Conjecture on the Physical Implications
of the Scale Anomaly

Christopher T. Hill
Fermi National Accelerator Laboratory
P.O. Box 500, Batavia,
Illinois 60510, USA

Invited Talk delivered at the Santa Fe Institute
on the Occasion of the Celebration of the
75th Birthday of Murray Gell-Mann.
July 23, 2005

(Dated: December 8, 2005)

Abstract

Murray Gell-Mann, after co-inventing QCD, recognized the interplay of the scale anomaly, the renormalization group, and the origin of the strong scale, $\Lambda_{QCD}$. I tell a story, then elaborate this concept, and for the sake of discussion, propose a conjecture that the physical world is scale invariant in the classical, $\hbar \to 0$, limit. This principle has implications for the dimensionality of space-time, the cosmological constant, the weak scale, and Planck scale.
I. A STORY

I arrived at Caltech as a beginning graduate student in the Fall of 1972 and immediately beamed myself up to the fourth floor of the Lauritsen Laboratory of High Energy Physics, to see what was going on and to catch a glimpse of the great men, Richard Feynman and Murray Gell-Mann. Feynman had recently attended a meeting in Chicago at which he had broken his knee by tripping on a street curb. He was in a cast, and mostly working at home.

On the first occasion of a seminar in the fall quarter, however, Murray Gell-Mann showed up. He was dressed in a blue suit, smoking a cigar, and began describing the results of his previous sabbatical year spent at CERN. There, working with Bill Bardeen, Harald Fritzsch and Heinrich Leutwyler, he had put together the defining elements of what we now call QCD, the Yang-Mills gauge theory of the strong interactions based upon quark color \[1, 2, 3\]. The theory had a long genealogy, emanating from the idea of O. Greenberg of parastatistics, and the Han-Nambu model. Gell-Mann et al. realized that the simplest assumption of a local gauge theory of a quark color degree of freedom, commuting with electromagnetism, made the most sense. They didn’t call it QCD then, but rather the “color octet vector gluon picture.” They had written down the lagrangian of the \(SU(3)_c\) Yang-Mills theory, had argued about its infrared problems and quark confinement, had counted the number of colors, \(N_c = 3\), from the anomaly governing \(\pi^0\) decay, and understood that something had to fix the short-distance behavior of the theory to make it consistent with the recent observations at SLAC of electroweak scaling, as anticipated by Bjorken.

The SLAC electroproduction data when interpreted in terms of Bj’s brilliant light-cone scaling hypothesis \[4\], represented the first “photographs” of quarks moving within the nucleon at short distances. Feynman had reinterpreted Bj’s hypothesis in the more popular language of the ‘parton’ model \[5\]. Something electrically neutral was seen to be carrying half the momentum of the proton in the infinite momentum frame, and this was the first evidence for glue.

Throughout the fall, 1972, and winter, 1973, at Caltech the color octet vector gluon picture was the main topic of discussion in Gell-Mann’s Wednesday afternoon seminar course. While Gell-Mann, et al., were coming at QCD from the infrared, where quarks are imprisoned within hadrons, the forces are strong and the quarks are “fat” because the chiral symmetries are broken, the SLAC data indicated that short-distance quarks were almost
massless, rattling around like little nuts in a tin can. During that period Murray said, (on more than one occasion):

“Some enterprising graduate student should figure out how to reconcile the color octet vector gluon picture with Feynman’s ‘parton’ model.”

This was, if properly interpreted, the most important homework assignment of the last third of the twentieth century. I do not know why he didn’t encode it slightly more explicitly, such as “go calculate my $\Psi(g)$ function” ... but such is the fog of research. The message was clear, however: if the color octet vector gluon picture was right, something had to happen to make quarks and gluons act like free particles at very high recoil energy.

On the east coast there were some “enterprising graduate students”: David Gross, David Politzer, and Frank Wilczek. I later heard a detailed account, over beers at the Athenaeum from David Politzer, as to what led to the discovery of asymptotic freedom in the spring of 1973. It involved some serendipity and the collective talents of a large number of people, as stated in his 2004 Nobel speech. The roles of Sydney Coleman, Eric Weinberg, and Feynman’s famous article from 1963 on gravitation, in which Yang-Mills perturbation theory is developed, were significant.

Murray, upon hearing of the news of asymptotic freedom, declared that he now considered the color octet vector gluon picture “established.” It was renamed Quantum Chromodynamics (QCD), and he turned to other issues, including a foray into gauge unification, inventing the neutrino mass seesaw mechanism (with Pierre Ramond and Dick Slansky), etc. Supersymmetry would soon arrive on the scene, and it became the exclusive topic of his seminar course for the remainder of my time at Caltech (from sometime in 1975 through 1977). Gell-Mann was an early champion of string theory.

At some point in the time frame of spring 1973 to winter 1975 an interesting event occurred. It happened a little out of the sequence of discussions in his course, and it was on the occasion that Murray lectured, so I would be inclined to date it as no later than the winter of 1975. Present were Harald Fritzsch and Peter Minkowski. I believe David Politzer was visiting, as Caltech was interested in recruiting him, but I am a little foggy on this, and the visitor may actually have been Heinrich Leutwyler. This is the story I want to tell.

