Nonlinear Sigma Models

In systems with spontaneous symmetry breaking from a full symmetry group $G$ to a subgroup $H$ which is the invariance of the vacuum, the appropriate mathematical description is in terms of nonlinear realizations.


Let the generators of $H$ be denoted by $V_i$ and the remaining generators of $G$ be denoted by $Z_j$. The $V_i$ and the $Z_j$ together form a complete set of generators of $G$, which may be taken to be orthonormal with respect to the Cartan-Killing metric $g_{ab} = f_{ad} f_{bd}$ (which is nonsingular for semisimple $G$).

The Lie algebra of $G$ can be written

\[
\begin{bmatrix}
[V_i, V_j] &= i f_{ij}^{k} V_k \\
[V_i, Z_j] &= i f_{ij}^{m} Z_m \\
[Z_e, Z_m] &= i f_{em}^{n} Z_n + f_{em}^{k} V_k
\end{bmatrix}
\]

For a compact group $G$, take the generators to be Hermitian, structure constants $f_{ij}^{k}$ real.

If $f_{em}$ in the third commutation relation vanish, then $G/H$ is a symmetric space and the algebra of $G$ admits an automorphism $V_i \rightarrow V_i, \ Z_j \rightarrow -Z_j$. [This is the case for chiral $SU(n) \times SU(n)$ groups: the parity operator induces an automorphism which changes the sign of the axial vector generators.]

A general group element $g \in G$ can be represented uniquely in the form $g = c^{i} s_{i} z_{i} h$, where $h \in H$ and the $s_{i}$ parametrize the coset space $G/H$. Acting on the left with $g_{0} \in G$, one can separate the product $g_{0} g$ in the same way:

\[
g_{0} g = c^{i} s_{i} z_{i} h_{1} \quad \text{and} \quad g_{0} c^{i} s_{i} z_{i} h = c^{i} s_{i} z_{i} h_{1}
\]

In general, $s_{i}$ and $h_{1}$ are functions of $g_{0}$ and $s_{m}$.

For $g_{0} = h_{0} \in H$, however, one has

\[
(ch_{0} c^{i} s_{i} z_{i} h) h_{0} = c^{i} s_{i} z_{i} h_{1}
\]
Since the $Z_e$ form a representation of $H$, this implies
\[ e^{i\xi \cdot z} = h_0 e^{i\xi \cdot z} h_0^{-1} \quad \text{and} \quad h' = h_1 h, \quad h_1 = h_0 \]
So in this case the transformation $\xi^m \rightarrow \xi'^m$ is linear.

On the other hand, for $\varphi_0 = e^{i\xi_0 \cdot z}$, one has
\[ e^{i\xi_0 \cdot z} e^{i\xi \cdot z} = e^{i\xi \cdot z} h_1 \]
which is a nonlinear and inhomogeneous transformation on the $\xi^m$.

For infinitesimal transformations, one has, where $I$ is the identity,
\[ e^{i\xi \cdot z} (\varphi_0 - 1) - e^{i\xi \cdot z} \int e^{i\xi \cdot z} = h_1 - I \]
Consequently, for $\varphi_0 = e^{i\xi_0 \cdot z}$ with $\xi^m$ infinitesimal, one has
\[ e^{i\xi \cdot z} (\varphi_0 - 1) e^{i\xi \cdot z} - e^{i\xi \cdot z} \int e^{i\xi \cdot z} = h_1 - I \]

On the LHS of this equation, one has only generators $V_i$ of $H$. So the coefficients of the $Z_e$ on the LHS must be set equal to zero. Using the algebra of the group $G$, this allows one to calculate the $\xi'^m$.

In the case of a symmetry-space $G/H$, where there is a $Z_e \rightarrow Z_e$, $V_i \rightarrow V_i$ automorphism, one also has
\[ e^{i\xi_0 \cdot z} e^{i\xi \cdot z} = e^{i\xi \cdot z} h_1 \quad \text{(Same $h_1$)} \]
This allows one to eliminate $h_1$ and obtain
\[ e^{i\xi_0 \cdot z} e^{i\xi \cdot z} e^{i\xi \cdot z} = e^{i\xi \cdot z} \]
Above, we have just considered the nonlinear transformations of the coset parameter $\xi^m$ in the abstract. They also give the transformations of nonlinearily transforming fields $\xi^m(x)$ under rigid $G$ transformations generated by $V_i$ or $Z_e$. In order to construct covariantly transforming differential quantities, one employs the Cartan-Maurer differential form $\xi = \omega_i dx^i$
\[ \omega_i = e^{i\xi \cdot z} \frac{d}{dx^i} e^{i\xi \cdot z} = \xi_i \cdot z + i \xi \cdot V. \quad \text{Dependence:} \quad \xi_i (x, z, V, x') \]
The forms $\omega_i$ and $V = \frac{d}{dx} \omega_i$ transform linearly under $H$. For coset generated transformations $\varphi_0 = e^{i\xi_0 \cdot z}$, on the other hand,
\[ \xi'_i = h_1 \xi_i (h_1)^{-1} + h_1 \frac{d}{dx} \omega_i (h_1)^{-1} \]
\[ i \xi'_i V = h_1 \xi_i V (h_1)^{-1} + i h_2 \xi_i (h_1)^{-1} \quad \text{(gauge field for $H'$)} \]
The transformation law of $\tilde{\pi}^m$ is therefore of the same form as for a linear $H$ transformation, except now with the group element $h_1 (y, g_o)$. Accordingly, $H$-invariant products of the $\tilde{\pi}^m$ are automatically invariant under the full group $G$. For example, one can construct a $G$-invariant Lagrangian kinetic term $-\frac{1}{2} \tilde{\pi}^m \tilde{\pi}^n g_{mn}$ with $g_{mn}$ the Cartan-Killing metric for $H$. For compact semisimple groups $G$ there is an appropriate normalization such that $g_{mn} = \delta_{mn}$, so then one has simply $-\frac{1}{2} \tilde{\pi}^m \tilde{\pi}^n$ as the rigidly $G$-invariant kinetic term for the $\tilde{\pi}^m(x)$.

The transformation of the $H$-valued part $\psi_1$ of $\psi$ is also important for the construction of $G$-invariant Lagrangians. Let $\psi$ be some additional field carrying a linear representation $R$ of the stability subgroup $H$. Thus

$$g_o : \psi \rightarrow \psi' = R(h_1) \psi$$

For $h_1 = h_0 \in H$, this is a standard linear realization of the rigid transformation $h_0 \in H$. For a transformation $g_o = e^{i \phi^a T_a}$, however, one has $h_1 (\phi^a(x), g_o)$ from the general nonlinear realization structure, and although $g_o$ will be taken to be a rigid ($x^a$ independent) symmetry transformation, $h_1 (\phi^a(x), g_o)$ acquires $x^a$ dependence from the dependence on $x^a$ of the nonlinearly transforming "Goldstone" field $\phi^a(x)$.

One can straightforwardly make $G$-invariant Lagrangian terms for $\psi$ simply by hooking up indices in a manifestly $H$-invariant fashion. For nonderivative terms, this poses no problem, but for derivatives the $x^a$ dependence of $h_1 (\phi^a, g_o)$ requires a refined construction. This is where the $V_a$ come in: $\partial_a \psi \rightarrow \partial_a \psi' = i h_1 (\partial_a \psi h_1 + h_1 \partial_a (h_1))$ is a transformation of gauge-field type, even for $x^a$ independent $g_o$.

Recall that $\psi_a (\phi, g_o)$ is not an independent gauge field in its own right, but is built from $\phi^a(x)$ and $\phi^m(x)$ in the Cartan-Maurer form.
Use $U(\gamma, \delta)$ to construct a covariant derivative for the $R(h_2)$ representation transformation of $\mathcal{H}$. For any representation $R$ of $\mathcal{H}$, one has $R(h_2) = R(h_2^+) R(h_2^-)$ for all $h_2, h_2' \in \mathcal{H}$.

So, defining $D_a Y = \partial_a Y + i R \left( \frac{\gamma_a}{\gamma} \cdot \nu \right) Y$ one finds

$D_a Y \rightarrow D_a (R(h_2) Y) = R(h_2^+) D_a Y + R(\partial_a h_2) Y$ and

$i (\frac{\gamma_a}{\gamma} \cdot \nu) Y \rightarrow i R \left( h_2^+ \frac{\gamma_a}{\gamma} \cdot \nu \cdot (h_2^-) \right) R(h_2^+) Y + R(\partial_a h_2) (h_2^-) R(h_2^+) Y$

$= i R(h_2^+) R \left( \frac{\gamma_a}{\gamma} \cdot \nu \right) Y - R(h_2^+) R(\partial_a h_2) R(h_2^-) R(h_2^+) Y$

$= i R(h_2^+) R \left( \frac{\gamma_a}{\gamma} \cdot \nu \right) Y - R(\partial_a h_2) Y$

So, putting things together $D_a Y \rightarrow R(\partial_a h_2) D_a Y$, transforming linearly in the $R$ representation of $\mathcal{H}$ for the element $h_2 (\gamma, \delta)$. In this manner, one also can construct $G$-invariant Lagrangians simply by picking up the $R$ representation indices to make invariants. Note that any $H$ invariant term constructed in this way using $\frac{\gamma_a}{\gamma}$ and $\frac{\gamma_a}{\gamma} \cdot \nu$ will yield a rigidly $G$-invariant term by this construction.

The construction of the covariant derivatives $\frac{\gamma_a}{\gamma} \cdot \nu$ and $\frac{\gamma_a}{\gamma}$ are made more explicit using the formula

$\frac{\gamma_a}{\gamma} \cdot \nu \left( \frac{\gamma_a \cdot \gamma}{\gamma} \cdot \nu \right) \left[ \left( 1 - e^{i A_{5/2}} \right) \frac{\gamma_a \cdot \gamma}{\gamma} \right] \frac{\gamma_a}{\gamma}$

where the operator $A_{5/2}$ is defined by $A_{5/2} X = \left[ \frac{\gamma_a \cdot \gamma}{\gamma} \cdot \nu, X \right]$ and functions of $A_{5/2}$ are given by their power-series expansions. Then it follows that $\left( \frac{\gamma_a \cdot \gamma}{\gamma} \cdot \nu \right) \left[ \left( 1 - e^{i A_{5/2}} \right) \frac{\gamma_a \cdot \gamma}{\gamma} \right] \frac{\gamma_a}{\gamma}$

where the ... terms are nonlinear. For systems such as the $SU(\frac{\gamma}{\gamma}) \times SU(\frac{\gamma}{\gamma}) / SU(\frac{\gamma}{\gamma})$ coset, the commutator of two $\frac{\gamma_a}{\gamma}$ generators gives a $\gamma$ generator and it is then useful to separate the odd from the even multiple commutators. Then one obtains $\frac{\gamma_a}{\gamma} \cdot \nu = -i \left[ \left( 1 - e^{i A_{5/2}} \right) A_{5/2} \right] \frac{\gamma_a}{\gamma}$

and $\frac{\gamma_a}{\gamma} \cdot \nu \left( \frac{\gamma_a \cdot \gamma}{\gamma} \cdot \nu \right) \left[ \left( 1 - e^{i A_{5/2}} \right) \frac{\gamma_a \cdot \gamma}{\gamma} \right] \frac{\gamma_a}{\gamma} = 0$

Using $\frac{\gamma_a}{\gamma}$ and $\frac{\gamma_a}{\gamma} \cdot \nu$, one can construct $G$-invariant Lagrangian terms $- \frac{1}{2} \frac{\gamma_a}{\gamma} \cdot \nu \frac{\gamma_a}{\gamma}$ and have consequently $G$-covariant field equations $(D_a \frac{\gamma_a}{\gamma}) \frac{\gamma_a}{\gamma} = 0$, i.e. $D_a \frac{\gamma_a}{\gamma} \cdot \nu + \frac{\gamma_a}{\gamma} \cdot \nu \frac{\gamma_a}{\gamma} \cdot \nu = 0$, which can also be written $D_a \frac{\gamma_a}{\gamma} \cdot \nu \frac{\gamma_a}{\gamma} \cdot \nu = 0$, and $[\frac{\gamma_a}{\gamma} \cdot \nu, \frac{\gamma_a}{\gamma} \cdot \nu \frac{\gamma_a}{\gamma} \cdot \nu] = 0$, using the representation generators $T_{ij} = -i \frac{\gamma_a}{\gamma} \cdot \nu$ for the $\frac{\gamma_a}{\gamma}$ rep carried by $\frac{\gamma_a}{\gamma}$. 
Was apply this general nonlinear realization formalism to the chiral symmetry effective field theory. A conventional way to write the nonlinear sigma model for $SU(3) \times SU(3)/SU(3)$ is in terms of the $SU(3)$ matrix

$$U(x) = \exp \left( 2i \frac{\pi^a R^a}{F_{\pi}} \right) = \exp \left[ \frac{i}{F_{\pi}} \begin{pmatrix}
\frac{1}{12} \pi^0 + \frac{\sqrt{6}}{6} \eta^0 & \pi^+ & K^+
\pi^- & -\frac{1}{12} \pi^0 + \frac{\sqrt{6}}{6} \eta^0 & K^0
K^- & K^0 & -\frac{\sqrt{2}}{18} \eta^0
\end{pmatrix}
\right]$$

where the $SU(3)$ generators $T^a = \frac{1}{2} \lambda^a$, with $\lambda^a$ the eight Gell-Mann matrices. $F_{\pi}$ is a constant with dimensions of energy, known as the pion decay constant. Fitting to data gives $F_{\pi} = 92$ MeV/$c^2$.

