UNIFICATION IN TEN DIMENSIONS

Edward Witten

Joseph Henry Laboratories
Princeton University
Princeton, New Jersey 08544

As an attempt to reconcile quantum mechanics with gravity, superstrings are not a new idea, but real phenomenology of superstrings became possible only upon the discovery of a generalized mechanism for cancellation of anomalies[1]. The discovery of a superstring theory with $E_8 \times E_8$ gauge group[2] has made the subject even more exciting, and by now there has been much exploration of the issues that are involved in attempting to make contact between superstrings and four dimensional physics[3−18] In attempting here a brief review of these matters, I will concentrate on a single theme. The theme will be to describe the similarities and differences between grand unification in four dimensions and unification in a higher dimensional Kaluza-Klein or superstring context. I will contrast four dimensional unification with higher dimensional unification in the following four areas: (i) the origin of flavor; (ii) the fine tuning problem and the nature of Higgs bosons; (iii) properties of the Yukawa couplings; (iv) magnetic monopoles and electric charges.

First we discuss the flavor problem, which is the modern version of Rabi's classic question 'Who ordered the muon?' The question is why observed fermions seem to fit into (at least) three generations which are identical as regards gauge quantum numbers. Why did nature choose to duplicate structure in this way?

To have any hope of understanding why there are three generations, we must first discuss what the really fundamental feature is of the generation structure that nature has chosen to duplicate. The most fundamental feature – unknown, of course, when Rabi originally posed the question – is that the quarks and leptons that make up a fermion generation transform in a so-called complex representation of the $SU(3) \times SU(2) \times U(1)$ gauge group. This means that the left-handed massless fermions transform in a representation V_L of the gauge group which is not equivalent to the representation V_R in which the right-handed fermions transform. While V_R and V_L are inequivalent, they are (by the CPT theorem) complex conjugates of each other, so the statement that V_L is not equivalent to V_R amounts to saying that V_L is not equivalent to its own complex conjugate or in other words that V_L is a complex representation of $SU(3) \times SU(2) \times U(1)$.

The fact that the fermion representation is complex means that the gauge interactions violate parity. It is often described by saying that there is a chiral asymmetry or a left-right asymmetry in the fermion quantum numbers. This left-right asymmetry is important not only because parity violation is interesting and important but also because the left-right asymmetry is the foundation of our understanding of why fermions that are light enough to be experimentally observeable exist at all.

As long as gauge symmetries are conserved, a left handed fermion can gain a mass only by pairing up with a right handed fermion of the same gauge quantum numbers (since a massive spin one half particle has two helicity states, which must have the same gauge charges). The fact that the left- and right-handed fermions transform differently under the gauge group means that they must remain massless as long as the gauge group is unbroken. What seems to happen in nature is that the observed quarks and leptons remain masssless down to a mass scale of a few hundred Gev at which the electroweak gauge group is broken down to a subgroup – electromagnetism. At that point, the fermions are in a real representation of the remaining gauge symmetries, and they can and do get masses, except possibly for the neutrinos. We do not understand why the scale of weak interaction symmetry breaking is so tiny compared to the Planck mass;

this is the gauge hierarchy problem. But we do at least understand the lightness of the quarks and leptons in terms of the lightness of the W and Z; it follows from the chiral asymmetry between left and right handed fermions.

Could it be that the left-right asymmetry in fermion quantum numbers is just an illusion? Perhaps there are mirror fermions at a Tev energy scale with $V+A$ couplings to the usual W bosons. (The mirror fermions could not be much heavier than a Tev, since their masses would violate $SU(3)\times SU(2)\times U(1)$, given that this is so for the usual fermions.) I think that the existence of mirror fermions is extremely unlikely. If they were discovered, we would lose our successful explanation of why the usual fermions are so light compared to the mass scale of grand unification or gravity; we would be faced with an embarrassing puzzle of why the usual fermions are significantly lighter than their mirror counterparts (much lighter, in the case of the first generation). Also, if mirrors exist, the intricate cancellation of $SU(3)\times SU(2)\times U(1)$ gauge anomalies among the fermions of a single generation is just an accident, since the mirrors would have canceled the anomalies anyway.