Murray that afternoon lectured about an apparent residual “puzzle” in understanding mass in QCD. Here we have “Exhibit A” for something that Coleman and E. Weinberg had
dubbed “dimensional transmutation” [11]. A dimensionless number, $g_0$, which is a definite numerical value of the coupling constant of QCD, is chosen at some arbitrary, very high energy scale, $M_0$. Then, according to the Gell-Mann–Low equation the coupling constant “runs” with scale $\mu$:

$$\frac{d \ln g}{d \ln \mu} = \Psi(g).$$

(1)

Gross, Politzer and Wilczek had shown to leading order that:

$$\Psi(g) = b_0 g^2 \quad \text{where} \quad b_0 = -\frac{1}{16\pi^2} \left( \frac{11}{3} N_c - \frac{2}{3} n_f \right),$$

(2)

where $n_f$ is the number of active quark flavors, e.g., $n_f = 6$ including up, down, ..., top, and $N_c = 3$ colors. Solving the renormalization group equation we find that the running coupling constant, $g(\mu)$ perturbatively blows up at a lower energy scale we call $\Lambda_{QCD}$:

$$\frac{\Lambda_{QCD}}{M_0} = \exp \left( \frac{1}{2b_0 g_0^2} \right).$$

(3)

Thus, a dimensionless quantity, $g_0$, is converted into a physical scale, $\Lambda_{QCD}$. The key point that makes this work, of course, is the fact that $b_0$ is negative and the theory is asymptotically free (note that the situation is reversed in QED, where $b_0$ is positive and the electric charge blows up in the far UV at what we call the “Landau pole”). This is the origin of mass in QCD.

It is important to realize that this has nothing directly to do with other mass scales in nature. This is a point that is often confused by graduate students, who intuitively think that some other scale, like the GUT scale or Planck scale, is somehow mysteriously feeding in to generate the subservient strong scale. In fact, even if there were no GUT scale or Planck scale at all, once we are told by an experimentalist what the (nonzero) value of $g_0$ is at any arbitrary value of energy $M_0$, we must still have the QCD scale. Such is the essential miracle of dimensional transmutation.

Gell-Mann had been concerned with the issue of the mass scale of the strong interactions over the previous 20 years. He had long considered the options for the generation of a mass scale in QCD, and had at one point believed that it was spontaneously generated, with the concomitant formation of a dilaton, the Nambu-Goldstone boson of a spontaneous symmetry breaking. In other words, he was long pondering the fate of scale invariance in QCD.

Any local field theory admits a scale current, $S^\mu$. Given the stress-tensor, $T_{\mu\nu}$, which is always conserved to yield Newton’s equations of motion, $\partial_\mu T^{\mu}_{\nu} = 0$, we can likewise construct
the moment of the stress tensor, \( S_\mu = x_\nu T_{\mu\nu} \). If we then compute the divergence of \( S^\mu \) we see that:

\[
\partial_\mu S^\mu = T^\mu_\mu
\]

the “trace” of the stress tensor. The trace involves the mass terms of the theory, and thus the trace is associated with the breaking of the conservation of the scale current \( S_\mu \). If the theory has no mass parameters, the trace of the stress tensor will be zero and scale symmetry will be an exact symmetry of the theory. If the theory has a mass scale, it represents a special scale in the theory, yields a nonvanishing trace, and the scale invariance is thus broken.

The strong interactions indeed have a mass scale and thus must have a nonzero trace for the stress tensor. The trace, however, could be a dilaton, and \( \partial_\mu S^\mu \sim \Lambda_{QCD} \partial^2 \sigma \). In this case the scale symmetry is really exact in the lagrangian of QCD, yet spontaneously broken by the vacuum. So, how exactly does it work for QCD? How is the scale current related to the dimensional transmutation and renormalization group origin of mass in QCD?

At the blackboard during his lecture Gell-Mann wrote down the classical form of the stress-tensor of a pure Yang-Mills theory,

\[
T_{\mu\nu} = \text{Tr}(G_{\mu\rho}G^\rho_\nu) - \frac{1}{4} g_{\mu\nu} \text{Tr}(G_{\rho\sigma}G^{\rho\sigma})
\]

Upon tracing this we find trivially:

\[
T^\nu_\mu = \text{Tr}(G_{\mu\nu}G^{\mu\nu}) - \frac{4}{4} \text{Tr}(G_{\mu\nu}G^{\mu\nu}) = 0
\]

Thus, classically the trace is zero and the scale current is conserved. Symmetries and their conserved currents are not renormalized by perturbative quantum effects. Pure gluonic QCD would thus appear to be scale invariant, and evidently must therefore contain the dilaton. Yet, no experimental data supports its existence. His lecture was inconclusive, and it ended up in a discussion in the lecture room consisting of Murray and a few of the audience participants, David Politzer, Harald Fritzsch, and Peter Minkowski, remaining in the seminar room discussing the puzzle.

It so happens that the morning of this lecture I had been reading papers of Chanowitz and Ellis [12], dealing with the canonical conformal anomaly. The authors were mainly addressing electrodynamics, and formulating how the conformal anomaly entered \( e^+e^- \) collider physics. The main point, relevant to Murray’s discussion, was that \( T^\mu_\mu \) generally has both a classical part, \( h^{(0)} \equiv 1 \) representing the defining classical input parameters, as well as an
anomalous part beginning at order $\hbar^{(1)}$, coming from quantum loops, which Chanowitz and Ellis had written as $(R/32\pi^2)F_{\mu\nu}F^{\mu\nu}$ (where $R$ is related to the cross-section for $e^+e^-\rightarrow \gamma \rightarrow$ hadrons, in electron collider experiments).

I brought one of the papers into the seminar room and sat down next to Peter Minkowski and listened. At a lull in the discussion, I nudged Peter, who took the paper and looked at it for a moment. Harald then took the paper and began to study it. Murray at this point noticed the commotion and grabbed the paper away from Harald and began examining it. After a minute or two Murray handed the paper back to me and said: “go make a copy of the first page of this paper.”

So I went down the hall to the Xerox machine across from Helen Tuck’s office, and returned momentarily with the copied first page. By the time I was back in the seminar room Murray was at the blackboard and had written the following equation on the board (with considerable input from Peter Minkowski, who may have suggested the $\Psi(g)$ factor):

$$\partial_\mu S^\mu = \Psi(g) \text{Tr}(G_{\mu\nu}G^{\mu\nu}).$$

He declared that the problem was solved, and we now understood the origin of mass in QCD and its fundamental connection to the $\Psi$ function of the renormalization group, and thus dimensional transmutation! For the next few weeks Murray seemed elated. One day he stopped by the doorway of my office, walked in, and said with a grin, “you realize, this is a very interesting result!”