The $SU(3) \times SU(3)/SU(3)$ chiral sigma model gives an account of the interactions of pions, kaons and the $\eta^0$ that works well phenomenologically after the large strange quark mass is incorporated via perturbation theory. To simplify the story, however, let us now exclude the third, strange, flavour and focus just on the $SU(2) \times SU(2)/SU(2)$ subsystem for the pions:

$$U(x) = \exp \left( 2i \frac{\eta^a R^a}{F_{\pi}} \right) = \exp \left[ \frac{i}{F_{\pi}} \begin{pmatrix}
\eta^0 & \sqrt{2} \pi^+ \\
\sqrt{2} \pi^- & -\eta^0
\end{pmatrix}
\right] \in SU(2)$$

where $R^a = \frac{1}{2} \lambda^a$ are the $SU(2)$ generators. Under $SU(2) \times SU(2)$, the pions transform as $U \rightarrow g_L U g_R$. The simplest invariant under $SU(2) \times SU(2)$ that one can write is

$$\mathcal{L}_{\text{hadron}} = -\frac{F_{\pi}^2}{4} \text{Tr} \left[ (\partial \Phi \partial \Phi) - \partial \mathcal{W} \partial \mathcal{W}^* \right]$$

where the electromagnetic coupling to $A_\mu$ has been retained in

$$D_\mu \Phi = (\partial_\mu - i q A_\mu) \Phi$$

for a scalar field with electromagnetic charge $q$. This explicitly breaks the $SU(2) \times SU(2)$ rigid symmetry, but is included for physical completeness. Setting $A_\mu$ to zero restores the chiral symmetry. Expanding $\mathcal{L}_{\text{hadron}}$ out to quadratic order yields normal scalar kinetic terms and electromagnetic couplings, followed by specific interactions

$$\mathcal{L}_{\text{hadron}} = -\frac{1}{2} \partial \mu \eta \partial^\mu \eta - \frac{1}{2} \partial \mu \partial^\mu \pi - \frac{1}{2} F_{\mu \nu} F^{\mu \nu} - \frac{1}{18} F_{\mu \nu \rho \sigma} F^{\mu \nu \rho \sigma} + \cdots$$

and so on.
In this non-renormalizable effective theory, terms with more derivatives are also expected, but these are contained by the $SU_c(2) \times SU_6(2)$ symmetry. The next possible terms have four derivatives:

$$S_4 = L_1 \left[ (\partial \Phi)^2 \Phi^{*2} \right]^2 + L_2 + \left[ \partial \Phi \partial \Phi^{*2} \partial \Phi \partial \Phi^{*2} \right]$$

with the coefficients determined from scattering experiments:

$$L_1 = 0.65 \quad L_2 = 1.89 \quad L_3 = -3.06$$

Since these terms have extra derivatives, their contributions at low energies will be suppressed at low energies by powers of $E^2 / F_0$ compared to the predictions of the leading two-derivative term, for which the $\frac{-i}{\pi}$ coefficient gives canonically normalized $\pm \partial^2 \Phi + \partial^2 \Phi^*$ kinetic terms.

Now let us see what can be said about the effect of quark mass terms in the underlying QCD theory. Chiral masses explicitly break the chiral $SU_c(2) \times SU_6(2)$ symmetry, with the consequence that the pions are not massless Goldstone bosons, but instead are light pseudo Goldstone bosons. With the quark mass terms as

$$S_m = -\frac{i}{2} q M q$$

with $M = (m_u \, m_d)$. Now we employ a trick to restore the chiral symmetry that is broken by $S_m$: assign a transformation property to $M$: $M \to q, M q$. This makes $S_m$ invariant, formally. When constants are treated as fields in this way, they are called spinors.

Although the low-energy theory of pions does not contain quarks, its details can depend on the mass matrix $M$. One can use the spinors $SU_c(2) \times SU_6(2)$ to constrain the effect of the quark mass matrix $M$ on the nonlinear chiral effective Lagrangian. The leading $SU_c(2) \times SU_6(2)$ term that can be added to the pion nonlinear sigma model is

$$S_m = \frac{V}{2} (M + M^T)$$

Expanding this, one obtains

$$S_m = \frac{V}{2} \left( m_u \, m_d \right) \left( m_u \, m_d \right)^T + \ldots$$
The coefficient $V^3$ is fixed by requiring the vacuum energy arising from $\langle \Delta \rangle$ to match that arising from $\langle \Delta \rangle$. For $\langle \Delta \rangle = \langle \Delta \rangle_0 = V^3$, one has $\langle \Delta \rangle = V^3 \langle m + \eta \rangle$, which matches the constant term in the expansion of $\langle \Delta \rangle$. Given that assignment, one obtains the pion masses from the quadratic term in $m$:

$$m^2 = \frac{V^3}{F_\pi^2} \langle m + \eta \rangle.$$

This is the Call-Mann-Oakes-Renner relation: the square of the pion mass scales linearly with the quark masses. For $V \sim V_{QCD} \sim 250$ MeV/$c^2$, $F_\pi = 92$ MeV, and $m = 140$ MeV, one obtains $m + \eta \sim 10.6$ MeV/$c^2$. This sort of value agrees with results from lattice QCD. So, although based just on leading approximations, the constraints of spontaneously broken chiral symmetry give rather good results in comparison with experiment and more elaborate phenomenological calculations.

Another relation that one can obtain from the effective chiral symmetry theory concerns interactions between pions and nucleons. Historically, interactions between pions $\pi^k$ and nucleons $N = (n)$ were described in terms of a Yukawa interaction $\mathcal{L}_{\pi NN} = -g_{\pi NN} (\bar{N} \gamma_5 \tau^n N) \pi^k$, with $g_{\pi NN}$ found to be close to 14. (Note: $\gamma_5$ coupling since $\pi^k$ are pseudoscalars.)

On the other hand, consider the most general $SU(2) \times SU(2)$ and parity invariant coupling between pions and nucleons:

$$\mathcal{L}_{\pi NN} = -i \bar{N} \gamma_\mu \partial_\mu N - i \frac{g}{2} \bar{N} \gamma_\mu \gamma_5 \tau^n \gamma_\mu \gamma_5 \gamma_5 N \pi^k.$$

Integrating by parts and using the nucleon Dirac equation $(\gamma_\mu \partial_\mu) N = 0$ yields an interaction $-\frac{m^2_{\pi NN}}{F_\pi} \bar{N} \gamma_5 \tau^n \gamma_5 \gamma_5 N \pi^k + \cdots$ leading to the identification $g_{\pi NN} = \frac{g_{\pi NN}}{F_\pi}$.

This is the Goldberger-Treiman relation.
The value of the pion-nucleon coupling constant $g$ is obtained from consideration of neutron decay. Obtain the Noether currents $j_{\mu}^p$ and $j_{\mu}^\pi$ from the SL(2) x SL(2) transformations $g_L = 1 + \frac{i}{2} (2\bar{x} + \bar{\beta}) \gamma \gamma$ and $g_R = 1 + \frac{i}{2} (2\bar{x} - \bar{\beta}) \gamma \gamma$ and let $j_{\mu}^p$ and $j_{\mu}^{\pi}$ become space-time dependent to calculate the currents. In the original QCD action, one has $\mathcal{L}_{\text{QCD}} = \int d^4 x \left( j_{\mu}^{p \pi} j_{\mu}^{p \pi} + \frac{1}{3} \bar{\beta} \cdot \bar{\beta} j_{\mu}^A j_{\mu}^A \right)$ while in the effective low-energy theory, one has $\mathcal{L} = \frac{1}{2} \left( \bar{\lambda} \cdot \Gamma \lambda + \bar{\bar{\lambda}} \cdot \bar{\Gamma} \bar{\lambda} \right)$ so from $\mathcal{L}_{\text{QCD}} = \int d^4 x \left( \bar{\lambda} \cdot \Gamma \lambda + \bar{\bar{\lambda}} \cdot \bar{\Gamma} \bar{\lambda} \right)$ one finds

\[ j_{\mu}^p = - \varepsilon_{\mu \nu \rho \sigma} \partial \bar{\lambda}^{(\nu)} \sigma^{(\rho)} + \ldots \]
\[ j_{\mu}^A = \frac{\lambda}{2} \left( \bar{\lambda} \cdot \Gamma \lambda + \bar{\bar{\lambda}} \cdot \bar{\Gamma} \bar{\lambda} \right) \]

When coupling to nucleons is included, these become

\[ j_{\mu}^p = - \varepsilon_{\mu \nu \rho \sigma} \partial \bar{\lambda}^{(\nu)} \sigma^{(\rho)} + \frac{1}{3} \bar{\beta} \cdot \bar{\beta} \lambda^A \lambda^A N \]
\[ j_{\mu}^A = \frac{\lambda}{2} \left( \bar{\lambda} \cdot \Gamma \lambda + \bar{\bar{\lambda}} \cdot \bar{\Gamma} \bar{\lambda} \right) + \frac{1}{2} \bar{\beta} \cdot \bar{\beta} \lambda^A \lambda^A N + \ldots \]

Nuclear matrix elements are

\[ \langle N(p,0) | j_{\mu}^p | N(p',0) \rangle = \frac{1}{2} \varepsilon_{\mu \nu \rho \sigma} \partial \bar{\lambda}^{(\nu)} \sigma^{(\rho)} + \frac{1}{3} \bar{\beta} \cdot \bar{\beta} \lambda^A \lambda^A N + \ldots \]
\[ \langle N(p,0) | j_{\mu}^A | N(p',0) \rangle = \frac{1}{2} \varepsilon_{\mu \nu \rho \sigma} \partial \bar{\lambda}^{(\nu)} \sigma^{(\rho)} + \frac{1}{2} \bar{\beta} \cdot \bar{\beta} \lambda^A \lambda^A N + \ldots \]

with $\bar{\lambda}(0) = 1$ and $\bar{\bar{\lambda}}(0) = 1$, $\lambda^A = \lambda^A \lambda^A N$ and

Then from the neutron decay rate, one finds $g = 1.269$

\[ \text{Anomalies} \]

We note that the classical QCD Lagrangian with just the three $u,d,s$ quarks had a $U(2) \times U(2)$ symmetry, or, keeping just the $u,d$ quarks, a $U(2) \times U(2)$ symmetry. Spontaneous symmetry breaking down to $SU(2) \times SU(2)$ (with $SU(2) = \mathbb{C} \times \mathbb{C}$, baryon number symmetry) gives a good description of linearly transforming baryon and meson multiplets other than the pions. The pions (and their kaon and $\eta$ partners when the $S$ quark is included) are interpreted as Goldstone bosons, or
pseudoscalar Goldstone bosons when quark masses are turned on. As mentioned earlier, however, the count of such apparent Goldstone modes is wrong for the $U(2) \times U(2)$ symmetry: in addition to the $\pi^a, a=1,2,3$ for the broken $\frac{1}{2} \sigma^a \bar{\tau}_5$ generators, there would be one more, for the $\bar{\tau}_5 \gamma_5$ generator, corresponding to $\chi$ chiral transformations $\bar{q}_i \to e^{\frac{i}{\hbar} \bar{\tau}_5 \gamma_5} \bar{q}_i \bar{q}_i \to (e^{\frac{i}{\hbar} \bar{\tau}_5 \gamma_5})^+ \bar{q}_i \to \bar{q}_i e^{\frac{i}{\hbar} \bar{\tau}_5 \gamma_5} \bar{q}_i$, for each $\mathbf{u}, \mathbf{d}$ quark type.

Although there is a candidate pseudoscalar meson, the $\eta'$, which would seem to fit into the 3-flavor $U(3) \times U(3)$ picture, it is not light ($M_{\eta'} = 958 \text{ MeV}/c^2$ as compared to $M_{\eta} = 135 \text{ MeV}/c^2$).

