Deconstruction, $G_2$ Holonomy, and Doublet-Triplet Splitting

Edward Witten

Institute For Advanced Study, Princeton NJ 08540 USA

We describe a mechanism for using discrete symmetries to solve the doublet-triplet splitting problem of four dimensional supersymmetric GUT’s. We present two versions of the mechanism, one via “deconstruction,” and one in terms of $M$-theory compactification to four dimensions on a manifold of $G_2$ holonomy.

October, 2001
1. Introduction

One of the central problems in four-dimensional Grand Unified Theories (GUT’s) is the splitting between standard model Higgs doublets and their color triplet partners. The problem persists in supersymmetric GUT’s, which will be the focus of the present paper.

In $SU(5)$ models, for example, the Higgs doublets can be naturally placed in chiral superfields $V, \bar{V}$ transforming as $5 \oplus \bar{5}$. $V$ has for its standard model content a possible Higgs doublet $H$ as well as a color triplet $Q$, while $\bar{V}$ has fields $\bar{H}, \bar{Q}$ transforming in the conjugate representations. To get standard model phenomenology, $H$ and $\bar{H}$ must be essentially massless at the GUT scale – receiving mass only at the electroweak scale. But $Q$ and $\bar{Q}$ have renormalizable couplings – related by $SU(5)$ to the couplings of $H$ and $\bar{H}$ that are needed to give mass to quarks and leptons – that mediate baryon number violating processes. $Q$ and $\bar{Q}$ must therefore obtain masses close to the GUT scale in order to obtain an even roughly reasonable proton lifetime. (Even if this is achieved, there are more obstacles to getting a realistic proton lifetime; they will be discussed in section 2.1.)

A variety of field theory solutions to the doublet-triplet splitting problem or fine-tuning problem have been proposed. For a brief review of some of the proposals up to 1995, see [1]. There also are more recent field theoretic proposals such as one based on strong supersymmetric dynamics [2].

Possible solutions to the problem also exist in the framework of string theory and higher dimensions [3]. In this context, unification only arises in some dimension greater than four and the unified group $G$ is broken down to the standard model (or an extension of the standard model that is phenomenologically viable at relatively low energies) in the process of compactification. A key ingredient in this approach is “gauge symmetry breaking by Wilson lines,” in which one aims, while compactifying from ten to four dimensions, to project the dangerous color triplets out of the low energy spectrum while leaving the Higgs doublets.\footnote{A generalization of symmetry breaking by Wilson lines is symmetry breaking by orbifolds [4], where the symmetry breaking is carried out using a discrete symmetry that does not act freely. In perturbative string theory, this is not usually used as a mechanism for GUT symmetry breaking, because typically the massless twisted sector modes would not be in complete $G$ multiplets and the successes of grand unification would not be preserved.}

For a review of Calabi-Yau compactification of the heterotic string, in which this mechanism can naturally be incorporated, see chapters 14-16 of [5]. For a more recent discussion of some stringy constructions, see [6]. Mechanisms of roughly this type have
lately come to be widely studied from a more phenomenological and bottom-up point of view [7-11].

In the present paper, we will be concerned with solutions to the fine-tuning problem that make use of discrete symmetries. In fact, some but not all field theory proposals for the fine-tuning problem and some but not all string theory proposals make use of discrete symmetries.\(^2\) The basic reason that discrete symmetries might be relevant to the fine-tuning problem, in a supersymmetric GUT-like theory, is as follows.

Suppose that we are given a discrete symmetry \(F\) under which the components \((Q, H)\) of the \(5\) transform as \((e^{i\alpha}, e^{i\beta})\), while the components \((\bar{Q}, \bar{H})\) of the \(\bar{5}\) transform as \((e^{i\gamma}, e^{i\delta})\). If \(e^{i(\alpha+\gamma)} = 1\), then this symmetry allows \(Q\) and \(\bar{Q}\) to get GUT scale masses, while if \(e^{i(\beta+\delta)} \neq 1\), then \(H\) and \(\bar{H}\) are massless. In fact, in this scenario, the “\(\mu\)-term,” an \(H\bar{H}\) term in the superpotential that is needed for supersymmetric phenomenology, violates \(F\) and can only arise at lower energies where (hopefully) \(F\) is spontaneously broken.