By continuously changing the parameters of a theory with unbroken gauge group $SU(3)\times SU(2)\times U(1)$, there is no way to disturb the left-right asymmetry in the quantum numbers of the massless fermions. It does not matter here whether the parameters that are being continuously varied are, say, unknown coupling constants in the underlying equations or artificial parameters that label theoretical assumptions. Because the left-right asymmetry only depends on qualitative facts about a theory, not on the details of how it is presumed to behave, it has been over the years a fruitful matter to think about in trying to understand grand unification both in four dimensions and in the context of higher dimensional theories. It is very likely that we will be able to predict the 'universality class' of the low energy world long before we can predict the details; since the left-right asymmetry depends only on the universality class of a theory with $SU(3)\times SU(2)\times U(1)$ gauge interactions, it will probably be determined on theoretical grounds long before we understand the Cabibbo angle or the mass of the electron.

There is another reason that the chiral asymmetry is important. Chiral asymmetry seems to be a delicate thing which is easily lost and less easily gained. Passing from a theory at one energy scale Λ with left-right asymmetry to an effective theory at a lower energy scale Λ', there are many ways that the chiral asymmetry of the original theory can be lost in going down to lower energies. For instance, chiral asymmetry is actually lost in nature in electroweak symmetry breaking at an energy of order one hundred Gev; it could easily have been lost in compactification from ten to four dimensions, if a certain slightly intricate chain of steps that I will mention later were not followed. While we know many ways to lose chiral asymmetry, we know of no ways to gain it in going from one energy scale to a lower one. To gain chiral asymmetry by a dynamical process starting with a non-chiral theory would require a mechanism for the binding of charged massless fermions, and I personally would consider such a mechanism strongly counter-intuitive. In fact, there are theorems which tend to show that spontaneous generation of chirality does not occur in the dynamics of non-chiral (vector-like) gauge theories.[19] If it is true that chiral asymmetry is easily lost but difficult or impossible to gain, then the chiral asymmetry that we observe in nature must be a trace of the most fundamental physical laws, whatever those may be. Indeed, along with general covariance and Yang-Mills gauge invariance, left-right asymmetry of fermion quantum numbers would appear to be one of the few really fundamental observations about nature.

Let us now survey some approaches to the problem of family replication, bearing in mind that the chiral asymmetry of the fermion families is one of the basic observations. Most work on the family problem in the context of conventional grand unification has been based on the idea that the unified gauge group G is larger than a minimal $SU(5), O(10)$, or E_6 unified group, large enough so that several generations of quarks and leptons fit into a single generation of G. The attempt to carry out this idea with G being a large $SU(N)$ group is a fascinating idea that runs into innumerable difficulties. An attempt which comes much closer to working is the idea[20] of taking G to be a large orthogonal group $O(4k+2)$ with fermions in the spinor representation of G. This idea of 'orthogonal family unification' beautifully fits several generations of fermions into an irreducible representation of G. The only thing which is really wrong with it is that it predicts at the same time an equal number of opposite chirality mirror generations, which cannot be superheavy since (given that the conventional fermions are light) their masses violate the gauge symmetries of the electroweak theory. Recently there have been heroic attempts to make a viable model in which the mirror fermions will be placed at the Tev mass region[21]. While it may turn out that that is how nature works, I consider it unlikely for reasons that I have already indicated. It is on the other hand a very striking fact that orthogonal family unification works just right in Kaluza-Klein theory without producing the mirror generations.

The basic idea behind this is the following. The Dirac equation in ten dimensions is $D\Psi = 0$, where D is the ten dimensional Dirac operator $D = \sum_{i=1...10}\Gamma^i D_i$, the $\Gamma^i, i = 1...10$ being the ten dimensional gamma matrices. Now, in fact, $D = D_4 + D_K$, where $D_4 = \sum_{i=1...4}\Gamma^i D_i$ is the four dimensional Dirac operator and $D_K = \sum_{i=5...10}\Gamma^i D_i$ is the Dirac operator of the compact Kaluza-Klein space. This means that the internal Dirac operator D_K is the 'mass' operator of the effective four dimensional theory. In fact, if $D_K\Psi = \lambda\Psi$, then the ten dimensional equation $D\Psi = 0$ reduces in four dimensions to $D_4\Psi + \lambda\Psi = 0$, so that Ψ will be observed in four dimensions as a fermion of mass λ. Now, the internal Dirac operator D_K acting on the compact space K will have a discrete spectrum, and the eigenvalues which are not zero will be of order $1/R$, R being the radius of K. Eigenvalues of order $1/R$ correspond to fermions of Planckian masses, which certainly would not have been observed to date. The observed quarks and leptons have masses that are essentially zero in Planck units, and they must correspond to zero eigenvalues of the internal Dirac operator.