What is happening is the phenomenon of anomalies: although the axial $U(1)$ is a good symmetry at the classical level, it is broken at the quantum level. Derive the Noether current for $U(1)$ by the standard procedure of letting the parameters $\beta \to \beta(x)$ become spacetime dependent. Then

$$\delta \left( \pi \eta \right) \Rightarrow \delta \left( \pi \eta \right) = \left[ \frac{\partial}{\partial \beta} \delta \beta(x) \right] \pi \eta$$

so the $U(1)$ axial current is $j^a_5 = \bar{q}_i \gamma^a \gamma_5 q_i$, summed over all quark types $q_i$. If one includes $-m_i q_i$ mass term for the quarks, use of the Dirac equation $\not{\partial} \phi = 0$ and $\not{\partial} \phi \gamma_5 \phi - m_i \phi = 0$ gives the broken conservation equation

$$\partial_j j^a_5 = [\partial_j \partial^\mu] \left( \phi \gamma^a \gamma_5 \not{\partial} \phi \right) = 2 m_i \bar{q}_i \gamma_5 \gamma_5 q_i,$$

so for $m_i = 0$ the $j^a_5$ is classically conserved.

To see what happens to this axial current at the quantum level, restrict the discussion to just one species of fermion, $q_i \to q$, and couple it to an electromagnetic field $A_i$, and also to a fictitious $B_5$ gauge field that couples to the $j^a_5$ current:

$$-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} (A) - \frac{1}{4} (B_5 - i \mathbf{A} + m) q_i \not{\partial} q_i = \partial_j j^a_5 q_i.$$  

No kinetic term is included for $B_5$: its role is purely to facilitate discussion of quantum matrix elements involving insertions of $j^a_5$.  

There are a variety of ways to evaluate the effects of $j^\mu$ insertions. The most straightforward way proceeds with Feynman diagrams:

$$
\begin{array}{c}
\text{B} \to \text{A}
\end{array}
$$

We are particularly interested in the case where $B_\mu = 2j_\mu$, i.e. when one evaluates matrix elements $\langle A I 2 j^\mu | A \rangle$ taken between external electromagnetic fields $A_\mu$. As we have seen, $2j^\mu = 2m_j \delta_{\delta q}$ for massive fermions.

The triangle diagrams for the insertion of $j^\mu$ look superficially linearly divergent $\sim \frac{1}{\epsilon^2}$, but in fact converge since external momenta $p^\mu p^\nu$ come out, leaving an ultraviolet convergent result, at worst $\sim \frac{1}{\epsilon}$. One finds

$$
\langle A I 2 j^\mu | A \rangle = \frac{e^2}{32\pi^2} \epsilon_{\mu\nu\lambda}^{\gamma} F_{\mu\nu}(A) F_{\lambda\gamma}(A).
$$

This suggests that

$$
\langle A I 2 j^\mu | A \rangle = \frac{e^2}{16\pi^2} \epsilon_{\mu\nu\lambda}^{\gamma} F_{\mu\nu}(A) F_{\lambda\gamma}(A) = \frac{e^2}{8\pi^2} F_{\mu\nu}^2 F_{\lambda\gamma}(A),
$$

where $F_{\mu\nu}^2 F_{\lambda\gamma}$. Note that the mass $m_j$ of the fermion field going around the triangle loop has cancelled out. "anomaly!"

The above result was very puzzling when it was recognized in the post-war 1940s (e.g. Weinberg, 1949). For zero fermion mass, the axial current is classically conserved, but the non-vanishing and in independent result gives non-conservation at the quantum level. Doing the calculation directly for $m=0$ massless fermions produced a host of other puzzles, not least how to obtain an electromagnetic gauge invariant result. Schwinger solved that problem in 1951. The modern understanding of anomalies was achieved through the work of Adler, Bell and Jackiw in 1969. The most elegant derivation of..."
anomalies was given by Fujikawa in 1979-1981 in terms of non-invariance of the quantum field theory path integral measure under chiral R transformations.

The physical implications of axial anomalies for rigid symmetries are varied. Recalling that the pseudoscalar \( 7_0 \) proton also couples to fermions through \( 7_0 \bar{q} \gamma_5 q \), the triangle diagrams give the dominant contribution to \( 7_0 \to 2\gamma \) (2 gamma ray) decay. The non-conservation of the axial \( U(1) \) current in the low-energy effective field theory explains why the \( 7_0 \) is not protected from a large decay in the same way as the pions (and, to a lesser extent, the kaons and \( 7_0^0 \)). So rigid axial anomalies can produce physically important results.

Anomalies can also occur for chiral coupled fermions in gauge theories, as well, and there they are definitely not welcome. Anomalies in Yang-Mills gauge currents from renormalizability and can produce strange fractional degrees of freedom. So a key aim in constructing a gauge theory such as the Standard Model is to ensure anomaly freedom. For left-handed spinor fields \( 4^m \) transforming under some symmetry as \( SU(\lambda)^m \), where \( T_a \) are generator of the symmetry group, a theory will be anomaly free if the anomaly coefficients

\[
A_{abc} = \text{tr} (T_a T_b T_c)
\]

vanish for all \( a, b \) and \( c \), where the trace is taken over all types of fermions. So, for example, one needs to sum over every quark of every flavour and every lepton in each generation. Fields such as the gluon in the Standard Model which are defined to be right-handed are included in the calculation of \( A_{abc} \) by listing their left-handed charge conjugates (like the left-handed \( \bar{c} \) quark). When the anomaly coefficients don't vanish, the anomaly is

\[
2J^{a} = \frac{A_{abc}}{64\pi^2} g^2 \epsilon^{\mu\nu\alpha\beta} F_{\mu\nu} F^{\alpha\beta}
\]
A consequence of the anomaly coefficient structure is that anomalies cancel for real or pseudoreal fermion representations $T_a$. A pseudoreal representation is one for which the generators $(i T_a)$ are real up to a similarity transformation, i.e. $T_a^\ast = -S T_a S^{-1}$ for some invertible matrix $S$. To see how this works, note that for Hermitian generators $T_a^T = T_a$, so

$$A_{abc} = \text{tr} \left( T_a \left[ T_b, T_c \right] \right) = \text{tr} \left[ \left( T_a T_b T_c \right)^T \right]$$
$$= \text{tr} \left( \left\{ T_c^T, T_b^T \right\} T_a^T \right) = \text{tr} \left( \left\{ T_a^\ast, T_b^\ast \right\} T_a \right)$$
$$= -\text{tr} \left( S^{-1} \left\{ T_c, T_b \right\} S^{-1} T_a S^{-1} \right) = -\text{tr} \left( \left\{ T_c, T_b \right\} T_a \right) = -A_{abc}$$

Since the $A_{abc}$ are totally symmetric.

As an application of this result, consider a case where fermion number is a good conserved quantity and where left- and right-handed fermions transform in the same $T_a$ of the group in question. (For this purpose, the right-handed fermions are considered in their right-handed forms, not charge-conjugated.) Then, assembling the total representation $T_a$ of left-handed fermions (right-handed now charge conjugated), one may write $T_a$ in block-diagonal form:

$$T_a = \begin{pmatrix} T_a & 0 \\ 0 & -T_a^\ast \end{pmatrix}$$

Since charge conjugation sends $i\epsilon^{abc} \to -i\epsilon^{abc}$, this total representation $T_a$ is consequently pseudoreal, since $T_a^\ast = -S T_a S^{-1}$ with $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Consequently, any symmetry which is left-right symmetric in this way is anomaly-free, for example in theories with just Dirac fermions that are given an overall symmetry representation $T_a$.

An important application of the existence of anomalies in the case of a rigid chiral symmetry is to $1^0$ decay. Recall that $U = \exp \left( \frac{i}{\alpha} \left( \gamma^\mu \gamma_5 \right) \right)$ and $U \to g_{\mu} U \gamma_5$, $g_{\mu} < 0$, with the axial (spontaneously broken) symmetries having $g_{\mu} = g_\pi^+$. 
Let \( q_i = \text{Exp}(i\frac{\gamma_5}{2} (0 \cdot 1)) = g_i^{+} \) so under this transformation

\[
\exp\left(\frac{i}{F_{\pi}} \pi^0 (1 \cdot 0)\right) \rightarrow \exp\left(i\frac{\gamma_5}{2}(1 \cdot 0)\right) \exp\left(-\frac{i}{F_{\pi}} \pi^0 (1 \cdot 0)\right) \exp\left(i\frac{\gamma_5}{2}(1 \cdot 0)\right)
\]

ie. \( \pi^0 \rightarrow \pi^0 + i\omega F_{\pi} \).

Recalling the electromagnetic couplings \( D_{\mu} = (\gamma_{\mu} - ig A_{\mu}) \) for charged fields of charge \( q_i \) in the chiral Lagrangian

\( \mathcal{L}_{\text{Chiral}} = \frac{-F_{\pi}}{4} \text{tr} \left[(D_{\mu}W)(D^{\mu}W)^\dagger\right] \), one notes that the chiral \(SU_{(2)} \times SU_{(2)}\) symmetry is actually broken by such couplings (it is also broken by small corrections \( m_u/m_c \approx \mu_{\text{initial}} \)) as one sees from the Gell-Mann - Nambu - Oakes relations. The only part of the \(SU_{(2)} \times SU_{(2)}\) symmetry that survives as a symmetry of the combined strong plus electromagnetic theory is the part that commutes with the \(u\) quark electric-charge matrix \( Q_{\text{EM}} = \begin{pmatrix} 2/3 & 0 \\ 0 & -1/3 \end{pmatrix} \). The surviving axial generator is

\( T_{3A} = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix} \).

Experimentally, the decay of the neutral pion \( \Pi^0 \) into photons is well described by an interaction term

\( f_{\text{EM}A} = \frac{e^2}{32\pi^2 F_{\pi}} \Pi^0 \gamma_{\mu} \epsilon F_{\mu \nu}(A) F^{\nu \rho}(A) \).

The problem with such an interaction term, however, is that it does not respect the \( U_3(1) \) transformation generated by the supposedly surviving \( T_{3A} \), under which \( \Pi^0 \rightarrow \Pi^0 + \omega F_{\pi} \).

Other interactions like \( f' = \frac{e^2}{32\pi^2 F_{\pi}} (\frac{\Lambda\Pi^0}{\Lambda^2}) \epsilon_{\mu \nu \rho \sigma} F_{\mu \nu}(A) F^{\rho \sigma}(A) \) would address that problem, but for \( \Lambda \sim 1/\alpha_{\text{EM}} \), they give decay rates that are too small for any value of \( \Lambda \) large enough to justify use of the effective pion theory.

What is happening, however, is an anomaly in \( U_3(1) \), specifically the anomaly arising from \( A_{\text{hands}} = A \). From \( A_{\text{abc}} = \text{tr} (T_a, T_b, T_c) \), one has \( A = 2N_c (\frac{13}{3})^2 - (\frac{17}{3})^2 \times 2 \),

where the final factor of 2 comes from inclusion of the charge-conjugated right-handed \( u_{\text{R}} \) and \( d_{\text{R}} \) quarks.
The Nc factor comes from the number of colours from the SU(3) strong interactions. In fact, this is a key point. The anomaly coefficient is \( A = 2Nc/3 \), so the violation of the anomalies \( U_{\mu\nu} \) with the Standard Model is \( J_{\text{SM}} = \frac{e^2}{2}\pi F_{\mu\nu} F^\mu_{\nu} \).

For \( Nc = 3 \), i.e., 3 colours, this gives \( J_{\text{SM}} = \frac{e^2}{3}\pi F_{\mu\nu} F^\mu_{\nu} \), which precisely reproduces the \( U_{\mu\nu} \) violation coming from the photon interaction \( J_{\text{SM}} = \frac{e^2}{32}\pi F_{\mu\nu} F^\mu_{\nu} \).

Now consider the dangerous potential anomalies in gauge currents. For consistency of the Standard Model, it is essential that such anomalies cancel. One needs to check the anomaly coefficients \( A(3,3,1) \), \( A(3,3,2) \), \( A(3,3,1) \), \( A(3,1,1) \), \( A(3,1,1) \) where \( A(3,3,3) \) denotes the anomalies involving three SU(3) generators, etc. \( A(3,3,3) \) is straightforward, since the SU(3) representations are all left-right symmetric, i.e., the left-handed \( \lambda_{\alpha} \) and right-handed \( \lambda_{2\alpha} \) and \( \lambda_{3\alpha} \) all carry fundamental 3 representations of SU(3). \( A(3,3,2) \) vanishes because they are all proportional to the trace of the SU(2) generators \( T_{\alpha} \), which are all traceless. \( A(3,3,1) \) are a bit more involved.

The SU(3) generators are \( T_{\alpha} = \frac{1}{2} \lambda_{\alpha} \), where the \( \lambda_{\alpha} \) are the Gell-Mann matrices, which satisfy \( \lambda_{\alpha} \lambda_{\beta} \lambda_{\gamma} = \frac{4}{3} \delta_{\alpha\beta\gamma} + 2 \delta_{\alpha\beta} \lambda_{\gamma} \).