I thought that this new observation would be quickly written up. Harald and Peter continued to discuss it, but Murray got drawn into travel and other pursuits, and it seemed to slip by, remaining only as an interesting observation. It was, of course, a reframing of something we already knew by a different name. I guess I assumed that those smart guys at Princeton and Harvard already completely understood this, and we just let it go. Peter Minkowski eventually wrote a nice (unpublished) article some years later \[13\]. In it he makes arguments similar to the original ones of Symanzik \[14\] making the context of QCD clear. The first published study in the literature focused on QED and obtaining the result of eq.(7) in perturbation theory is that of Adler, et al., in 1977 \[15\]. The main point in all of this is that the scale anomaly is a general statement of the breaking of scale invariance. The Ward identities of the anomalous current divergence yield the renormalization group equations when applied to the renormalized 3-point and 2-point functions of the theory.
Murray Gell-Mann, I now believe, was perhaps the first person to clearly see the full connection to the scale anomaly that, via quantum mechanics, yields the strong interaction scale in the real world. This connection is not well known to the general community to this day. Even well-known and talented theoretical physicists I have met are unaware of, and even resistant to, the fact that the strong scale, most of the mass of the proton, comes mysteriously from quantum mechanics itself.

Therefore, befitting this occasion I would like to expand on this notion. I would like propose an expansive conjecture, even if it is merely tentative and operational, \textit{i.e.}, merely for the sake of discussion – a “talking point.” This is a style of discussion that Murray always advocated (provided it is otherwise sensible). The scale anomaly must surely have deeper and even more profound implications for physics than the remarkable origin of the strong scale. Let’s begin, however, by summarizing how it works for the strong interaction.

II. WHAT DOES IT MEAN?

Anomalies are intrinsic effects of quantum mechanics on the symmetry structure of a field theory.

Quantum mechanics instructs us to relate energy to frequency and momentum to wave-number through $\hbar$. We are also instructed to compute coherent sums of amplitudes, square them, sum over final states, and interpret the result as a probabilistic observable. If we set $\hbar$ to zero, the Feynman tree-diagrams now describe the motion of classical waves of photons and waves of electrons with frequencies and wave-numbers that interact nonlinearly. In the $\hbar = 0$ limit we can in principle directly measure the wave amplitudes and we no longer compute squares of amplitudes, and sum over final states.

Thus, at order $\hbar$ we have loop diagrams, the onset of true quantum effects. Most of the symmetries present at the classical level of a theory carry through into the quantum theory. Anomalies are exceptions. They are particular loop effects that modify the conservation laws of the classical theory. These effects are fundamental and cannot be renormalized away.

Let’s consider momentarily the axial anomalies. For example, a theory of a single Weyl spinor (\textit{e.g.}, a purely left-handed relativistic spinor) coupled to a photon does not exist because the electric current of the spinor will have an anomaly, hence the electric current is not conserved. The left-handed spinor’s current anomaly in the theory $S = \int d^4 x \bar{\psi}_L (i\hat{\phi} - \psi_L)$
\( \gamma_{\mu} A^{\mu} - M_0 \psi_L \) is given in eq.(44) of Bardeen’s paper \[17]:

\[
\partial_{\mu} J^\mu = -\frac{1}{48\pi^2} \epsilon_{\mu\nu\rho\sigma} F^{\mu\nu} \tilde{F}^{\rho\sigma} \tag{8}
\]

where \( J^\mu = \bar{\psi}_L \gamma_\mu \psi_L \) and \( \psi_L = (1 - \gamma^5) \psi_L / 2 \). Current conservation is essential, since the equation of motion of electrodynamics is:

\[
\partial_{\mu} F^{\mu\nu} = e j^\nu. \tag{9}
\]

where \( F^{\mu\nu} \) is the antisymmetric field strength tensor. Owing to the antisymmetry:

\[
\partial^\nu \partial_{\mu} F^{\mu\nu} = 0 = e \partial^\nu j^\nu. \tag{10}
\]

The current must thus be conserved if it is to be the source term for an electromagnetic field. Hence, a single charged Weyl fermion cannot consistently couple to the photon through an electromagnetic current because of the anomaly. The anomaly destroys the symmetry of gauge invariance, the symmetry that leads to current conservation by Noether’s theorem.

There are many ways to solve this problem. Electrodynamics chooses to make the electron a “Dirac particle.” A Dirac fermion is a pair of Weyl spinors for which the anomalies in the vector current cancel between the pair. The anomaly then harmlessly resides in the ungauged axial vector current. The point is that electric charge conservation requires minimally a Dirac fermion, such as the electron in QED.

In general, if a conservation law must be enforced, as in the case of a gauge theory which makes no sense without strictly conserved currents, then anomalies in those current must be absent. We all know how this works for the standard model, where it controls the quark-lepton generation structure seen in nature. Axial anomalies have a fundamental topological significance. The anomaly in even \( D \) is related to the Chern-Simons term of odd \( D + 1 \), since \( D \) can be viewed as the boundary of \( D + 1 \). The Chern-Simons term has a quantized coefficient to maintain gauge invariance in \( D + 1 \), which implies that, apart from the coupling constant renormalization in the anomaly prefactor, the anomaly itself is not renormalized. This is explicitly verified in the \( D \) dimensional perturbation theory, which is the content of the Adler-Bardeen theorem \[18\]. Axial anomalies are thus intrinsically topological objects.

The scale anomaly is the quantum violation of the conservation of the scale current. It is generally not topological (though it can be linked to the axial anomaly in supersymmetric
theories). It encodes the fundamental way in which scale invariance is broken by quantum mechanics. Of course, we can have fully quantum mechanical field theories that remain scale invariant. This requires that there is a special value of \( g \), called \( g^* \) such that \( \Psi(g^*) = 0 \). Then we say that \( g^* \) is a conformal fixed point.