Now, why would a Dirac operator have zero eigenvalues? Some of the basic facts about this question are known to physicists from instanton days. A Dirac operator can have zero eigenvalues for topological reasons, the number of zero eigenvalues being determined by the values of suitable topological invariants. Now, in the case at hand there is a separate Dirac equation for each choice of $SU(3) \times SU(2) \times U(1)$ quantum numbers, and the number of massless multiplets in a given $SU(3) \times SU(2) \times U(1)$ multiplet is simply the number of fermion zero modes for the corresponding Dirac operator on K. The family problem – the question of why there are several massless multiplets of given $SU(3) \times SU(2) \times U(1)$ quantum numbers even though the microscopic theory began with a single irreducible multiplet – is in this context simply the question of why a suitable Dirac operator has not one but several zero eigenvalues. But this is not an unusual behavior for Dirac operators; to cite a relatively familiar example, a color $SU(3)$ instanton acting on a single multiplet of fermions in the adjoint representation of $SU(3)$ would have six zero eigenvalues. We are thus led to the program[22–32] of relating the number of generations to topological invariants of K.

If we do manage to find several fermion generations, corresponding to several zero modes of the internal Dirac operator, what chirality will these have? To answer this, note that four dimensional chirality is measured by the product $\Gamma^{(4)} = \Gamma^1\Gamma^2...\Gamma^4$ of the usual four gamma matrices. On the other hand, chirality in the sense of the Dirac equation on K is measured by the internal chirality operator $\Gamma^{(K)} = \Gamma^5\Gamma^6...\Gamma^{10}$. And chirality in the ten dimensional sense (before compactification, so to speak) is measured by the product $\Gamma^{(10)} = \Gamma^1\Gamma^2...\Gamma^{10}$ of all ten gamma matrices. Now, it is just a fact of life of ten dimensional supergravity that in supergravity theories that have elementary gauge fields (otherwise, it is impossible to get chiral fermions [26]), the ten dimensional fermions have definite chirality, say $\Gamma^{10} = +1$. Since $\Gamma^{(10)} = \Gamma^{(4)} \cdot \Gamma^{(K)}$, the fact that $\Gamma^{(10)} = 1$ means that $\Gamma^{(4)} = \Gamma^{(K)}$. This wonderful equation says that chirality as measured by four dimensional experimentalists (who, in effect, measure $\Gamma^{(4)}$) coincides with chirality as understood by observers on K studying the internal Dirac operator D_K. A zero mode of D_K of $\Gamma^{(K)} = +1$ will give rise to a massless fermion in four dimensions of $\Gamma^{(4)} = +1$; a zero mode of $\Gamma^{(K)} = -1$ will give rise to a massless fermion of $\Gamma^{(4)} = -1$. To obtain left-right asymmetry in the effective four dimensional world requires that the Dirac operator on K has (for given $SU(3)\times SU(2)\times U(1)$ quantum numbers) more zero modes with one eigenvalue of $\Gamma^{(K)}$ than with the opposite eigenvalue. This is not an unlikely behavior at all; it amounts to saying that the Dirac operator on K has a non-zero character- valued index. These ideas in the Kaluza-Klein context were developed in [22,24,26].

