Naturalness and Top-Down BSM
Lecture 1

MITP Summer School 2018

Nathaniel Craig

Department of Physics,
University of California,
Santa Barbara, CA 93106

No warranty expressed or implied.
Suggestions, clarifications, and comments are welcome.

Contents

1 Naturalness

What are the natural sizes of parameters in a quantum field theory? The original notion is the result of an aggregation of different ideas, starting with Dirac's Large Numbers Hypothesis ("Any two of the very large dimensionless numbers occurring in Nature are connected by a simple mathematical relation, in which the coefficients are of the order of magnitude unity" [1]), which was not quantum in nature, to Gell-Mann's Totalitarian Principle ("Anything that is not compulsory is forbidden." [2]), to refinements by Wilson and 't Hooft in more modern language. In any event, for simplicity we will refer to this aggregate notion of naturalness as **Dirac naturalness**:

> In a theory with a fundamental scale Λ, given an operator O of the form
>
> $$\mathcal{L} \supset c_O O \tag{1}$$
>
> with scaling dimension Δ_O, the natural size of the coefficient c_O in natural units is
>
> $$c_O = \mathcal{O}(1) \times \Lambda^{4-\Delta_O} \tag{2}$$

This has the flavor of mere dimensional analysis, but it is reinforced by the nature of quantum corrections in QFT.

Of course, we have many examples of QFT which appear to violate this expectation. This leads to a refined notion of naturalness, due primarily to 't Hooft[1], which we will refer to as **technical naturalness**:

> Coefficients can be much smaller than their Dirac natural value if there is an enhanced symmetry of the theory when the coefficient is taken to zero. In this case, the natural size of the coefficient c_O is
>
> $$c_O = \mathcal{S} \times \mathcal{O}(1) \times \Lambda^{4-\Delta_O} \tag{3}$$
>
> where $\mathcal{S}$ is a parameter that violates the symmetry in question.

[1] Though certainly anticipated by Gell-Mann, who in the sentence after articulating the Totalitarian Principle notes "Use of this principle is somewhat dangerous, since it may be that while the laws proposed in this communication are correct, there are others, yet to be discussed, which forbid some of the processes that we suppose to be allowed."

The origin of this is fairly transparent: if the parameter $\mathcal{S}$ is zero, then there is an enhanced symmetry of the theory. Quantum corrections respect symmetries of the quantum action, and so radiative corrections will not regenerate c_O. If the symmetry is violated by nonzero $\mathcal{S}$, then there is a selection rule: radiative corrections must be proportional to the symmetry violation. We can formalize this at the level of spurion analyses, familiar from the chiral Lagrangian in QCD.

The two notions of naturalness are clearly on different footings. Ultimately, we expect Dirac naturalness to hold in all underlying field theories. Technical naturalness is itself a sort of halfway-house – while it allows us to understand how a small parameter can be radiatively stable, we are still left wanting an explanation for how the small parameter came about in the first place. But technical naturalness gives us the ability to understand how hierarchies observed in the infrared can be protected against radiative corrections that would otherwise spoil them.

These two notions of naturalness have been borne out countless times in nature, and provide a successful characterization of many of the parameters in the Standard Model. Two classic examples are the proton mass and flavor hierarchies.

1.0.1 The proton mass

This was the problem that originally motivated Dirac, and his own answer was wildly off the mark. Dirac understood that there was a mass scale associated with gravity, $M_{Pl} \sim 10^{19}$ GeV, as well as a mass scale associated with the proton, $m_p \sim 1$ GeV, and wished to understand why $m_p \ll M_{Pl}$.

Although the true explanation eluded Dirac, we now understand it to be a beautiful triumph of naturalness criteria. The answer is that the proton mass is dynamically generated by confinement, which in turn arises from the logarithmic evolution of a dimensionless coupling, which itself is the manifestation of a violation of symmetry – in this case, (classical) conformal symmetry. This phenomenon, known as *dimensional transmutation*, explains the existence of exponentially different scales.