Conformal field theories are of wide-ranging interest throughout physics. As an example, in string theory the Weyl symmetry of the world sheet is a \( D = 2 \) conformal symmetry, and it must be anomaly free since it should not matter how one places coordinates or a metric on the world sheet. The string target space dimensionality, \( i.e., \) the dimensionality of spacetime for consistency with string theory, is selected by this criterion. Moreover, the top quark Yukawa coupling constant in the Standard Model is remarkably close to a nontrivial conformal quasi-fixed point value \[19\]. The asymptotic freedom of QCD implies that the theory evolves toward the conformal fixed point, \( g^* = 0 \), as we rescale into the far ultraviolet.

We are presently interested in the generation of physical mass by the scale anomaly. Let us display the leading factor of \( \hbar \) explicitly, and collect together the Gell-Mann-Low equation:

\[
\frac{d \ln g}{d \ln \mu} = \hbar \Psi(g),
\tag{11}
\]

and the scale anomaly equation:

\[
\partial_\mu S^\mu = \hbar \Psi(g) \text{Tr}(G_{\mu\nu}G^{\mu\nu}).
\tag{12}
\]

Let us now list some observations about this system.

\( i \) The Gell-Mann–Low Renormalization Group is Equivalent to the Scale Anomaly

The Gell-Mann–Low renormalization group is one of the earliest instances of recursion in the scientific literature \[6\]. The equation tells us how the theory continuously morphs from one scale to another, but in a way that is independent of the scale itself, depending only upon the structure of the theory at any given scale, which dictates the form of \( \Psi(g) \), and its coupling constant \( g \). Thus, as we move up or down in energy, the theory makes a copy of itself with a new value of \( g \) in a self-similar fashion.

The scale anomaly and the Gell-Mann–Low equation are essentially equivalent. The scale anomaly may be viewed as slightly more general than the RG equation since the renormalized Ward Identities of the 3-point function can be massaged into the form of the RG equation.
Nonetheless we can reverse the procedure and even “derive” the scale anomaly from the Gell-Mann–Low equation on the back of an envelope if we hand-wave a bit. Consider a scale transformation of the lagrangian:

$$L = -\frac{1}{2g^2} \text{Tr} G_{\mu\nu} G^{\mu\nu}. \quad (13)$$

Treat $g$ as a running coupling constant and let $Q = \int d^3 x S_0$ be the scale charge. If we commute $Q$ with an operator $Y(x)$ we generate an infinitesimal scale transformation $dY/d\ln \lambda$, where $Y \to \lambda^d Y(\lambda x)$. Thus, given that the classical trace of the stress tensor is zero, we have $[Q, \text{Tr} G_{\mu\nu} G^{\mu\nu}] = 0$. This can be preserved at loop level provided we sweep all scale dependent renormalizations into $g^2$. However, we then have that:

$$[Q, L] = \frac{\partial}{\partial \ln \lambda} L = -\frac{\partial}{\partial \ln \lambda} \left( \frac{1}{2g^2} \right) \text{Tr} G_{\mu\nu} G^{\mu\nu} = -2\hbar \Psi(g) L. \quad (14)$$

Passing to canonical normalization the last term is $\hbar \Psi(g) \text{Tr} G_{\mu\nu} G^{\mu\nu}$ which is the usual scale anomaly.

(ii) The Scale Anomaly Naturally Generates Hierarchies

Let’s reemphasize the fact that the mass scale generated in QCD comes from quantum mechanics. It does not “trickle down” from some higher energy scale, such as $M_{\text{Planck}}$. However, we can imagine that there is physics at the Planck scale, e.g., string theory, that actually sets the value of the strong coupling constant. For example, perhaps some enterprising graduate student will one day show that $\alpha_s(M_{\text{Planck}}) \approx 1/4\pi^2 + \ldots$. Then the renormalization group automatically runs the coupling into the infrared and establishes a large hierarchy of $\Lambda_{\text{QCD}}/M_{\text{Planck}}$. If the theory is successful, the correct value of $\Lambda_{\text{QCD}}/M_{\text{Planck}} \approx 10^{-20}$ is predicted by:

$$\frac{\Lambda_{\text{QCD}}}{M_{\text{Planck}}} = \exp \left( -\frac{1}{8\pi \hbar b_0 \alpha_s(M_P)} \right) + \text{(higher order corrections)} \quad (15)$$

It is a stunning aspect of the renormalization group that a twenty order of magnitude hierarchy can naturally be generated by a normal perturbative input parameter, $\alpha_s(M_{\text{Planck}}) \sim 10^{-1}$. In fact, we know of no other way to do it. The detailed explanation of the origin of the hierarchy between the weak scale, $G_F^{-1/2}$, and the Planck scale, remains a mystery. It is tempting to believe that a similar mechanism may underlie the electroweak hierarchy.
The Custodial Symmetry of the Hierarchy is Classical Scale Invariance!

In the 1970’s ‘t Hooft proposed a general rule concerning hierarchies. Namely, when we have a hierarchy of two physical quantities, such as \( a/b \ll 1 \), then in the limit that \( a/b = 0 \) there will always be a “custodial symmetry” that maintains the special value \( a/b = 0 \) to all orders of perturbation theory.

For example, consider the \( e^-\tau \) mass hierarchy in the standard model, \( m_e/m_\tau \sim 10^{-4} \). In the limit that \( m_e/m_\tau = 0 \) we have the \( U(1)_L \times U(1)_R \) chiral symmetry (modulo a harmless anomaly) of the electron. This chiral symmetry is then maintained to all orders of perturbation theory, and the nonzero value of \( m_e \) is not regenerated.