The SU(3) generators \( \lambda_{\alpha} \) are traceless, so the dyson \( \lambda_{\alpha} \) part gives zero, while the trace over colours of the other part gives \( 4/3 \).

Consequently, the \( A(3,3,1) \) coefficient is proportional to the trace over all left-handed coloured fields of the \( U_{\mu\nu} \) by exchange sum \[ \sum \text{quarks} y = 3(2y_{u} + y_{d} + y_{c}) = 3[2(\frac{1}{6}) + (-\frac{1}{3}) + (\frac{1}{3})] = 0 \]
where the overall factor of 3 is the number of generations and the 2 on \( y_{u} \) comes from the two SU(2) flavours.

\( A(3,1,1) \) with \( (X,Y = 2n+1) \) vanish because \( \lambda_{\alpha} = 0 \) for SU(3).
For $A(2,2,2)$ recall that the only non-trivial SU(3) maps in the Standard Model are doublets, generated by $\tau_u - t_x^2$. The Pauli matrices all satisfy $\tau_a^* = -i \sigma_a \tau_a$, i.e. this representation is pseudoreal, so $A(2,2,2) = 0$. For $A(2,2,1)$ note that the Pauli matrices satisfy $\tau_3 \sigma_3 = \frac{1}{2} \tau_3$, and the factor of $\frac{1}{2}$ is cancelled by summing over doublets. So $A(2,2,1) = \sum_{\text{doublets}} \tau_3 = 3(3 \tau_3 + \tau_3 \tau_3) = 3 (3 \frac{1}{2} + \frac{1}{2}) = 0$

where the factor of 3 on $\tau_3$ comes from colours. $A(2,1,1)$ involves the trace of a single Pauli matrix, which vanishes.

Finally, $A(1,1,1)$ is proportional to the sum over all left-handed fermions of the cube of the hypercharge:

$$A(1,1,1) = 2 \sum_{\text{all fermions}} \tau_3 = 2 \left( 2 \tau_3^2 + \tau_3 \right) = 6 (2 \frac{1}{2} + \frac{1}{2}) = 6 \left( \frac{1}{2} + 1 + \frac{1}{2} \right) = 6 \left( \frac{1}{2} + 1 + \frac{1}{2} \right) = 0.$$

This rather non-trivial anomaly cancellation hinges upon a number of physical issues. The equality of proton and electron charge absolute values depends on having $\gamma_{ee} = 1$, while the neutrality of the neutron requires $\gamma_{ee} = 1/3$. Moreover, if $\gamma_{ee}$ were different from 1, this would give rise to a shift in the identification of the electric charge generator, and this shift would then disturb the electric charge cancellation for the neutrinos. So anomaly cancellation is quite essential for the consistency and physical correctness of the Standard Model.

The QCD $\Theta$ parameter

In discussing the adjustments to the generic form of the Standard Model action needed to make the quark mass matrix real, positive and diagonal, we have used chiral transformations that are exactly of the form of the SU(N) x SU(N) effective low-energy theory's rigid approximate symmetries. So we need to be concerned about the effects of anomalies on these mass-matrix-aligning transformations.
The impact of such fermionic field redefinitions is most clearly seen from Fujikawa's approach to anomalies. For a rigid chiral transformation of a fermionic field $\psi \rightarrow e^{i\theta \gamma^5} \psi$, the classical action may be invariant but the path integral measure is not:

$$\mathcal{D}\Psi \mathcal{D}\bar{\Psi} \rightarrow \mathcal{D}\Psi \mathcal{D}\bar{\Psi} \exp(-i\frac{g^2}{2\pi} \varepsilon_{\mu\nu} a^\mu \bar{\varepsilon} \gamma^5 \bar{\Psi} \gamma^a \gamma^5 \gamma^a \psi)$$

where the $F^{a\mu\nu}_\tau$ are Yang-Mills field strengths for any gauge symmetries under which the fermionic field $\psi$ is changed and $g$ is the corresponding change. Note that the coefficient of $w$ agrees with the formula $A_{\alpha\beta\gamma} = \text{tr}((T_\alpha T_\beta T_\gamma))$, e.g. for colour $SU_c(3)$. This is a mixed anomaly involving a rigid axial $\gamma^5$ with generator $T_\alpha \gamma_5$ and two $SU_c(3)$ generators $T_a = \frac{1}{2} \lambda_a$, $T_c = \frac{1}{2} \lambda_c$. Using again $\varepsilon_{\mu\nu} a^\mu \bar{\varepsilon} \gamma^5 \bar{\Psi} \gamma^a \gamma^5 \gamma^a \psi = \frac{2g}{2\pi} \varepsilon_{\mu\nu} (\chi a^\mu \bar{\varepsilon} \gamma^5 \bar{\Psi} \gamma^a \gamma^5 \gamma^a \psi)$ and noting that $\text{tr}(a^\mu a^\nu a^\gamma) = \frac{3}{2} \delta^{\mu\nu} \delta_{\alpha\beta} \delta_{\gamma\delta}$, one has

$$A_{\alpha\beta\gamma} = \text{tr}((T_a T_b T_c)) \frac{g}{2\pi} \varepsilon_{\mu\nu} (\chi a^\mu \bar{\varepsilon} \gamma^5 \bar{\Psi} \gamma^a \gamma^5 \gamma^a \psi)$$

Then for $\text{tr}(T_a)$ note that $T_a \rightarrow e^{i\theta T_a}, \bar{T}_a \rightarrow e^{-i\theta T_a}$ so

$$A_{\alpha\beta\gamma} \rightarrow e^{i\theta \gamma^5} A_{\alpha\beta\gamma} e^{-i\theta \gamma^5}$$

Recall that for a spacetime dependent parameter $W(x)$ in the transformation $\Psi \rightarrow e^{i\theta \gamma^5 \gamma^a W}$, one has $\delta\mathcal{L} = \frac{i}{2} \varepsilon_{\mu\nu} \delta W(x)$, which one can rewrite as

$$-\frac{i}{2} \varepsilon_{\mu\nu} \delta W(x)$$

So, subsequently taking $w = \frac{g}{2}$ spacetime independent,

one finds agreement with the anomalous variation of the path integral integral $e^{i\theta \gamma^5 \gamma^a W(x)} e^{i\theta \gamma^5 \gamma^a \psi}$. For chiral transformations rotating between multiple generations,

$$\psi_i \rightarrow e^{i\theta \gamma^5} \psi_i, \bar{\psi}_i \rightarrow e^{i\theta \gamma^5} \bar{\psi}_i$$

the angle $\theta$ is given by

$$\delta (\mathcal{L} + \mathcal{L}) = i e^{i\theta} \delta \mathcal{L}$$

where $\delta \mathcal{L} = \frac{g}{2\pi} \varepsilon_{\mu\nu} (\chi a^\mu \bar{\varepsilon} \gamma^5 \bar{\Psi} \gamma^a \gamma^5 \gamma^a \psi)$, corresponding to the absence of an anomaly.

The inclusion of a term $\mathcal{L} = -\frac{1}{8} g^2 \varepsilon_{\mu\nu\rho\sigma} a^\mu F_{\rho\sigma}$ in the Lagrangian at the classical level would have no effect on physical phenomena, because $\frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} a^\mu F_{\rho\sigma} = \frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} K^a_{\mu\nu} a^a$ is a total derivative, where $K^a_{\mu\nu} = 2 \varepsilon_{\mu\nu\rho\sigma} (a^a_{\rho\sigma} A^a_{\rho\sigma} + \frac{1}{2} f^{abc} A^a_{\rho\sigma} A^b_{\rho\sigma} A^c_{\rho\sigma})$ is the Chern-Simons current. This total derivative structure clearly is what is required by the $-\frac{1}{8} g^2 \varepsilon_{\mu\nu\rho\sigma} a^\mu F_{\rho\sigma}$ change in the Lagrangian.
\[ \Theta \text{ angle terms: } \alpha = -\frac{\Theta_3}{32\pi^2} \text{ for } \alpha_{\mu\nu} \text{ are invariant under charge conjugation } C \text{ because they are quadratic in } F^\mu_{\nu}, \text{ but they violate } \mathbb{P} \text{ and } \mathbb{CP}, \text{ and consequently violate } T. \text{ The key test for non-invariance under } \mathbb{P} \text{ is the presence of a single } \mathbb{G}^\mu_{\nu} \text{ tensor, which has the effect that in each term in the sum } \alpha_{\mu\nu} F^\mu_{\nu}\text{ there is one sign flip, either because } \phi \rightarrow -\phi \text{ or because } A^\mu_{\nu} \text{ flips sign (recall that under } C, \text{ gauge fields have sign flips or not according to the symmetry/antisymmetry of } T^\alpha, \text{ while under } \mathbb{P}, \text{ } A^\alpha_{\mu\nu} \rightarrow -A^\alpha_{\mu\nu}(x^9-x^1)) \]

Once one has admitted the possibility of CP non-invariant terms in the SM Lagrangian, there are three possible \( \Theta \) angle terms:

\[ \theta_{\text{tree}} = -\frac{\Theta_2}{32\pi^2} \sigma^{\mu\nu} G^{\mu\nu} - \frac{\Theta_7}{32\pi^2} W^{\mu}_{\nu} W^{\nu}_{\mu} - \frac{\Theta_8}{32\pi^2} B_{\mu} B^{\mu} \]

However, not all of these have observable consequences; at least in the original Standard Model without neutrino masses. Precisely because there are anomalous variations of the fermion path integral measure under rigid chiral transformations, there is ambiguity in where the effects of these-\( \Theta \) angle terms will be encountered. Revisit the transformations made in order to diagonalize the \( W \) and Q quark mass matrices. We wrote the original Yukawa coupling matrices

\[ Y_{\text{W}} = U_{\text{W}} a_{\text{W}}^{\text{m}}, \quad Y_{\text{Q}} = U_{\text{Q}} a_{\text{Q}}^{\text{m}}, \quad U_{\text{W}} = U_{\text{W}}^\dagger U_{\text{W}}^{\text{m}}, \quad U_{\text{Q}} = U_{\text{Q}}^\dagger U_{\text{Q}}^{\text{m}} \]

So the quark mass matrices can be diagonalized by first making \( U_{\text{W}}^\dagger \) and \( U_{\text{Q}}^\dagger \) chiral rotations on the RH quarks only, which does change \( \Theta_3 \) and \( \Theta_7 \), since the RH quarks couple to \( Q_{3L} \) and \( L_{3L} \), followed by non-chiral rotations (\( U_{3L} = U_{3L}^{\text{m}}, \quad U_{3L}^{\dagger} a_{3L}^{\text{m}}, \quad U_{3L}^{\dagger} \)), then \( \Theta_8 = U_{\text{Q}}^{\dagger} a_{\text{Q}}^{\text{m}}, \quad U_{\text{Q}}^{\dagger} = U_{\text{Q}}^{\dagger} a_{\text{Q}}^{\text{m}}, \quad U_{\text{Q}}^{\dagger} = U_{\text{Q}}^{\dagger} a_{\text{Q}}^{\text{m}} \)), the non-chiral rotations changing none of the \( \Theta \) angles.
The phase induced by the $U^{\text{unr}}$ and $U^{\text{det}}$ rotations that changes $\Theta_3$ is 
\[ \text{arg det}(U^{\text{unr}} U^{\text{det}}) = -\text{arg det}(D^{\text{unr}} D^{\text{det}})/(D^{\text{unr}} D^{\text{det}})) \]
\[ = \text{arg det}(y^{y_w} y^{c_d}) \] since $D^{\text{unr}}$ and $D^{\text{det}}$ are diagonal, real matrices. Suppose, on the other hand, one wants instead to eliminate $\Theta_3$. That would be achieved by an inverse set of transformations $U^{\text{unr}}, U^{\text{det}}$, but those would then restore a phase - arg det($U^{\text{unr}} U^{\text{det}}$) = arg det($U^{\text{unr}} U^{\text{det}}$) into arg det($y^{y_w} y^{c_d}$) for the Yukawa terms, thus making the quark mass matrix complex. Moving a phase back and forth in this way is just a matter of basis choice, and cannot affect physical quantities, which are invariant under such change of variables. The basis-independent quantity is $\bar{\Theta}_3 = \Theta_3 + \text{arg det}(y^{y_w} y^{c_d})$.

For the $SU(2)$ and $SU(1)$ angles, the situation is different. Since the RH quarks do not couple to $SU(2)$, $\Theta_2$ is unaffected by the $U^{\text{unr}}, U^{\text{det}}$ transformations, but it can be removed by transformations of the LH quarks. Consequently, one may make LH quark transformations to remove $\Theta_2$, thus moving its effect into the Yukawa couplings, but then make compensating transformations on the RH quarks to remove those effects without disturbing the new zero $\Theta_2$. Similarly, for $\Theta_1$, one can make LH neutrino transformations to set it to zero and since there are no Yukawa couplings to the LH neutrinos after symmetry breaking, one doesn’t have to worry about $\Theta_1$ moving into the Yukawa terms. So only the strong interaction $\bar{\Theta}_3$ is of physical relevance in the minimal Standard Model. (Metals can be different after inclusion of neutrino masses, but the resulting physical effects would be very small.)