There are many ways to implement these ideas. One approach which was originally proposed in an *ad hoc* way in [26] but recently proved [5] to be very natural and attractive in the context of $E_8 \times E_8$ superstrings is the following. Consider an $SO(16)$ gauge group in ten dimensions with fermions in the positive chirality spinor (or **128**) of $SO(16)$. After compactifying to four dimensions, we have on the compact six manifold K a spin connection which is a connection on the tangent bundle of K and so is a gauge field of $SO(6)$ (or perhaps a subgroup thereof). If we 'embed the spin connection in the gauge group' by setting the gauge fields of an $SO(6)$ subgroup of $SO(16)$ equal to the spin connection, then $SO(16)$ is broken down to a subgroup – which generically is $SO(10)$. Now, $SO(10)$ is essentially the only orthogonal group which is suitable for four dimensional grand unification because it has a complex representation, the **16**, which is just right for accomodating a standard generation. With other orthogonal groups one runs into the chirality problem I mentioned above, though one might try to deal with this problem along the lines of [21]. It was for this reason that in [26] an $SO(16)$ gauge group rather than some other orthogonal group was taken as the starting point. After breaking $SO(16)$ to $SO(10)$ by 'embedding the spin connection in the gauge group,' study of the Dirac equation shows that we indeed get chiral fermions in the **16** of $SO(10)$, the number of generations N_{gen} being related to one of the most basic topological invariants of K, namely its Euler characteristic. (In the superstring context, the precise relation turns out to be [5] $N_{gen} = (1/2)\cdot\chi(K)$, where $\chi(K)$ is the Euler characteristic of K, which in six dimensions can be any even number. In the model of [26] the number of generations was twice as large.) Choosing K to have a suitable Euler characteristic, one can well obtain several generations in this way, so that the flavor question *per se*, the question of how to get a multiplicity of generations starting with a unified underlying framework, is no mystery. What is unsatisfying is that several aspects of the construction are rather artificial. The gauge group and fermion representation that we started with and the choice of embedding the spin conection in the gauge group were all simply chosen to get the standard model in four dimensions after compactification. It is therefore very satisfying that [5] recent developments in superstring theory give a natural justification for these seemingly arbitrary choices. The group E_8 has a maximal $SO(16)$ subgroup, and the adjoint representation of E_8 contains a **128** of $SO(16)$, so these ingredients have been practically forced on us by recent developments about anomaly cancellation. Also, embedding the spin connection in the gauge group turns out in the superstring context to cancel various anomalies and to give (if K obeys certain conditions stated shortly) a solution of the equations of motion of the theory, without generating a cosmological constant in four dimensions. This is in fact the only presently known way to obtain compactified solutions of the superstring equations, and it is one of the few examples in any Kaluza-Klein theory of any sort in which a realistic candidate vacuum state does indeed obey the hoped-for equations. Actually, the metrics which seem to give solutions of the equations of the theory are Ricci flat Kahler metrics which are the so called metrics of $SU(3)$ holonomy. Their existence was conjectured by Calabi [33] and proved by Yau [34]. The choice of such a metric leads first of all to unbroken $N = 1$ supersymmetry in four dimensions, a property which may well be desireable, and second to a four dimensional gauge group which is E_6 rather than $SO(10)$. Just as $SO(10)$ is the one orthogonal group which is suitable for unification in four dimensions, E_6 is the one suitable exceptional group [35]. Compactification from ten to four dimensions on manifolds of $SU(3)$ holonomy turns out to lead in four dimensions to a theory with chiral fermions in the **27** of E_6 (which is, of course, the right representation [35]), the number of generations still being half the Euler characteristic. Actually, the logic I have followed here in sketching these matters was the reverse of that in [5]; there unbroken $N = 1$ supersymmetry was taken as the initial requirement, and the other features were deduced as consequences. Yet another possible line of development, sketched in the last section of [5] and pursued further in [36] begins with the requirement of a conformally invariant two dimensional sigma model as the starting point and deduces the other properties from this. It is very satisfying that similar conclusions can be reached from so many different starting points.[37].

This completes what I will say about the family problem. Now we turn to the question of Higgs bosons. If K is not simply connected there is a simple mechanism for grand unified symmetry breaking which involves no ingredients that aren't present anyway. (Some of the relevant issues were first considered in [38], the main difference being that when the fundamental group of K is finite, as in many cases of interest, one is led to topological questions, such as those sketched below, rather than to the dynamical issues considered in that paper.) Let γ be a non-contractible loop in K, beginning and ending at a point x. Let $U(x) = Pexp \oint_\gamma A \cdot dx$, A being the E_6 or $O(10)$ gauge field that is still unbroken after embedding the spin connection in the gauge group. For various reasons, such as a wish to keep unbroken $N = 1$ supersymmetry or a simple wish to obey the equations of motion, it is desireable for A to be a pure gauge locally, with zero field strength. If γ is non-contractible, this does not require $U = 1$. If $U \neq 1$, then in many ways $U(x)$ is a field like any other. In particular, if $U \neq 1$, then $O(10)$ (or E_6) is broken to the subgroup that commutes with U. It is possible in this way to get various more or less realistic gauge groups of rank five or rank six, such as $SU(3) \times SU(2) \times U(1) \times U(1)$ or $SU(3) \times SU(2) \times U(1) \times U(1) \times U(1)$ [7,12,13]. A rank four group cannot be obtained in this way, so there is a prediction of at least one new gauge interaction beyond the standard model. If K has $SU(3)$ holonomy, a rank five group can emerge [7] if and only if the fundamental group of K is non-abelian; otherwise there must be two new gauge interactions.

In many ways, U is similar to a Higgs boson ϕ in the adjoint representation of the gauge group, $U \sim expi\phi$. But there are essential differences. One difference arises if $\pi_1(K)$ is finite, which is so for many choices of K of current interest (for instance, it is true for all manifolds of $SU(3)$ holonomy). For instance, suppose $\pi_1(K) = Z_n$. Then n circuits of the curve γ considered before make a contractible loop, so it is necessarily so that $U^n = 1$. This means that the eigenvalues of U are quantized; they are n^{th} roots of unity.

