TOEs, fingers and the nose on your face

By James E. Beichler

Unification and the basic food groups

The greatest long-term trend in all of science can be found in the attempted unification of the laws of nature that comprise the science of physics. In general, science progresses by several methods. Among these are included both the synthesis of pre-existing concepts and the explanation of newly discovered phenomena. While the synthesis of accepted theories and concepts forms the basis of unification, the explanation of newly discovered phenomena also follows the pattern of unification. When they are first discovered, an attempt is made to explain new phenomena by older accepted theories. However, if they cannot be so explained then new theories are developed so that science can cope with the new phenomena. Later, the new theories are unified with the older theories, either as addendums to the older theories once they are expanded, or they are unified through the development of a still newer theory that incorporates both the old and new concepts in a more comprehensive model of nature.

Examples of both of these synthesizing processes abound in the history of physics and science, so much so that unification seems to be a major, if not the primary task of theoretical physics. The development of thermodynamics offers an excellent example of this synthesis process. Thermodynamics was born in the 1840s when James Joule unified two independent branches of Natural Philosophy, the kinetic theory of matter, which explained heat and Newtonian mechanics. This unification seemed inevitable since Hermann von Helmholtz and other scientists came to the same conclusions independent of Joule's groundbreaking unification. Joule's unification is highly significant in the history of science because it forced the emergence of physics as an academic discipline separate from Natural Philosophy.

On the other hand, the development of the electromagnetic theory, which occurred over more than a century of physical research, involved the discovery of new phenomena that could not be explained within the Newtonian paradigm. Although many scientists were involved with the development of the concept of electromagnetism and made significant contributions to the theory, the two most important were Michael Faraday, who laid the experimental foundations of electromagnetism between 1820 and 1850, and James Clerk Maxwell, who rendered Faraday's theory into a mathematical model in the 1860's after appropriate modifications and additions. What had originally been two separate branches of scientific enquiry, electricity and magnetism, had been unified into a single and comprehensive electromagnetic theory. The history of science for that century long process is virtually littered with new discoveries of electrical phenomena. However, the story of electromagnetism did not end at that moment in history.

While electromagnetic theory explained many phenomena dealing with the propagation of light waves and successfully predicted still more phenomena, problems rapidly arose with phenomena that were associated with both Newtonian mechanics and electromagnetism. Primarily, those phenomena that showed evidence of an interaction between the smallest particles of matter and electromagnetic waves defied explanation by either Newtonian mechanics or electromagnetic theory. These phenomena included the spectral lines of elements and compounds as well as black body radiation. In still another area where these two paradigms came into contact, the necessity of the luminiferous aether for the mechanical propagation of light waves, both theories fell apart as demonstrated by the Michelson-Morley and similar experiments. It was from these and similar failures of the two theories that two new unifications evolved which altered the course of physics, the developments of quantum theory and special relativity at the turn of the last century.

In the opening years of the twentieth century, quantum theory and special relativity were successful in unifying mechanics and electromagnetism at a very high price for classical physics. Fundamental changes in physics became more and more evident as each new theory progressed beyond its original formulation. The quantum theory of 1901 developed into a system of quantum and wave mechanics by 1927 and special relativity expanded into general relativity by 1916. In these forms, each new theory came to represent essentially incompatible aspects of reality. Quantum mechanics relies upon the discrete nature of reality while general relativity portrays the continuous nature of reality as represented by the concept of the field. After the 1920s, quantum mechanics became the dominant theory in modern physics for several decades. Einstein never fully accepted the Copenhagen Interpretation of quantum mechanics and spent the remaining decades of his life in opposition to mainstream physics while searching for a unified field theory that would unite the electromagnetic and gravitational fields within a single field model. He hoped that the quantum would literally appear as a byproduct of the mathematics modeling his unified field. In this endeavor, very few physicists came to the aid of Einstein.

In the meantime, most physicists accepted the Copenhagen Interpretation of the quantum and sought to unify physics according to their own model of reality. This line of thought culminated in such concepts as the quantum field theory (QFT) and quantum electrodynamics (QED), but these theories were never totally successful in their unification of the quantum and special relativity while gravitation theory has never been incorporated into the quantum model. The lack of success in uniting even special relativity and quantum mechanics was well recognized by the founders of QFT.

> The ambitious program of explaining all properties of particles and all of their interactions in terms of fields has actually been successful only for three of them: the photons, electrons and positrons. This limited quantum field theory has the special name of *quantum electrodynamics*. It results from a union of classical electrodynamics and quantum theory, modified to be compatible with the principles of relativity. (Guillemin, 176)

As Guillemin has testified in his history of quantum theory, QED is only "compatible" with the principles of relativity. It does not provide a framework for the unification of special relativity and the quantum. This same idea has also been confirmed by Julian Schwinger, one of the founders of quantum electrodynamics, who summed up the situation in 1956.

> It seems that we have reached the limits of the quantum theory of measurement, which asserts the possibility of instantaneous observations, without reference to specific agencies. The localization of charge with indefinite precision requires for its realization a coupling with the electromagnetic field that can attain arbitrarily large magnitudes. The resulting appearance of divergences, and contradictions, serves to deny the basic measurement hypothesis. We conclude that a convergent theory cannot be formulated consistently within the framework of present space-time concepts. To limit the magnitude of interactions while retaining the customary coordinate description is contradictory, since no mechanism is provided for precisely localized measurements. (Schwinger, xvii)

Schwinger clearly acknowledged in this statement that QED, the primary form of a quantum field theory, has reached a specific limit whereby it cannot be judged without reference to an outside framework of space-time. The prevalent framework of space-time, also referred to as the "customary coordinate description," at this juncture of history is that supplied by the theories of relativity, so Schwinger obviously believed that QED had not yet been unified with special relativity. Special relativity just forms a limiting condition for the mathematical model of quantum field theory that does not indicate that they have been unified in a single theory. It would further seem that the real unification to which scientists subscribe is between quantum mechanics and GR, since both describe the motion of matter in space-time while the unification with GR would certainly include an implied unification with special relativity.

New scientific advances in the 1960s and thereafter have brought GR to the forefront of physical research even as old philosophical problems which have plagued the Copenhagen Interpretation of quantum mechanics were revealing more cracks in the prevalent quantum paradigm. During the last few decades, these later developments have produced a climate of change within theoretical physics, which has resulted in a renewal of Einstein's search for a unified field theory. However, nearly all of the modern attempts at unification follow first from quantum theory rather than beginning from field theory as represented in special relativity or general relativity. Most scientists treat relativity as something to add onto or incorporate into for unification model have been proposed with such colorful names as supergravity, Grand Unification, superstrings and finally the 'Theory of Everything' (TOE). Even accepting the possibility that a TOE might exist marks a drastic change of attitude within the scientific community. But the problem of unifying the discrete and continuous aspects of the physical world has never been resolved in spite of attempts to do so from both the quantum and field approaches to the quantum perspective. Under these circumstances, more recent attempts to base a future theory of physics on Einstein's hoped nature. This dichotomy is represented indirectly within modern physics by such concepts as the wave/particle duality of both matter and light.

Although the philosophical problems presented by the differences between the discrete and continuous are not universally recognized in the physics community, a few physicists have been brave enough to question the established norms of modern physics in this regard. In his book *A Unified Grand Tour of Theoretical Physics*, Ian D. Lawrie has confirmed the uneasiness felt by physicists although he has not clearly defined the cause of his concerns beyond stating the modern physicists "do not properly understand what it is that quantum theory tells us about the nature of the physical world" even though "there are respectable scientists who write with confidence on the subject." Evidently, "the conceptual basis of the theory is still somewhat obscure." (Lawrie, 95) Mendel Sachs is far more straightforward with his criticisms. Sachs has noted two distinct and separate strains of scientific progress within modern physics.

> The compelling point about the *simultaneous* occurrence of these two revolutions (relativity and the quantum) is that when their axiomatic bases are examined *together*, as the basis of a more general theory that could encompass explanations of phenomena that require conditions imposed by both theories of matter (such as current 'high energy physics"), it is found that the widened basis, which is called 'relativistic quantum field theory', is indeed logically inconsistent because there appear, under a single umbrella, assertions that logically exclude each other. (Sachs, 1988, 236-237)

Sachs is, of course, referring to the logical and mutually exclusive nature of the quantum (the discrete) and the field (the continuous). He does little to hide either this fact or his criticism of the shortcomings of present day physics. Sachs has concluded that "neither the quantum theory nor the theory of relativity are in themselves complete as fundamental theories of matter," (Sachs, 256) due to the fact that they represent incompatible fundamental concepts of the discrete and continuous aspects of nature.

These philosophical problems have physical counterparts within the mathematical model of the singularity. GR falls apart at just the point where the continuous field of gravity meets the physical boundaries of the discrete particles whose curvature creates gravity, where the curvature of the space-time metric becomes so extreme that it becomes infinite. This singularity not only occurs at the heart of elementary particles, but also in black holes and the Big Bang, which theoretically created our universe. On the other hand, quantum mechanics deals with singularities in a different manner, although no more successfully than GR deals with them. Quantum mechanics utilizes a rather artificial method known as renormalization to deal with singularities, or divergences as they are called, and then ignores the problems created by the divergences.

In QED, each particle is associated with a field, so there are as many fields as there are different particles. This situation gives rise to an unpleasant expansion of the concept of field which some have criticized. (Popper, 194) Yet far more serious problems exist in QED. At the point where the different fields interact, one would expect to find the action and reaction of the particles as caused by forces in the classical sense of the term. However, at the point where the fields interact there are mathematical divergences rendering the masses of elementary particles infinite and undefined. Using perturbation methods these divergences or infinities can be renormalized to yield definite answers. Therefore, localizing and defining a point particle in QED amounts to using an artificial mathematical method for no other physically valid reason than that the method yields finite results that can be experimentally verified.

While many scientists do not see this procedure as a problem since its predictive power makes QED one of the most successful theories ever developed in science, the artificial nature of renormalization is at the very least philosophically unsatisfying and unsettling to other scientists and scholars. The method is considered at least *ad hoc*, but otherwise a necessary evil at present. Karl Popper was very critical of this shortcoming of QED.

> Moreover, the situation is unsatisfactory even within electrodynamics, is spite of its predictive successes. For the theory, as it stands, is not a deductive system. It is, rather, something between a deductive system and a collection of calculating procedures of somewhat *ad hoc* character. I have in mind, especially, the so-called '*method of renormalizatio*n': at present, it involves the replacement of an expression of the form 'lim *log x* - *lim log y*' by the expression 'lim (*log x* - *log y*)'; a replacement for which no better justification is offered than that the former expression turns out to be equal to 4 - 4 and therefore to be indeterminate, while the latter expression leads to excellent results (especially in the calculation of the so-called Lamb-Rutherford shift). It should be possible; I think, either to find a theoretical justification for the replacement or else to replace the physical idea of renormalization by another physical idea - one that allows

> us to avoid these indeterminate expressions. (Popper, 194-195).

While this criticism was leveled several decades ago, shortly after QED was first developed, it is still a valid criticism. QED is considered theoretically inconsistent because of these and other problems in spite of its great successes. It is only a theory of electron interactions and does not unify electromagnetism with any other of the basic forces in nature.

During the 1960s, Quantum Chromodynamics (QCD), which is a theory of nuclear interactions, followed QED. QCD unifies the Yang-Mills field, which describes the nuclear forces of binding between neutrons and protons in the atomic nucleus, with the Standard Model of quarks. QCD does not share the inconsistencies that plagued QED, but does require renormalization methods similar to those in QED. Renormalization methods therefore remain a general characteristic of the quantum field theories and QCD suffers from some of the same criticisms that plague QED. It is hoped by quantum theorists that newer methods and developments in quantum field theory will eventually justify or replace renormalization, rendering quantum field theory more palatable to its critics.

> Their hope is that by including gravitational interactions in the existing formulations of quantum field theory systems, it will be possible to construct a finite theory without any infinite renormalizations. They thus hope to avoid the consistency problem of the renormalization theory. (Schweber, 603)

Yet these problems persist. A large part of the work in quantum field theory still involves finding proper renormalization procedures that yield workable solutions. If a mathematical method of renormalization works, it is adopted as long as its results are experimentally confirmed, even if there is no physical reason for its success. This is not the best situation for physics, but physicists have accepted the method in principle and moved on to other theories.

The infinite masses of elementary particles that resulted in QED and QCD before renormalization can be effectively compared to the singularity problems of GR. However, such a comparison has not been common within either physics or the philosophy of science. In spite of the lack of recognition of this 'coincidence,' physics has continued to progress toward unification. It would seem that the fact that two fundamentally different approaches to physical reality, the continuous field and the discrete quantum, lead to the same inconsistency would be an important clue to identifying the problem of their unification. Since this clue has gone unnoticed, unification has proceeded along other lines. Both quantum physicists and relativity physicists dance around the problem of the singularity/divergence even though this problem should represent the object of the main thrust toward unifying the two perspectives of physical reality.

Unification using a fifth dimension

The first recognized attempt to use a fifth dimension for the unification of gravitation and electromagnetic fields was made by Theodor Kaluza. Shortly after Einstein's development of GR, a few scientists expressed their dissatisfaction with the artificial nature by which electromagnetic forces were imported into the field equations. They thought that both gravitation and electromagnetism should arise from a single geometric structure, the 'unitary' or unified field, but GR did not fully represent that field. In GR, the structure of the space-time continuum is modeled by the equation

$$R_{ij} - \tfrac{1}{2} g_{ij}R = -kT_{ij} \ ,$$

where R_{ij} is the contracted Christoffel tensor or the Ricci tensor, g_{ij} is the metric tensor, R is the curvature scalar, k is the gravitational constant and T_{ij} is the stress-energy (or matter) tensor. This equation accounts for gravitational attractions between material bodies, but does not explicitly include the electromagnetic field.

