the Physics. It is important to note that the change is not at all like a Lorentz transformation of inertial reference axes. We have an independent change in components at each point in space- time. This is because f is a function of space and time.

Now this situation is quite similar to the case of General Covariance, where invariance under arbitrary coordinate change amounted to independent Lorentz transformations at each point. Furthermore, as in the case of General Covariance this invariance of Electromagnetism also helps to fix the interaction of electromagnetic fields with charged matter. In Quantum Mechanics, matter is described by a wavefunction ψ We now insist that the combined equations of the A fields and the ψ field are such that they remain invariant under the change in λ fields already mentioned, and a simultaneous change.

$$\psi'(\mathbf{x}, t) = e^{if(\mathbf{X},t)}\psi'(\mathbf{x}, t)$$

Thus the phase of the wavefunction is simultaneously changed in a space-time dependent manner. This Principle has now been established to a great level of accuracy. It also reproduces correctly the interaction of classical (macroscopic) currents and charges with electromagnetic fields.

At first it is difficult to see the geometrical analogy of the Gange Invariance prescription. IIt becomes necessary to expand our set of geometrical concepts. First of all we need to imagine that the space of f is a circle. That is f takes values from 0 to π.

Note that this is consistent with the use we are putting f to. In the equations of A fields we only need derivatives of f, so multiples of 2π can be added without changing anything. Similarly, in the ψ transformation, it is clear that f enters only modulo 2π.

Now expand your notion of "space-time" as follows. Imagine at every point in space-time, a circle corresponding to the space of f. Thus in fig.s 4 and 5, just as we have shown a local inertial frame at a typical point P, so should we have a local copy of the space of f-function. The question is, where on the circle does the zero value of f lie? A given set of A fields and the f field corresponds to a particular choice of the placement of zeros. The statement of invariance is that if we now shifted the zero at point P by an amount f where f is an arbitrary smooth function of $\mathbf{x}$ and t then the Physics does not change.

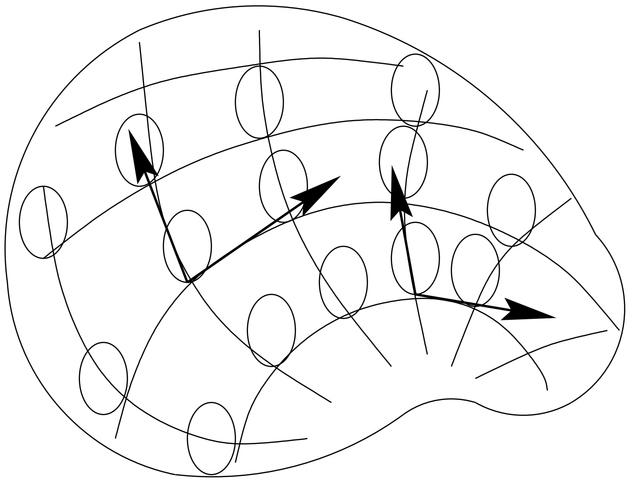

Figure 6: Ordinary space of fig.s 4 and 5 augmented with an internal space of f-functions

We have to thus imagine in our space-time not only the possibility of choosing a local inertial frame, but also a particular value of the f-function, called the Gauge Function. The choice of zero of the f is analogous to the choice of space-time orientation of an inertial frame of reference. Just as in the case of Gravity, a change of coordinates amounts to a new choice of frames of reference at each point, similarly in Electromagnetism two different observers' Gauge Functions can differ by an additive function of space and time.

Although there is a clear analogy between General Covariance and Gauge Invariance, there is a fundamental difference. Even in the absence of Gravity, we need to choose a frame of reference. But in the absence of Electromagnetism, the f-function is completely superfluous. Even in the absence of Gravity, space-time remains, and motion occurs. But the abstract space of the f-function is meaningless without Electromagnetism. For this reason, the two kinds of symmetry principles are classified as "external" (General Covariance), and "internal" (Gauge Invariance)

3.3 Application to nuclear forces:

It is now known that neutrons and protons are not fundamental particles but made up of the strongly interacting quarks. The forces among the neutrons and the protons inside an atomic nucleus are therefore a complicated reside left over after the quarks are bound up. Thus the binding energies of the nuclei and their excitations are very difficult to compute. However the strong nuclear force itself, acting between quarks, is tractable, especially at high energies.

In addition to the strong nuclear force, there is a weak nuclear force, about a million times weaker. This is the one responsible for the decay of an isolated neutron. However, the empirical facts to which the theory could be applied were not yet discovered. During the 1960's, in separate proposals, Glashow, Weinberg and Salam applied the Yang-Mills theory to the weak nuclear force.

The remarkable discovery from mid-1960's onwards is that both the strong and the weak nuclear forces are gauge forces. The symmetries associated with them are a generalisation of the case for Electromagnetism. In particular, one needs *twelve* f-functions, eight for the strong, and three other such functions for the weak force. It also turns out that the electromagnetic force and the weak nuclear force are inter-related, that is, their f-functions can mix.

If we now wish to make a geometrical picture, we have to modify geometry the way we did in going from fig.s 4, 5 of sec. 3.1 to fig. 6 in sec. 3.2. There we proposed attaching a circle, the space of the f-function, to each point of the ordinary space-time. Now we must attach to each point an agglomerate of three spaces: one for the eight f-functions of the strong force, one for the three others that describe the symmetries of the weak force, and one for an f-function, which after "mixing" with one of the f's of the weak force will produce Electromagnetism. Each of these spaces will be either a circle, a solid sphere or an eight dimentaional generalisation of a sphere.