Now consider the total derivative term such as
\[ -\frac{g_3^2}{2\pi F} \bar{G}_{\mu\nu} \Gamma \Gamma^\nu \] can have physical effects. Quantum mechanical tunneling processes can be calculated using solutions.
to Euclideanized field theory, obtained after a Wick rotation $t = -i\rho$ of the time coordinate. After this Wick rotation, the Euclidean path-integral phase becomes $e^{-S_E}$, where $S_E$ is the Euclidean action. For Yang-Mills theories, $S_E = \frac{i}{4} \int d^4x \, F_{ij}^a F_{ij}^a$, $s_j = 1 \ldots d$. Solutions to the Euclidean field equations derived from $S_E$ extremality can contribute to tunneling amplitudes if $S_E$ is finite when evaluated at such solutions. Finite Euclidean action $S_E$ requires $F_{ij}^a = 0$ as one goes to infinity, but $F_{ij}^a$ can be non-vanishing provided the Euclidean space is $d = 4$ dimensional.

The restriction to $d = 4$ dimensions comes about from a scaling argument. In order for $S_E[A]$ to be extremal, the field equations $\nabla_i F_{ij}^a = 0$ must be satisfied. Moreover, these are preserved by transforming $A_i^a(x) \rightarrow A_i^a(x) + \xi^a \nabla_i \phi^a$ (under which $F_{ij}^a(A) \rightarrow R^{-2} F_{ij}^a(A(\xi)) \chi^a = \phi^a$, and $\nabla_i = \phi^a R^{-2} \nabla_i$) in other words the Yang-Mills equations are scale invariant. But $S_E[A^a] = R^{-d+4} S_E[A]$ so for $d = 4$, there can be no non-trivial extrema of $S_E[A]$ unless $S_E[A] = 0$. But if $S_E[A] = 0$, then $F_{ij}^a = 0$ everywhere. By a gauge transformation one can then make $A_i^a = 0$ everywhere as well.

For $d = 4$, however, one can find "instanton" solutions with $F_{ij}^a$ non-vanishing except at infinity. In order to have finite $S_E[A]$, $A_i^a(x)$ can vanish as slowly as $\frac{1}{|x|}$ provided the field approaches a pure gauge configuration $A_i^a \rightarrow g^a(x) \partial_i \phi^a$ where $g^a(x)$ is a direction-dependent element of the gauge group $G$. Moreover, $A_i^a(x)$ is unaffected if one replaces $g^a(x)$ with $g_0 g(x)$ for any fixed $g_0 \in G$.

Then by choosing $g_0 = g^a(\hat{x})$ for some specific direction $\hat{x}$, one arranges that $g^a(\hat{x}) = I$ (the identity element of $G$). Each gauge field with finite $S_E[A]$ thus defines a mapping from the Euclidean unit sphere $|\hat{x}| = 1$ to the group manifold of $G$, with $\hat{x}$ mapped to the identity element of $G$. 
The set of classes of such topologically distinct mappings $S_{d-1} \rightarrow G$ with one point of $S_{d-1}$ mapped onto a fixed element of $G$ is $\pi_{d-1}(G)$, the $(d-1)$th homotopy group of the group manifold. For $d=4$, one is therefore interested in $\pi_3(G)$; this homotopy group is non-trivial for any semisimple Lie group $G$.

Using the fact that $0 \leq \int (F_{ij}^a - \frac{1}{2} \epsilon_{ijk} F_{kl}^a) (F_{ij}^a + \frac{1}{2} \epsilon_{ijk} F_{kl}^a) d^4x$ where $\epsilon_{ijk}$ is the Euclidean tensor, $\epsilon_{1234} = 1$, and noting that for $F_{ij}^a = \frac{1}{2} \epsilon_{ijk} F_{kl}^a$, one has $\frac{1}{2} F_{ij}^a = F_{ij}^a$, it follows that

$$0 \leq 2 \int F_{ij}^a \cdot F_{ij}^a - 2 \int F_{ij}^a \tilde{F}_{ij}^a \quad \text{so} \quad \text{Se} \geq \frac{1}{2} \int F_{ij}^a \tilde{F}_{ij}^a.$$ 

The lower bound is thus reached if and only if $F_{ij}^a$ is either self-dual or anti-self-dual and $\tilde{F}_{ij}^a = \pm \frac{1}{2} \epsilon_{ijk} F_{kl}^a$. Note that such self-dual or anti-self-dual field strengths automatically satisfy the $D_i F_{ij}^a = 0$ field equation because $D_i F_{ij}^a = 0$.

The integrated quantity $\int d^4x F_{ij}^a \tilde{F}_{ij}^a$ is a topological invariant known as the $\frac{e^2}{32\pi^2}$ winding number of the Euclidean field configuration, the integral of the Chern-Pontryagin density. The winding number so expressed is an integer $\in \mathbb{Z}$.

Often, one sees it given in terms of differently scaled field strengths $F_{ij}^a = gF_{ij}^a = \tilde{g}^a \tilde{A}_{ij}^a - \frac{1}{2} \epsilon_{ijk} \tilde{A}_{ij}^a \tilde{A}_{kl}^a$ with $\tilde{A}_{ij}^a = gA_{ij}^a$, in terms of which $\text{Se} = \frac{1}{32\pi^2} \int d^4x F_{ij}^a \tilde{F}_{ij}^a$.

Solutions with $F_{ij}^a = \tilde{F}_{ij}^a$, which extremize the Bogomolnyi inequality $\text{Se} \geq \frac{1}{2} |F_{ij}^a \tilde{F}_{ij}^a|$ were found by Belavin, Polyakov, Schwartz and Tyupkin (Phys. Lett. 59B (1975), 259):

$$A_i(x) = \left( \frac{\bar{r}^2}{r^2 + \bar{r}^2} \right) g_1(x) \theta, \quad g_1(x) = \left( \frac{x^4 + \bar{x}_i x_i^2 - r^2}{r^2} \right)^{\frac{1}{2}},$$

where $\bar{r}$ is an arbitrary scale factor and $g_1(x)$ is an element of an $SU(2)$ subgroup of the gauge group $G$ with $g_1(x) = \left( \begin{array}{c} x^4 + \bar{x}_i x_i^2 - r^2 \\ r^2 \end{array} \right)$, $\bar{x}_i = \frac{1}{2} x_i^2$, $SU(2)$ generators.

This is the original instanton solution, and has winding number $V = 1$. It is centered at $x^i = 0$ in the Euclidean $\mathbb{R}^4$, hence the "instanton" name. The Euclidean action for this
Solution is $S_{E}A_{D} = \frac{8\pi^2}{g^2}$, corresponding to $\int_{V} F_{\mu\nu} F^{\mu\nu} = 32\pi^2 v$

with $v = 1$.

Having found an instanton solution with winding number $v = 1$, it is then clear that there exist solutions with any $v \in \mathbb{Z}$.

For instance, solutions with $v = N$ can be approximately constructed by superposing $N$ solutions with $v = 1$ taken with centers far apart, so the nonlinearities of the YM field equations can be ignored. A solution with $v = -1$ can be obtained by replacing $g_{i}(x)$ with $g_{i}^{-1}(x)$; this then leads to $v = -N$

Solutions by superposition.

Asymptotically as $|x| \to \infty$, the instanton solutions approach pure gauge configurations (e.g. $iA_{i} \to g_{i}^{-1}(x) A_{i} g_{i}(x)$

for the $v = 1$ BPST solution), and consequently $F_{\mu\nu} \to 0$ as $x \to \infty$.

Looking back to the Minkowski space theory, we can think of an instanton as giving a transition between an $F_{\mu\nu} = 0$ solution as $t \to -\infty$ and a different $F_{\mu\nu} = 0$ solution as $t \to +\infty$; in other words a "vacuum to vacuum" tunneling. In the quantum theory, one needs to sum over all histories in the path integral and this will thus necessarily include the effects of all $v \in \mathbb{Z}$ instantons.

The weighting of these various contributions can be shown on general field theoretical grounds (especially cluster decomposition) to be of the form $e^{i\int F_{\mu\nu} F^{\mu\nu}/8\pi}$, which is thus equivalent to the inclusion of $\int A_{\mu} F^{\mu\nu} / 16\pi$ since upon returning to Minkowski space one finds $\int (g_{i}(x) A_{i} g_{i}^{-1}(x))_{\mu\nu} = -i F_{\mu\nu}$ (k=1,2,3) and $\varepsilon_{\lambda\sigma} = -1$ so $v = -\frac{g^2}{32\pi^2} \int d^{4}x F_{\mu\nu} F^{\mu\nu}$ in the Minkowski space theory.

For a winding number $v$ instanton, the Euclidean action is $S_{E}A_{D} = \frac{8\pi^2}{g^2} v$, so the path integral is weighted by $\exp[\frac{1}{4\pi} S_{E}] = \exp\left[ -\frac{g^2}{32\pi^2} v \right]$. As $g \to 0$, this exponential
vanishes strongly enough to kill any powers of $g$ that occur in perturbative corrections: $g^{-n} \exp \left[ -\frac{8\pi \hbar n}{g^2} \right]$ and all its derivatives with respect to $g$ vanish at $g=0$. Such contributions in instanton backgrounds are accordingly highly non-perturbative; they are never encountered in any finite order in perturbation theory.

Despite the non-perturbative nature of the instanton contributions, one can obtain information about their effects using the chiral non-linear sigma model effective theory abstracted from QCD. We had $\mathcal{U}(g^2) = \exp \left[ (\pi^a \pi^a) / 8 \right]$ where $\pi^a = \frac{i}{2} \sigma^a$, and $\pi^n$ are the pion fields. Including the mass term, the chiral effective theory Lagrangian we took to be

$$L_{\text{mass}} = -\frac{F_{\pi}^2}{4} \text{tr} \left[ (\pi^a \pi^a)^2 / 8 \right] + \frac{V_3}{2} \text{tr} [M^2 \mathcal{U} + \mathcal{U}^2 M]$$

where $V_3 = \langle \mathcal{U} \rangle = \langle \pi^3 \rangle$ and $M$ is the QCD quark mass matrix. We saw earlier that expanding the mass term and comparing the vacuum energy arising from it to that arising from the QCD level quark mass terms in the presence of the $\langle \pi^3 \rangle$ condensate. This led to the Gel-Mann-Oakes-Renner relation $F_{\pi}^2 m_{\pi}^2 = V_3 (\text{tr} M^2)$. Now consider the impact of the invariant $\bar{\Theta}$ parameter. One may make chiral transformations on the up and down quarks to remove the $\bar{\Theta}$ coefficient from the $F_{\pi}^2$ term in the QCD Lagrangian, at the price of having its effects appear in the Yukawa couplings, i.e. in condensate $\langle \bar{q} \gamma_5 q \rangle$. After electroweak symmetry breaking, this leads to complex quark masses $\mathcal{M}$, one finds

$$\mathcal{M} = \left( \begin{array}{cc} m_u & \bar{e} \bar{\Theta} \\ 0 & m_d \end{array} \right)$$

for the up and down quark mass matrix. This has the effect of making the vacuum energy depend on $\Theta$:

$$E(\bar{\Theta}) = V_3 (\bar{e} \bar{\Theta} m_u - m_d) \cos(\frac{1}{2} \bar{\Theta}) = F_{\pi}^2 m_{\pi}^2 \cos(\frac{1}{2} \bar{\Theta})$$

using the Gel-Mann-Oakes-Renner-Oakes relation. This indicates that the parameter $\bar{\Theta}$ controls the energies of distinct vacuum states, called $\bar{\Theta}$-vacua.

Another consequence of the QCD $\bar{\Theta}$ parameter is that the neutron acquires an electric dipole moment proportional to $\bar{\Theta}$.

Then, from the experimental bound on the neutron EDM, one obtains a bound $\bar{\Theta} < 5 \times 10^{-10}$. Why so small?
The Strong CP problem concerns the extreme smallness of $\theta_3$ despite the relatively large amount of CP violation arising in the weak interactions of the Standard Model.

One conceivable solution would be that one of the quarks is massless. Then one could eliminate $\theta_3$ in a fashion similar to the elimination of $\theta_1$ by a chiral rotation without effect on the quark masses. The chief candidate for masslessness might be the $u$ quark. There are several problems with this suggestion, however. One is that there is no symmetry protecting $m_u = 0$, so $m_u$ would just have to be tuned to be small instead of having $\theta_3$ to be small. So proposing that $m_u = 0$ just moves the fine-tuning problem around. Another problem is that it conflicts with lattice-gauge theory calculations, which give a non-zero value for $m_u$ with a large statistical significance.