Many scientists had long maintained that matter itself is electrical in nature and electrical forces structure space-time, a view that dates back more than a century. So the conjecture that matter is no more than curved space-time is unsatisfying to those scientists. They assumed that electromagnetism must play as important a role in the structure of space-time as gravitation. However, the best that could be said on this subject within the context of GR was that electromagnetism could be added to the defining equation of the space-time structure by the inclusion of an electromagnetic term, E_{ij}, such that

$$R_{ij} - \tfrac{1}{2} g_{ij}R = - k\,[T_{ij}+E_{ij}] \ .$$

When approached in this manner, the addition of the electromagnetic term seems rather artificial. Nor does it lead to simple solutions for charged particles in combined electromagnetic and gravitational fields although an uncharged particle can be accounted for in the combined field.

The first attempt to unify these fields was made by Hermann Weyl in 1918. He sought to keep the four-dimensional space-time continuum intact while developing a more comprehensive geometry to deal with space-time. On the other hand, Kaluza suggested that a fifth dimension could be added to the space-time continuum to account for both fields simultaneously. The earliest mention of his work came in a letter to Einstein in 1919. (Raman, 212) Einstein encouraged Kaluza to continue working on his theory and further develop his physical model, (Middleton, 2) but nothing was made public regarding Kaluza's theory until 1921 when his paper "Zum Unitätsproblem der Physik" was finally published. The published theory was simple and straightforward.

In general, when a Riemannian rank two tensor is four-dimensional it can be characterized by ten components. T_{ij} is such a tensor and represents the metric structure of our space-time continuum. The ten components of T_{ij} describe the motion of material bodies within this metric. The addition of electromagnetism to this metric field structure would require four more components than exist for the four-dimensional configuration, for a total of fourteen components. So the four dimensional configuration cannot include electromagnetism. However, the Riemannian structure of our space-time can still be saved and an adequate number of independent components found by increasing the number of dimensions to five. This dimensional increase yields fifteen independent components, one more than is needed to describe the combined field. Kaluza was the first to see this as a possible answer to the problem of developing a unified field structure. Both electromagnetism and gravitation could exist on an equal footing within the geometric structure of a four-dimensional space-time continuum embedded in a fifth dimension as specified by Kaluza. At the time of his initial development of the five-dimensional model, quantum mechanics had not yet been developed and the whole issue of the direction which quantum theory would progress was in question, so there was not any need to incorporate the quantum into the theory at that time. Nor was there any hint that there might be a fundamental problem between the quantum and relativistic views of physical reality. The only purpose of Kaluza's theory was to unify gravity and electromagnetism within a common field because he saw such a unification as the primary problem in physics.

Kaluza's model yielded a geometrical representation of the generally covariant form of Maxwell's electrodynamics. By introducing the fifth dimension, he immediately raised two problems: (1) Since only fourteen variables are necessary for the combined field, the fifteenth term must be omitted or its affect nullified, and (2) Since every indication implies that our world is only four-dimensional, a five-dimensional assumption would necessitate an explanation of the absence of evidence that the fifth dimension exists. These two problems are not characteristic of Kaluza's theory alone, but form the central points of contention for any theory that utilizes a hyper-dimensional structure in a physical description of the world. These same two problems also define the major lines along which Kaluza's theory has been extended by some scientists and criticized by still others.

Kaluza added no physical significance (Bergmann, 1976, 254; Einstein and Bergmann, 683) to his five-dimensional hypothesis, but merely used it as a tool. In so doing, he had a great deal of latitude in overcoming both problems and was able to deal with them by assuming that the field variables, $\gamma_{\mu\nu}$, were independent of the fifth coordinate. They need only depend on the four coordinates of the space-time continuum when a suitable coordinate system was chosen. This choice also had the consequence of allowing a somewhat "atrophied" (Hoffman, 403) fifth dimension and thus a somewhat less generalized theory. Kaluza's mathematical model of the fifth dimension forced cylindricity on extensions in the fifth direction by guaranteeing that a vector in the fifth dimension, A^{μ}, satisfies the Killing equation. By further requiring that "the lines to which the A^{μ} are tangents - the 'A- lines' - have to be geodesics," (Bergmann, 1976, 258-259; Einstein and Bergmann, 686) the various A-lines in the fifth dimension were shown to be equal as well as constant. The norm of A was also constant throughout all of space and not only along the A-lines. Points in our four-dimensional space-time were extended into the fifth dimension along the A-lines.

Under these conditions, Kaluza based his five-dimensional structure on a special coordinate system defined by the metric,

$$d\sigma^2 = \gamma_{\mu\nu} dx^{\mu} dx^{\nu} \ ,$$

where μ and ν range from zero to four. This metric structure represented a five-dimensional manifold in which a four-dimensional continuum cut each of the A-lines that extended into the fifth dimension only once. The distance along any A-line could then be used to derive a value of $\gamma_{00} = +1$, where γ_{00} is the field variable in the fifth direction, thus normalizing all other components of the field. This particular structure, based on the cylindrical condition and referred to as an A-cylindricity, had the dual effect of guaranteeing that there would be no physical evidence of the fifth dimension while reducing the number of variables from fifteen to fourteen, solving both of the problems incurred by the addition of another dimension.

There are no intuitive guidelines or experiences on which to base this model of space-time, so other guidelines must be taken into account to derive a model of the fifth dimension. The cylindrical condition allows the four dimensions of space-time to be independent of the fifth dimension. After a manner of speaking, the cylindrical condition thus serves to explain why there is no physical evidence of a fifth dimension. All observables associated with physical phenomena are four-dimensional and thus independent of the fifth dimension. The cylindrical condition also limits the kind of coordinate transformations possible, allowing only those which lead to covariant field equations (Tonnelat, 1966a, 7) since the fifth coordinate must play a spatial role. The special role played by the fifth coordinate is evident under the cut-transformation, whereby anti-symmetrical derivatives of the A-curve vanish while the anti-symmetrical derivatives of A_{μ} remain allowing a correlation with the magnetic field. The cylindrical condition is necessary to the successful unification of the electromagnetic and gravitational fields in Kaluza's theory.

In fact, Kaluza discovered two transformations that would leave the equations invariant and thus preserve the unique character of the system. Invariance under transformation is a necessary property for both the gravitational and electromagnetic fields. The two transformations that leave the components invariant are the "four-transformation," over four dimensions, and the "cut-transformation." The field variables on which these transformations act can be grouped into three general classifications corresponding to the results of the transformation process; the γ_{mn}, γ_{0m} and γ_{00}, where m and n vary from 1 to 4 representing our four-dimensional space-time. The γ_{mn} correspond to the sixteen components of a matrix representing the four-dimensional space-time continuum as in GR. They reduce to ten independent components that describe gravitation. Under the four-transformation, the γ_{mn} act as a four-tensor and under the cut-transformation they are invariant. The γ_{0m} correspond to the eight components (or four + four) in a five by five matrix that represents the mixed terms of normal space-time and the fifth dimension.

$$\begin{pmatrix} \gamma_{mn} & \gamma_{m0} \\ \gamma_{0m} & \gamma_{00} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \text{GR} & \text{EM} \\ \text{EM} & +1 \end{pmatrix}$$

Under four-transformation, the γ_{0m} act as a four-vector, while under the cut-transformation they vary by an additive term. The variation due to the additive term allowed the introduction of the electromagnetic four-vector into the new space-time structure. The γ_{0m} were equated to the electromagnetic potentials φ_m since "this corresponds to the fact that the electromagnetic potentials are defined only up to additive terms which are gradients of an arbitrary function." (Einstein and Bergmann, 687)

The final term, $\gamma_{00,}$ is purely fifth dimensional. It was set equal to +1 in Kaluza's original theory. This term is invariant and constant under both transformations and thus proved to be effectively removed from the space-time structure, as we perceive it. In this way, the dependence of the field structure on electromagnetism as well as gravitation was reflected in the metric, which defines space-time. Electromagnetism was wholly incorporated into the new field structure, thus completing the correlation to the electromagnetic field while all of the components of the new unified field were accounted for. Kaluza's theory was simple, elegant, reproduced both the electromagnetic and gravitational fields from within a unified metric structure and yielded the geodesic equations for both charged and uncharged particles within the combined fields, but his findings were not without controversy.

Within the context of what he was trying to accomplish, the unification of electromagnetism and gravitation into a single hyper-field structure, Kaluza was moderately successful. However, within the context of a mounting tide of criticisms, the success of the quantum theory and the discovery of new forces in nature which had not been foreseen in the early 1920s, the success of Kaluza's theory was short-lived and generally overlooked for almost fifty years. Since Kaluza placed no physical significance in his fifth dimension, using the concept only as a mathematical tool with which to derive his goal of unification, he opened his theory to excess criticism. It was severely criticized by some for going too far just by introducing the fifth dimension and by others for not going far enough and adding some physical significance to the fifth coordinate once it was used. Other criticisms were leveled concerning the variable representing the purely five-dimensional characteristics of space-time, γ_{00}. Setting this variable equal to +1 seemed either unnecessary or unwarranted. Both A-cylindricity and the correlations drawn by Kaluza between the field constants in his model and the known field constants of electromagnetism also seemed artificial in some respects. Thus his theory seemed rather *ad hoc*. (Graves, 257) Kaluza's theory was further criticized for not making any predictions that would allow it to be tested. It merely reproduced electromagnetism without expanding or adding to the already existing Einstein-Maxwell equations. Yet all of these criticisms are indirectly concerned with the question of the reality of the fifth dimension. So, the only real criticism of his theory can be posed in the simple question, Why add a fifth dimension when all physical evidence implies a four-dimensional space-time?

After two decades of pursuing his own extension to Kaluza's theory, Einstein's final comment to this question was given in the second appendix to the fourth edition of *The Meaning of Relativity*. He stated that any hyper-dimensional theory could only be considered a valid theoretical option when it could be shown why all empirical data leads to a strictly four-dimensional world. (Einstein, 1956, 166) In other words, in the absence of any observational or experimental evidence of the existence of a fifth dimension, scientists could only accept the hypothesis if there were an overwhelming reason to do so. Any scientist developing a five or higher-dimensional theory must not only contend with this problem, but also justify the basic assumption of the higher dimension by demonstrating that this hypothesis and only this hypothesis can account for observed natural phenomena.

The cylindrical condition is an extremely important component in Kaluza's theoretical model. In subsequent extensions of his theory, at least in those developed prior to the 1960s, the imposition of the cylindrical condition was an important point of criticism and constituted a major weakness in these theories. Yet the special status, or perhaps the peculiarity of the fifth component of the field, is revealed in this condition. For this reason, the condition was interpreted by some as being too restrictive or merely an 'additional' condition which was neither necessary nor justified. (Tonnelat, 1966a, 8) It was therefore thought possible that a condition less stringent than the cylindrical condition could be used to obtain the same results for the fourteen equations describing a combined field, while leaving the fifteenth equation intact to describe other field phenomena. So modifying the cylindrical condition was the first method of choice for extending Kaluza's theory during the early years. This approach would render the five-dimensional theory more general in its application to the physical world.

In the projective theories, such as the theory developed by Einstein and W. Mayer in 1931, the cylindrical condition was interpreted quite naturally as a projective condition that demonstrated the purely auxiliary role of the five-dimensional space. (Tonnelat, 1966a, 7) The cylindrical condition was also thought to lead to a mere codification within a five-dimensional formalism, such that it was a mathematical convenience rather than a physical characteristic of space. In that case, it was assumed that the five-dimensional space is real or that there is a true five-dimensional geometry that can describe space-time, rather than a geometrical (mathematical only) formalism representing space-time. In this instance, the cylindrical condition could be modified or dropped altogether. Einstein, Peter G. Bergmann and V. Bargmann in 1941 as well as J. Podolanski in 1949 took this approach to the problem. In their theories, the extra dimension, or dimensions in the case of Podolanski's theory, was considered to be real, but of special structure. Instead of a cylindrical condition, the theory of Einstein, Bergmann and Bargmann used a fifth dimension which was closed with respect to the four dimensions of normal space-time. Podolanski took another path and solved the problem by assuming a "laminated structure" such that all the points in a given layer correspond to a given point in the four-dimensional space-time (Tonnelat, 1966b, 403) continuum. Under such conditions, these theories were able to answer those criticisms that attacked the fifth dimension dependence on the cylindrical condition.

On the other hand, Kaluza's theory has become popular and gained a new respectability within the scientific community since the 1970s. The cylindrical condition has now become an important factor in the theory's newest incarnation rather than a point of criticism. In the latest extensions to Kaluza's model, the extent of the cylindricity is uniquely small limiting the fifth dimension to the domain of the quantum world. So the fifth 'contracted' or 'compacted' dimension is not perceptible in the common world of the four-dimensional space-time. It is also beyond the experimental capabilities of science at this particular point in history so it is not even possible to detect the fifth component of space-time in any manner at present. Oskar Klein first developed this particular extension of Kaluza's theory in 1926 and is today a part of the much grander model of physical reality known as the 'superstring' theory. In his original theory, Klein altered the cylindrical condition in order to accommodate the new developments quantum mechanics.

Klein's interpretation of the Kaluza theory

Klein has been credited with both the formalization of Kaluza's theory and several attempts to extend the theory into the domain of the quantum. His name has been so closely associated with Kaluza's theory that some scientists and authors have given

Klein partial credit for the original theory and refer to it as the Kaluza-Klein theory. This synthesis is so complete that some authors have given Kaluza credit for innovations that Klein made to the theory and vice versa. Kaluza's theory seems to have lain fairly dormant until Klein's first exposition of it appeared, so it is generally thought that Klein was instrumental in popularizing the theory. This interpretation of the historical events seems all the more probable since the majority of physicists were dealing with developments in quantum theory during the early to middle 1920s and would have ignored Kaluza's modification of GR since it offered nothing new to their interests in the microscopic domain of the quantum. However, Klein's alterations and extensions of Kaluza's work into the quantum domain rendered the theory more relevant to events occurring in the rest of the world of physics. To be sure, Kaluza's theory would have seemed inconsequential against the onslaught of quantum mechanics in the early 1920s. Most physicists would have ignored it until Klein related it to the newly forming concepts of quantum and wave mechanics whose interpretations within the larger framework of science had not yet been fully developed.