To understand why this might be desireable, recall that one of the key mysteries in grand unified theories is the so-called fine tuning problem. This is the question of why the energy scale of weak interaction symmetry breaking is so much smaller than that of unification. The problem arises because eigenvalues of Higgs boson fields are continuously variable parameters, which depend continuously on the values of unknown coupling constants. A massless weak doublet can arise only if Higgs eigenvalues have special values, and conventionally this is artificial. But, as I have just explained, in the context under discussion here the relevant eigenvalues are naturally quantized and can take only discrete values. On this grounds alone it is not too surprising that one finds that there are possibilities for solving the fine tuning problem [7,12,13], though I will not review the details of this here.

Other special features of grand unified symmetry breaking by Wilson lines deserve mention. This procedure does not disturb the classical field equations, assuming that these were obeyed at $U = 1$, and does not induce a cosmological constant (at least not in the classical approximation) or break supersymmetry (if this is otherwise unbroken). There is no need to postulate Higgs bosons or a Higgs potential or any other ingredient that is not present automatically. Also, symmetry breaking by Wilson lines is more or less topological in nature (for instance, it involves discrete choices if the fundamental group of K is finite) so there is a chance that it can eventually be understood and predicted by general, qualitative arguments – which, when available, are almost always more convincing and satisfying than dynamical arguments. Many other topological approaches to grand unified symmetry breaking could be considered, but generally speaking the others would disturb the successful predictions concerning quark and lepton quantum numbers. (We have seen that unlike the situation in conventional grand unified theories, the choice of vacuum configuration in a higher dimensional theory determines the fermion quantum numbers; this was the key to getting the right chiral structure in the first place.) Symmetry breaking by Wilson lines is unusual as a topological approach to grand unified symmetry breaking that does not have harmful implications for the fermion quantum numbers.

Now we turn to a discussion of gauge and Yukawa couplings. Again we begin with some mathematical preliminaries. If $\pi_1(K) \neq 0$, then K has a 'covering space' K_0 with $\pi_1(K_0) = 0$. If the theory is formulated on K_0 the number of generations is $N_0 = (1/2) \cdot |\chi(K_0)|$. On K the number of generations is smaller, being in fact $N = (1/2) \cdot |\chi(K)|$. In passing from K_0 to K many or most of the quark and lepton states are 'lost.' To quantify the effect of this, let me describe symmetry breaking by Wilson lines in another way. On K_0 we have a discrete symmetry group G that acts freely; K is the quotient K_0/G. This means that for any $x \in K_0$ and $g \in G$, x and gx are considered equivalent as points in K. An ordinary scalar field ψ on K is the same thing as a field $\psi(x)$ on K_0 that obeys $\psi(gx) = \psi(x)$. This condition means that ψ 'lives' on K, not K_0. When we introduce symmetry breaking by Wilson lines, we must make a slight modification in this requirement. If U is the Wilson line corresponding to g (taken in whatever representation of the gauge group ψ is in) then the appropriate requirement is $\psi(gx) = U\psi(x)$. States that exist on K_0 but do not obey the condition just stated do not 'survive' when the theory is formulated on K rather than K_0.

The important point here is that, in general, the surviving states are not states that on K_0 were the $O(10)$ or E_6 partners of one another. If a given quark obeys $\psi(gx) = U\psi(x)$, its lepton partner will (typically) have different U and so will **not** obey this condition. This has the following consequence. In this theory as far as counting states is concerned the particles appear to form representations of a grand unified group. There are as many u quarks as d quarks or neutrinos or charged leptons. But they are not states that on K_0 were $O(10)$ or E_6 partners. This has an immediate beneficial consequence. Since the physical fermions are not $O(10)$ or E_6 partners of one another, there are no simple group theory relations among Yukawa couplings. This is a desireable state of affairs, since such relations (such as $m_d = m_e$, and its generalizations) are generally not in agreement with experiment. On the other hand, the observed gauge bosons of the four dimensional world all began as elements of a single $O(10)$ or E_6 representation (the adjoint representation) so their couplings are related to one another by the usual group theoretical formulas. As a consequence, the Georgi-Quinn-Weinberg relations among the gauge couplings will hold; these relations, of course, are quite successful. It should definitely be counted as a success of higher dimensional theories that – in this way – the group theory relations among gauge couplings are preserved but the (superficially analogous) group theory relations among Yukawa couplings are violated.