The most favorably discussed idea for solving the Strong CP problem is the introduction of a hypothetical particle field called the axion $a$. This field is assumed to be a Goldstone boson for an approximate axial symmetry involving at least one of the quarks. If this symmetry were exact, the axion would have a shift symmetry $a(x) \rightarrow a(x) + c$, for constant parameter $c$.

Such a symmetry is called a Peccei-Quinn symmetry after its original authors. The idea has subsequently been developed by Weinberg and Wilczek. Suppose that the axion has an effective Lagrangian

$$L_{\text{axion}} = -\frac{1}{2} \partial_a \phi^3 \frac{a}{\text{GeV}} + \left( \frac{\alpha}{g_{\phi}^2 g_{a}^2} \right) \epsilon_{\mu
u} \epsilon^{\mu
u} \phi$$

where $\phi = a$. The constant $\alpha$ is an axion decay constant, and $\epsilon_{\mu
u}$ is the antisymmetric tensor of the quarks. This produces a
phase $\exp \left( \frac{i}{2} \left( \tilde{\Theta}_3 - \tilde{a}_3 \frac{\alpha}{g_s^2 f_a} \right) \right)$. In the mass matrix $M = \left( \begin{array}{cc} m_u & 0 \\ 0 & m_d \end{array} \right)$.

Consequently, the energy of the vacuum is

$$E(\tilde{\Theta}, \tilde{a}) = V^2 \left( m_u m_d \right) \cos \left( \frac{1}{2} \left( \tilde{\Theta}_3 - \tilde{a}_3 \frac{\alpha}{g_s^2 f_a} \right) \right)$$

$$= \frac{M^2}{2} m_u^2 \cos \left( \frac{1}{2} \left( \tilde{\Theta}_3 - \tilde{a}_3 \frac{\alpha}{g_s^2 f_a} \right) \right)$$

using Gell-Mann-Ramge-Oakes relations.

So the minimum of vacuum energy occurs for $\tilde{a}_3 = g_s^2 f_a \tilde{\Theta}_3$, at which value the phase of the quark mass matrix vanishes, and the strong CP violating effects are cancelled out.

Proposing a new particle species in this way to cause the CP violating $\tilde{\Theta}_3$ angle to relax and cancel out the strong-interaction CP violation is not cost-free, however. As with every extension of a successful theory, one has to ensure that new physical effects arising from the extension are not in conflict with observation. The mass $m_a$ of the axion may be obtained from $\frac{2}{\tilde{a}_3} \frac{\partial}{\partial \tilde{a}_3} E(\tilde{\Theta}, \tilde{a}) = \frac{g_s^2 f_a}{4 \pi^2} \tilde{a}_3^2 = m_a^2$.

So $m_a \approx \frac{g_s^2 f_a}{2 \pi} \tilde{a}_3$, inversely proportional to $f_a$.

If $f_a \approx 1$ TeV, this yields a mass $\approx 10$ keV, while if $f_a \approx 10^6$ GeV, it yields a mass $\approx 10^{-4}$ eV.

Astrophysical bounds, particularly from red giants, require $f_a > 10^6$ GeV, while cosmological bounds require $f_a < 10^{12}$ GeV (too many axions would overclose the universe). Consequently, the axion should be very weakly coupled with a mass in the range $10^{-4} \text{ eV} < m_a < 10^{-2} \text{ eV}$.

The possibility that the axion couples not only to the strong interaction $g_s$ but also to $F_{\mu \nu}$ in QED means that in a strong electromagnetic field an axion could convert into a photon. This is the basis for a search by the ADMX experiment at the University of Washington, predicated on the supposition that our galaxy has a dark matter halo composed largely of axions.
Grand Unification

The Standard Model very successfully accounts for the known electroweak phenomena, and with a simple extension to include right-handed gauge singlet neutrinos it can also accommodate neutrino masses as well. However, being based on the non-simple gauge group $\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)$, it has three independent coupling constants $g_3, g_2$ and $g_1$. Had the gauge group been a simple group, there would be a single coupling constant, clearly a more complete unification of the fundamental forces.

It is remarkable how close this comes to working, in the Georgi-Glashow model, which is based on the gauge group $\text{SU}(5)$. This group can be made to break spontaneously down to $\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)$ by use of a Higgs field $\Phi$ transforming in the adjoint $(24)$ representation of $\text{SU}(5)$. One can construct a potential for a renormalizable theory (i.e. maximum dimension 4) from the three available $\text{SU}(5)$ invariants $V = -m_i^2 \text{tr} (\Phi^i) + a \text{tr} (\Phi^4) + b \text{tr} (\Phi^2)^2$. Note that the vacuum value of $\Phi$ can always be made diagonal $\langle \Phi \rangle = \text{diag}(\lambda, \lambda, \lambda, \lambda, \lambda)$ with $\sum \lambda = 0$ by an appropriate basis choice. Choosing the coefficients $m_i^2$, $a$ and $b$ appropriately, one can arrange to have a vacuum value $\langle \Phi \rangle = \text{diag}(2, 2, 2, -3, -3)$, around which the excitations of the $\Phi$ field all have non-negative (mass) values.

The mass term for the $\text{SU}(5)$ gauge field $A_\mu$ is obtained from the kinetic term for the Higgs field $-\frac{1}{2} \text{tr} [[[A_\mu, \Phi]^T, \Phi]]$. For $\Phi = \langle \Phi \rangle$, i.e. from $\frac{g_2}{2} \text{tr} [[[A_\mu, \Phi]^T, \Phi]]$. So the $A_\mu$ fields remaining massless, which belong to the unbroken subgroup $\text{SU}(3)$, belong to the $\text{SU}(5)$ generators that commute with $\langle \Phi \rangle$. 
Write the SU(5) generators in the defining 5x5 fundamental representation as

\[
(T^I)_A = \begin{pmatrix}
  a_1 & a_4 & a_5 & b_1 & b_2 \\
  a_4^* & a_2 & a_6 & b_3 & b_4 \\
  a_5^* & a_6^* & a_3 & b_5 & b_6 \\
  b_1^* & b_2^* & b_3^* & c_1 & c_3 \\
  b_2^* & b_4^* & b_6^* & c_3 & c_2 \\
\end{pmatrix}
\]

The \( T^I \) generate infinitesimal transformations of \( \mathbb{R}^5 \) by \( S_u = i \epsilon_{IJK} T^I T^J \) which preserve hermiticity of \( \mathbb{R}^5 \) formal \( \mathbb{R}^5 \).

Given the \( \langle \phi \rangle = \nu, \text{diag}(2,2,2,-3,-3) \) vacuum value of \( \mathbb{R}^5 \), the "a" 3x3 and "c" 2x2 diagonal blocks clearly commute with \( \langle \phi \rangle \) and so remain unbroken. These correspond to the SU(3) and SU(2) SM factor groups.

There is one more unbroken generator, which is proportional to \( \langle \phi \rangle \) itself. In its conventional SM guise it is \( \mathcal{Y} = \text{diag} \left( -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{2}, \frac{1}{2} \right) \), where the normalization is chosen so as to give the correct hypercharge assignments for quarks and leptons.

One needs to be careful with the normalization of \( \mathcal{Y} \) however. For any simple compact Lie group, there is a conventional choice of generators \( T^I \) so that in each reducible or irreducible representation \( R \) they satisfy the normalization condition \( \text{tr}(T^I T^J) = C_R S_{IJ} \). \( N_R \) is the Dynkin index of the representation \( R \). For the fundamental representation \( F \) of SU(\( N \)), \( C_F = \frac{1}{2} \) by convention. \( C_R \) then determines \( N_R \) for other representations. For example, for the adjoint \( A_{N-1} \) rep., \( C_A = N \) for SU(\( N \)); this is related to the Cartan-Killing metric \( g_{IJ} = -\frac{1}{N} \epsilon^{IJK} f^{JL} f^{K} = \delta_{IJ} \) for compact semi-simple groups.

For the SU(3) and SU(2) subgroups corresponding to the "a" and "c" blocks, the standard normalization is achieved by taking \( T^{3g}_{(3)} = \frac{i}{2} \lambda^g \), for which \( \text{tr}(\lambda^g \lambda^g) = 2 \), and \( T^{3g}_{(2)} = \frac{i}{2} \sigma_3 \), for which also \( \text{tr}(\sigma_3 \sigma_3) = 2 \). When it comes to \( \mathcal{Y} \), however,
The traditional normalization does not agree with the $N_f = \frac{3}{2}$ convention for SU(5) fundamental representation generators. Instead, one has $tr(Y_i) = \frac{3}{4} + \frac{7}{4} = \frac{5}{2}$. So in order to incorporate the $Y$ generator into a conventionally normalized SU(5) set, one needs to define $\hat{Y} = \frac{3}{5} Y$, i.e.

$$\hat{Y} = \text{diag}(\frac{1}{115}, \frac{1}{115}, \frac{1}{115}, \frac{3}{120}, \frac{3}{120}).$$

This rescaling has an impact on the hypercharge coupling constant as well: the SM covariant derivative $D^\alpha = \partial^\alpha - ig_\alpha Y^\alpha$, needs to become $D^\alpha = \partial^\alpha - ig_\alpha \hat{Y}^\alpha$, with $g_\alpha = \frac{3}{5} g_\alpha$. Accordingly, at energy scales above the scale of SU(5) spontaneous breaking, one will have $g_5^2 = g_3^2 = g_2 = g_1 = \frac{3}{5} g_1$, where $g$ is the SU(5) coupling constant.

Of course, there are more SU(5) gauge fields than just the $8+3+1 = 12$ standard-model gauge fields for the SU(3) x SU(2) x U(1) SM subgroup. SU(5) has 12 extra generators, corresponding to the off-diagonal "b" block. With respect to the SM (SU(3), SU(2))y subgroup, the $X_{\alpha\alpha}'$s transform as $(3,2)_{C/6}$ (or as $(3,2)_{C/2}$ with respect to (SU(3), SU(2))y). Effects of such fields have not yet been seen, and so a successful grand unified theory needs to arrange for symmetry breaking from SU(5) down to SU(3)xSU(2)xU(1) at a sufficiently high mass scale in order to make the effects of the $X_{\alpha\alpha}'s$ unobservable to date. This would be analogous to the SU(2)xSU(2) → U(1) breaking in the Standard Model, which would have to occur at lower (≈100 GeV/c²) mass scales. It is precisely the occurrence of new, unseen phenomena that poses the key challenge to any proposal of larger unification: confirmation such as the discovery of the W± and Z0 leads to triumph, but contradiction can kill the proposal.
The quarks and leptons of the Standard Model need to be assembled into representations of SU(5), as well, and the remarkable thing about the Georgi-Glashow model is how precisely this works, with no predictions of "exotics" that would have to be explained away. The fundamental 5 representation of SU(5) decomposes as $5 \rightarrow (3,1)_{-2/3} \oplus (1,2)_{1/6}$ under SU(3)$_c$ x SU(2)$_L$. Label the 5 representation components as $(f_1, f_2, f_3, h_1, h_2)$ (written as a row). The traceless structure of the SU(5) 9 generator requires that the corresponding hypercharges satisfy $3y_f + 2y_h = 0$. What can these spinor fields be? Standard Model fermions transforming as $(3,1)$ under SU(3)$_c$ x SU(2)$_L$ can only be found among the RH quarks: either $U_{3R}$ or $Q_{3R}$. The $(1,2)$ partner fermions under SU(5) must have the same RH spinor structure, since Lorentz and internal symmetries remain in a direct product. So one needs to find RH spinor fields in the $(3,1)$ and $(1,2)$ SM reps with $y_f = -\frac{3}{2}y_f$. Looking at the available SM species, the combination that works is to let the $f_3$ triplet be the charm quark, which have hypercharge $y_f = -\frac{1}{3}$, and to let the $h_2$ doublet be $(\bar{L}_{10}a) = E_{ab}(L_{10})^b$, which have $y_f = +\frac{1}{2}$.