Klein saw within the five-dimensional hypothesis a vehicle for introducing the quantum into the space-time continuum rather naturally, as well as a way to account for the atomicity of electric charge. He first equated the geodesic in the fifth dimension to the periodicity of the electric potential N. This implied a quantum of action, while a conjugate momentum in the fifth dimension was fixed to account for the positive and negative electrical charges. By forming the five-dimensional Lagrangian of a particle in a combined electromagnetic and gravitational field, and then differentiating it with respect to the velocity along the fifth component, he established a relationship within the field yielding the charge-to-mass ratio of the electron. This allowed the conjugate of the fifth coordinate to appear in a manner analogous to the way that matter and momentum were conjugates in our normal four-dimensional space-time.

The periodicity that he introduced into the fifth dimension also allowed Klein to make an association between a function in the fifth dimension and Schrödinger's wave function, Ψ. Klein further derived a fundamental length of $l_o = (2k)^{1/2}$ (hc/e), where k is Einstein's gravitational constant, and h, c and e are Planck's constant, the velocity of light and the electron's charge. This configuration gave Klein's fundamental length a value of 0.8×10^{-30} centimeters. (Klein, 1926, 516) He later proposed "to relate the fifteenth quantity γ_{00} with the wave function Ψ, which characterizes matter, in order to achieve a formal unity between matter and field" (Klein in Mehra, 53) and thereby further cement the relationship between Schrödinger's wave mechanics and the five-dimensional framework.

Klein later admitted that his first theoretical attempts were not satisfactory and for the next decade he published nothing more dealing with them. (Klein in Mehra, 80) In 1939, he developed new extensions to his earlier theories of a grand unification between quantum and field theories by incorporating the newly discovered "mesotonic" forces of

Yukawa into his theory. Klein's "mesotonic" forces are better known today as the strong nuclear force. Klein reasoned that the mathematical treatment of the Yukawa potential was analogous to his earlier mathematical treatment of the five-dimensional framework. He wrote that the "direct and general way it expresses the fundamental conservation and invariance theorems seems to make this representation a natural starting point for a general quantum theory comprising also the charged fields, which are supposed to correspond to the mesotons." (Klein, 1939, 79) His newly extended theory included the construction of a new Lagrangian containing, in addition to the gravitational and electromagnetic components, the spinor as a tensor field component. By using a variational principle, actions between protons, neutrons and electrons, explained by the interactions of neutrinos such as theorized in Yukawa's theory of nuclear forces, were found.

In 1947, the theory was further extended when Klein developed field equations for free mesons and derived the wave equation for nucleons. To achieve this end, he had replaced the assumption that the field quantities were independent of the fifth coordinate by an assumption that they were periodic functions of a length in the fifth direction with a period of l_0. This development introduced the indeterminacy "which would exclude the use of the fifth dimension in any geometrical sense, and had the practical meaning that particles of given charges have naturally coherent wave functions, as is always assumed." (Klein, 1947, 3) A new fundamental length was also introduced which was equal to the product of his older fundamental length, l_0, and a constant equal to e^2/hc.

Klein quickly became dissatisfied with his 1947 theory. He thought that his theory had "such features that it should hardly be taken literally." (Klein, 1956, 59) So he made one last attempt to include nuclear forces in his five-dimensional framework by deriving, via the same periodicity function, "a theory of more physical aspect, whereby charge invariance appears as a part of a natural generalization of gauge invariance." (Klein, 1956, 59) As a further consequence of his concept of a fundamental length, Klein calculated that a particle, approximating a quantum in a linear wave equation and with a wavelength approaching zero, would have a gravitational self- energy approaching the kinetic energy corresponding to its volume. In this manner, he hoped to do away with the remaining divergences of the electron theory. This more stringent generalization of the theory from the quantum theory point of view, implied possible states of matter with a multiple charge. When all is considered, Klein's theoretical work could best be characterized as a continuing attempt to save the basic tenets of Kaluza's space-time framework while keeping pace with the advances being made in atomic and sub-atomic physics. Portions of Klein's work live on in today's most advanced theoretical research where the Kaluza-Klein theory is the foundation upon which the theory of superstrings has been developed.

The super theories

During the 1960s, Steven Weinberg and Abdus Salam used weak gauge symmetry to account for the masses of W and Z particles through a spontaneous breaking of gauge symmetry. This method allowed the unification of the electromagnetic and weak forces without depending upon the same type of renormalization process that was necessary in QED to prevent infinite masses.

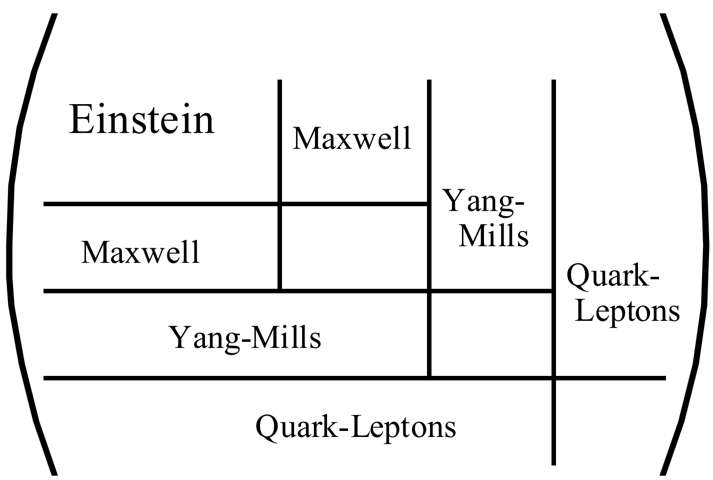
> These encouraging successes have led to the belief that the weak and electromagnetic forces are really two aspects of a unified electroweak force. However, ... perhaps amalgamated is a better word than unified. The crucial element in this success was the formulation of the theory in terms of gauge symmetries, and this has encouraged the theoretical examination of a variety of other gauge theories for the description of the strong and gravitational forces, and their eventual unification with the electroweak force. (Davies and Brown, 56)

With the success of this 'electroweak' theory, emphasis in the theoretical research of quantum field theories changed from inventing renormalization methods that had no physical basis but gave finite and interpretable solutions to applying the correct gauge symmetries. Unlike renormalization in QED, which has no physical counterpart, symmetries are a common characteristic of physical bodies and systems so renormalization in the Weinberg-Salam model became physically acceptable. Capitalizing on this success, the next wave of unification theories was based upon the various symmetries inherent in nature.

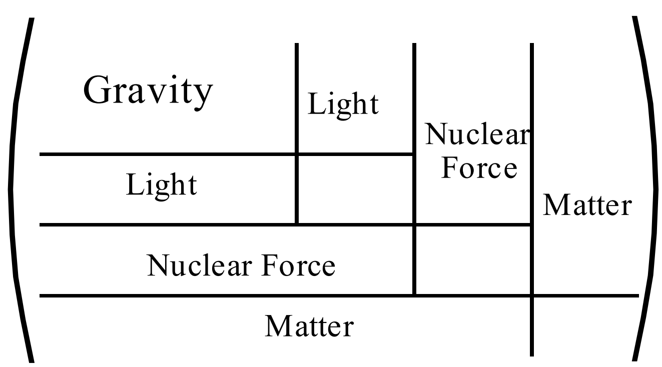
Grand unification theories (GUTs) and supergravity were both developments of the 1970s. GUTs were attempts to unify the electroweak theory with QCD, thus unifying electromagnetism, the weak and strong forces, by embedding the gauge symmetries of each of the individual theories within a larger all embracing gauge group. Unfortunately, the GUTs that were developed predicted the existence of magnetic monopoles and an extremely large but finite half-life for protons. In the ensuing years, neither of these predictions has been verified by experiment or observation, so the GUTs have been seriously hampered. The GUTs did not include the force of gravity and GR within their framework, so they were not TOEs in today's sense of the name, but the concept of supersymmetries was used during the same period of time to include gravity within the GUT framework. This class of theories is known as supergravity.

In the older quantum field theories, gravitational forces were always mediated by particles called gravitons. The graviton has never been observed in nature. The new supergravity theories predicted that not only the graviton acted as the conveyor of gravity, but a new particle called the gravitino should also exist in that role. Like gravitons, the gravitino was very weakly interactive with matter and would therefore be very difficult to detect in nature. In spite of this shortcoming, supergravity did introduce a new (or perhaps old and forgotten) concept into the search for a TOE. The geometrical structure of space-time could be greatly simplified if the unified force of supergravity was recast within an eleven-dimensional framework.

This discovery gave a new impetus to the search for hyperspatial theories of unification and physicists rediscovered the Kaluza-Klein theory in the early 1980s. When the theory of supergravity was rewritten as an eleven-dimensional Kaluza-Klein theory, all of the forces of nature were reduced to nothing more than different forms or adjuncts of a single gravitational field. Since the supergravity theory is an extended Kaluza-Klein theory, it can be represented as a matrix with the first four indices (variables or dimensions) representing the space-time of GR in the upper left hand corner, followed by electromagnetism for the fifth index and then the Yang-Mills field and the quark-lepton view of matter.



or



(Kaku, 146-147)

The extra dimensions had no physical meaning within the original supergravity theory. However, in the Kaluza-Klein modification they were interpreted as real physical dimensions that were rolled up in such minute proportions that they were effectively unobserved as well as unobservable in nature. The extra dimensions were associated with various abstract gauge symmetries that were independent of the minuscule size of the extra dimensions. Unfortunately, the supergravity theories suffered from a rather crucial and perhaps fatal flaw. The weak force violates a special type of left-right mirror symmetry, referred to as violating parity. This property is called chirality and can be shown to exist only in odd-dimensioned spaces. This property therefore requires any unified field theory that includes the weak force to use a framework with an odd number of spatial dimensions plus one time dimension. Such a configuration would yield a total number of dimensions that is even. The eleven-dimensional space-time continuum of supergravity is odd, so it clearly does not fulfill this requirement.

Meanwhile, lurking in the shadows of theoretical physics was an answer to this latest predicament. Even before the advent of the supergravity theories, the concept of strings had been introduced into the physics of quantum fields. The quantized motion of a vibrating string was first used by Gabrielle Veneziano to model hadrons. Then John Schwarz and André Neveu discovered a second group of strings for modeling fermions. When QCD theory was introduced, the string model was all but abandoned by its advocates. Yet Schwarz and Joel Scherk continued to develop the string model. While strings did not seem to correspond to any of the known elementary particles found in nature, they did have properties similar to gravitons, which suggested that they might be ideal for a TOE. No other quantum field theory had been able to account for gravitons and the gravitational field. Still, GUTs and supergravity overshadowed string theory and few in the scientific community paid it any attention. But string theory did profit from the prior development of these theories since they popularized and legitimized the use of hyper-dimensional models in physics and brought supersymmetries to the attention of physicists. Scientists no longer ignored theories, which assumed space-times of dimensions greater than four, and the symmetries seemed to justify their application.

Supersymmetry is deeper and more powerful than the normal symmetries of space and time. Its most endearing feature is that "it provides a geometrical framework within which fermions and bosons receive a common description." (Davies and Brown, 44) However, this supersymmetry requires the addition of five more dimensions of space to Kaluza's five-dimensional space-time. The concept is not without problems, since there is no "unequivocal confirmation in nature" of such supersymmetries. (Davies and Brown, 47) Yet the application of supersymmetry yields startling results. Simple string theory utilizes the supersymmetry to unify the four forces of nature within a single common geometrical framework and is thus called superstring theory. In essence, the supersymmetry allows matter and radiation to be combined. Within this framework hyper-dimensional strings can represent all elementary particles. Schwarz, Michael Green and Edward Witten developed the superstring theory in the early 1980s.

Superstring theory postulates a space-time continuum of either ten or twenty-six dimensions. Only these configurations give reproducible and understandable results. These extra dimensions are every bit as real as the normal three dimensions of space, but we cannot detect or otherwise perceive the higher dimensions because they are curled up or contracted to Planck length sizes, about 10^{-33} centimeters.

> These fundamental strings possess a tension that varies with the environment in which they reside and this tension becomes large enough to shrink the loops of string to approximate points at the low energies we witness in the universe today. ... The enticing aspect of the string theories has been the unexpected discovery that the requirement of finiteness and consistency alone should prove to be so constraining. (Barrow, 31-32)

This restriction guarantees that the strings are of such minute size that they are virtually impossible to detect. Such small sizes can only be reached experimentally by energies far greater than any that science can even dream of at this time, let alone reach in high energy physics laboratories. With the direct detection of strings so far out of the experimentalist's grasp, their existence cannot be directly confirmed. So, belief in the validity of the theory must depend on its 'beauty,' simplicity and logical structure. It is hoped that the mathematics will eventually yield predictions that will confirm the theory at much lower, and thus attainable, energies. However, even this path presents a problem since mathematics is not yet advanced enough to solve the problems associated with the superstring theory. Only approximate solutions to the superstring model exist. Many scientists consider superstrings a theory of the next century that happened to fall into this era accidentally. (Witten, 102; Kaku, 160)

Quantum field theory assumes that particles are non-extended mathematical points in space. This practice led to the infinite masses and meaningless divergent expressions that required renormalization. In this practice, quantum field theory is not alone. In both quantum mechanics and classical mechanics, in fact in all other branches of physics, particles are portrayed as points without internal structure. Science has generally shied away from questions concerning the interior portion of elementary particles. The mathematical point assumption is no longer necessary in the superstring model. Superstring theory replaces these points with one-dimensional curves called strings. One advantage of discarding points for strings is the disappearance of divergences, but there are other advantages such as the explanation of anomalies. Speaking from his own experience, Schwarz explains that he and his colleagues were surprised at this result in superstring theory. When "the quantum corrections to gravity for string theory" were made, he and his colleagues began "to get numbers that did make sense, numbers that were given by finite expressions." (Schwarz, 75) Simply put, the theory overcomes many of the major problems inherent in earlier quantum field theories such as supergravity and GUTs.