The essential idea is that the QCD coupling, like all couplings in the Standard Model, runs as a function of scale, giving rise to a renormalization group equation of the form

$$\frac{\partial \alpha_3}{\partial \ln \mu} = -7\frac{\alpha_3^2}{2\pi} + \ldots \tag{4}$$

where $\alpha \equiv g^2/4\pi$. We can solve the RGE at one loop, starting from couplings defined at a fundamental scale (taken to be, e.g., M_{Pl}) down to some lower scale

μ:

$$\frac{1}{\alpha_3(M_{Pl})} - \frac{1}{\alpha_3(\mu)} = \frac{7}{2\pi} \ln \left(\frac{M_{Pl}}{\mu} \right) \tag{5}$$

This tells us that, starting from a finite value of g, it will eventually diverge in the infrared. Although there is no rigorous proof, we understand this to be associated with confinement in QCD. At one loop, we can take the scale of confinement Λ_{QCD} to be the scale at which the coupling diverges, in which case

$$\frac{1}{\alpha_3(M_{Pl})} = \frac{7}{2\pi} \ln \left(\frac{M_{Pl}}{\Lambda_{QCD}} \right) \Rightarrow \Lambda_{QCD} = M_{Pl} e^{-\frac{2\pi}{7}\frac{1}{\alpha_3(M_{Pl})}} \tag{6}$$

Lo and behold, we observe a new scale that is exponentially far from the fundamental scale, where the exponential difference owes to the gentle, logarithmic evolution of a dimensionless coupling.

As the proton acquires most of its mass from confinement, $m_p \sim \Lambda_{QCD}$, we see there is a Dirac natural explanation for $m_p \ll M_{Pl}$: all parameters can take a Dirac-natural size at the scale M_{Pl}, and a new scale is generated dynamically.

It's worth mentioning Dirac's approach to the problem. He observed that

$$\frac{Gm_p^2}{\hbar c} \sim 5 \times 10^{-39}, \tag{7}$$

the problem at hand. He also noted that the Hubble age of the universe was about

$$T \frac{m_p c^2}{\hbar} \sim 10^{42} \tag{8}$$

and that the mass of the universe to its visible limits was about

$$\frac{M}{m_p} \sim (10^{40})^2 \tag{9}$$

To him this suggested that there was a causal connection between dimensionless constants and powers of T. Since T changes in time, that also implies that fundamental constants change in time, e.g., that G evolves as $1/t$, and that M evolves as t^2. He proceeded to develop an elaborate theory of cosmology around this idea. This is not a cautionary tale against taking naturalness too seriously, but rather a cautionary tale about taking care to understand what naturalness is telling us.

It's amusing to note that Dicke (1961) pointed out that questions about the age of the universe could only arise if conditions were right for the existence of life, with the specific criteria that the universe must be old enough so that some stars completed their time on the main sequence and produced heavy elements, and young enough that some stars were still undergoing fusion. Working these out in terms of fundamental units, Dicke found the upper and lower bounds essentially lead to Dirac's equation – but rather than resulting from time variation of fundamental parameters, are anthropic in nature. Weinberg (1989) points to this as the first use of anthropic arguments in modern physics, about which more will be said soon.

1.0.2 Flavor hierarchies

A more subtle example is provided by the flavor hierarchies of the Standard Model. In the Standard Model, we see large hierarchies in fermion masses, e.g.

$$\frac{m_e}{m_t} \sim 10^{-5} \qquad \frac{m_\nu}{m_t} \sim 10^{-11} \tag{10}$$

Of course, in the Standard Model fermion masses are generated by electroweak symmetry breaking, so that these hierarchies emerge from hierarchies of Yukawa couplings.

These numerical hierarchies are not Dirac natural, but are technically natural. In the limit that the Yukawa couplings are taken to zero, there is an enhanced symmetry of the Standard Model, namely a $U(3)^5$ flavor symmetry. This corresponds to a $U(3)$ symmetry for each type of left-handed Weyl fermion, although it is often more conveniently decomposed into the following symmetries:

$$SU(3)_Q \times SU(3)_U \times SU(3)_D \times SU(3)_L \times SU(3)_E \tag{11}$$

$$\times U(1)_B \times U(1)_L \times U(1)_Y \times U(1)_{PQ} \times U(1)_E \tag{12}$$