Recall that the $L_{10}$ leptons transform as $(1,2)_{-1/2}$, so their charge conjugates $\bar{L}_{10}$ are RH spinors transforming in the $(1,2)_{+1/2}$. However, in SU(2)$_L$ one can lower an index with $E_{ab}$, so $(\bar{L}_{10}a) = E_{ab}(L_{10})^b$ are RH spinors transforming in the $(1,2)_{+1/2}$ as required. Note that the $(y_f, y_h)$ changes hit precisely the traditional $Y = -y_f y_h (1 + 1/2)$. We still need to find an SU(5) covariant way to incorporate the RH $Q_{1R}$ in the $(1,1)_{-1}$, the RH $U_{3R}$ in the $(3,1)_{2/3}$ and the LH $Q_{1L}$ in the $(3,2)_{-1/6}$. Consider the $Q_{1L}$ first. Since they transform in the (SU(3)$_c$ x SU(2)$_L$) rep 3 x 2, with both SU(3) and SU(2) indices, they cannot be lodged in a single-index SU(5) representation. Try instead an antisymmetric two-index
$\text{SU}(5)$ tensor representation $\chi_{[AB]}$. This is a $5^{\frac{4}{2}} = 10 \oplus$ dimensional representation of $\text{SU}(5)$:

$$\chi_{ABm} = \left( \begin{array}{l}
\varepsilon^{\overline{a}\overline{b}c}(U_{\overline{m}})_{\overline{a}} - Q_{\overline{l}\overline{m}c} \vspace{5pt} \\
Q_{\overline{l}\overline{m}c} \overline{c}_{ab}(ER_{m})_{a} \end{array} \right)$$

every component of $\chi_{ABm}$ is a LH spinor.

See how the hypercharge assignments work out for this $\text{SU}(5)$ representation. For a transformation matrix $M_{A}^{B} \in \text{SU}(5)$, $\chi_{AB}$ transforms as $\chi_{AB} \rightarrow M_{A}^{E} M_{B}^{F} \chi_{EF}$, so for a transformation $M_{A}^{B} = \exp \left( i \frac{\lambda}{\sqrt{2}} \right)_{A}^{B} = \exp \left( i p \gamma_{5} \right)$ with $p = \frac{1}{2} \overline{\sigma}$ rescaled parameter and the traditional $\gamma_{5} = \text{diag}(-\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, 1, 1)$ generator, the hypercharges corresponding to the A and B indices simply add, so for the components of $\chi_{ABm}$ we find traditional $\gamma_{5}$ hypercharge assignments

$$\left( \begin{array}{c}
-\frac{1}{2} - \frac{1}{2} = -\frac{3}{2} \\
-\frac{1}{2} + \frac{1}{2} = 0 \end{array} \right)$$

which are the correct $\gamma_{5}$ hypercharge assignments for the indicated SM LH fermions.

(Recall that $U_{m}$ has $y = \frac{2}{3}$ so $(UR_{m})_{a}$ has $y = -\frac{3}{3}$ and $E_{m}$ has $y = -1$ so $(ER_{m})_{a}$ has $y = +1$.)

Thus, assembling the roster of SM spinor fields written all as LH spinors, and so change conjugating the SU(5) 5 rep. $Y_{m}^{A} = (d_{m\overline{a}}, E_{ab}(L_{m})_{a}^{b})^{B}$ to give an SU(5) $\overline{5}$ representation $(Y_{m})^{A} = (d_{m\overline{a}}, E_{ab}(L_{m})_{a}^{b})$, all the LH Standard-Model fermions fit precisely into SU(5) representations $5 \oplus \overline{5} \oplus 10$ for each $m = 1, 2, 3$ generation.

Of course, neutrino masses can be incorporated via the see-saw mechanism by including gauge singlet right-handed neutrinos in each generation. Changing conjugating them for the purpose of making a purely left-handed spinor roster, the SU(5) content then becomes $5 \oplus 10 \oplus 1$. 
With just the adjoint $24$ Higgs $\Phi$ taking the
$\langle \Phi \rangle = v, \text{diag} (2,2,2,-3,-3)$, one directly arrives
at a system with massless $\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)_R$ gauge fields
plus massive vector fields $X^a$ with masses of order $g v$.
After eating the $\Phi$ components corresponding to the broken
"b" generators of $\text{SU}(5)$. This, of course, can't be the
full story because one hasn't worked in the symmetry
breaking structure of the Standard Model itself. In
order to do this, one needs to include more $\text{SU}(5)$ Higgs
fields. This can be done by including an $\text{SU}(5) \subset \text{Hyp}$
$H_\alpha$. Respecting renormalizability, the general $\text{SU}(5)$-invariant
potential containing terms of dimensions $\leq 4.5$
$$V(\Phi, H) = V(\Phi) + V(H) + \lambda_1 (r \Phi^2)(H^+H) + \lambda_2 (H^+\Phi^2 H)$$
with
$$V(\Phi) = -m^2 r(\Phi^2) + a r(\Phi^4) + b [r(\Phi^2)]^2$$
$$V(H) = -m^2 (H^+H) + \lambda (H^+H)^2$$
Strictly speaking, one should start from $V(\Phi, H)$ and
analyse the full symmetry breaking pattern down to just $\text{SU}(3) \times \text{U}(1)_R$.
It is easier to present the symmetry-breaking structure as
a two-stage process, however:
$$\text{SU}(5) \overset{H}{\rightarrow} \text{SU}(3) \times \text{SU}(2) \times \text{U}(1)_R \overset{H}{\rightarrow} \text{SU}(3) \times \text{U}(1)_R$$
In the first stage, for $a > 0$ and $b > -\frac{3}{2} a$, one has
$\langle \Phi \rangle = v, \text{diag} (2,2,2,-3,-3)$ with $v^2 = m^2/(14a + 60b)$. After
this first stage of symmetry breaking, the $X^a$ ($3,2$) vectors
develop a mass $m^2_{X^a}$ after eating the "b" block $\Phi$ Higgs.
Meanwhile, the $\text{SU}(5) H$ Higgs breaks into a $(3,1)$ colour triplet $h_\alpha$
and a flavour doublet $h_d$. These acquire mass terms
$$m_{h_\alpha}^2 = -m^2 + (30 \lambda_1 + 4 \lambda_2) v^2$$
$$m_{h_d}^2 = -m^2 + (30 \lambda_1 + 9 \lambda_2) v^2$$
In order for the second, $\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)_R \rightarrow \text{SU}(3) \times \text{U}(1)_{em}$,
stage of symmetry breaking to occur at lower energy $v \approx 2.46 \text{GeV}$,
one needs to have $m_{h_d} \leq v^2$. If that is the case, $h_d$ will
Derive down to energies of order $v_2$ while the "superheavy" particles such as the $X_i^\pm$ with masses of order $v_i$ decouple. The relevant physics of the surviving "light" particles will then be determined by an $SU(3) \times SU(2) \times U(1)$ invariant Higgs potential

$$V_{eff}(h^0) = -m^2 h^0 h^0 + \lambda (h^0 h^0)^2$$

which is just the ordinary Higgs potential for the Standard Model. Then $h^0$ develops a VEV $v_2 = (m^2/\lambda)^{1/2}$, which should be about $246 \text{ GeV}$. A proper treatment of the combined system $V(\Phi, H)$ would yield a $<\Phi>$ VEV, that is slightly shifted, together with other small corrections of order $v_2/v_1$. The $X_{11}$ and $X_{12}$ massive vectors also develop a mass difference of order $v_2$.

At the $v_1 \approx M_{EW}$ scale, $SU(5)$ invariance imposes the relation between coupling constants $g_5 = g_3 = g_2 = \sqrt{\frac{3}{8}} g_1$ gives $\sin^2 \theta_W = \frac{g_1^2}{g_5^2 + g_3^2} = \frac{3}{8} = 0.38$ as compared to the experimental value $\sin^2 \theta_W = 0.231$. Even worse agreement is found for the strong coupling $g_5^2$, which is much larger than $g_3^2$ or $g_2^2$. What has not yet been taken into account in this discussion is the evolution of couplings with scale, which we now need to consider. At energy scales much less than $M_{EW}$, there will be radiative corrections proportional to $\ln(M_{EW}/E)$.

The Running of Coupling Constants

At tree level in quantum field theory, one has vertices like $g_5 q \bar{q}$ with coupling constant $g_5$. Radiative corrections to such interactions involve loop diagrams with the same external lines, like $\frac{g_5^2}{4\pi^2}$ These generally are divergent and require renormalization of the coupling constant $g_5$. In the process of renormalization, one needs to introduce a regulator to make such ultraviolet divergent diagrams finite.
so they can meaningfully be handled. The most straightforward regulator is simply a cutoff, i.e., a maximum scale \( \Lambda \) for \( \mathcal{F} \) in momentum-space integrals. There are other, more Lorentz-covariant regularization techniques, such as dimensional regularization \( \mathcal{F} \rightarrow \mathcal{F}^{1-n} \) or \( \gamma \)-function regularization, or Pauli-Villars regularization. All of them, in one way or another, involve the introduction of a reference subtraction scale \( \mu \).

In the process of renormalization, a renormalized coupling \( \alpha_\mu \) is expressed in terms of an unrenormalized coupling \( \alpha_0 \) and the regulator (e.g., the cutoff \( \Lambda \)), referred to the subtraction scale \( \mu \). For example, in electrodynamics one has

\[
\alpha_\mu^2 = \alpha_0^2 - \frac{\alpha_0^4}{12\pi^2} \ln \left( \frac{\Lambda^2}{\mu^2} \right),
\]

from corrections to the 1-loop level.

When amplitudes are rewritten in terms of \( \alpha_\mu \) and other renormalized quantities like the electron mass, the \( \Lambda \) dependence of the divergent diagrams cancels against the \( \mu \) dependence introduced by rewriting expressions originally involving \( \alpha_0 \) in terms of \( \alpha_\mu \). Subsequently, one may take the limit \( \Lambda \rightarrow \infty \) and obtain non-divergent results, treating \( \alpha_\mu \) as a finite coupling constant (which then needs to be determined by experiment or relation to other couplings). Although the renormalization procedure removes the ultraviolet divergences, it does leave a footprint of the procedure through dependence on the reference subtraction scale \( \mu \).

Considering once again the example of electrodynamics, if one compares the values of \( \alpha_\mu^2(\mu) \) and \( \alpha_\mu^2(\infty) \), one finds (again including just the leading 1-loop corrections)

\[
\frac{1}{\alpha_\mu^2(\mu)} = \frac{1}{\alpha_\mu^2(\infty)} + \frac{\alpha_0^4}{12\pi^2} \ln \left( \frac{\mu^2}{\Lambda^2} \right).
\]

In consequence of renormalization, \( SU(N) \) gauge coupling constants \( g_N \) evolve with scale \( \mu \) according to the
\[
\frac{\partial \beta g^2}{\partial \mu} = \beta(g^2).
\]

Writing \( \beta = -\frac{3g^3}{4\pi} \), one finds the following contributions from generic spin-\( j \)\(^1\) spin-\( j \)\(^2\) and spin-1 fields going around the one-loop diagram for SU(N) gauge symmetry with the standard normalization:

\[
b_N^L = \frac{1}{12\pi} (11C_A - 4n_1^L C_1 + n_2^L C_2) - n_3^L C_3.
\]

where \( n_1^L \) is the number of fermions in the \( i \)th representation, \( n_2^L \) is the number of scalars in the \( i \)th representation, and \( C_A \) is the Dynkin index of the representation \( R \). Recall that for SU(N), \( C_A = N \) for the adjoint representation and \( C_F = \frac{N}{2} \) for the fundamental representation. For U(1), the results are different owing to the difference in normalization of the traditional \( j \) generator. To get \( b_j \), compare to electrodynamics, for which \(-\frac{2}{e^2(u)} \frac{\partial \beta}{\partial \mu} = -\frac{2}{24\pi^2} \) where the final factor of \( 2 \) on the RHS arises from counting \( L+R \) projections of the electron field, each having \(\text{tr}(J^2) = 1\).

Analogously, for U(1) one has \( b_y = \frac{g_3}{24\pi^2} \text{tr}(J^2) = -\frac{1}{24\pi^2} \) so \( b_y = -\frac{1}{24\pi^2} \text{tr}(J^2) \).

Leaving aside contributions from the Higgs scalars, which are small compared to the contributions from the massless vectors and spinors, in the Standard Model one has

\[
\mu \frac{2}{\partial \mu} g_3^2(\mu) = -\frac{g_3^2(\mu)}{4\pi^2} \left( \frac{11}{4} - \frac{n_1^L}{3} \right)
\]

\[
\mu \frac{2}{\partial \mu} g_2^2(\mu) = -\frac{g_2^2(\mu)}{4\pi^2} \left( \frac{11}{6} - \frac{n_1^L}{3} \right)
\]

where \( n_1^L \) is the number of generations

\[
\mu \frac{2}{\partial \mu} g_1^2(\mu) = -\frac{g_1^2(\mu)}{4\pi^2} \left( -\frac{5n_1^L}{9} \right)
\]

(since \( \text{tr}(J^2) = \frac{10}{3} n_1^L \)).

At the \( \Lambda_{\text{QCD}} \) scale where SU(5) should be restored, one would have \( g_3^2(\Lambda_{\text{QCD}}) = g_2^2(\Lambda_{\text{QCD}}) = \frac{8}{5} g_1^2(\Lambda_{\text{QCD}}) \).