In superstring theory, fermions are particles of matter and bosons are the particles that interact between material particles as the forces of nature. In the case of gravity, superstring theories succeed where all of the quantum field theories have failed, allowing a unification of gravity with the other forces within in a single field theory. In fact, gravity is of fundamental importance to the superstring theory.

> The most remarkable feature of string theory, as we have emphasized, is that Einstein's theory of gravity is automatically contained in it. In fact, the graviton (the quantum of gravity) emerges as the smallest vibration of the closed string. While GUTs strenuously avoided any mention of Einstein's theory of gravity, the superstring theories demand that Einstein's theory be included. (Kaku, 157)

Witten's characterization of the role of gravity in superstring theory is still stronger. The theory is especially attractive to scientists because gravity is "forced upon them" by the theory itself rather than being incorporated from outside the theoretical framework. "All known consistent string theories include gravity, so while gravity is impossible in quantum field theory as we know it, it's obligatory in string theory." (Witten, 95)

The superstring framework also has intuitive appeal, which is something that quantum field theories lack. The analogy with ordinary vibrating strings is quite a strong tool for physicists. With this analogy, they have been able to develop a picture of what occurs on the sub-quantum level of physical reality. Particles are actually little loops of string, like "lassos," which oscillate about as they move through space. As time moves forward, these loops describe "something which is rather like a tube going through space and is called a 'world-sheet.' That is the trajectory of a particle according to the superstring idea." (Ellis, 152) The "world-sheet" of superstrings corresponds to the 'world-line' in the space-time framework of special relativity, which further cements the relationship between relativity theory and superstrings.

While following their more classical trajectory through space and time, the string is also vibrating (non-classically) in the higher dimensions of space. The modes of a string's vibration determine the particular characteristics of the particle as well as its type and class. According to Schwarz

> When you have a string it can oscillate and vibrate in different ways - rotate and so forth - and each of these different modes of vibration or oscillation can be thought of as describing a particular type of particle. So one can think of the electron as one mode of vibration, and a quark as another mode of vibration, and a graviton as yet another (Schwarz, 79)

Steven Weinberg agrees with this assessment and further adds that modes of vibration may even explain strange quarks and other esoteric particles in modern physics.

> The strings are vibrating in all these extra dimensions and that leads to a lot of different modes. It is the extra dimensions (or other extra physical variables) which produce the many different modes. In fact, that's one of the encouraging things about string theory. Because of that it's natural to find multiple generations of particles, not just the lowest generation with the light quarks and the electrons, but also the next generation which includes the strange quarks and the muons and so on ... (Weinberg, 217)

However, all of the particles with which physicists are familiar, such as electrons, protons, and neutrons, correspond to the lowest frequency mode of vibration. Other modes, representing higher frequencies with higher energies, are not normally seen or detected in nature. "The next mode would be hopelessly too heavy to have ever been seen." (Weinberg, 217)

And yet the common vibrational modes do not exhaust all of the possible actions of a superstring that could be utilized to determine any individual particle's physical properties. For example, the electric charge of an elementary particle could arise from some type of unspecified non-vibrational action. "In fact, what we call electric charge would be some sort of collective property of the string as a whole and if the string oscillated in different ways then it would seem to have a different electric charge." (Ellis, 154) In this respect, the basic electric charge on particles would not be a quantity that is just experimentally determined without any apparent connection to the rest of the universe. The fundamental electric charge would have a deeper hidden meaning that is related to the motion of the string. Specific characteristics of quarks could also be attributed to these non-vibrational actions of the superstrings.

> This superstring, in addition to oscillating in space, rather like a traditional violin string, also has certain internal degrees of freedom which you can't really visualize in terms of simple oscillations of space, and the actual difference between, say, an up quark and a down quark would presumably be some sort of a combination of these internal properties and these oscillations in space. (Ellis, 153)

There is very little in physical reality that cannot be attributed to either the various motions of superstrings or their existence.

Even the mathematical points that represent empty space may well turn out to be no more than superstrings.

> The notion of a string is inseparable from the space and time in which it's

> moving, and therefore if one has radically modified one's notion of the particle responsible for gravity, so that now it's string-like, one is also forced to abandon at some level the conventional notions of the structure of space and time. When I say at some level, the level I'm talking of is at these incredibly short scales associated with the Planck distance. (Green, 125)

In relativity theory, space and time only exist relative to the bodies with mass that constitute the universe. In the general expression of this idea, matter either curves space-time or matter reduces to the curvature of space-time. So, if curvature is the true reality and is associated with material particles and thus superstrings, it is not a far stretch of the imagination to speculate that relative points of space are themselves superstrings. Each and every point of space could be no more than an object of higher dimensionality that is curled up in a little ball or loop. The group of all such loops would constitute a "stringy space-time" which is an "approximation" of a far "richer structure" of superstrings. (Green, 131) All of physical reality seems to be covered within this notion. Superstrings have become the material particles, the forces between material particles as well as the relative point positions between particles, leaving little else to exist in the physical world. Thus, the superstring theories are TOEs, quite literally, if they can fulfill their promise.

Superstring theories are at present the best contenders for a TOE, if indeed such a theory is possible. At least physicists are now thinking in terms of a theory that covers everything in nature. So they believe a TOE is a distinct possibility in the near future. It would be prudent, then, to ask just what one might expect of a theory that seems to cover 'everything' in its wake. Davis and Brown have considered just this question.

> What should we expect from a truly satisfactory TOE? First, it should explain why physicists observe the various elementary particles that they do, and correctly predict all of their key properties such as mass, electric charge, magnetic moment and so on. Second, it should faithfully describe all the interactions between the particles, which means that it should account for not only the four fundamental forces of nature, but also their relative strengths. Calculations with the theory also ought to yield precisely the observed values of the various inter-particle scattering amplitudes, decay rates, branching ratios, etc. In short, the theory should account for all the measured parameters of particle physics. In addition to this, it should provide an explanation for the geometry and topology of space-time, such as the number of perceived dimensions, and offer a convincing account of how the universe came into existence.
>
> But this is not all. A TOE should also *unify* physics. (Davies and Brown, 5)

Their criteria are simple. A TOE must explain (1) matter, (2) the forces affecting matter, and finally (3) the space-time framework of matter as well as unify the quantum and relativity theories. Although they have placed the unification of physics last in their own wish list, it should be placed first. The foremost task in theoretical physics should be to unify the quantum and relativity, the discrete and continuous aspects of physical reality and nature. Whether or not a theory of 'everything' is even possible under these or any criteria is still an arguable point as John Barrow has pointed out. (Barrow, 230-231, 282) But unification in physics is an essential task independent of a theory that describes all in nature.

It is also questionable whether superstrings can fulfill this notion of 'everything.' The problem of defining and understanding space and time is quite formidable. Faraday ran into a similar problem a century and a half earlier when he tried to conceive the 'continuous' electric and magnetic fields between particles. He sidetracked the problem of continuity by talking about the 'contiguous' points of charge in space that carry the electromagnetic field. Historians and philosophers of science still argue about the meaning of Faraday's use of the word 'contiguous.' There are also modern analogies to this problem. A few decades ago, quantum theorists also speculated about the discrete nature of both time and space, but no new science ever came from this speculation. The modern superstring theories carry the same stigma. At best, they can only speculate about the actual points of space as 'curled up dimensions.' However, this view is no solution to the discrete/continuous debate. It merely forestalls the debate to a later point in time and a much smaller unit of discreteness.

Beyond the already stated problems with superstrings, two more problems are inherent in these last speculations. (1) If space is no more than little 'loops' of curled up higher dimensions, then what are those little 'loops' moving within when they follow trajectories through time? And this leads to the next problem. (2) What then is time? A theory of everything should make some definitive statement about the nature of time, but superstring theory does not seem to do so. These are only some of the fundamental problems with the theory of superstrings and they are not the only problems. There still remain the obvious difficulties that are overlooked or shunted aside by stating that superstrings are a theory of the future. The makers of superstring theories can safely defer confirmation of their theory to late in the next century if not later, which seems to forestall any attempt at falsification of the theory. And finally, there is one last problem that has rarely been mentioned within the context of superstrings.

During the past two decades, another important trend has developed in physics. This trend is to define or discover the relationship between consciousness and physics and solve the mind/body dichotomy. Coincidentally, these questions have been raised within the same historical time frame as the change in scientific attitude toward the acceptance of the possibility of a TOE. This coincidence would seem to indicate that the development of a TOE and the discovery of a role in physics for mind and consciousness are connected, but superstring theorists have not adequately addressed these questions. So far, superstring theory is a purely physical theory without room for consciousness so it cannot, at this time, be used to make any statement concerning the mind/matter paradox. These criticisms serve to emphasize the fact that there is surely more room in modern physics for contending theories of unification. There are certainly other options for unification without the premature declaration that superstrings represent the last word in physics and the ultimate TOE.

Universality and the single field theory
a. Philosophical arguments for an extra dimension

Supergravity, GUTs and superstring theories have popularized, sanitized and legitimized the concept of higher physical dimensions of space. However, the concept should stand on its own merits without reference to these modern interpretations. The question of why a fifth dimension (and/or higher dimensions) should be adopted in science is of far greater significance than any one theory. The concept should be considered independent of these or any single physical theory. Indeed, it is a question that has plagued science for several centuries, but which gained more immediacy only after the popularization of the new non-Euclidean geometries during the middle of the nineteenth century. In answer to this question, there are several simple logical arguments. Although they offer no absolute proof of the existence of a higher dimension of space, they do lend some credence to the possibility.

There is a very simple and straightforward argument in favor of higher dimensions that dispenses with all but the simplest of mathematical forms. It is a very well known fact that a mathematical point has no dimensions. A line is one-dimensional, a surface is two-dimensional and a solid is three-dimensional. However, these geometrical figures are mathematical abstractions. A real point is not dimensionless just as a real physical line is not one-dimensional. A real physical line must have a thickness so it is at least a two-dimensional object. A two-dimensional surface must have a thickness in a third dimension to be assured of physical existence. But these are still abstractions. Physical reality has at least one more dimension than the corresponding mathematical model of reality. Real physical objects are represented mathematically by three-dimensional geometries. Therefore, by extrapolation, a real physical object must have another dimension to be physical. Real physical bodies must be four-dimensional; they must have four spatial dimensions.

This argument actually strikes at the heart of the problem for quantum field theories and other forms of physical theories. Superstring theory has the advantage over previous theories because it essentially adds one more dimension to the physical (and dimensionless) points that past theories assumed. This procedure is not without precedent in the history of physics, as most people should know from their fundamental physics course. A little more than a century ago, statistical methods were employed in the kinetic theory of matter (and thus thermodynamics) to successfully derive the Ideal Gas Law.

The derivation assumed that gas molecules were perfectly elastic and dimensionless point particles and were thus unaffected by short-range forces. These molecules had no volume. These properties allowed the derivation of the correct and well-known gas law. However, when these criteria were abandoned and the dimensionless molecules were given a small but finite volume which allowed the interaction of forces at close range, the kinetic theory of matter was not only able to explain the transition of gas to liquid and liquid to solid, but also describe the structure of matter within each of the three different phases. In similar manner, the one-dimensional string of superstring theory has replaced the dimensionless physical points of mechanical theories thereby explaining gravity as represented by the space-time curvature. Superstring theories need that extra dimension just as a three-dimensional object (as perceived) necessitates a four-dimensional extension (which is not normally perceived). But even the superstring theory is an admitted approximation.

Our fundamental concept of space was developed prior to Newton with the work of Francesco Patrizi and Pierre Gassendi. Newton synthesized the views of Patrizi, Gassendi and others into a new concept of absolute and relative spaces to be used in the support of his mechanical system. Patrizi's was the first concept of space that bore a resemblance to Newton's and our own intuitive concept of space. He considered two cases for space. First, that space was a container and second that space extended beyond the boundaries of the material universe. He also countered the Aristotelian argument that space was not a 'thing.' If space was 'no-thing,' it could not be considered a legitimate subject for mathematics or science. Since empty space was 'nothing,' it could not be subdivided mathematically for analysis.

Space was irreducible in the physical sense, differing from mathematical space and therefore impervious to physical abstraction. Of course, that may or may not be true today. Modern science has developed the relativistic viewpoint that accepts space as a 'nothing,' even though it has properties. Modern science has reduced space to the relationship of positions between real physical objects. Given today's concept of symmetries, the 'nothing' called space has still more properties than twenty years ago. But in the older view, even an infinite number of dimensionless points could not constitute an extended space. Quantum mechanics overcame this problem by adopting probabilities and uncertainties. A probability distribution describing a particle can be continuous across an extended volume of mathematical points of space because it has no physical reality. In wave mechanics, the wave function is continuously extended across all of space and has been interpreted in quantum mechanics as corresponding to the probability function. The physical wave function allows the mathematical probability distribution its physical presence.