We can then think of the yukawas as spurions for breaking the symmetry. For example, the up- and down-type Yukawa couplings break the $SU(3)_q^3 \equiv SU(3)_Q \times SU(3)_U \times SU(3)_D$ global symmetry, while the lepton yukawas break the $SU(2)_\ell^2 \equiv SU(3)_L \times SU(3)_E$ symmetry. We can track the symmetry breaking by treating the yukawas as fields transforming in definite representations of the global flavor symmetry, whose vacuum expectation values spontaneously break the symmetry. In this sense the yukawas are "spurion fields" for the broken symmetry. Qua spurions, the various yukawas transform as

$$Y^u \sim (3, \bar{3}, 1)_{SU(3)_q^3} \tag{13}$$

$$Y^d \sim (3, 1, \bar{3})_{SU(3)_q^3} \tag{14}$$

$$Y^e \sim (3, \bar{3})_{SU(3)_\ell^2} \tag{15}$$

Consequently, radiative corrections to the yukawa couplings are proportional to these spurions. Any numerical hierachies in the spurions are therefore radiatively stable.

Of course, we would still like an explanation for the origin of the numerical hierarchies – why the yukawas might have hierarchical values to begin with – but this can be accomplished by model-building at some fundamental scale at which the yukawas are generated. Once the hierarchies are generated, they persist into the infrared.

There are, however, three parameters of the Standard Model (coupled to gravity) that do not satisfy naturalness criteria. In order of increasing dimensionality, they are: the Strong CP problem, the electroweak hierarchy problem; and the cosmological constant problem. It's worth noting that "problem" is probably the wrong term here; "puzzle" is more appropriate. They aren't really problems in the sense that nothing breaks down in the theory itself if there is no deeper explanation; they are just suggestive that a deeper mechanism might be at play. In any event, for reasons that will become clear, we will discuss them out of their dimensional ordering, beginning with Strong CP and ending with the electroweak hierarchy problem.

2 The Strong CP Problem

We know that CP is not a symmetry of the Standard Model, being broken by the weak interactions. But there is another potential source of CP violation that is not, as yet, observed. The QCD Lagrangian in principle contains a term of the form

$$\mathcal{L} \supset -\theta \frac{g_s^2}{64\pi^2} \epsilon^{\mu\nu\alpha\beta} G^a_{\mu\nu} G^a_{\alpha\beta} = -\theta \frac{g_s^2}{32\pi^2} G^a_{\mu\nu} \widetilde{G}^{a\mu\nu} \tag{16}$$

This θ term is P- and T-odd, hence CP-odd. You can see this by analogy with electrodynamics, where $F_{\mu\nu}F^{\mu\nu} = -2(E^2 - B^2)$ and $F_{\mu\nu}\widetilde{F}^{\mu\nu} = -4E \cdot B$. Since E is P-odd and T-even, while B is P-even and T-odd, so the former combination is P- and T-even, while the latter is odd under both.

While this can be written as a total derivative, for non-abelian theories we are not entitled to discard boundary terms due to the existence of instantons. (Although most treatments of the Strong CP problem elaborately discuss instanton physics, here we will sidestep it in favor of the chiral lagrangian. It suffices only to note that they prevent us from ignoring field configurations at infinity.) So we have to contend with the possible physical consequences.

Both the physical substance of the θ term and implications thereof are easier to work out if we make use of the anomalous $U(1)_A$ symmetry of the strong interactions. You have likely encountered this before; of the $U(3) \times U(3)$ symmetry of the light quarks in QCD, this $U(1)$ factor is anomalous with respect to the strong interactions. As a result, we have the freedom to move the θ parameter around somewhat in the Standard Model Lagrangian.

As a reminder, a single *massless* Dirac fermion has two independent global symmetries,

$$\psi \to e^{i\alpha}\psi \qquad \psi \to e^{i\beta\gamma_5}\psi$$

which we are accustomed to thinking of as $U(1)_V$ and $U(1)_A$ global symmetries.