From a field theory point of view, it can be difficult, depending on one’s assumptions, to get a discrete symmetry with the necessary properties. It is generally not true in GUT’s that a discrete symmetry of the low energy theory must commute with the GUT group \(G\); it might be the product of a discrete symmetry that “normalizes” the standard model subgroup of \(G\) (conjugates it to itself) times an ordinary discrete symmetry that commutes with \(G\). However, if \(G = SU(5)\), an element of \(G\) that normalizes \(SU(3) \times SU(2) \times U(1)\) is actually contained in \(SU(3) \times SU(2) \times U(1)\). This means that, modulo a standard model gauge transformation, a discrete symmetry in a four-dimensional \(SU(5)\) model actually commutes with \(SU(5)\). A discrete symmetry that is the product of a standard model gauge transformation and a symmetry that commutes with \(SU(5)\) leaves the \(H\bar{H}\) term in the superpotential invariant if and only if the \(Q\bar{Q}\) term is invariant. (Both terms are invariant under the standard model gauge group, and a discrete symmetry that commutes with \(SU(5)\) does not distinguish them either.) So such a discrete symmetry cannot solve the fine-tuning problem.

Things are no different in four-dimensional GUT’s based on the other standard simple GUT groups such as \(SO(10)\) and \(E_6\). The reason is that each of these groups, with the usual standard model embedding, contains a unique \(SU(5)\) subgroup \(G’\) that contains \(SU(3) \times SU(2) \times U(1)\), and the above argument can be carried out using \(G’\).

---

\(^2\) Discrete symmetries were not used, for example, in the doublet-triplet splitting mechanism proposed in [3].
As explained in [12], in the context of a four-dimensional $SU(5)$ model, mixing the $5$ and $\bar{5}$ with additional $SU(5)$ representations does not change this conclusion, but if one starts above four dimensions, one can readily get discrete symmetries of the desired kind.

This is a benefit of having extra dimensions. However, it has recently been pointed out [13,14] that some higher-dimensional setups can be “deconstructed,” or simulated by a four-dimensional model in which, roughly speaking, the extra dimensions are replaced by a lattice (which may have a very small number of lattice points). In section 2, following this lead, we deconstruct one version of the higher-dimensional approach to the doublet-triplet splitting problem. In its minimal form, this involves beginning with an $SU(5) \times SU(5)$ gauge theory, with the standard model diagonally embedded in the product of the two $SU(5)$’s. Such structures have been considered in many papers on deconstruction such as [13,15]. By starting with $SU(5) \times SU(5)$ rather than $SU(5)$, the constraints on discrete symmetries are relaxed, and it is readily possible to find discrete symmetries that can solve the fine-tuning problem.

In section 3, we present a higher-dimensional version of the same mechanism. In fact, we describe how discrete symmetries that can naturally split triplets from doublets can arise in the context of $M$-theory compactification to four dimensions on a manifold of $G_2$ holonomy. This is a natural way to obtain a four-dimensional model with $\mathcal{N} = 1$ supersymmetry, and since it can be dual to heterotic string compactification on a Calabi-Yau threefold, it is fairly clear that it must be possible to express some of the mechanisms for doublet-triplet splitting that are familiar for the heterotic string in the language of compactification on $G_2$ manifolds. We do this in section 3. Some of the ingredients of this construction have appeared in previous papers [16,17].

In fact, the approach sketched in section 3 was worked out first. Deconstruction was attempted following a question raised by Hsin-Chia Cheng, when this work was presented in a seminar at the University of Chicago.

**Similarities And Differences Of The Two Approaches**

The deconstructed theory presented in section 2 is not technically a unified theory by some definitions, since the $SU(5) \times SU(5)$ gauge theory has two independent gauge couplings (there is no symmetry exchanging the two factors). However, the diagonal embedding of the standard model ensures that most of the familiar consequences of grand unification, such as the SUSY-GUT prediction for $\sin^2 \theta_W$ and constraints on the quantum
numbers of quarks and leptons, do hold. By contrast, the $M$-theory model unifies not just the gauge fields but also the Higgs fields, standard model fermions, and gravity.