But the collapse of the wave function must occur at a real extended position in space, so the collapse of the wave function marks deterioration from a mathematical continuity to a discrete physical reality. Dimensions have been created out of nondimensions to replace non-extended mathematical points with extended three-dimensional physical points. This process is highly questionable, so the probability distribution is deemed physically real instead of mathematically real. This would imply that the probability interpretation of the wave function is erroneous. The mathematical continuity of the probability distribution merely overlaps the physical continuity of the wave function; they are not the same thing. Nor is there a concept of space in quantum mechanics that rivals the space-time of relativity. The probability distribution merely mimics the continuity of space, which presents a major paradox for quantum field theory. Since the probability distribution only corresponds to the wave function, without being the wave function, a new interpretation of wave mechanics is demanded. The wave function has a far richer physical structure than just a probability distribution, as suspected by Einstein, Erwin Schrödinger and Louis deBroglie. It must have a real physical interpretation since it 'collapses' to a real physical point in space. Others have noted the discontinuity problem of the 'collapse of the wave packet' as a serious problem for quantum mechanics, but never within this particular context.

This demand coincides with the need for another dimension of space. The present interpretation of both quantum and wave mechanics depends on mathematical points of space whereas superstring theory has demonstrated that a complete description of the forces of nature requires an extra dimension that the dimensionless point cannot contribute. A real string is not a one-dimensional object even if its mathematical model is one-dimensional. To use an old analogy, these are all cases of scientists and mathematicians mistaking the finger pointing at the moon for the moon itself.

These issues are related to the second philosophical argument for a higher dimensioned physical reality. It is modeled on Gödel's theorem. Kurt Gödel demonstrated that a mathematical system could not be proven true or untrue from within that system; or rather its logical consistency could not be decided. A mathematical system depends on stated fundamental axioms, so any proof of consistency (truth) within that system also depends on the same fundamental axioms. If the system is 'proven' to be non-consistent, then the axioms are false, but then so is the 'proof' that was used to demonstrate the non-consistency. A mathematical system is therefore denied falsifiability in the Popperian sense when the issue of consistency is determined from within the system. All arguments within the system are based upon the same axioms or rules that they are trying to disprove, so the consistency (truth) of the system can only be judged from outside of the system. The consistency proof of a mathematical system must rely on a larger and more general set of axioms than those used to establish the system that it is testing.

In a similar manner, a physical theory can never be proven absolutely true in our physical world. It is a philosophical impossibility to prove physical truth. That is why we have theories that can change and grow instead of absolute physical laws that never change. Yet a physical theory can be proven valid (although not true) by falsification. A theory must be falsifiable to determine its validity. Our physical world is a logically consistent system based upon the physical equivalents of axioms. The equivalents are the conservation laws, symmetries, other principles and various rules of nature that exist under different names. Our universe may be quite complex, but it is nonetheless a system that makes it possible to mimic physical reality with a mathematical model. As a logical system, physical reality is also subject to Gödel's theorem even though it would be too difficult a task to find all of the physical axioms upon which the system is built. It is the primary function of science to discover these physical axioms and develop a model of reality from them. To discover all of the 'axioms' of this natural system is the first step toward developing a TOE.

As a self-consistent system (we must take it on faith that the universe is self-consistent or science fails), physical reality must follow Gödel's theorem. So physical reality must rely on something outside of itself merely to exist. Applications of Gödel's theorem in fields other than mathematics are not all that uncommon, including in the field of physics and the search for a TOE.

> The reason why mathematics is so successful in describing the way the world works is because the world *is* at root mathematical. Any limitations to mathematical reasoning, like those uncovered by Gödel, are thus not merely limitations on our mental categories but intrinsic properties of reality and hence limitations upon any attempt to understand the ultimate nature of the Universe. (Barrow, 248)

Our physical reality is part of a larger logical system whose existence verifies the existence of our four-dimensional space-time continuum. At least one higher dimension is indicated by the previous arguments as providing this larger required physical system. A higher-dimensional world, or rather a higher-dimensional extension of our world, is necessary for the existence of our four-dimensional physical reality.

In a rather strange sense, the criteria of a TOE even imply this. According to Davies and Brown:

> The ultimate TOE would, ideally, need no recourse to experiment at all! Everything would be defined by everything else. Only a single undetermined parameter would remain, to define the scale of units with which the elements of the theory are quantified. This alone would be fixed empirically. (In the ultimate case, experiment merely serves to define a measurement convention. It does not determine any parameter in the theory) (Davies and Brown, 7)

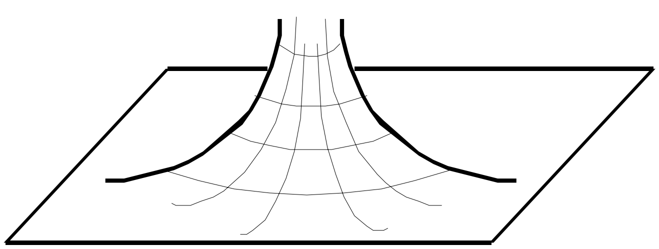
In an off-handed way, they have confirmed that Gödel's theorem applies to our physical reality. The ultimate TOE that they envision would not be falsifiable so there would be no way to validate the theory. Even then, it couldn't explain everything. There would have to be one independent variable (or axiom in mathematical terminology) to justify existence, just as there would have to be an independent axiom within a larger more comprehensive system to prove the consistency of a less complete mathematical system. However, the TOE that Davies and Brown foretell is not possible. They require a single independent variable, as verified by experiment, to define the "scale of units." But how could this be? What could it define the scale of units relative to, since there is nothing outside nor nothing more than the 'everything' covered by the theory. Their parameter could only gauge or scale measurements relative and internal to the system and would therefore be dependent upon the system, not independent of the system as assumed. What is needed for the system that is covered by their TOE is a larger more comprehensive system, another higher embedding dimension than their theory assumes.

At a more mundane level, the best scientific and practical reason for adopting a fifth or higher dimensions is inherent in the superstring theories, although not unique to them. Higher dimensions offer more degrees of freedom for explaining the more paradoxical physical properties of four-dimensional space-time. But does this mean that hyper-dimensional theories are more expedient for the theorist or does it mean that there are real directions of space other than length, breadth and width that are perpendicular to all three at the same time? If extra dimensions are to have any physical meaning, they must be real orthogonal extensions of our four-dimensional space-time. Otherwise, the extra-dimensions reduce to mathematical gimmicks that only serve to explain (or explain away) unwanted or misunderstood physical properties and parameters. Whatever the case may be, Einstein's requirement still remains the best guide for the adoption of hyperspatial theories. No such theory can be taken seriously unless it can explain why we cannot 'experience' the higher dimension or dimensions.

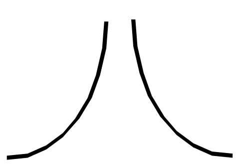
While these reasons for adopting a higher dimensional space are straight forward, they do not exactly demand such a hypothesis. There is, however, one problem in modern physics that has been grossly overlooked yet has a direct effect on the physicists' model of reality within this context. It demands a higher dimension for physical space-time. Scientists have long ignored the gravitational forces within the domain of single atoms, which means that they have also ignored space-time curvature within the confines of the atom. Gravitational forces within the atom are so small compared to the strong forces within the nucleus and the electromagnetic forces outside the nucleus, that they have been considered inconsequential. Yet the mass of electrons, protons and neutrons is an essential element of any calculated quantities within the atom and mass is related to space curvature according to GR. Theories of the nucleus, which have never been totally successful, have always depended upon quantum explanations even though the nucleus can be represented by singularities in the space-time continuum of relativity theory. It would seem that when 'everything' is taken into account, neither the gravitational forces within the atom nor the curvature associated with the nucleus could be ignored if quantum field theory is ever to offer a 'complete' description of nature. Any theory that claims to represent a unification of the quantum and relativity, whether a TOE or not, must address this apparent paradox.

Both the questions and solutions regarding this paradox are centered on the concept of the singularity in GR. There is a strange parallel between the problems raised by singularities within the gravitational field and the convergent infinities at the point location of particles in quantum mechanics. Since both theories fall apart under the same extreme conditions, within the interior of elementary particles, one would suspect that the interior of material particles could offer a point of connection between the two theories as well as a point of unification for the concepts of continuity and the discrete nature of matter. The singularity of GR is a discrete disruption of the smooth flow of continuity described by the field. So, understanding what happens to the space-time continuum within the boundary of elementary particles offers the best hope of unification in physics. On the other hand, a look at how physics treats other singularities in the gravitational field offers the best hope of solving the problem.

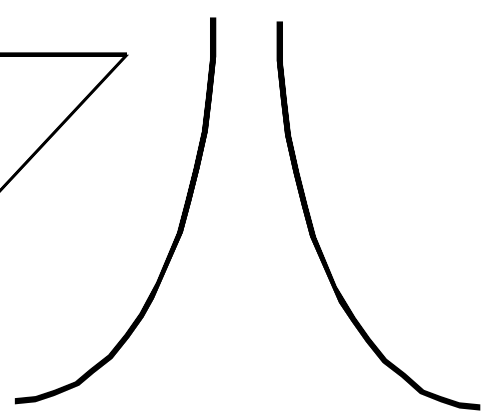
In particular, the same mathematical singularities that are used to model very massive bodies such as black holes are also used for elementary particles. Yet these are distinctly different cases. What are the physical differences between the singularities of particles and those representing very massive bodies? Mathematically, there may be no qualitative differences, but physically there must be a difference. Massive bodies are, at most, a collection of elementary particles crowded in close proximity with their surfaces in contact whereas it can be assumed that the interior of elementary particles is continuously curving. The difference is difficult to understand, but can be portrayed graphically.



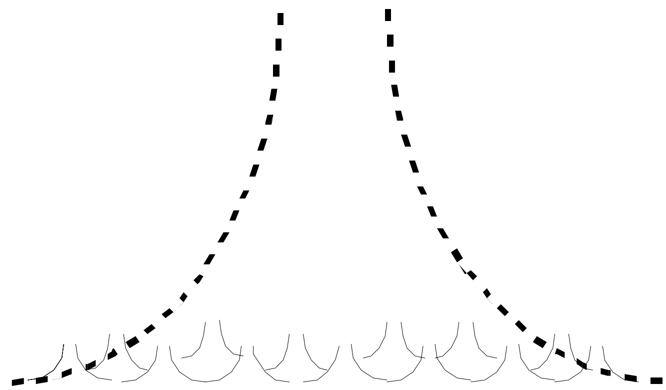
A singularity in space-time



An elementary particle
is a small singularity



The black hole is just
a larger singularity



However, a black hole is actually a
large group of tightly packed elementary particles or small singularities.

There is a grave discrepancy between these two views of a black hole, or for that matter any large accumulation of matter which creates a singularity in GR. How can the individual particles add together to give a mathematically singular entity? These diagrams depict a 'real physical curvature' in a higher dimension.

Within the context of mathematics, curvature is a property of a space or manifold, regardless of whether it is an embedded space or not. The term 'curvature' has a specific and understandable meaning as an intrinsic property of space just as it does of an extrinsic property. However, within the above context, a 'real physical curvature' refers to a 'physicalist' concept of curvature that 'requires' a higher embedding space or manifold. Under these circumstances, the accepted mathematical model of GR favors neither a higher embedding dimension, which implies an extrinsic space curvature, nor an intrinsic space curvature, which requires no higher dimension. The curvature of physical space-time has traditionally been treated as an intrinsic property of the four-dimensional manifold even though either the theory or the mathematical model does not require that treatment. Since the original development of GR, scientists have merely assumed that curvature is either an intrinsic property of the space-time continuum or a purely mathematical property describing space-time rather than a feature of physical space-time itself requiring a higher embedding dimension. In the opinion of John C. Graves, GR does not stipulate the necessity of the intrinsic case of a four-dimensional curvature over the extrinsic case evident by manifolds of higher dimensions. He has written that

> Mathematically it is probably simpler to deal with flat spaces of more than four dimensions than with non-Euclidean four-spaces. And if the notion of higher dimensions with regard to physical space seems incomprehensible, one might say the same about curved space that cannot be embedded. ... Finally, there is no obvious explanation within GR of why space-time

> should have four dimensions, no more and no less - the formalism itself makes no reference to dimensionality. (Graves, 192)

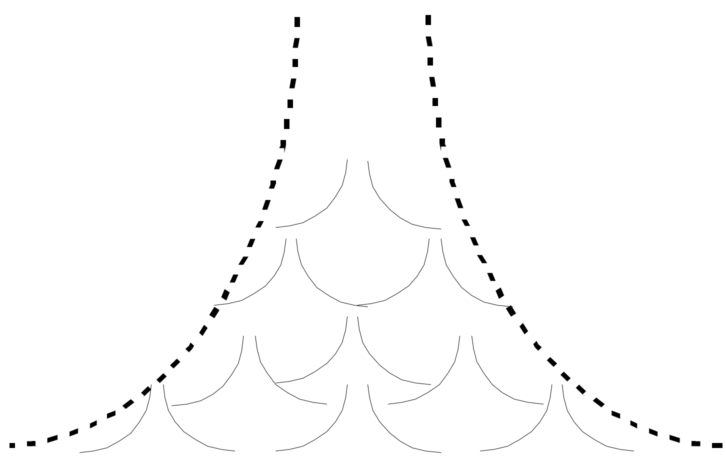
Even though an extrinsic curvature of the space-time continuum would necessitate at least a real five-dimensional manifold in which the four-dimensional space-time continuum is embedded while an extrinsic curvature does not.