In terms of component Weyl fermions $\psi_{R,L} = \frac{1}{2}(1 \pm \gamma_5)\psi$ these are just the transformations

$$\psi_L \to e^{i(\alpha-\beta)}\psi_L \qquad \psi_R \to e^{i(\alpha+\beta)}\psi_R$$

Now the Noether currents associated with these two symmetries are

$$J^\mu = \bar{\psi}\gamma^\mu\psi \qquad J^{\mu 5} = \bar{\psi}\gamma^\mu\gamma^5\psi$$

In QCD we have three light quarks, so we can think of rotating them independently, and also into each other. This is a $U(3)\times U(3)$ global symmetry, which we often split into $SU(3)_V \times SU(3)_A \times U(1)_V \times U(1)_A$. In terms of a flavor triplet of Dirac spinors $q = \begin{pmatrix} q_u \\ q_d \\ q_s \end{pmatrix}$, we can write the two $SU(3)$ transformations transformations as

$$\begin{pmatrix} q_u \\ q_d \\ q_s \end{pmatrix} \to e^{i\alpha_a\tau^a + \gamma_5\beta_a\tau^a} \begin{pmatrix} q_u \\ q_d \\ q_s \end{pmatrix}$$

where the set of transformations parameterized by α^a with $\beta^a = 0$ is the diagonal subgroup isospin, and those parameterized by β^a with $\alpha^a = 0$ are axial rotations.

The corresponding currents, packaged in terms of the flavor triplet of Dirac fermions q, are

$$J^{\mu a} = \bar{q}\tau^a\gamma^\mu q \qquad J^{\mu 5a} = \bar{q}\tau^a\gamma^\mu\gamma^5 q \qquad J^{\mu V} = \bar{q}\gamma^\mu q \qquad J^{\mu A} = \bar{q}\gamma^\mu\gamma^5 q$$

In particular, under the $U(1)_A$ transformation of a light quark

$$q \to q' = e^{i\alpha\gamma_5/2} q \tag{17}$$
$$\bar{q} \to \bar{q}' = \bar{q} e^{i\alpha\gamma_5/2} \tag{18}$$

from the Fujikawa (path integral-based) interpretation of anomalies, one can show that $S[q,\bar{q}] \to S[q',\bar{q}']$ but there is a non-trivial transformation associated with the path integral measure for the quarks

$$\int [dq][d\bar{q}] \to \int [dq'][d\bar{q}'] e^{-\alpha \int d^4x \frac{g_s^2}{32\pi^2} G^a_{\mu\nu} \widetilde{G}^{a\mu\nu}} \tag{19}$$

due to the axial anomaly of QCD. Among other things, this means that we can trade the θ parameter for a complex phase in the quark masses, or visa versa.

Alternately, you can think of this as arising from a term in the effective action linking $G\tilde{G}$ and the axial current $J^{\mu 5} = \bar{q}\gamma^\mu\gamma^5\psi$ coming from the traditional triangle diagram.

One implication of this is that any axial rotations we might do in the course of eliminating phases from quark yukawa couplings (enroute to putting them all in the CKM phase) would lead to a shift in θ. That is to say, in general diagonalizing the quark mass matrix M gives us complex eigenvalues $m_i e^{i\alpha_i}$. We can remove the phase by the axial rotations

$$q_i \to e^{i\alpha_i\gamma_5/2} q_i \tag{20}$$
$$\bar{q}_i \to \bar{q}_i e^{i\alpha_i\gamma_5/2} \tag{21}$$

but this just shifts the θ term proportional to $\arg\det M = \sum \alpha_i$. We therefore often talk about the Strong CP problem in terms of the CP-violating phase associated with the total combination $\bar{\theta} = \theta + \arg\det M$.

Among other things, the interchange helps us to build confidence that θ is physical. If the quarks were massless, we could change θ without consequence, rendering it unphysical. But being massive, the quarks do not let us eliminate the phase, and we must still contend with the consequences.

To see what θ does physically (per Crewther, DiVecchia, Veneziano, Witten '79 [3]), it is convenient to move θ to the quark masses. We will actually want to be a little smart about this, in the following sense: we can just do the $U(1)_A$ rotation on all three quarks. But it will be useful to imagine doing both a $U(1)_A$ rotation and an $SU(3)_A$ rotation at the same time. For the time being, we will keep track of this freedom by doing an axial rotation of the form

$$q \to e^{-i\theta T\gamma_5/2} q \tag{22}$$

where T is some 3×3 diagonal matrix that we are going to normalize to $\mathrm{tr}T = 1$ and will otherwise fix momentarily. Then for small θ this gives us

$$-\theta \frac{g_s^2}{32\pi^2} G^a \tilde{G}^a \to -\bar{q}_R M e^{i\theta T} q_L + \text{h.c.} \to -i\theta \bar{q}_R M T q_L + \text{h.c.} \tag{23}$$

where we've just expanded to leading order in θ.