There now seem to be too many new members to the act, but the power of the gauge principle is that one does not need twelve different coupling constants, one for each f-function. We need only three coupling constants, one for each of the spaces described above.

3.4 Symmetry principles and unification:

The fact that the nuclear and Electromagnetic theories all rely on the Gauge Invariance principle, and that the latter in turn is closely related to General Covariance has raised the hope that perhaps these are various pieces of the same theory. Perhaps the differences arise due to specific limiting conditions and if only we know how to transced these, there would be a single theory with a unique symmetry principle and requiring only coupling constant. Specifically, it is proposed that at extremely high energies of interaction, or requivalently, under explorations of space-time regions at a super-microscope scale, the limiting conditions disappear.

The earliest attempts along these lines were those of T. Kaluza and ăO. Klein ă In sec. 3.2 we said that we expand ăour ănotion ăof Ŋ space by attaching to each point a copy of the space of the gauge function. This latter space happens to be a circle. Kaluza ăand Ŋ Klein ăproposed that this circle actually constitutes ăan ăextra Ŋ– fifth –ăădimension ăof ăspace-time ăbut ăwith ăătwo important Ŋ differences. One that this dimension is not open or unending like the ordinary dimensions, but is closed up to make a circle. Second, that the scale of distances on this circle is so tiny that in all experiments so far, we do not perceive a "motion" into that dimension. The remarkable fact is that the Gauge Invariance principle can now be derived as a consequence of General Covariance on the five dimensional space-time.

The Kaluza-Klein proposal can be generalised to include the gauge function spaces of the nuclear forces. Then all one would need is General Covariance on a space-time of some large number of dimensions, eleven according to one proposal. These ideas however do not work because there are some fundamental facts predicted from them which conflict with observations. Alternatively one has to modify the idea in some complicated and inelegant way.

Through the 1970's and 1980's , several deep and elegant proposals have been made for unification, but I shall mention only two more in support of a point to be made in the concluding section. One is the Grand Unification proposal of Georgi and Glashow. It proposed that the space of the nuclear gauge functions is 24-dimensional and that only one coupling constant is sufficient to explain all the observed gauge forces. The theory failed because the proton did not display the required miniscule extent of instability towards decay into lighter particles. A more ambitious and far reaching proposal is the superstring theories. These exploit intricate symmetries of two dimensional spaces to explain all the observed symmetries. It also has the potential to explain certain other things physicists never expected to be able to explain. This includes the dimensionality of the space-time we live in. It also unifies Gravity and nuclear forces elegantly. Superstring theories are yet being explored. The absence in them of any verifiable predictions makes their relevance of Physics all the harder to explain.

4 Implications for Philosophy:

As indicated in the Introduction, the aim of Philosophy of Science ought to be a scientific study of the methods scientists use. Although finished bodies of knowledge the area looks chaotic. There are rules of thumb applicable within certain domains and there are quaint violations of the rules in certain others and so on. In organising these into a theory, scientists rely on intuition and on tentative ground-rules which are themselves external to scientific knowledge. In this sense the working rules are metaphysical.

The symmetry paradigm described above is first such a metaphysical principle that has come to be reasonably well articulated and actively pursued. I hope that it proves to be of some interest to the Philosophers.

I shall now briefly menion the "insider" philosophical views on this principle.

In the classical age (which includes Einstein) there was a sense of the divine attached to the metaphysical view-points. This tendency is still very strong among individual members of the theoretical physics community. Simplicity and beauty are considered hallmarks of correct theories. The cleansing sociological process of science is however never deceived by such mystical statements. Beauty means mathematical elegance. Those not trained in mathematics will agree, and even more so those who suffered mathematical training against their will, that such beauty lies strictly in the eyes of the beholder. As for simplicity, some anthropocentric position can easily be constructed that says that had the advancement and training in scientific thinking not reached a certain point the theory would hardly look simple. Conversely, that which does not appear simple at a particular epoch is unlikely to be discovered in that form. A realistic approach to what physicists consider beautiful is best expressed in the words of S. Weinberg, "We consider theories beautiful in the same way that we may consider beautiful a horse that wins us a prize".

The world we see has certain unique unchanging features, for instance its dimensionality, the dimensions of the gauge function spaces, the kind of fundamental spin-$1/2$ particles we have and so on. If there is a simple grand principle which fits all of this and explains most of it, its predictions must be pretty much unique. What has been found so far is that strongly stated principles with unique predictions have failed - example of Grand Unified Theory of Georgi and Glashow; or the principle leaves open the possibility of so many worlds that ours has to be viewed as an accidental outcome, this being the case of the superstring theories. Thus, principles we consider beautifuland simple can also fail or give non unique world.

It is unlikely however that this anigma can be resolved at the philosophical level. It is dear that unification of all forces including Gravity is not possible unless we have additional clues about Physics at that unknown and unexplorable energy scale called the Planck scale. Similarly, the key principle of thee unification of nuclear forces can not be decided unless higher energy experiments such as those being planned now are carried out.

In conclusion, I have presented here what may be called the metaphysics of the state of the art. The presentation is brief and selective and precision of mathematical language has been sacrificed to facilitate accessibility. I hope that inaccuracies and omissions here will be rectified by persual of many excellent and lucid books being written on these matters.