On the other hand, it could also be argued that GR accounts for an intrinsic curvature alone. In other words, the Riemannian curvature tensor could be considered an inherently intrinsic object of four-dimensional space-time. According to the mathematician J.J. Stoker,

> In fact, throughout this chapter, as its title indicates, the inner or intrinsic geometry of surfaces will be studied extrinsically, and for two reasons:
> 1. It is both interesting and important for its own sake to characterize intrinsic properties extrinsically.
> 2. Such a procedure, as was remarked earlier, can be carried out in a fashion that furnishes clues for the generalization to Riemannian geometry in a manifold of any dimension without the necessity of embedding in a Euclidean space.
>
> A purely intrinsic treatment of Riemannian geometry is desirable for obvious reasons, and also for a reason rooted in physics: it would seem rather strange to treat Einstein's general theory of relativity, which is basically the Riemannian geometry of a certain four-space, by first embedding it in a higher dimensional Euclidean space, since Einstein's object was to investigate the character of the actual space in which we live - and to find that it is not Euclidean. (Stoker, 153)

Stoker's purpose is not to develop a 'physics' of space-time, but to investigate an interesting mathematical problem or model. However, he does relate his studies to physics in the form of Einstein's GR and in so doing points out that using a higher-dimensional geometry to model physical space-time would indeed seem "rather strange." Many physicists share this opinion, but it is not a requirement of the mathematics used to model space-time in GR. On the other hand, Charles Misner, Kip Thorne and John A. Wheeler have pointed out several methods of deriving "Einstein's field equation" as found in GR. Among these methods are included two cases based upon higher-dimensional embedding manifolds. One case proceeds from considering the "physics on a space like slice or hyperspace of simultaneity" using electromagnetism as the model. The second method proceeds from considering an infinitely dimensioned super space composed of points, each of which describes "a complete three geometry ... with all of its bumps and curvatures." (Misner, Thorne and Wheeler, 419-425) In each of these cases, the curved space-time of GR displays both intrinsic and extrinsic characteristics.

The normal interpretation of GR utilizes only an intrinsic curvature that cannot be easily portrayed or explained, as Graves has stated above. Embedding space-time in higher dimensions allows a greater versatility in accounting for physical phenomena that occur within the normal four-dimensional space-time continuum. Again, according to Graves, one advantage to using this extra-dimensional embedding structure is that it "allows us to define new concepts which might help characterize our geometric structures in an especially revealing way." (Graves, 193) But this advantage has been lost since theoreticians have traditionally adopted the intrinsic model of curvature as adequate to describe physical reality. Why then would embedding dimensions be either desirable or necessary? Kaluza's answer would be 'so that electromagnetism could be unified with the gravitational field.' On the other hand, if the curvature of the individual particles would somehow be additive in the higher dimension, the structural differences between different physical types of singularities could be accounted for. In the case where space-time is strictly four-dimensional and the curvature intrinsic, the additive effect of curvature could not be so easily explained. However, a real fifth dimension displaying a real physical curvature could easily account for the additive nature of the curvature.



The more particles that are packed tightly together, the greater the overall curvature into the higher dimensions. If space-time had a minute 'thickness' in the fifth direction, acting as a four-dimensional sheet, then the curvature outside of the physical boundaries of individual particles could add together in the fifth direction when the particles are in close proximity to one another. The closer material particles are packed together, up to the point where their physical surfaces come into contact, the greater the additive affect of their individual curvatures. In other words, if the fifth dimension is real and the four-dimensional space-time continuum is like a sheet, then individual particles in close proximity to one another can ride up 'higher' along their neighboring particles' curves in the fifth direction. The closer the proximity of particles, the higher they are pushed into the fifth direction in the geometric center of the overall body. The overall curvature of an extended material body is a consequence of its density rather than the total mass. The additive affect requires a higher embedding dimension to explain the curvature of collections of particles.

The philosophical points of argument beg the question whether they have physical validity or not which is not true for the physical arguments. The philosophical points of view are scientifically valid only in so far as they agree with nature. As for providing evidence for developing or providing clues to a new theory, they can only create a little smoke from smoldering, but no fire. The physical arguments are different. They hold more weight when developing or deciding between competing theories. Yet philosophical arguments have the power to either advance the cause of theories or destroy them. These arguments can still be used as scaffolding to support a theory, but only as ammunition for the acceptance or perpetuation of a new theory. They can help interest scientists and scholars in an idea and they can create attitudes, climates of opinion and advertise a theory, all of which are instrumental steps in the acceptance of new theories. They can also destroy or retard the acceptance of a new theory as they did in the case of Kaluza's theory and could thus act as a detriment to science.

Undoubtedly, Kaluza had a very good idea, as modern versions of his theory seem to indicate. But his theory was ignored by all but a handful of scientists until the past two decades. His theory successfully duplicated the Einstein-Maxwell equations, a fact which cries for some answer to why it was so unpopular. It created a little heat and perhaps some sparks, but never a fire under the scientific community of his day. The only reason that can be found for this mystery is the fact that the philosophical arguments against Kaluza's theory were overwhelming when considered against the background of the newly developed quantum theory coupled with the lack of perceiving the higher dimension upon which he based his model. The philosophical arguments provided above indicate the existence of a fifth embedding dimension to our space-time continuum and the physical arguments strengthen the case even more, justifying the adoption of Kaluza's model, with appropriate modifications.

b. The basic assumption and its consequences

Today's climate of change is more conducive to new and innovative concepts, although any new ideas must still have some scientific basis. This fact is demonstrated no more clearly and thoroughly than with the success of the multi-dimensional concept in the theory of superstrings. In the latest applications of higher dimensions, it has been assumed that Klein's interpretation of Kaluza's theory is absolute and Kaluza's cylindrical condition describes minute cylinders of curled-up space-time. Kaluza seems to have been somewhat vague on this issue himself.

Although all our previous experience hardly provides any

> suggestion of such an extra world-parameter, we are certainly free to consider our space-time to be a four-dimensional part of an R_5; one then has to take into account the fact that we are only aware of the space-time variation of quantities, by making their derivatives with respect to the new parameter vanish or by considering them to be small as they are of higher order ("cylinder condition"). The fear that by this condition the introduction of the fifth dimension would be revoked is unwarranted, because of the linkage of world-parameters in the three-index symbols. (Kaluza, 62)

Kaluza's "cylinder condition" is only one of two options that fit the physical facts and even then details of the cylinder's properties are not forthcoming. At this point, Kaluza only requires that the new parameter, a measurement in the fifth direction, be small. This requirement does not necessarily guarantee the cylinder itself will be of a minute size, such that the cylinder will have a small diameter and circumference. The measurement could be small and the cylinder of a comparatively larger circumference if there were other restrictions on the measurement independent of the "cylinder condition."

Klein seized upon one particular interpretation of Kaluza's theory. He assumed that both the measurement in the fifth direction and the circumference of the cylinder were extremely small and equal, so that the cylinder itself could account for the whole extension in the fifth direction of space-time. His assumption allowed him to both explain why the fifth direction is undetectable and relate the cylinder condition to the quantum of action. One would think that the only possible interpretation of Kaluza's theory is that cylindricity is so small that the fourth dimension of space, or the fifth dimension of space-time, cannot be perceived or otherwise detected. But Kaluza's original theory only required that the A-lines, extensions of three-dimensional spatial points in the fifth direction, be of equal and constant length. These criteria are equally valid for space that is closed and Riemannian (a sphere) in the fifth dimension as well as cylindrical on a large macroscopic scale rather than the scale of a Planck length. Kaluza's basic theory could also be modified such that the "cylinder condition" would undergo a change of scale or cylindricity could be discarded altogether for a Riemannian curvature in the fifth direction. Einstein and Peter Bergmann in 1938 as well as Podolanski in 1947 introduced theories of this type. Podolanski assumed a laminate space with six dimensions while Einstein and Bergmann discarded the cylindrical condition and adopted a closed Riemannian space as the embedding structure for space-time.

Discarding the cylindrical condition with a minuscule size again invokes the question of why we cannot perceive or detect the fifth dimension. Yet the minuscule width of the cylinder, which has been discarded, could alternately be considered the 'effective width' of a four-dimensional sheet of space-time. All matter and material contact could be confined to this sheet so that the extra component of space would still be undetectable. In this manner, our four-dimensional space-time could be portrayed as a thin sheet whose spatial points are fully extended in a fifth dimension and loop around the closed figure. While the 'effective width' of the space-time sheet in the fifth direction is of minuscule extent, the loop in the fifth direction could be any length. This model assumes the reality of the fifth dimension and thus our space-time actually curves into the fifth dimension. Curvature in the fifth direction could then be regarded as wrinkles in the four-dimensional sheet.

This notion is not new and, in fact, predates GR by nearly a half century. The first interpretations of non-Euclidean geometry were wholly physical. In the late nineteenth century, it was generally thought that if mathematical space could be non-Euclidean, then physical space could also be non-Euclidean. Great debates over the true nature of space followed and some astronomers attempted to measure the curvature of space through parallax observations. William K. Clifford offered the first physical theory of the type described although his primary goal was to develop a model of Maxwell's electromagnetic theory in a four-dimensional space and include gravity and other forces later. Clifford's was the first true attempt to derive a TOE in the more modern sense of the concept. In 1870, he stated that matter is no more than curved space and matter in motion is only the time varying changes in the curvature of space. A decade later, Edwin Abbott wrote a popular book about beings in a two-dimensional world that experienced contact with three-dimensional beings for the first time. The book was titled *Flatland* and has become a classic of the period. But Abbott was just explaining the ideas expressed in Clifford's model of space to the common people of Britain and the world.

Clifford's own model extended well beyond the few short paragraphs that he presented before the Cambridge Philosophical Society in 1870 even though this particular presentation is normally all of his theory that is cited in others' work. There are strands of his theory stretching throughout his mathematical and philosophical writings. Enough is available from these other sources to reconstruct a fairly accurate model of the theory of space curvature that he was trying to develop even though he never published his theory in any one source. Clifford developed whole new forms of geometry based upon his biquaterions to demonstrate that dynamics in the three dimensions of space plus time reduce to kinematics in a four-dimensional space plus time which is characterized by an 'elliptical' Riemannian geometry. (Beichler, 1996) His theory sounds very much like a modern TOE.

Charles Hinton also popularized and extended the intuitive model of a four-dimensional space through several books and articles over a period of two and a half decades. The model of space first proposed by Hinton was a three-dimensional sheet of ether in which atoms were embedded. The complete structure was curved within a fourth dimension. The material atoms were likened to threads passing through the sheet from outside the three dimensions of the sheet, the points of intersection representing the individual atoms. (Hinton, 1980, 16-20) In this model, Hinton could only account for some of the fundamental properties of matter. In another of his essays, he introduced "twists" as mechanical models of electrical activity. (Hinton, 1980, 36-37, 74-75)

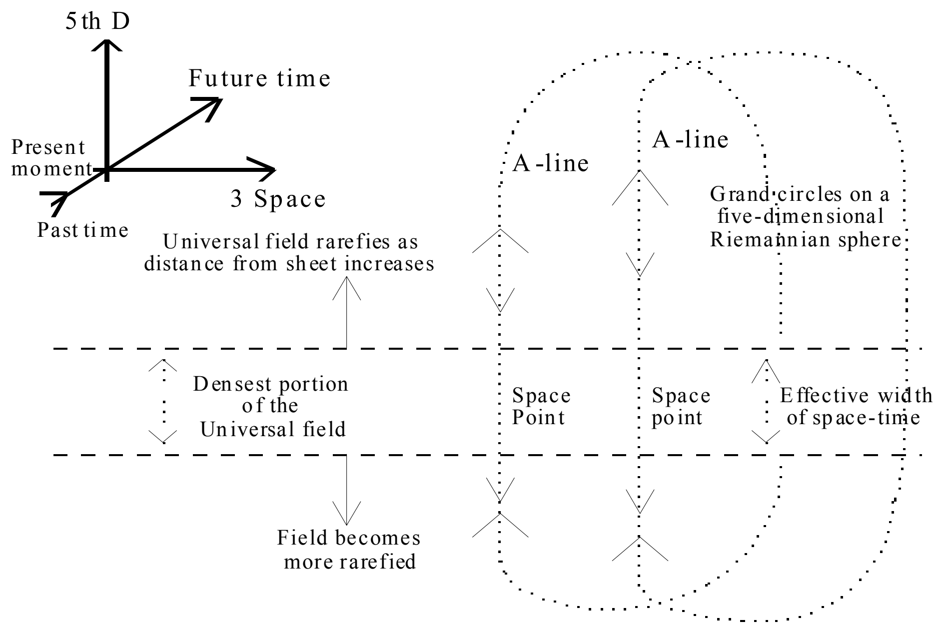
Aspects of Hinton's theory resemble portions of today's superstring theory. His notion of threads intersecting the three dimensional sheet of space are similar to superstrings and the twists that he later introduced to explain electricity are similar to the non-vibrational modes of superstrings which are suspected of yielding other physical properties. However, the notion of twists was not new with Hinton. The "twist" had been developed by Clifford as the fundamental physical point of space, again in a manner similar to the superstring theories of our own era. W.W. Rouse Ball developed a theory of gravitation very similar to Hinton's in the 1890s. Ball noted the similarity between his and Hinton's theories and stated that they were parallel developments, surrendering priority for the idea to Hinton. (W.W. Rouse Ball, 1891, 21) Hinton's later work overlapped the beginning of the Second Scientific Revolution and Einstein's first work on special relativity, but there is no evidence that Hinton's and Einstein's ideas ever came together in any form, that they influenced one another or that they even knew of each other's work.

Many scientists and scholars prior to World War I supported these ideas, although their supporters formed a small minority relative to the larger scientific and academic communities. Worried about the null affect of the luminiferous ether, as supported by the Michelson-Morley experiments, the astronomer Simon Newcomb theorized that our three-dimensional space consisted of two parallel sheets of three-dimensional ether separated by a small distance in a four dimensional space. Physical events occurred between the two sheets. (Newcomb, 1888/1891, 514-515) Robert S. Ball, another astronomer and a follower of Clifford, tried to detect the curvature of space (R.S. Ball, 1882, 519) and developed the mathematical "theory of screws" to model the higher-dimensional space. In these and the other instances, scientists assumed a real four-dimensional space characterized by curvature as opposed to the flat Euclidean space of Newtonian physics. In these earlier models, a more intuitive and less analytical concept of geometry was applied rather than the analytical tensor calculus that Einstein later used to model GR. The tensor calculus had either not yet been developed or was in its early stages of development when these theories were proposed. In fact, these geometric models of physical space and similar ideas formed part of the impetus to develop the new analytical systems of geometry. These earlier notions of curved four-dimensional space were overwhelmed by the advances in quantum and relativity theory during the early twentieth century and all but forgotten.