We cannot make \( b_0 \propto -(11N_c - 2n_f) \to 0 \) by tuning integers \( n_f \) and \( N_c \). The issue of taking the limit must actually apply in pure QCD with \( n_f = 0 \), and \( N_c \to 0 \) is then meaningless. Even if we could engineer \( b_0 = 0 \) at order \( \hbar \) we would still have running at order \( \hbar^2 \), etc. The “Banks-Zaks fixed point” attempts to cancel the \( O(\hbar) \) against the \( O(\hbar^2) \) term, but this largely is a model building tool, employed \( e.g. \), in walking technicolor (see \[24\]), but is not a useful mathematical lever for taking the limit. We could take \( \alpha_s(M_P) \to 0 \), the conformal fixed point, but then we lose the interactions of QCD at all scales.

If we would like to have a limit in which the structure and interactions of the theory are intact the only parameter at our disposal to vary remains \( \hbar \). We see that in the classical limit \( \hbar \to 0 \) the hierarchy \( \Lambda_{QCD}/M_{Planck} \to 0 \) (assuming \( M_{Planck} \) is held fixed!)

Thus, we can view the custodial symmetry of the strong hierarchy as a classical limit in which all anomalies are turned off and the scale symmetry is exact.

III. A CONJECTURE

The notion of a classical symmetry as the custodial symmetry of a quantum mechanical hierarchy seems to be arguably at work in QCD. We now wish to amplify this notion, and we begin by suggesting a talking point conjecture:

All mass scales in physics are intrinsically quantum mechanical and associated with scale anomalies. The \( \hbar \to 0 \) limit of nature is exactly scale invariant.

In fact, since the notion of classical scale invariance as the custodial symmetry of the QCD mass scale makes sense, then it may seem absurd that other mass scales be intrinsically
classical.

On the other hand, this limit may make no sense for many theories, including some of our favorites. String or M-theory has a classical input parameter and we would be challenged to find a quantum mechanical origin for this. However, string theory is intrinsically quantum mechanical and the limit $\hbar \to 0$ may be meaningless. In particular, theories that are defined by duality in a fundamental way may not have a meaningful classical limit. In the modern parlance, “duality” refers to physical states that have properties (masses, charges, etc.) that scale as $1/\hbar$, and are usually topological, and often in one-to-one correspondence with states whose properties may scale as 1 or $\hbar$. Thus, magnetic monopoles, whose magnetic charges scale as $1/e\hbar$ are not compatible with the $\hbar \to 0$ limit. However, we’ll argue below that, since these objects are intrinsically topological and are associated with the mass scales of spontaneously broken symmetries, objects like the 't Hooft-Polyakov magnetic monopoles unwind into nothingness as these symmetries are restored in the scale invariant limit.

It isn’t easy to contemplate all of the things that happen if we try to take $\hbar \to 0$ in reality. What we really mean, in a weaker sense, by this conjecture is the statement that:

$$T_\mu^\mu = \mathcal{O}(\hbar)$$

(16)

There are no classical or $\mathcal{O}(1)$ parameters in the trace of the stress tensor in a perturbative power series expansion in $\hbar$. String theory, in a larger sense, intrinsically involving quantum mechanics may be viewed as consistent with this hypothesis.

Let us now briefly summarize the naive consequences of this conjecture. The arguments will all be hand-waving, and there would be much more to do to develop this hypothesis beyond what I will present in this brief discussion.

(i) A selection rule explaining why we live in $D = 4$?

Consider any one of the Yang-Mills elements of the standard model, e.g., QCD, in $D$ dimensions. The classical stress tensor has the same form, but taking the trace yields:

$$T_\mu = \text{Tr} G_{\mu\nu} G^{\mu\nu} - D \frac{3}{4} \text{Tr} G_{\mu\nu} G^{\mu\nu}.$$  

(17)

This must be zero by hypothesis, hence $D = 4$. 

12
It is well-known that the couplings constants in $D \neq 4$ carry mass dimension. The Gell-Mann–Low equation is modified, having the classical part:

$$\frac{d \ln g}{d \ln \mu} = \frac{1}{2}(D - 4) + \hbar \Psi(g)$$

The first term on the rhs leads to power-law running of the coupling constant. Our conjecture, however, provides a simple selection rule for the dimensionality of space-time. The classical part of the above equation must vanish. And, it agrees with experiment.

Does this imply that extra compact dimensions cannot exist? They can, but they would have to be quantum mechanical in origin and would have an effective radius associated with the scale provided by a scale anomaly. We do know a simple way to construct fake effective extra dimensions in this manner. This can be done by way of “deconstruction,” [21], which is a lattice of QCD-like theories wired together in a sequence so that an effective tower of KK-modes appears in the spectrum. It is really none other than a Wilson style lattice description of an extra dimension, where the Wilson links are chiral fields coming from QCD-like condensates. Such a system could be consistent with our conjecture, being classically scale invariant. In the $\hbar \to 0$ limit the theory would fall apart into a large number of classical massless Yang-Mills theories. In fact, since there is no geometrical principle at work here, the renormalization group itself must serve to define the lattice of an effective quantum extra dimension. This has solutions that are geometrical, but also fractal ones as well [22], where the number distributions of KK-modes with energy can have a nonintegral exponent:

$$N(E) = \left(\frac{E}{M}\right)^{h(D-4)}$$

This bears some resemblance to QCD with $(D - 4)/2 \sim \Psi(g)$ and $\Lambda_{QCD} \sim M$, since it is the pattern structure of a renormalization group solution. [The “renormalization group as a substitute for geometry” seems to me to be yet another hypothesis worthy of further consideration elsewhere (see [22]).]

(ii) Infrared cancellations, topology and tunneling.