The two types of model differ in a few other interesting ways. In $M$-theory models, it is believed (there is not a complete proof of this) that the discrete symmetries are always anomaly-free (they may be spontaneously broken by the transformation law of an axion). In the context of deconstruction, opinions may differ about whether an anomalous discrete symmetry should be considered technically natural, but at any rate it would be phenomenologically viable to try to solve the fine-tuning problem using such a symmetry.

In the deconstructed model, gauge anomalies must cancel separately in each $SU(5)$ factor of the gauge group. Gauge anomaly cancellation is a less severe constraint in the $M$-theory approach, since there is an anomaly inflow mechanism [16] (analogous to anomaly inflow for $D$-branes [18]) that can shift the anomaly from one factor to the other. Anomaly inflow, since it involves Chern-Simons-like couplings, appears difficult to deconstruct, but see [19] (which appeared on hep-ph a few days after the original version of the present paper). Finally, like most perturbative heterotic string models [20], the models derived from $M$-theory generally have superheavy unconfined color singlet particles with fractional electric charges, and reciprocally a larger quantum of magnetic charge than would be expected in a four-dimensional GUT. The deconstructed models obey conventional quantization of electric and magnetic charge.

After submission of the original version of the present paper, I learned of [21], which presents a construction similar to that in section 2.1.

2. $SU(5)' \times SU(5)''$ And Deconstruction

2.1. Direct Construction Of The Model

We start with a gauge theory in four-dimensions in which the gauge group is the product $G = SU(5)' \times SU(5)''$ of two copies of $SU(5)$. We suppose that the standard model group is diagonally embedded in the product of the two factors. The hypercharge group $U(1)_Y$, for example, is the diagonal subgroup of $U(1)'_Y \times U(1)''_Y$, the product of the hypercharge groups of the two $SU(5)$'s.

We assume that, in addition to the standard model being unbroken, a discrete global symmetry group $F' \cong \mathbb{Z}_n$ is unbroken at the GUT scale. We take $F'$ to be a diagonal product of an ordinary global symmetry $F = \mathbb{Z}_n$ (which commutes with $G$) and the $\mathbb{Z}_n$ subgroup of $U(1)'_Y$. In what follows, we pick a fixed generator of $F'$. An explicit and fairly
simple set of Higgs fields that can break $G \times F$ to $SU(3) \times SU(2) \times U(1) \times F'$ will be given in section 2.2.

Given this low energy structure, it is straightforward to solve the doublet-triplet splitting problem. We suppose that the Higgs bosons, whose expectation values will ultimately give masses to quarks and leptons, consist of multiplets $V, \tilde{V}$ transforming under $SU(5)' \times SU(5)''$ as $(5, 1) \oplus (1, 5)$. $V$ decomposes under the standard model as $(Q, H)$ and $\tilde{V}$ decomposes as $(\tilde{Q}, \tilde{H})$; here (as in the introduction) $H$ and $\tilde{H}$ are standard model Higgs fields and $Q, \tilde{Q}$ are colored partners.

$V$ is neutral under $U(1)'_Y$, so in this multiplet $F'$ acts as an ordinary global symmetry. Hence, under the generator of $F'$, $V$ transforms as

$$\left(\begin{array}{c} Q \\ H \end{array}\right) \rightarrow e^{i\alpha} \left(\begin{array}{c} Q \\ H \end{array}\right),$$

for some $\alpha$. But on the $(1, 5)$, $F'$ acts as the product of a global symmetry and a $U(1)'_Y$ gauge transformation, so the transformation under the generator of $F'$ is

$$\left(\begin{array}{c} \tilde{Q} \\ \tilde{H} \end{array}\right) \rightarrow e^{i\gamma \tilde{Q}} e^{i\delta \tilde{H}}.$$

Here $e^{i\gamma}$ and $e^{i\delta}$ are arbitrary $n^{th}$ roots of 1, depending on the choice of $F$ charge of the $(1, 5)$ as well as the precise diagonal subgroup of $F \times U(1)'_Y$ we have chosen for $F'$. Now it is clear how to solve the doublet-triplet splitting problem; we merely choose the charges so that $e^{i(\alpha + \gamma)} = 1$ but $e^{i(\alpha + \delta)} \neq 1$. Then a $Q\tilde{Q}$ term in the superpotential is $F'$-invariant, but $F'$ forbids an $H\tilde{H}$ term.