The only assumption made in this new theory is the existence of a real five-dimensional extension of our space-time continuum. All other aspects of this model come from physical theories that are already accepted by the scientific community such as GR and Kaluza's five-dimensional model. Although dimensions higher than the fifth may well exist, only the fifth dimension is necessary to explain four-dimensional phenomena at this time. Beyond the assumption of a real fifth dimension, the rest of the theory is derived from describing phenomena and using experimental and perceptual observations to determine the type, shape and characteristics of the model that will explain how physical events occur. This method may be criticized as an *ad hoc* fitting of the model to observed phenomena, but it is no more *ad hoc* than Newton's explanation of 'how' gravity works rather than 'why' gravity works.

The first order of business is an explanation why the fifth direction cannot be perceived or detected, fulfilling Einstein's requirement. Kaluza's model requires that the extension of every point in four-dimensional space-time in the fifth direction be of equal and constant length, but he also suggests that the measurement along the extended lines be small. So, in this model the A-lines have a very long length (of macroscopic proportions) while the four-dimensional sheet, which is our physical world, has a small quantum sized measurement along those lines. The A-lines are orthogonal to the sheet at every point. The sheet has a 'practical' thickness in the fifth dimension, which is designated as the 'effective width' of the sheet even though each point actually extends beyond the sheet along an orthogonal line. Four-dimensional phenomena are restricted to occur within this 'effective width' of space-time. So the fifth dimension cannot be observed, normally perceived or detected through normal experimental procedures.

The fifth dimension has no material reality, the key word being 'material' referring to the existence of matter, but is a perfectly continuous field of varying densities. The 'effective width' of the sheet consists of the densest portion of the field that effectively constitutes the four-dimensional continuum. The portion of the five-dimensional field outside of the sheet is quite rarefied and becomes considerably less dense the farther away from the sheet along the A-lines. The A-lines curl around the fifth dimension, which is closed, and reconnect at the same point from which they originated on the other side of the sheet, and through the 'effective width.' In other words, the A-lines are continuous and unbroken in the fifth direction, just as they would be in the minute cylinders posited by the Kaluza-Klein theory.



Although space-time has an effective width in the fifth dimension, each point extends into the fifth dimension and eventually curves back on itself along a grand circle on a sphere.

The fifth dimension exhibits Riemannian curvature in the large, just as the four-dimensional space-time portion of the five-dimensional world exhibits Riemannian curvature in the large as described by GR. The 'effective width' of the sheet is of the order of Klein's original fundamental length (l_o) of about 10^{-30} centimeters. The effective width is constant throughout all of space (for all practical purposes) since it is relative to all of the matter in space.

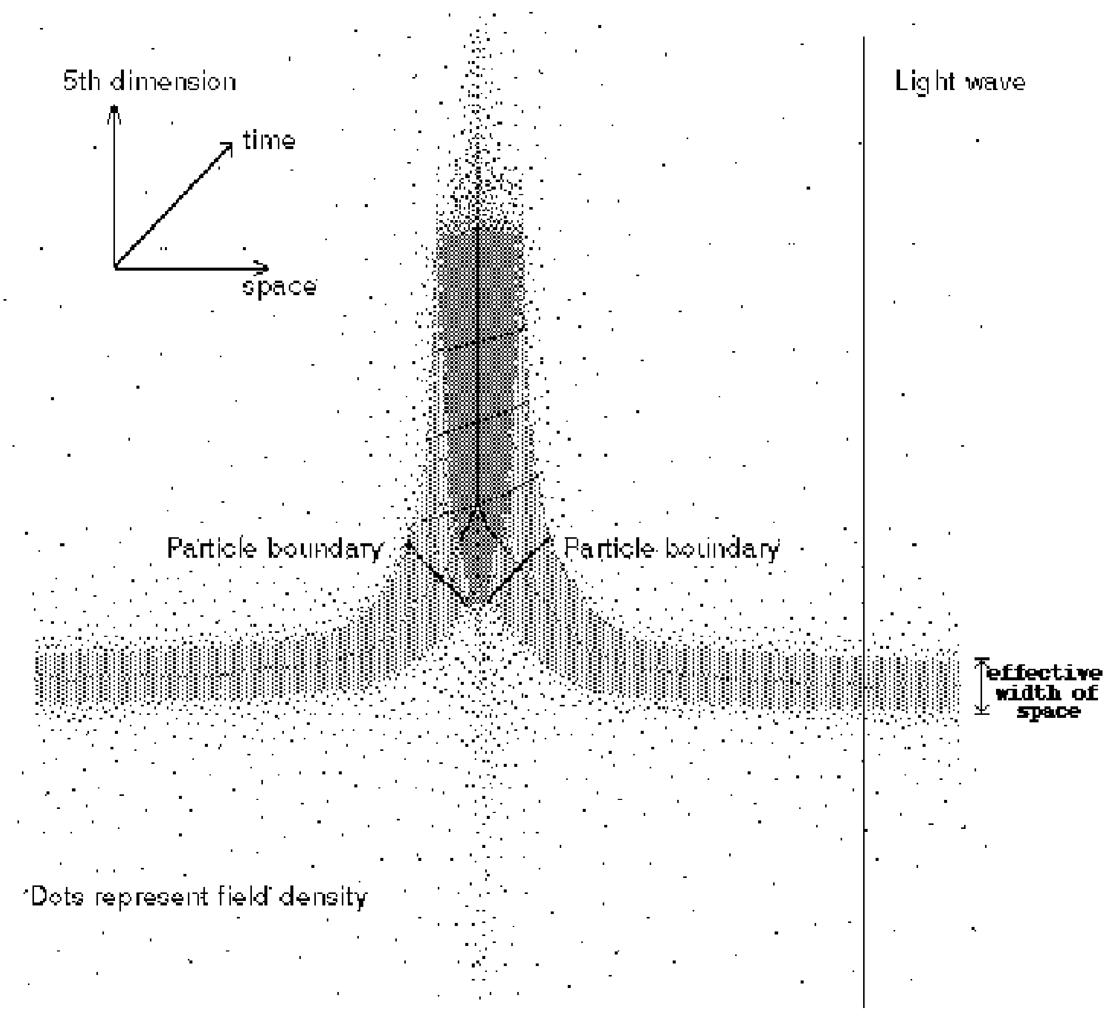
The A-lines can be seen as 'propensities' in Karl Popper's sense of the word (Popper, 69-71) or perhaps a better analogy could be found in Faraday's 'lines of force.' They are not physical lines, just as mathematical points have no physical reality. Light waves lay along the A-lines in their fifth component extension, giving the A-lines their closest approximation to a physical reality. The wave/particle duality of light waves thus reduces to which portion of the light wave along an A-line is interacting with the rest of the world. The portion of the light wave which is particulate, the photon, is just that part of the light wave (or A-line) which cuts across the 'effective width' of the sheet while the portion of light (along the A-line) which extends across the rest of the fifth dimension, outside of the sheet, is pure wave and exhibits wave interactions with the rest of the world. Wave/particle duality is thus a common feature of the five-dimensional space-time.

Since electromagnetic waves coincide with the A-lines across the fifth dimension, the magnetic permeability μ_o and electric permittivity ε_o of free space must be related to the five-dimensional field, but active within the four-dimensional sheet. The permittivity of free space is an interstitial connectivity constant which acts across three-space between contiguous points of space-time in the sheet. It therefore acts in a direction perpendicular to the fifth dimension and the A-lines. The permeability is also an interstitial connectivity constant of the field and it acts in a direction perpendicular to the A-lines, but it acts torsionally between contiguous points within the sheet. These two field constants guarantee that the maximum speed of light, all electromagnetic waves and the speed limit to matter itself in free space is constant. As field constants, these two quantities provide the connectivity between contiguous points of space and time in the direction of the normal four-dimensions. The connectivity of space-time is thus related to the speed of electromagnetic waves through empty or free space and limits the speed in that medium to $c = (\mu_o\varepsilon_o)^{-½.}$

The universal gravitational constant G is also a connectivity constant, but it acts between material particles rather than points in space. It extends across five-space from particle to particle in a direction parallel to the average macroscopic 'flatness' of the space-time sheet. The sheet is not 'flat' over macroscopic distances, but the curvature is so large that it approaches 'flatness.' G is constant and unvarying since it acts across five-space, whereas the permittivity and permeability act within the space-time sheet and are thus affected the local presence of matter. Permittivity and permeability vary from their values in 'free space' due to the local presence of matter, or more accurately they vary from their values in non-curved space-time due to the presence of local variations in curvature.

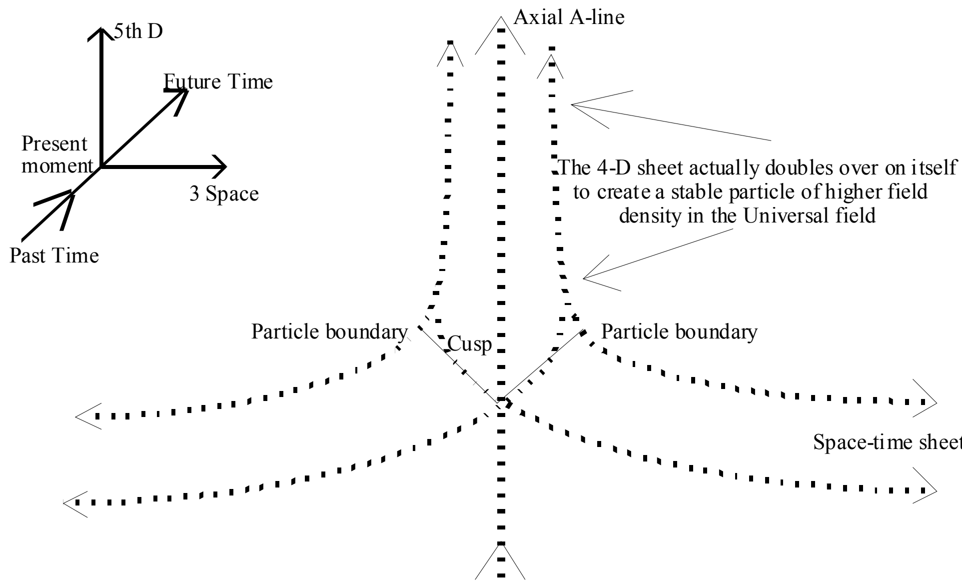
In Kaluza's theory, the purely five-dimensional component of the field, γ_{00}, is constant and equal to +1. Setting this field variable equal to one normalizes all other components of the field, so this is, and will remain, a normalization constant in the fifth direction in the new model. As such, it guarantees that the amount or quantity of the field (its average linear density) along the A-lines in the fifth direction is constant. The value of +1 also guarantees that a particle exists somewhere along a specified A-line when the wave function collapses such that the particle has a probability of 1 (or 100%) of existing along the A-line within the 'effective width' of the sheet. Like electromagnetic waves, the wave function of a particle lies along the A-lines. The wave function is the particulate equivalent of an electromagnetic wave, but unlike an electromagnetic wave it is characterized by extension in the directions of normal space. This requirement automatically combines the probability distribution of a quantum mechanical point particle with the wave function in the five-dimensional model. The significance of these requirements will become apparent as the characteristics of the model are developed further and material particles are explained in more detail.

With the assumption of a real fifth dimension, the curvature of the space-time continuum described by GR becomes a real curvature as exemplified by the extrinsic nature of the four-dimensional sheet. Elementary particles are curves of the four-dimensional sheet extending into the fifth dimension.



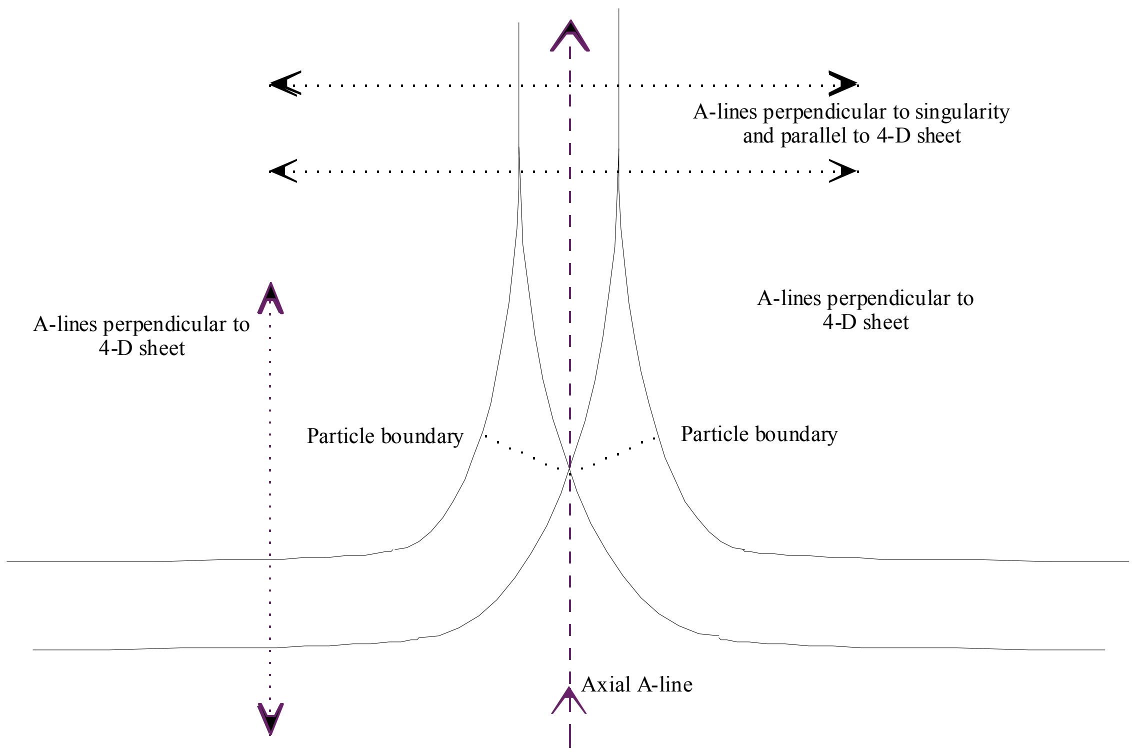
However, when a sheet with a definite thickness curves into a higher dimension, it folds on itself, doubles over or buckles forming a denser area called a cusp extending from the underside of the curve at the point of folding through the center of the doubled sheet. The exact form of the high-density pocket in the curve depends on the degree of the curvature.