Another interesting point is the following. The usual X and Y bosons of grand unification – the $SU(5)$ partners of the photon – have $U \neq 1$ and do not exist. Of course, particles with the same quantum numbers and the same order of magnitude of mass and coupling of the X and Y bosons will exist and will mediate proton decay, but since the proton lifetime scales like the fourth power of the mass and the minus two power of the coupling of the heavy bosons, the relation between unification scale and proton lifetime might differ by a couple of orders of magnitude from what is conventionally calculated. One might almost say that minimal $SU(5)$ could be alive and well in ten dimensions!

I might also note that in discussions of proton decay, an important question is to determine the branching ratios. In particular, when the proton decays, is it more likely to emit an electron or a muon? This question is conventionally a question about X and Y boson couplings which in turn (since those are gauge bosons) amounts to the question 'which charged lepton is the $O(10)$ or E_6 partner of the d quark?' In higher dimensional unification, the latter question has no answer, as I have just explained, and the X and Y bosons do not exist. The question about branching ratios in proton decay is still meaningful, of course, but the ingredients in answering it will be somewhat different.

Turning now to our third topic, if Yukawa couplings do not obey simple relations of a group theoretical origin, what relations will they obey? Before stating the answer, I would like to note that it should come as no surprise that one has, potentially, a great deal of predictive power for the Yukawa couplings. In fact, the light fermions and Higgs bosons are all zero modes of suitable wave operators on the compact Kaluza-Klein space K. The Yukawa couplings are certain cubic terms that arise in expanding the exact ten dimensional theory in powers of the light fields, and they can be computed by taking a suitable product of wave functions and integrating it over K. Thus, if one knew everything about K (including its metric) one would simply determine the zero mode wave functions by solving the relevant wave equations on K (numerically, if need be) and then one could determine all the Yukawa couplings by evaluating a certain integral (which arises in expanding the ten dimensional action in powers of the light fields). While the procedure just stated is perfectly sound in principle, it presupposes an amount of information about K that will probably not be available at least for a very long time. The question really should be 'what simple relations among Yukawa couplings are there that can be deduced in a general way?' analogous to group theory relations in standard grand unification. The answer to this is that the Yukawa couplings obey relations of a topological origin. A preliminary discussion of this was made in [39] and stronger results were recently obtained in [40]. Topological relations emerge because the particular integrals that arise in evaluating Yukawa couplings have a topological interpretation, in terms of the so-called cohomology ring of K. While that description holds in limiting low energy field theory, it can be shown [37] that at least the original relations derived in [39] also hold in string theory, at least to all finite orders in sigma model perturbation theory.

What I find promising about this is the following. In the quark and lepton mass matrices, certain elements seem to be zero or very small. For instance, certain of the fermions are extremely light, and in the the Fritzsch form of the mass matrix, certain matrix elements are taken to be zero. The most straightforward approach to trying to get some matrix elements to vanish is to assume suitable global symmetries. However, attempts to explain the observed form of the fermion mass matrices via global symmetries have always led to difficulties. At least in simple approaches, symmetries that forbid unwanted elements of the mass matrices seem to also forbid elements of the mass matrices that are not zero or small in nature. Higher dimensional unification may ultimately shed a completely new light on this problem since the unwanted elements of the mass matrix could well vanish for topological reasons without unwanted consequences for other matrix elements. Although models with realistic mass matrices are probably still far way, it is already possible to say something about Yukawa couplings in toy models. For instance, in [7] it was determined which Yukawa couplings were zero in the four generation model introduced in [5]. (It was not necessary to use the full force of topological reasoning since a relatively simple tool which might be referred to as 'pseudosymmetries' sufficed in that case.) The result was that of the four generations, two obtained tree level masses and two were massless at tree level. Various Yukawa couplings vanished at tree level, and interestingly the vanishing Yukawa couplings were such that (like the Yukawa couplings in the real world) the pattern would be difficult to explain via global symmetries. Yukawa couplings in another toy model were recently studied in [10].