Now we want to follow this term down to low energies, into the chiral Lagrangian.

In case you need a refresher: we hypothesize that the QCD vacuum spontaneously breaks $SU(3)_V \times SU(3)_A \to SU(3)_V$. This leads to 8 goldstone bosons, so at low energies the effective theory of QCD should be described by the physics of these goldstones. And indeed, we see eight light pseudoscalars: the three pions, the eta, and the four kaons. We organize them into a goldstone matrix,

$$\Sigma(x) = \exp\left[2i\frac{\pi^a \tau^a}{f_\pi}\right] = \exp\left[\frac{\sqrt{2}i}{f_\pi}\begin{pmatrix} \frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0 & \pi^+ & K^+ \\ \pi^- & -\frac{1}{\sqrt{2}}\pi^0 + \frac{1}{\sqrt{6}}\eta^0 & K^0 \\ K^- & \bar{K}^0 & -\sqrt{\frac{2}{3}}\eta^0 \end{pmatrix}\right] \tag{24}$$

These goldstones transform in the adjoint of $SU(3)_V$ and with a shift under $SU(3)_A$. Here $f_\pi = 93$ MeV.

The idea is then to write down the most general set of interactions of the Σ field allowed by the $SU(3)\times SU(3)$ symmetry, as the goldstones nonlinearly realize the genuine symmetry of the theory. This is the chiral Lagrangian. (More generally, one can show that for a group G broken to a subgroup H, any H-invariant Lagrangian constructed out of the goldstones of the coset G/H can be expressed in terms of a G-invariant Lagrangian constructed of Σ fields.)

However, the symmetry is not exact; it is broken by quark masses, by a small amount compared to Λ_{QCD}. Therefore pions are not exact goldstones, but rather pseudo-goldstones. We can include the effect of this small breaking by promoting the mass matrix into a spurion – i.e., promoting it to a field and assigning it transformation properties of the form necessary to preserve the symmetry. This spurion can then also appear in the chiral Lagrangian in any way consistent with the symmetries. The leading operator involving the mass that we can write down is

$$\mathcal{L} \supset \frac{1}{2} f_\pi^2 \tilde{\Lambda} \mathrm{tr}(M\Sigma) + \text{h.c.} \tag{25}$$

where M is the mass matrix for the three light quarks and $m_\pi^2 = (m_u + m_d)\tilde{\Lambda}$ is determined by matching terms with QCD.

We can now introduce the effects of a complex mass into the chiral Lagrangian by taking $M \to M e^{i\theta T}$ everywhere we usually do in the chiral Lagrangian. so e.g.

$$\mathcal{L} \supset \frac{1}{2} f_\pi^2 \tilde{\Lambda} \mathrm{tr}(M e^{i\theta T}\Sigma) + \text{h.c.} \to \frac{1}{2} i\theta f_\pi^2 \tilde{\Lambda} \mathrm{tr}(MT\Sigma) + \text{h.c.} + \dots \tag{26}$$

$$= -\theta f_\pi \tilde{\Lambda} \mathrm{tr}(MT\pi^a \tau^a) + \text{h.c.} + \dots \tag{27}$$

where in the first line we're truncating all the higher order terms in θ and the CP-conserving mass term for the pions. In the second term we've expanded the pions and kept the linear term.

Naively, it looks like we've just created a huge problem. If M is real, the leading contribution from the mass in the chiral Lagrangian is a mass term for the pseudo-goldstones, as we expect. However, if M is complex, the term linear in the goldstones no longer vanishes! This creates a source term for pions, so the goldstones now mix with the vacuum. We can still compute in this theory, but it's a mess; we must constantly keep track of this effect.

We can take care of this from the beginning by making a wise choice for T by ensuring QCD vacuum alignment, i.e., that the CP-violating perturbation has zero amplitude to create a Goldstone (π, K, η) out of the vacuum. We can do this by taking

$$T = \frac{M^{-1}}{\mathrm{tr} M^{-1}} = \frac{1}{m_u + m_d} \begin{pmatrix} m_d & & \\ & m_u & \\ & & 0 \end{pmatrix} + O(m_u/m_s) + O(m_d/m_s) \tag{28}$$

because $\mathrm{tr}\pi^{\mathrm{a}}\tau^{\mathrm{a}} = 0$. (The denominator is fixed by choice of normalization.)