In such cases, these denser areas of the curve correspond to the four-dimensional portion of the singularities that have occurred in GR and the infinite divergences that plague the quantum theories. The physical boundary of the particle in space-time must extend across the 'effective width' from the point of the cusp and perpendicular to the sheet. The cusp is the densest part of the field. It marks the point where the two sides of the sheet fold together and overlap each other.



The particle boundary actually marks the outer edge of the singularity portion of a particle with respect to the normal four-dimensional space-time continuum. However, the singularity disappears where it extends into the fifth dimension perpendicular to the sheet since field density decreases with increasing distance in the fifth direction from the four-dimensional sheet.

Since the singularity is essentially perpendicular to average flat portions of the space-time sheet and composed of a doubled or overlapping portion of the sheet something strange occurs. Each point of the overlapped sheet is still associated with A-lines, even the doubled over portion which is perpendicular to normal space-time. So, the A-lines associated with points along the axial A-line within the singularity radiate outward from the singularity parallel to the four-dimensional space-time sheet and extend through the fifth dimension. A-lines extending from the particle itself reach out across the fifth dimension to connect with all other material points throughout the universe. This structure represents what is called 'entanglement' of the wave functions in normal quantum theory. This means that there is a superstructure to the universe whereby all material particles are interconnected outside of normal four-dimensional space-time continuum.

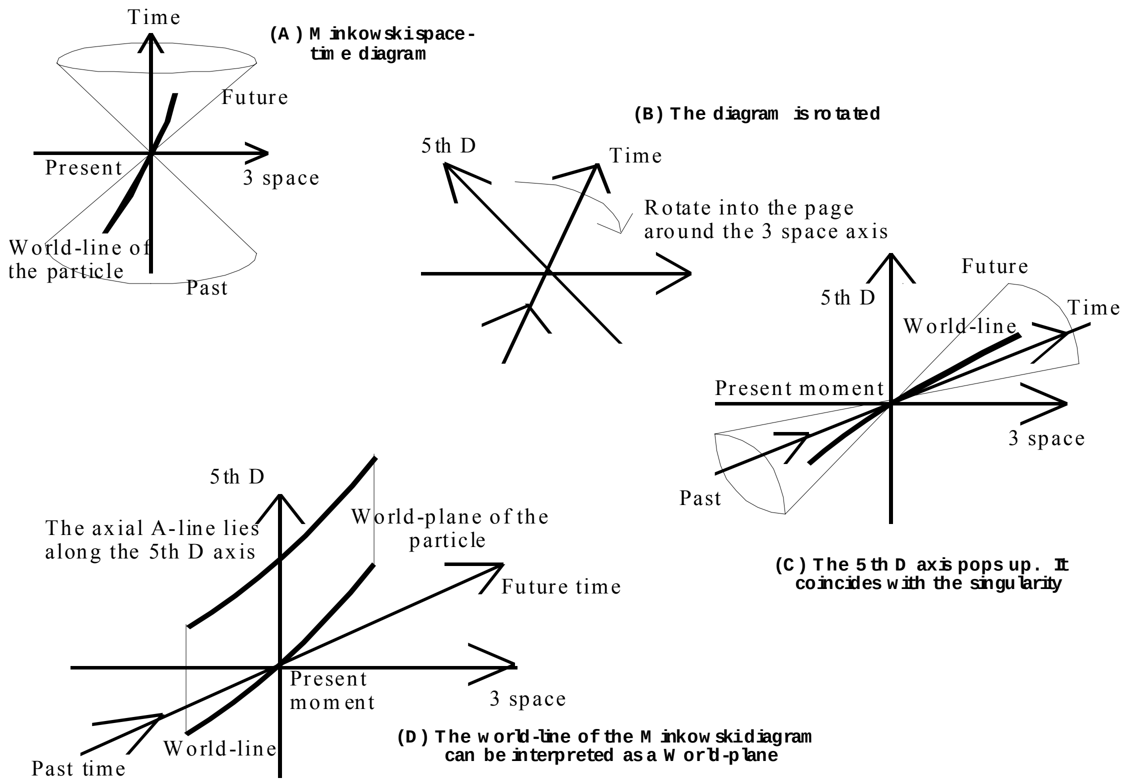


Since the A-lines have equal and constant length, the A-lines emanating outward from material particles must wrap around five space and force the spherical four-dimensional space-time to have the same overall structure in the large scale as the spherical five-dimensional space-time, or vice versa. The total length of the A-lines depends on the overall large-scale curvature of the four-dimensional universe and the fifth dimension is Riemannian with constant curvature of the same scale as the four-dimensional Riemannian universe.

The mass of an individual particle is finite, due to the curvature up to the particle boundary, even though the curvature becomes perpendicular to the sheet as a singularity. The mathematical singularity is not the same as a physical singularity, but models the physical singularity because the particle follows the A-lines that come around through the fifth dimension Riemannian sphere and close on themselves. The particle is symmetric (it appears to be spherical) in three space around a central or axial A-line which marks the point of the cusp in the center of the particle. This axial A-line marks the densest part of the field. The axial A-line is continuous around the grand circle of the Riemannian five-dimensional sphere which defines the particle across the fifth dimension. The probability distribution usually associated with the wave function of a particle corresponds directly to this axial A-line. The wave itself marks a point in space or space-time which collapses to a physical particle which is extended symmetrically in five space about the axial A-line. The particle is a reality independent of the probability distribution and the wave function, but the particle can be represented by the wave function in quantum considerations.

The idea of a ‘collapse of the wave function’ has caused philosophical problems for both quantum mechanics and quantum theory in general because it creates a discontinuity of action. The discontinuity is a mathematical artifact not a physical reality because the probability corresponds to the wave function but it is not the wave function itself. The discontinuity occurs where the abstract mathematical point designated by quantum theory devolves into a physically real point in space-time. This model distinguishes between the two. The abstract mathematical point of quantum theory corresponds to the A-line’s intersection with a geometrical plane at the center of the sheet. The wave function itself corresponds to the singularity surrounding the axial A-line of the particle which is physically real. So the wave function can be considered the five-dimensional ‘volume’ of the particle. The physical boundaries of the particle mark the classical physical reality as related to and described by Newton, Faraday, Maxwell and Einstein and their respective theories. These four theories are unified with both quantum mechanics and wave mechanics in this single model of an elementary particle as signified by a five-dimensional component. The five-dimensional component of this model is the ‘hidden variable’ that was first suggested by David Bohm.

Bohm’s physical model progressed from the ‘hidden variable’ concept to a hypothetical underlying reality, which he called the ‘quantum potential.’ The sheet and its corresponding A-lines in this model correspond to Bohm’s ‘quantum potential.’ But the ‘wave collapse’ of classical quantum mechanics also marks a progression in time, a strictly dynamical view of the universe. The above diagrams do not adequately depict this time progression, which lies outside of the flat surface of the paper. The normal method of diagramming relativistic progression in time is via Minkowski space-time diagrams whose components are the space and time axes, the light cone marked by a past and future extending above and below the origin, and a world-line tracing the particle’s history (the past, present and future of a particle’s existence). The world-line is the path of events that occur during the particle’s lifetime. The five-dimensional model of a particle can likewise be viewed in a new diagram that is no more than the space-time diagram turned on its side. When rotated through ninety degrees into the paper and turned on its side, the light cone of the past emanates out in front of the page and the light cone of the future stretches out behind it.



As a particle moves into the future at a constant rate of time, it follows a line along the axis perpendicular to the axial A-line and the three-space axis. The axial A-line representing the particle becomes an axial A-plane along the particle's 'world-line.' This axial A-plane is a diagrammatic representation of the history or lifetime of a point particle. But the reality of the particle is an axial-sheet. In other words, the world-plane has an 'effective width' which defines the sheet of the four-dimensional space-time continuum although its width diminishes the further away from the four-dimensional sheet that one goes in the fifth direction.

The particle is restrained from moving at the speed of light or faster by μ_o and ε_o, the connectivity (interstitial) constants of space-time. So the particle's future path is limited to an existence between the imaginary lines making up the light cone. As a particle moves into the future, it deconstructs the universal field at the moment of the present, but the particle must reconstitute its curvature in the sheet at the next moment in the future. The reconstitution is made from the future portion of the sheet. In other words, the axial A-line marking the particle's fifth component has a projection into the future along the time axis, which represents the state of non-motion if the particle is at rest or, a state of constant velocity if the particle is moving. As long as the reconstitution does not vary from the projected state of motion of the axial A-line, there is essentially no change in the field. This is the physical representation of Newton's first law of motion.

To vary a particle from its projected path into the future, an abstract A-plane or 'world-sheet' that is not physically real until the moment passes, the particle must accelerate. This action causes a curvature change relative to the projected axial A-line or A-plane. The change of curvature from moment to moment in time representing an acceleration corresponds to a deconstruction of the four-dimensional sheet curvature of the moment past to the newly reconstituted curvature of the moment future. This process changes the relative field with respect to the rest of the universe as time progresses into the future. The deconstruction/reconstitution process of the spatial change of position in time is a disturbance of the field that is commonly called inertia.

The whole field of the universe must react/interact (Newton's third law of motion) with a particle that changes its projected state of motion or relative projected position as time progresses into the future. The process of interaction between the whole field and the changing A-line at a moment causes inertia (Mach's Principle). The action necessary to cause the universal field's interaction (inertia) is the force (in the Newtonian sense) and is proportional to the amount of change of position, relative to time, also called acceleration. Thus we have $F = ma$, Newton's second law of motion. In other words, the universal field resists changes and the inertia of a material particle is the response of the whole field (which in a sense is the universe) to change. In this manner, the connectivity of the points in the field undergoes a change that is constrained by the permeability and permittivity constants. The universal field cannot react to the change or acceleration at a greater rate than constrained by these field constants within the sheet, so matter is limited by the constants in the form of the speed of light.

The Newtonian concept of inertia as a resistance to a change in motion is thus explained. Inertia is the resistance of space-time to changes in the deconstruction/reconstitution process. The amount of destruction/reconstitution necessary to effect a change is directly proportional to the curvature which gives the particle mass; the greater the curvature, the greater the deconstruction/reconstitution necessary during the process of change of position. Since gravitational acceleration is just the reaction of a particle to space-time curvature near a mass associated with the curvature, there is a direct relationship between gravitational mass and inertial mass. Inertial and gravitational masses are mathematical equivalents although they represent different concepts and are thus philosophically different quantities. This explanation places Einstein's equivalence principle on a whole new footing.

The whole process of deconstruction/reconstitution, which corresponds to the classical concept of motion, also corresponds nicely to Bohm's concept of the implicate and explicate orders. The implicate is the extension of the sheet both backward and forward in time, the past and the future. The explicate is the actual portion of the axial A-plane or the 'world-sheet' which separates the past and future as the present moment. When the implicate becomes the explicate, the next future moment is reconstituted as the present and the moment just past by is deconstructed. The implicate consists of both the past history of all the individual world-sheets that are coupled together or 'entangled' to create the universe as well as the future portion of all world-sheets, as bounded by the present which is the explicate.

At this point, the general development of the five-dimensional model of the universe is complete. It could be argued that this model is non-mathematical since neither mathematical formula nor equations have been offered, but the theory is completely mathematical. This theory and its model have added no new analytical mathematics other than the mathematics already used in Newton's theory of motion, Einstein's GR, Maxwell's electromagnetic theory, Schrödinger's wave mechanics and other theories discussed. The elucidation of this model also implies a direct relationship to superstring theory. The world-line of a particle can be approximated as the trajectory of the cusp moving through time. The cusp can be then be equated to a superstring which is vibrating in the fifth dimension along the axial A-line of the particle. The axial A-plane, also termed the world-plane, can be equated to a 'membrane' as described in the most recent mathematical models and extensions to superstring theory. These recently developed mathematical entities, superstrings and membranes, are only mathematical approximations to the real physical particles as portrayed in this model.

The mathematical analysis of this theory, in the form of equations and formulas, has already been expressed in these other theories. This model provides the "rich structure" for which other scientists have been searching in vain. The most surprising feature of this theory and model is the obvious reduction of all other theories to a single theory with only the assumption of a real fifth dimension and the accompanying answers to questions about how a real fifth dimension could work given our present physical 'laws' and theories. Once this model is adopted, the problem will not be finding answers to explain physical phenomena. The central problem of this theory is how to pose the proper questions to ask nature within the context of this theory. The answers become self-evident once the proper questions are asked. The answers are as simple as the nose on your face.

Go to PART II

Go to Bibliography

Last Updated 6 June 1999

E-mail comments and suggestions to Jim Beichler, editor, *YGGDRASIL*, at
jebco1st@aol.com

Top of Page

YGGDRASIL Home Page