Now let us consider how to incorporate quarks and leptons in this model. There are many choices, as the standard model quantum numbers of quarks and leptons could originate from either or both of the two $SU(5)$’s. We consider two illustrative models:

**All Quarks And Leptons From The First Factor**

In our first model, we assume that all quark and lepton quantum numbers arise from the first $SU(5)$. So the quarks and leptons arise from three copies of $(10, 1) \oplus (\bar{5}, 1)$. $F'$ acts by ordinary $G$-invariant global symmetries on these multiplets. We assume that the $(10, 1)$’s all transform by multiplication by $e^{i\sigma}$, with a common $\sigma$, and likewise the $(\bar{5}, 1)$’s all transform by multiplication by $e^{i\tau}$ for some $\tau$. 

5
Let us see what constraints are required by phenomenology. To give masses to up quarks, we want \( H(10, 1)^2 \) superpotential couplings; for this, we need

\[ e^{i(\alpha + 2\sigma)} = 1. \]  

(2.3)

To give masses to down quarks and charged leptons, we want \( \tilde{H}(10, 1)(1, \bar{5}) \) interactions; for this, we need

\[ e^{i(\delta + \sigma + \tau)} = 1. \]  

(2.4)

The experimental observation of neutrino masses strongly suggests that an \( H^2(\bar{5}, 1)^2 \) coupling is allowed; in the context of GUT’s, this leads to neutrino masses of roughly the observed magnitude. For this coupling to be allowed, we need

\[ e^{2i(\alpha + \tau)} = 1. \]  

(2.5)

But we do not want to allow a \( H(\bar{5}, 1) \) mass term, so we want

\[ e^{i(\alpha + \tau)} = -1. \]  

(2.6)

We can solve these equations in terms of an arbitrary angle \( \sigma \):

\[ \begin{align*}
\alpha &= -2\sigma \\
\tau &= 2\sigma + \pi \\
\delta &= -3\sigma + \pi.
\end{align*} \]  

(2.7)

Finally, and of great importance, to get a realistic proton lifetime, we need additional restrictions. We want to avoid renormalizable couplings \((10, 1)(\bar{5}, 1)^2\) that violate baryon number, so we want \( e^{i(\sigma + 2\tau)} \neq 1 \), or in terms of the above solution

\[ 5\sigma \neq 0. \]  

(2.8)

Moreover, it is highly desireable to avoid \((10, 1)^3(\bar{5}, 1)\) terms in the superpotential, which lead to dimension five baryon nonconserving operators. It is difficult for a GUT-like model to generate such terms and have a sufficiently long-lived proton; for a recent account, see [22]. So we want \( e^{i(3\sigma + \tau)} \neq 1 \), or in terms of the above solution,

\[ 5\sigma + \pi \neq 0. \]  

(2.9)

\[ ^3 \text{All equations for these angles are of course understood mod } 2\pi. \]
One interesting feature of this model is that although the $H(10, 1)^2$ Yukawa couplings that give masses to up quarks can arise from ordinary $G$-invariant cubic terms $V(10, 1)^2$ in the superpotential, the $\tilde{H}(10, 1)(\bar{5}, 1)$ couplings that give the down quarks and charged lepton masses cannot so arise, given that $\tilde{H}$ transforms as $(1, \bar{5})$. These latter couplings must instead be induced from unrenormalizable superpotential couplings of the general form $\tilde{V}(10, 1)(\bar{5}, 1)\Phi$, where $\Phi$ is constructed from fields (introduced in section 2.2) whose expectation values break $G \times F$ to the low energy subgroup $SU(3) \times SU(2) \times U(1) \times F'$, which does allow the $\tilde{H}(10, 1)(\bar{5}, 1)$ couplings. Because they arise from unrenormalizable couplings, it is natural to expect the down quark and charged lepton masses to be less than the up quark masses. For the third generation, this is true, as the bottom quark and tau lepton are much lighter than the top quark. If the ratios $m_b/m_t$ and $m_{\tau}/m_t$ are really to be obtained this way, the cutoff scale characterizing unrenormalizable interactions cannot be too much bigger than the GUT scale, since after all $m_b/m_t$ and $m_{\tau}/m_t$ are only moderately small. Moreover, the fact that the first two generations are so light compared to the top quark would presumably require some further mechanism.