Starting from such SU(5) compatible couplings and integrating the renormalization group equations down to scale \( \mu \),
one has
\[ \frac{1}{g_3^2(\mu)} = \frac{1}{g_3^2(M_{\text{cut}})} - \frac{1}{8\pi^2} \left( 11 - \frac{4n_g}{3} \right) \ln \left( \frac{\text{M}_{\text{cut}}}{\mu} \right) \]
\[ \frac{1}{g_2^2(\mu)} = \frac{1}{g_2^2(M_{\text{cut}})} - \frac{1}{8\pi^2} \left( \frac{22}{3} - \frac{4n_g}{3} \right) \ln \left( \frac{\text{M}_{\text{cut}}}{\mu} \right) \]
\[ \frac{1}{g_1^2(\mu)} = \frac{1}{g_1^2(M_{\text{cut}})} - \frac{1}{8\pi^2} \left( \frac{20n_g}{9} \right) \ln \left( \frac{\text{M}_{\text{cut}}}{\mu} \right) \]

Subtract the second of these from the first:
\[ \frac{1}{g_3^2(\mu)} - \frac{1}{g_2^2(\mu)} = -\frac{11}{24\pi^2} \ln \left( \frac{\text{M}_{\text{cut}}}{\mu} \right) \]

Subtract \( \frac{3}{2} \) of the third of these from the second:
\[ \frac{1}{g_2^2(\mu)} - \frac{3}{5g_1^2(\mu)} = -\frac{11}{12\pi^2} \ln \left( \frac{\text{M}_{\text{cut}}}{\mu} \right) \]

Then use \( g_1 = \frac{e}{\cos\Theta_w} \), \( g_2 = \frac{e}{\sin\Theta_w} \) together with
\[ \frac{1}{g_3^2(\mu)} - \frac{1}{g_3^2(\mu)} = \frac{1}{2} \left( \frac{1}{g_3^2(\mu)} - \frac{3}{5g_1^2(\mu)} \right) \] to deduce
\[ \sin^2\Theta_w = \frac{g_3^2}{g_2^2 + g_1^2} = \frac{1}{\alpha} + \frac{5\alpha^2}{9g_3^2(\mu)} \]
and
\[ \ln \left( \frac{\text{M}_{\text{cut}}}{\mu} \right) = \frac{4\pi^2}{11e^2\mu} \left( 1 - \frac{8\alpha^2}{3g_3^2(\mu)} \right) \]

The appropriate scale \( \mu \) at which to evaluate these relations is \( \mu \approx \text{M}_2 \), which is the typical energy of processes used to measure \( \sin^2\Theta_w \). Extrapolation of \( g_3 \) from lower energy data gives \( g_3^2(\text{MeV})/4\pi = 0.118 \pm 0.006 \) and \( \alpha^2(\text{MeV})/4\pi = 0.128 \).

These then give \( \sin^2\Theta_w^{\text{meas}} = 0.283 \) and \( \text{M}_{\text{cut}} \approx 1.1 \times 10^5 \text{eV} \).

The SU(5) based prediction of \( \sin^2\Theta_w \) is quite close to the measured value \( \sin^2\Theta_w = 0.231 \). However, in the years since 1974, the accuracy of both measurements and calculations has improved to the point where it is now clear that there is not a precise agreement. A more serious problem arises from proton decay, however.
Including an SU(5) singlet to accommodate RH neutrinos to generate neutrino masses via a Seesaw mechanism, the SU(5) fermion representations are 5 ⊕ 10 ⊕ 1. A natural question is whether this representation content can fit nicely into a single representation of a larger grand unification group. Even before the Georgi-Glashow SU(5) model, Howard Georgi realized that the 16 dimensional Spin(10) representation of Spin(10) decomposes under SU(5) x U(1) as

16 → 10, 5, 15, giving a way to obtain the correct SU(5) representations for the roster of left-handed SM fermions of a given m = 1, 2 or 3 generation. Having all the SM fermions fit into an irreducible representation is clearly attractive. For the vector fields, the adjoint of Spin(10) → adjoint of SO(10) decomposes as 45 → 24, 10, 10, 10, 10, 10, 10, 10.

A key task for Spin(10) grand unification is to ensure that the extra vectors lying outside the SM SU(3) x SU(2) x U(1) all become sufficiently massive to prevent any unseen physical effects being in conflict with observation.

Proton decay

The problem of physical predictions in conflict with observation comes to the fore in the SU(5) Georgi-Glashow model. The adjoint 24 of SU(5) contains the "b" block of generators, which under SU(5) → SU(3) x SU(2) x U(1) symmetry breaking give rise to massive Xμν vector fields. As we saw in the Standard Model itself, integrating out these vector fields will generate dimension 6 terms in the effective action (in particular, (SU(5)4 terms) with an effective coupling constant \( \frac{g_s^2}{M_n} \), where \( M_n \) is the mass of the Xμν vector field.
The X_{u,b} vector bosons carry both SU(3)_c colour and SU(2)_L


group and hence traditionally-normalized T hypercharge \(-\frac{1}{6}\):

\[ (3, \frac{1}{3}) \text{ rep.} \]

Consequently, there can be interactions like

\[ X_u \overline{u} e \rho^+ \nu^- \text{, i.e. } X_u \overline{u} \, u \rho^+ \nu^- (y^2 + y^2 + 1 = 0) \]

and \( X_d \overline{d} e^+ \rho^- \nu^0 \text{, i.e. } X_d \overline{d} \, d \rho^+ \nu^- (y^2 - y^2 - 1 = 0) \).

Using these interactions together with \( X_{u,b} \) propagators which


have leading structure \( y_{\mu
\nu} \overline{X}_a \overline{X}_b / M_X \), one can have
diagrams like

\[ \begin{array}{c}
\text{\begin{tikzpicture}
\node (mu) at (0,0) {\( \nu \)};
\node (d_1) at (-2,1) {\( \overline{u} \)};
\node (d_2) at (2,1) {\( u \)};
\node (d_3) at (4,0) {\( \nu \)};
\node (d_4) at (-4,0) {\( \overline{u} \)};
\draw[->] (d_1) -- (mu);
\draw[->] (d_2) -- (mu);
\draw[->] (d_3) -- (mu);
\draw[->] (d_4) -- (mu);
\end{tikzpicture}}
\end{array} \]

which give rise to processes like \( u \overline{u} \rightarrow d e^+ \)

can be the left diagram as \( \begin{array}{c}
\text{\begin{tikzpicture}
\node (mu) at (0,0) {\( \nu \)};
\node (d_1) at (-2,1) {\( \overline{d} \)};
\node (d_2) at (2,1) {\( d \)};
\node (d_3) at (4,0) {\( \nu \)};
\node (d_4) at (-4,0) {\( \overline{d} \)};
\draw[->] (d_1) -- (mu);
\draw[->] (d_2) -- (mu);
\draw[->] (d_3) -- (mu);
\draw[->] (d_4) -- (mu);
\end{tikzpicture}}
\end{array} \).

This process changes
electric charge (\( E_{\mu} : \frac{2}{3} + \frac{2}{3} \rightarrow \frac{2}{3} + 1 \)). It violates baryon
number (\( B : \frac{2}{3} + \frac{2}{3} \rightarrow -\frac{1}{3} + 0 \)) and violates lepton number
(\( L : 0 + 0 \rightarrow 0 - 1 \)), but it conserves \( B - L \) (\( B - L : \frac{2}{3} - \frac{1}{3} - 1 \)).

The \( u \overline{u} \rightarrow d e^+ \) process can give rise to proton decay;
\( p \rightarrow u \overline{u} d \rightarrow d e^+ e^+ \rightarrow \pi^0 e^+ e^+ \) with the \( \pi^0 \) subsequently
decaying into a pair of \( \gamma \)-rays, as we have seen: \( \pi^0 \rightarrow \gamma \gamma \).

Detailed analysis shows that the proton lifetime in the
minimal SU(5) Georgi-Glashow model can't be more than
10^{31} years. In Spin(10) models, the lifetime can be pushed up,
but by arranging specific cancellations for the most dangerous
processes. Proton decay is accordingly a key vulnerability for
grand unified models.

On the observational front, the best current limits
come from the Super Kamiokande Lab, located 1000 meters
underground in the Mozumi Mine near the city of Hida
in Gifu Prefecture, Japan. The detector consists of 50,000 tons
of ultra-pure water watched by over 13,000 photomultiplier tubes.
The Super Kamiokande detector contains $4.5 \times 10^{33}$ protons. If one of the many protons in the water tank were to decay, the decay products would be detectable. However, the detector has found no evidence for proton decay in data spanning over 17 years. From this data, a bound on the proton lifetime has been obtained: $t_{\text{proton}} > 5.9 \times 10^{33}$ years. This kills the minimal SU(5) Georgi-Glashow model.

There is a proposal to build a detector 10 times larger than the Super Kamiokande detector—the project is known as Hyper Kamiokande. Construction is expected to begin in 2018, with observations to start in 2025.

Crossing coupling constant trajectories

Another way to view the prospects for unification of forces beyond the Standard Model is from the bottom up. Instead of assuming a particular unified model based on some gauge group achieving unification of couplings at an energy scale like $10^{15}$ GeV, consider just the known evolution of $SU(3) \times SU(2) \times U(1)$ couplings in the Standard Model. For the coupling constants $g_3, g_2$ and $g_1 = \sqrt{3} g_2$, let $\alpha = g_2$.

One has the renormalization-group evolutions $\frac{1}{\alpha(\mu)} = \frac{1}{\alpha_i(\mu_0)} + b_i \ln \left( \frac{\mu^2}{\mu_0^2} \right)$ for $\alpha_i, \alpha_2$ and $\alpha_3$ as $\mu$ increases away from the $\mu_0 \approx M_Z$ scale. In a log plot, these coupling-constant trajectories are just straight lines. If there were a unified theory based on some simple gauge group (e.g., SU(5) or Spin(10)), there would have to be some scale $\mu$ where $\alpha_1, \alpha_2$ and $\alpha_3$ cross at a point, just as happens in SU(5) or Spin(10) models. The problem that emerges then from current data and calculational expertise is that the Standard Model coupling trajectories fail to cross at a point. They almost cross at a point, but miss, leaving a small triangle.
Plotting the $\frac{1}{\alpha_3}$, $\frac{1}{\alpha_2}$ and $\frac{1}{\alpha_1}$ trajectories versus $\log_2(\mu)$ gives

The implication of this failure to cross at a point is that something else has to happen in between the $M_Z \sim 90 \text{ GeV}$ and the anticipated $10^5 \text{ GeV}$ grand unification scale. Moreover, one would like whatever new feature emerges to address as much as possible the naturalness problem of having to finely tune the theory so as to obtain the vast hierarchy of scales. In the SU(5) Georgi-Gelashov model, the naturalness problem consisted in the need to have $M_Z \sim 246 \text{ GeV}$ for SM symmetry breaking while $M_1$, which governs the $SU(5) \rightarrow SU(3) \times SU(2) \times U(1)$ symmetry breaking, corresponding to $M_U \sim 10^5 \text{ GeV}$. The most intensely studied additional feature of the unification picture is supersymmetry, which introduces a "superpartner" for every current SM species. So one proposes the existence of selectrons, sneutrinos, gauginos, Higgsinos, ... These change the renormalization coefficients $b_1, b_2, b_3$ in a very useful way, providing the scale of supersymmetry breaking is not too large. One hopes for $M_{\text{ SUSY}} \sim 1 \text{ TeV}$; in that case the coupling constant trajectories naturally do cross at a point.

Moreover, supersymmetric SU(5) and Spin(10) models manage to have sufficiently long proton lifetimes to be consistent with current data.
Some projected proton lifetimes for various models:

<table>
<thead>
<tr>
<th>Model</th>
<th>years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal SU(5) (George-Glashov)</td>
<td>$10^{30} - 10^{31}$</td>
</tr>
<tr>
<td>Susy SU(5) (MSSM)</td>
<td>$\sim 10^{34}$</td>
</tr>
<tr>
<td>Minimal non Susy SO(10)</td>
<td>$\leq 10^{35}$ max</td>
</tr>
<tr>
<td>Susy SO(10)</td>
<td>$10^{32} - 10^{35}$</td>
</tr>
<tr>
<td>Flipped SU(5) MSSM</td>
<td>$10^{35} - 10^{36}$</td>
</tr>
</tbody>
</table>

Flipped SU(5) has gauge group $(SU(5) \times SU(1))/Z_5$

- fermions in $\tilde{5}$ for leptons L and $\nu$ quarks
- 10 for Q quark doublet, $d^c$ and RH neutrino $N$
- 1 for charged leptons $e^c$

Higgs are just in the $10s$.

SM SU(5) is in the $SU(4)$ factor of the GUT group together with
- a diquark $(\bar{15} \bar{15}, \frac{1}{2} - \frac{1}{2})$ from the SU(5)