It is required that the probabilisitic quantum measurement theory no longer hold in the $\hbar = 0$ limit. Dispersion relations relate probabilistic observables to (imaginary parts of) loop diagrams, hence as $\hbar \to 0$ so too go the probabilistic observables. Moreover, familiar
infrared cancellations in QED between collinear emission tree amplitudes, which are $O(\hbar^0)$ in amplitude, are known to produce $O(\hbar^1)$ singularities in the integral over phase space. In the full quantum theory these cancel against infrared $O(\hbar^1)$ loops \[23\]. Thus, we may view the $\hbar \rightarrow 0$ limit, which eliminates the loops, as also forcing us to abandon the probabalistic interpretation of quantum mechanics with the sum over final states. This isn’t surprising since the classical theory selects a unique outgoing state for any particular incoming state. Any singularities are now artifacts of the quantum calculational method and no longer part of the physics. The sum of tree diagrams yield the perturbation series of the Fredholm theory of the nonlinear wave theory for any particular incoming classical state.

Note that WKB tunneling is also suppressed, faster than any power, as $\hbar \rightarrow 0$. Thus instantons (tunneling) never occur in the classical theory. As mentioned above, states involving duality, typically topological objects like monopoles, skyrmions, disappear in the $\hbar \rightarrow 0$ limit as a consequence of scale invariance. These typically involve the topological current, $\epsilon_{\mu\nu\rho\sigma} \text{Tr}(U^\dagger \partial^\nu UU^\dagger \partial^\rhoUU^\dagger \partial^\sigma U)$ where $U^\dagger U = 1$. Without the latter unitary constraint there is no conserved topological current, and in most cases, such as monopoles and skyrmions, this arises from a nonlinear sigma model in which a field satisfies $\Phi^\dagger \Phi = v^2$. Thus in the $\hbar \rightarrow 0$ limit, $v \rightarrow 0$, and therefore a scale invariant world won’t contain as much topology. The dual toological states furthermore help to lock scale invariance to the $\hbar \rightarrow 0$ limit, enforcing our conjecture.

Of course, these are naive considerations. As we’ve noted, in theories in which a symmetry, such as the inversion symmetry of $M$-theory, is fundamental and maps dual states into anti-duals, etc., the $\hbar \rightarrow 0$ limit may be meaningless.

\[ (iii) \text{ The cosmological constant is classically zero.} \]

The cosmological constant is a term in the stress tensor of the form $T_{\mu\nu} = \Lambda g_{\mu\nu}$. Our conjecture requires $T^\mu_\mu = 0$, hence $\Lambda = 0$. This does not, of course, solve the problem of the cosmological constant, but it eliminates the usual classical aspect of it. The situation is akin to supersymmetry, in which exact SUSY also implies vanishing vacuum energy.

One thing is fairly clear: whatever zeroes $\Lambda$, leaving a possible ultra-small residual component with $\Omega \approx 0.6$, must be mechanism that can be understood within the context of any given scale anomaly. Thus, for example, we must be able to understand the zeroing
mechanism within the context of QCD alone.

Indeed, the scale anomaly of QCD will produce a cosmological constant as:

\[
\Lambda \sim \frac{1}{4} \langle 0 | \Psi(g) \text{Tr}(G^{\mu\nu}G_{\mu\nu}) | 0 \rangle \sim \Lambda_{QCD}^4.
\]  

One potential remedy for this would be the presence of an dilaton, and we return to this below.

(iv.a) The QCD scale is generated by quantum mechanics.

This is the basis of the conjecture.

(iv.b) The conjecture may be testable in the weak interactions

The original Weinberg-Susskind idea of Technicolor was largely motivated by the naturalness of the QCD hierarchy. It would produce the weak scale by way of a scale anomaly in the manner of QCD. Something like Technicolor would confirm the conjecture.

The first forays into this, i.e., the various incarnations of Technicolor in which fermion masses are treated as small, have largely failed. Technicolor models that do not reckon with the heaviness of the top quark (and even the charm and bottom quarks) are largely ruled out. Conversely, models that treat the top quark as part of the dynamics have enjoyed some success, and are still viable [24]. It is certainly not true that these models are ruled out, and they are in principle completely compatible with SUSY. Indeed, most recent efforts to ameliorate the fine tuning problems inherent in the MSSM are pushing that theory in a less perturbative direction with more emphasis on the third generation. Moreover, SUSY breaking is generally approached as a dynamical mechanism, consistent with the trace anomaly conjecture.

There are four classes of viable dynamical weak scale models and some are testable soon (possibly at Tevatron, surely at LHC). One is to combine Technicolor with a dynamics (Topcolor) that can accommodate the third generation heavy masses. This is built out of precursory models of Eichten and Lane in which multi-scales of condensates are involved. These models are thus complex [25]. These models offer lots of targets to experiment. Perhaps most intriguing would be something like a b-tagged dijet mass excess in the \( W + 2j \) channel at the Tevatron. Since Run I we have seen a modest excess there, but nothing is
yet conclusive.

My personal favorite model is the Top Seesaw \[20\]. This is motivated by the fact that if the top quark weighed in at 600 or 700 GeV, then there would be absolutely no doubt that we have a natural strong dynamics at the weak scale, and we would expect dimensional transmutation generated it. In fact, we can easily engineer such a model using Topcolor. Chivukula, Dobrescu, Georgi and I considered implementing the Gell-Mann–Ramond–Slansky seesaw \[10\], which adds new ingredients that can rotate the physical mass down to its 175 GeV observed value. This then lifts other objects (the \(\chi\) quarks) up to the few TeV scale. The additional ingredients must all have dynamically generated scales, but this appears possible and yields various heavy PNGB’s.

The Top Seesaw was preprinted in 1998 and was DoA (Dead on Arrival) as the LEP S-T error ellipse at that time was disconcordant with the theory by about 5-\(\sigma\). To our delight, in 1999 LEP recalibrated its beam energy, and the error ellipse lurched to the “northwest” on the plot. The Top Seesaw was thus ruled in at the 2-\(\sigma\) level. I like to say that we predicted this, but pundits say that we fine-tuned our theory. This strikes me as slightly acausal.