Finally, the spectrum of the model as we have presented it so far cannot be the whole story, since it is anomalous. The only chiral multiplet so far introduced that carries $SU(5)'$ charges is the Higgs multiplet $\tilde{V}$, transforming as $(1, \bar{5})$. The couplings of this field are anomalous. Likewise, $SU(5)''$ couples to three anomaly-free copies of $(10, 1) \oplus (\bar{5}, 1)$ as well as a Higgs multiplet $V$ transforming as $(5, 1)$; its couplings are again anomalous. A simple way to cancel the anomalies is to add additional fields $\tilde{S}, S$ transforming as $(\bar{5}, 1) \oplus (1, 5)$; their $F$ quantum numbers should be restricted to avoid various undesirable couplings. If one believes that $F$ should be anomaly-free for naturalness of the model, then the $F$ quantum numbers of $\tilde{S}, S$ must be further constrained. Purely for phenomenological purposes, however, gauge anomalies in $F$ would not lead to trouble.

It is interesting to speculate that the fields $\tilde{S}, S$ might play the role of “messenger fields” in gauge-mediated supersymmetry breaking (for surveys see [23,24]), communicating to the standard model fields the occurrence of supersymmetry breaking in a hidden sector. For this, there might be singlets $T$ whose expectation values violate supersymmetry and $F'$ and which have superpotential couplings $T\tilde{S}\tilde{S}$. Actually, since the color singlet and color triplet components of $S$ transform differently under $F'$ (and there is no such splitting for $\tilde{S}$), one would need different $T$ fields transforming differently under $F'$ to couple to the color singlets and triplets in $S\tilde{S}$. 

7
For a second model, which we will not develop as fully, we consider one possibility among many to use the distinct gauge theoretic origin of the two different Higgs fields to constrain quark and lepton masses. Since the top quark is much heavier than the charm or up quark, we might assume that the top originates from a $(\mathbf{10}, 1)$ while the charm and up quarks originate from two copies of $(1, \mathbf{10})$. Then the top quark gets a mass from renormalizable $V(\mathbf{10}, 1)^2$ couplings, while the charm and up masses originate from unrenormalizable interactions. Since the bottom quark and tau lepton are much heavier than the analogous particles in the first two generations, one might similarly suppose that the bottom quark arises from a $(1, \mathbf{5})$ (so that it can get its bare mass from a renormalizable coupling $\tilde{V}(\mathbf{1}, \mathbf{10}) (1, \mathbf{5})$) and the others from two copies of $(\mathbf{5}, 1)$. This spectrum, including the Higgs fields, is fortuitously anomaly-free so we do not need additional fields analogous to $S, \tilde{S}$ of the first model.

A problem with this model is that it will be hard to generate masses for all down quarks and charged leptons. In fact, one down quark mass and one charged lepton mass would have to come from a coupling $\tilde{H}(\mathbf{5}, 1)(1, \mathbf{10})$. Because different components of the $(1, \mathbf{10})$ transform differently under $F'$, while there is no such splitting for the $(\mathbf{5}, 1)$, this coupling cannot give a mass to both a down quark and a charged lepton, no matter what $F'$ charges we assume.

2.2. Interpretation Via Deconstruction

Next we will explain in what sense the above model can arise via deconstruction. First, let us explain what manifold is being deconstructed. We let $D_0$ be a two-dimensional disc. We can triangulate it as in the figure, with one vertex $P$ in the center and $n$ vertices $Q_1, \ldots, Q_n$ on the boundary.

The space we want to deconstruct is not $D_0$, but rather a space $D$ obtained by imposing the following equivalence relation: two points in $D_0$ that are on the boundary are equivalent if they differ by a $2\pi/n$ rotation of the boundary. Thus, an equivalence relation is imposed only on the boundary. If $n = 2$, $D$ is an unorientable manifold, the real projective plane $\mathbb{RP}^2$. Its deconstruction was described in [25] in the discussion of “spider web theory space.” The case $n = 2$ would not quite work for us, at least in the first model presented above, because (2.9) and (2.8) could not be obeyed. The deconstruction, however, is similar for $n > 2$, though $D$ is not a manifold (but a singular topological space) for $n > 2$. In fact, the triangulation or deconstruction of $D$ is very simple. There are only