The last topic I wish to discuss concerns the allowed values of electric and magnetic charge in theories that are unified only in some higher dimension [41]. In any theory that includes electromagnetism, one can define at spatial infinity the $U(1)$ Dirac monopole gauge field. Monopoles exist if this can be extended throughout all space without encountering a singularity. This is impossible if the gauge group is $U(1)$ or more generally if it is any group such as $SU(3) \times SU(2) \times U(1)$ which has a $U(1)$ factor. However, many years ago 't Hooft and Polyakov showed that in four dimensional unified theories one can always 'unwrap' the monopole, obtaining in conventional grand unified theories states whose magnetic charge equals the Dirac quantum. But what if unification in $SU(5)$ or $O(10)$ or E_6 occurs not in four dimensions but only in ten dimensions? In this case we cannot unwrap the monopole in the four dimensional unified gauge group, because there is none. We must try to unwrap it in the **ten** dimensional gauge group. The topological problem is completely different, and there is no reason to expect the answer to be the same. To be precise, what must be done to determine whether the monopole exists is the following. We begin with the Dirac monopole gauge field not on S^2 but on $S^2 \times K$ (here S^2 is the 'sphere at infinity' and K is the Kaluza-Klein space) and we try to extend it over $R^3 \times K$ (R^3 being physical three space) without singularity. Can this be done? It turns out [41] that there is a non-trivial obstruction. When the dust settles, the picture that emerges is the following. Magnetic monopoles always exist, but the minimum allowed value of magnetic charge is typically larger than it would be in conventional grand unified theories. Generically, if for instance $\pi_1(K) = Z_n$, the minimum possible magnetic charge is not the Dirac quantum $(2\pi/e)$ but is rather n times larger.

This is purely a field theoretic result, which emerges upon solving the topological problem I just described. However, the result that the minimum allowed value of magnetic charge is not $(2\pi/e)$ but $n \cdot (2\pi/e)$ is surprising at first sight, and one may ask what is the 'physical' reason for this result. It has no simple 'physical' explanation in field theory, as far as I know, but upon coupling to string theory a simple 'physical' explanation emerges in a dramatic way. I have been tacitly assuming that grand unified symmetry breaking was carried out by Wilson lines (otherwise, in fact, the whole topological problem must be reexamined), so implicit in the whole discussion is the fact that the fundamental group of K is non-trivial; in fact, we took it to be Z_n. This being so, in a theory that only has closed strings, there are stable, superheavy states in which a closed string wraps around a non-contractible loop in K. If we are incredibly lucky, we might one day observe such particles in cosmic rays; but that is another story. In any case, quantization of string theory in the 'winding' sectors shows that these modes have electric charge e/n precisely when the minimum allowed value of magnetic charge is $n \cdot (2\pi/e)$. I should stress, perhaps, that the winding states of electric charge e/n are color singlet, unconfined states. Their existence 'explains' why the minimum magnetic charge would be n times larger than the Dirac quantum. The remarkable thing about these results is that the magnetic charge is determined by geometric-topological methods which seem to have nothing to do with string theory, but which somehow 'know' that one day this problem will be coupled to the theory of closed strings.

In summary, unification in higher dimensions preserves the standard successes of four dimensional unification. The light fermions fit in representations of $O(10)$ or E_6, and the Georgi-Quinn-Weinberg relations are valid. On the other hand, there are a few interesting differences from the standard framework. The Yukawa couplings do not obey $O(10)$ or E_6 relations but obey instead other relations of topological origin. The Higgs expectation values are quantized, giving perhaps the possibility of an insight into the hierarchy problem. There are particles (albeit superheavy) whose electric charges could not arise in any representation of the grand unified group; and conjugate to this some of the usual magnetic monopoles are missing. There is every reason to think that higher dimensional theories, and especially superstring theory, will give us the opportunity in coming years to rethink some of the issues which have been left open by the successes of conventional grand unification and which have fascinated and puzzled us so much.

REFERENCES

1. M. B. Green and J. H. Schwarz, Phys. Lett. **149B** (1984) 117.
2. D. J. Gross, J. A. Harvey, E. Martinec, and R. Rohm, Phys. Rev. Lett. **54** (1985) 502.
3. E. Witten, Phys. Lett. **149B** (1984) 351.
4. P. Frampton, H. van Dam, and R. Yanamoto,UNC preprint (1984).
5. P. Candelas, G. Horowitz, A. Strominger, and E. Witten, Nucl. Phys. **B258** (1985) 46. See also the articles by G. Horowitz and A. Strominger in this volume.
6. M. B. Green, J. H. Schwarz, and P. West, Cal Tech preprint, 1984.
7. E. Witten, Nucl. Phys. **B258** (1985) 75, Princeton preprints, 1985.
8. M. Dine, V. Kaplunovsky, M. Mangano, C. Nappi, and N. Seiberg, IAS preprint, 1985.
9. R. Nepomechie, Wu Yong-Shi, and A. Zee, Univ. of Washington preprint, 1985.
10. K. Pilch and J. Schellekens, Stony Brook preprints, 1985.
11. S. Cecotti, J.-P. Derendinger, S. Ferrara, L. Girardello, and M. Roncadelli, CERN preprint.
12. J. Breit, B. Ovrut, and G. Segre, Univ. of Pa. preprint (1985).
13. A. Sen, Fermilab preprints, 1985.
14. M. Dine, N. Seiberg, R. Rohm, and E. Witten, IAS preprint, 1985.
15. V. Kaplunovsky, Princeton preprint, 1985.
16. M. Dine and N. Seiberg, IAS preprint, 1985.
17. J. P. Derendinger, L. E. Ibañez, and H. P. Nilles, CERN preprint (July, 1985), to appear in Nucl. Phys. B.