Little Higgs theories provide another interesting approach that attemptsto treat the Higgs boson as a Nambu-Goldstone boson. Here the strong dynamics occurs at \(\sim 10\) or 20 TeV, making a chiral condensate, leading to “mesons” (Nambu-Goldstone bosons) that are typically weak singlets (like \(\eta\)), doublets (like \(K\)) and vectors (like \(\pi\)) \[27\]. Then at the \(\sim 1\) TeV scale there is top dynamics, similar to the Top Seesaw in structure that generates a Coleman-Weinberg potential in which \(K\) develops a VEV and becomes the Higgs boson. These models are thus semi-natural, postponing the new strong dynamics upward in energy by an order of magnitude. They are challenging to engineer because one must forbid an induced \(\pi^0\) VeV (T-parameter constraint) and still have the global minimum of the Coleman-Weinberg potential at \(v_{\text{weak}} \ll 1\) TeV. Evidently only larger chiral symmetries with certain tensor representations suffice (what Georgi refers to as “kissing mexican hats”).

An intriguing renaissance of dynamical models using the neutrino sector, largely due to Appelquist and Shrock, is also quite appealing \[28\]. The role of neutrino mass in dynamical models is an interesting contemporary issue. Thus the weak scale may involve some interesting new dynamical physics.

(v) Scale invariant gravity in the \(h \rightarrow 0\) limit?
Here I will briefly mention an older, non-stringy approach to gravity, which is largely motivated by analogy to QCD. This attempts to directly derive the Planck scale in a manner similar to the strong QCD scale. Whether this is string theory in some kind of disguise is unclear, if not dubious. It in fact it suggests that the continuum of space-time proceeds to much shorter distances than the Planck scale. There is, however, a dramatic phase transition in physics within these theories at the Planck scale.

Indeed, there is an extensive literature of “quadratic gravity,” beginning with the conformally-invariant gravity of Weyl [29], and subsequent incarnations due to Adler [30], Mansouri [31], Tomboulis [32], et al., as well as detailed analyses of renormalizability by Stelle [33], et al. These seem optimally related to our present conjecture, though much of the interest in this venue halted with the rise of string theory in the early 1980’s.

The key question is how the $\sqrt{-g}M_{Planck}^2 R$ term could be generated dynamically by a quantum scale anomaly? Classical scale invariance is essentially the logic underlying some of the aforementioned work, e.g., in particular Adler’s formulation [30]. He implements the classical scale invariance by choosing dimensional regularization as a definition of quantum loops. One can sketch a scenario based upon these papers, in particular the nice work of Tomboulis [32].

The method here is to imitate QCD. Since in QCD, at high energies or in the $\hbar \to 0$ limit, we have scale invariance, therefore by analogy we seek a starting point for pure gravity that is scale invariant at high energies. Such a theory can be built of scale invariant terms that define “quadratic gravity,”

$$\frac{1}{h_1^2} \sqrt{-g} R^2 + \frac{1}{h_2^2} \sqrt{-g} R_{\mu\nu} R^{\mu\nu} + \frac{1}{h_3^2} \sqrt{-g} R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$$

(21)

The $\sqrt{-g} R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$ can be written in terms of the first two operators if we set the Gauss-Bonnet invariant to zero. The Gauss-Bonnet term is a topological index that takes on discrete values for metrics that are not continuously deformable to flat-space, (much like the Pontryagin index, $\text{Tr} \tilde{G} \tilde{G}$ of Yang-Mills). For simplicity we do not include it. Moreover, if one demands an identically zero trace for the classical gravitational stress-tensor, (derived by differentiating eq.(21) wrt $g_{\mu\nu}$), then one is led to a unique lagrangian which involves the Weyl tensor, $\sqrt{-g} C_{\alpha\beta\gamma\delta} C^{\alpha\beta\gamma\delta}$. This can be written, following Tomboulis [32], as:

$$\frac{1}{h^2} \sqrt{-g} (R_{\mu\nu} R^{\mu\nu} - \frac{1}{3} R^2)$$

(22)
This is taken to be our high energy theory, at scales well above the Planck mass. There is no $\sqrt{-g} M^2_P R$ term yet, as such will be generated in analogy to the QCD mass scale through a scale anomaly.

Now, there are immediate apparent problems at the outset with this theory. If we view the metric as the fundamental degree of freedom, this Lagrangian has quartic derivatives, and the graviton propagator has the form, $\sim 1/p^4$. Such a quartic propagator can be viewed as the limit:

$$\lim_{m^2 \to 0} \frac{1}{m^2} \left( \frac{1}{p^2} - \frac{1}{p^2 + m^2} \right)$$

The second term on the rhs has a negative residue. Thus the theory can be considered as the limit of a massless graviton and a massive ghost field as the ghost mass becomes small. The addition of the $M^2_P \sqrt{-g} R$ term would modify the propagator as $\sim 1/(p^2 - p^4/M^2_P)$, thus lifting the ghost mass, but not changing the negative norm of the ghost states. Ghosts, of course, spoil unitarity if they are asymptotic in- or out- states in the $S$-matrix. Thus, our high energy gravity theory as a quantum theory is somewhat sick.

This may be a harbinger of other nonperturbative sources of unitarity violation in gravity. For example, the creation by quantum fluctuations in the vacuum, e.g., the formation of mini-black-holes is also a putative unitarity problem for gravity. Tomboulis, exploiting old ideas of Lee and Wick, argues that this problem can be ameliorated if the massive ghost is sufficiently unstable that it doesn’t really enter asymptotic in or out states \[32\]. However, the problem is then swapped for fluctuations in causality at the Planck scale.

Irrespective of ghosts, it has been shown by several authors that this theory is renormalizable \[33\]! The $1/p^4$ sufficiently softens the UV structure of Feynman amplitudes and the scale symmetry, broken only by the trace anomaly, enforces a multiplicative renormalization of the action.