So we are dealing with a perturbation of the form

$$\delta\mathcal{L}_{CPV} \approx -i\bar{\theta}\frac{m_u m_d}{m_u+m_d}\bar{q}_R q_L + \text{h.c.} = -i\bar{\theta}\frac{m_u m_d}{m_u+m_d}\bar{q}\gamma_5 q \tag{29}$$

What does this do? We should follow it down to low energies. In the pion-nucleon effective Lagrangian, we now expect terms of the form

$$\mathcal{L} = \pi^a \bar{N}\tau^a(i\gamma_5 g_{\pi NN} + \bar{g}_{\pi NN})N \tag{30}$$

where the first term is the usual CP-conserving one, and the second one is due to CP violation. (The operator $\bar{q}i\gamma_5\tau^a q \sim \pi^a$, so this is one place we expect to see the CPV perturbation appear.)

We compute $\bar{g}_{\pi NN}$ by working out the matrix element,

$$\langle \pi^a N_f|\delta\mathcal{L}_{CP}|N_i\rangle = -\theta\frac{1}{f_\pi}\frac{m_u m_d}{m_u+m_d}\langle N_f|\bar{q}\tau^a q|N_i\rangle \tag{31}$$

which simply made use of the soft pion theorem for matrix elements of an operator O,

$$\lim_{q\to 0}\langle \pi^a(q)B|O|A\rangle = -\frac{i}{f_\pi}\langle B|[Q_5^a, O]|A\rangle \tag{32}$$

where $Q_5^a = \int d^3x\bar{q}\gamma_0\gamma_5\tau^a q$ is a charge of the axial part of the $SU(2)\times SU(2)$ flavor symmetry. This (dimensionless) matrix element can be estimated in terms of baryon and quark masses.

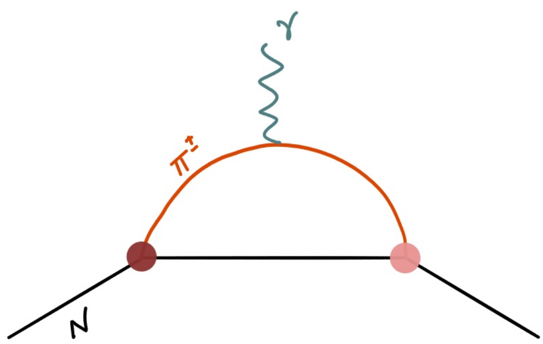

Figure 1: Loop diagram contributing to the neutron EDM.

Now that we have this CP-violating nucleon-nucleon-pion coupling, there is a one-loop diagram (shown in Figure 1) involving external neutrons with a loop of pions (to which we can attach a photon), where one vertex is the CPV one and the other is a standard one. We cut off the IR and UV divergences of this loop at m_π and the neutron mass, respectively, which gives us a neutron EDM of order

$$\begin{aligned} d_n &= \frac{g_{\pi NN}\bar{g}_{\pi NN}}{4\pi^2 M_N}\ln(M_N/m_\pi) \qquad (33) \\ &= \frac{g_{\pi NN}}{4\pi^2 M_N}\frac{-\bar{\theta}}{f_\pi}\frac{m_u m_d}{m_u+m_d}\langle N_f|\bar{q}\tau^a q|N_i\rangle \ln(M_N/m_\pi) M_N \qquad (34) \end{aligned}$$

This evaluates to $d_n \sim 5\times 10^{-16}\theta\, e$ cm, whereas the experimental bound is $|d_n| \lesssim 3\times 10^{-26}\, e$ cm. This implies $\theta \lesssim 10^{-10}$, whereas we might have expected it to be $\mathcal{O}(1)$. This is a violation of our naturalness expectations by ten orders of magnitude, with no obvious reason. Nothing goes wrong in the theory, or our universe, if θ obtains a value much larger than we observe. So this suggests a mechanism may be at work that renders θ natural.