18. I. Bars, USC preprint 85/015 (1985); I. Bars and M. Visser, USC preprint (1985); I. Bars, D. Nemeschansky, and S. Yankielowicz, SLAC preprint (August, 1985).

19. C. Vafa and E. Witten, Comm. Math. Phys. **95** (1984) 257.

20. M. Gell-Mann, P. Ramond, and R. Slansky, in **Supergravity**, ed. P. van Nieuwenhuysen et. al. (North-Holland Publishing Company, 1979); F. Wilczek and A. Zee, Princeton report, 1979 (unpublished), Phys. Rev. **D25** (1982) 553. A closely related idea can be attempted in the context of a four dimensional theory with E_8 gauge group, and has somewhat similar successes and difficulties; see I. Bars and M. Günaydin, Phys. Rev. Lett. **45** (1980) 859.

21. J. A. Bagger, S. Dimopoulos, E. Masso, and M. H. Reno, SLAC preprint (1984).

22. E. Witten, Nucl. Phys. **B186** (1981) 412.

23. G. Chapline and R. Slansky, Nucl. Phys. **B209** (1982) 461

24. C. Wetterich, Nucl. Phys. **B223** (1983) 109

25. S. Randjbar-Daemi, A. Salam, and S. Strathdee, Nucl. Phys. **B214** (1983) 491.

26. E. Witten, in the Proceedings of the 1983 Shelter Island Conference, ed. N. Khuri et. al. (MIT Press, 1985).

27. D. Olive and P. West, Nucl. Phys. B217 (1983) 248.

28. G. Chapline and B. Grossman, Phys. Lett. **125B** (1984) 109.

29. P. H. Frampton and K. Yamamoto, Phys. Rev. Lett. **52** (1984) 2016.

30. S. Weinberg, Phys. Lett. **138B** (1984) 47

31. P. H. Frampton and T. W. Kephardt, Phys. Rev. Lett. **53** (1984) 867

32. I. G. Koh and H. Nishino, Trieste preprint ICTP/84/129

33. E. Calabi, in **Algebraic Geometry and Topology: A Symposium in Honor of S. Lefschetz** (Princeton University Press, 1957) p. 58.

34. S.-T. Yau, Proc. Natl. Acad. Sci. **74** (1977) 1798.

35. F. Gursey, P. Ramond, and P. Sikivie, Phys. Lett. **60B** (1976) 177.

36. P. Candelas, G. Horowitz, A. Strominger, and E. Witten, to appear in the Proceedings of the Argonne-Chicago Symposium on Anomalies, Geometry, and Topology; C. G. Callan, Jr., D. Friedan, E. Martinec, and M. J. Perry, Princeton preprint, 1985. See also the articles of D. Friedan, E. Martinec, and S. Shenker in this volume.

37. See E. Witten, 'New Issues in Manifolds of $SU(3)$ Holonomy,' to appear in Nucl. Phys. B, for a discussion of more general vacuum states based on Calabi-Yau manifolds in which (while keeping unbroken supersymmetry) E_8 is broken to $SO(10)$ or $SU(5)$ rather than E_6. The construction involves the use of stable holomorphic vector bundles over the Calabi-Yau manifold other than the tangent bundle. A discussion of these more general vacuum states would take us too far afield.

38. Y. Hosotani, Phys. Lett. **129B** (1983) 193.

39. A. Strominger and E. Witten, to appear in Comm. Math. Phys.

40. A. Strominger, Santa Barbara preprint, 1985.

41. X. G. Wen and E. Witten, to appear in Nucl. Phys. B. See also my article, 'Topological Tools in Ten Dimensional Physics' in this volume.