There are three natural avenues. The first is to have a massless quark, since then θ is unphysical as it may be removed entirely by redefinitions of the massless quark (note all of our above expressions are proportional to products of all the quark masses). The second is to invoke technical naturalness, i.e., P or CP are exact symmetries at a high scale, broken spontaneously to give the CKM phase but not regenerate large θ. The third is to relax the value of θ.

The first option is ruled out by lattice data, which badly disfavors a massless quark. So let's briefly explore the second and third options.

2.1 Spontaneous P/CP Violation

Given what we've learned about technical naturalness, the most transparent option is to render the theta angle small by reference to technical naturalness: make P or CP good symmetries in the UV, broken spontaneously at some scale to give the known CP violation seen in the CKM matrix.

This is not an entirely trivial exercise. We have seen that the physical strong CP angle is the combination of the quark mass term phase and the intrinsic QCD phase,

$$\bar{\theta} = \theta + \arg\det M = \theta + \arg\det[Y_u Y_d] \qquad (35)$$

In the second step we've written it in the form we recognize prior to electroweak symmetry breaking: the phase lives in the Yukawa couplings of up-type and down-type quarks.

The challenge for a technically natural approach is thus to explain why

$$\arg\det[Y_u Y_d]$$

is small, but Jarlskog combination, which picks out the phase in the CKM matrix,

$$\arg\det[Y_u Y_d - Y_d Y_u]$$

is not. Naively, these two things are almost identical. But there are models that do just that. The most common route, known as the Nelson-Barr mechanism (due to Nelson [4] and Barr [5], separately in 1984), starts with CP as a UV symmetry and breaks it via the vevs of some complex scalars, which accumulate a relative phase. These scalars couple to Standard Model quarks with the assistance of some additional vector-like quarks. The couplings are engineered in such a way as to guarantee that $\arg\det[Y_u Y_d] = 0$ but the CKM phase is nonzero.

However, these models require a fair bit of exposition to understand in detail, so for illustration so let's explore a different possibility in which CP violation is always allowed. This originates with Barr, Chang, and Senjanovic (1991) [6]. In its simplest form, imagine extending the Standard Model with parity,

$$SU(3)_c \times SU(2)_L \times U(1)_Y \to SU(3)_c \times SU(2)_L \times SU(2)_R \times U(1)_Y \tag{36}$$

and adding an extra "mirror" copy of the Standard Model matter fields charged under $SU(2)_R$ (i.e., for $Q_L, u_R, d_R, L_L, e_R, H_L$ in the SM, add $Q_R, u_L, d_L, L_R, e_L, H_R$ with identical $SU(3) \times U(1)_Y$ quantum numbers, but $SU(2)_R$ charges instead of $SU(2)_L$ charges). Now the theory has an extended parity symmetry under which $P : SU(2)_L \leftrightarrow SU(2)_R$ and $L \leftrightarrow R$ on all the field labels. We can think of this as usual parity, $P : (t, \vec{x}) \to (t, -\vec{x})$, supplemented by action on the bosonic field labels. The θ term is odd under this parity, as with usual parity, and so is forbidden in the UV where the parity is good.

The parity then obviously requires the Yukawa couplings

$$y_L^u \bar{Q}_L u_R \tilde{H}_L + \text{h.c.} + \ldots \tag{37}$$

$$y_R^u \bar{Q}_R u_L \tilde{H}_R + \text{h.c.} + \ldots \tag{38}$$

(understood as 3×3 matrices in flavor space, and dots denote the down-type and lepton yukawas) to obey $y_L = y_R$. Then the 6×6 matrix in (flavor $\otimes$ LR) space for, e.g., the up-type quarks is

$$\mathcal{M} = \begin{pmatrix} 0 & y v_L \\ y^\dagger v_R & 0 \end{pmatrix} \tag{39}$$

where the rows are in Q_R, u_R and the columns in $\bar{Q}_L, \bar{u}_L$. This implies that $\det \mathcal{M} \sim \det y^\dagger y$ is real, so $\arg\det \mathcal{M} = 0$ even though y is complex.

Unlike Nelson-Barr models, parity violation is not needed to allow for a complex CKM phase; it's already allowed, and simply mirrored by a parity phase for the mirror fields. But parity violation is still required by observation: we don't see the additional $SU(2)_R$-charged states degenerate with their Standard Model counterparts. So we need $\langle H_R \rangle \gg \langle H_L \rangle$, which we can obtain by introducing some P-odd field ϕ that couples as $\phi(|H_L|^2 - |H_R|^2)$. When ϕ gets a vev this modifies the potential for H_L, H_R in such a way that we can have $\langle H_R \rangle \gg \langle H_L \rangle$ and decouple the parity partners. Of course, this violates the parity spontaneously, so θ can now be generated radiatively at lower scales. However, this is at least a three-loop effect and highly suppressed, so it can lead to a consistently small value of θ in the IR.

You might think that you can decouple the parity partners by an arbitrary amount, but this isn't so. The vev of ϕ can't be too big, because we have every reason to expect that Planck-scale physics generates operators like

$$\frac{1}{32\pi^2} \frac{\phi}{M_{Pl}} G\tilde{G} \tag{40}$$

Avoiding a reintroduction of the strong CP problem bounds $\langle \phi \rangle \lesssim 10^{-10} M_{Pl}$ (Berezhiani, Mohapatra, Senjanovic '93 [7]). Since at most $\langle H_R \rangle \sim \langle \phi \rangle$, this implies $\langle H_R \rangle \lesssim 10^{-10} M_{Pl}$, which puts the parity partners of first-generation fermions beneath 10 TeV.

2.2 Axion

To understand axion solutions, it's useful to begin with the θ dependence of the QCD vacuum energy, which we can figure out by again doing an axial rotation to eliminate θ from in front of $G\tilde{G}$. As a reminder, this gives us

$$\mathcal{L} \supset -\bar{q}_R M e^{i\theta T} q_L + \text{h.c.} \tag{41}$$

where T is the 3×3 matrix encountered earlier. If we just take care to keep the full exponential dependence on θ, rather than expanding, the leading piece (contribution to the vacuum energy) in the chiral Lagrangian is just

$$\mathcal{L} \supset \frac{1}{2} f_\pi^2 \tilde{\Lambda} \mathrm{tr} M e^{i\theta T} 1 + \text{h.c.} \tag{42}$$

If we ignore isospin violation and take $m_u = m_d$, this expression is particularly simple:

$$\mathcal{L} \supset m_\pi^2 f_\pi^2 \cos(\theta) \tag{43}$$

though you can work out the full result including isospin breaking from the above.

Now the basic idea is simple: if we can introduce a pseudoscalar field a that couples to $G\tilde{G}$ like θ, i.e.,

$$\delta\mathcal{L} = -\frac{g^2}{32\pi^2}\left(\theta + \frac{a}{f_a}\right) G\tilde{G} \tag{44}$$

then the total effective CP violating angle is $\theta + \langle a \rangle / f_a$ and the vacuum energy becomes

$$E(a, \theta) = -m_\pi^2 f_\pi^2 \cos(\theta + a/f_a) \tag{45}$$

This has a minimum at $\langle a \rangle = -\theta f_a$ where the total effective CP violating angle is set to zero, solving the strong CP problem.

We'll see how to accomplish this next lecture.

References

[1] P. A. M. Dirac, "New basis for cosmology," *Proc. Roy. Soc. Lond.* **A165** (1938) 199–208.

[2] M. Gell-Mann, "The interpretation of the new particles as displaced charge multiplets," *Nuovo Cim.* **4** (1956) no. S2, 848–866.

[3] R. J. Crewther, P. Di Vecchia, G. Veneziano, and E. Witten, "Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics," *Phys. Lett.* **88B** (1979) 123. [Erratum: Phys. Lett.91B,487(1980)].

[4] A. E. Nelson, "Naturally Weak CP Violation," *Phys. Lett.* **136B** (1984) 387–391.

[5] S. M. Barr, "Solving the Strong CP Problem Without the Peccei-Quinn Symmetry," *Phys. Rev. Lett.* **53** (1984) 329.

[6] S. M. Barr, D. Chang, and G. Senjanovic, "Strong CP problem and parity," *Phys. Rev. Lett.* **67** (1991) 2765–2768.

[7] Z. G. Berezhiani, R. N. Mohapatra, and G. Senjanovic, "Planck scale physics and solutions to the strong CP problem without axion," *Phys. Rev.* **D47** (1993) 5565–5570, `arXiv:hep-ph/9212318 [hep-ph]`.