Tomboulis has proposed an intriguing idea that simply imitates the renormalization group behavior of the theory in a manner similar to QCD. The idea is to incorporate matter, in particular a large number $N$ of Weyl fermions, added to the lagrangian,

$$\sqrt{-g} \sum_i^N \bar{\psi}_i D \psi_i$$

where $D$ is a gravitational (or otherwise) covariant derivative, which is implemented in the vierbein formalism. In the large $N$ (fermion bubble approximation) limit we can obtain the
Gell-Mann–Low equation for the gravitational parameter, $h$. From Tomboulis we have \[32\]:

$$
\frac{d \ln h}{d \ln \mu} = b_0 h^2 \quad b_0 = -\frac{hN}{160\pi^2}
$$

(25)

Remarkably, fermion loops cause the theory to be asymptotically free, an essential completion of the analogy with QCD. The RG running implies that there will thus be a scale anomaly, which we can likewise infer from Tomboulis’ work:

$$
\partial_\mu S^\mu = \frac{hNh^2}{160\pi^2}(R_{\mu\nu}R^{\mu\nu} - \frac{1}{3}R^2)
$$

(26)

(here we have passed to canonical normalization of the graviton; I do not know what the $\psi$ function for the pure gravity part of the theory is; this theory may always be asymptotically free due to the attractive nature of gravitation).

This picture is quite QCD-like. It implies that there will be a scale, generated by dimensional transmutation, at which $h$ blows up, given a measurement of $h = h_0$ at some arbitrary higher energy scale $M_0$:

$$
\frac{\Lambda_{Planck}}{M_0} = \exp(-160\pi^2/Nh_0^2)
$$

(27)

The dynamics of the theory is thereby modified at the scale $\Lambda_{Planck}$. Though a leap of faith, one naturally expects that normal gravity is generated at this scale. I say leap of faith, because the matching may be complicated, and one would expect some analogue vestige of quark-gluon confinement to occur – the low energy graviton may be a boundstate of confined high-energy degrees of freedom. Adler \[30\] gives explicit formulae for the induced cosmological constant and dynamically generated Planck mass. Naturally, we expect $M_{planck} \sim \Lambda_{Planck}$ and $\Lambda_c \sim \Lambda_{Planck}$ in this scheme.

Perhaps one should not be a priori bothered by the unitarity issues, as they may in fact be subtle effects of quantum gravitation that have not been seen by low energy experiment. The issue is in part cosmological. It is not hard to see that Weyl gravity has a scale invariant Friedman-Robertson-Walker solution in a scale invariant radiation dominated universe (as would be the case for $N$ Tomboulis fermions). The Hubble constant scales as $H = \dot{a}/a \sim 1/t$. This is presumably the solution at high energies, above the phase transition into normal Einstein gravity at which $M_{Planck}$ forms. Thus, it would be of interest to revisit the unitarity issue, i.e., the pole structure about this or other background geometries.

Perhaps another way to circumvent the problem of unitarity violation is to view this as a theory in which the Christoffel symbol, rather than the metric, is the dynamical degree
of freedom. Mansouri proposed [31] that the Christoffel symbol can be viewed as the Yang-Mills gauge field of a local GL(4,R) symmetry. The Christoffel symbol is related to the metric in the usual way. The relationship is not a dual relationship, but does bear slight resemblance to the relationship of the axion to the Kalb–Ramond field. One path might be to interpret $\sqrt{-g(x)}$ as the trace of a Wilson line that parallel transports a boundary metric $g_{\mu\nu}(0)$ from 0 to $x$ using the Christoffel symbol, $g_{\mu\nu}(x) \sim \exp(\int_0^x \Gamma^\mu_{\nu\rho} dx^\rho)$. This allows one to write an action that depends only upon $\Gamma^\mu_{\nu\rho}$, but has become a nonlocal theory of Wilson lines coupled to local fields. Perhaps this somehow revisits string theory and holography. It would be interesting if we could interpret the high energy theory as purely affine, and that there is no fundamental metric. Then gravity may be “emergent.”

I have selected this scenario to illustrate a gravitational mimic of QCD, and have foregone addressing the very large number of residual issues. I have no idea if this picture can be reconciled with string theory in some fundamental way to argue that the classical string scale is also emergent. It seems that it has some interesting things to say, and is worthy of further elaboration. various questions arise, e.g., can inflation occur at scales above $M_{\text{Planck}}$?

(vi) The dilaton?

We see the characteristic problem of a cosmological constant emerging in any strong dynamics when $\hbar \neq 0$ [34]. It seems to me that the only resolution to this must involve the existence of an additional degree of freedom, a dilaton, $\sigma$, which participates in the scale anomaly itself:

$$\partial_\mu S^\mu = (\text{all scale anomalies}) - f_D D^2 \sigma - V'(\sigma) \quad (28)$$

This is analogous to the axion in the axial anomaly:

$$\partial_\mu A^\mu = (\text{all axial anomalies}) - f_a \partial a - V'(a) \quad (29)$$

Such a dilaton implies that the world is truly scale invariant when $\hbar \neq 0$. It in some sense implies that $\hbar \neq 0$ is a spontaneous breaking of a classical world.

Such a dilaton can be introduced “evanescently” by way of dimensional regularization by incorporation of factors in loops like $\exp(\epsilon \sigma / f_D)$. This maintains quantum scale invariance as one goes to $\epsilon = D - 4$ dimensions. This will generate an equation such as eq.(28) with an induced kinetic term for the dilaton as well. The “evanescent” dilaton is purely quantum
mechanical, associated only with loops and its couplings absent in the classical theory. It may represent local small changes in space-time dimensionality, and even lead to a dynamical interpretation of $\hbar$ itself.

The point is that gravity is supposed to see the entire rhs or eq. (28) (while, e.g., the proton only sees the QCD scale anomaly from which it gets its fixed mass). The rhs of eq. (28) is now zero by the dilaton equation of motion. By analogy to the case of axions, eq. (29) governs the putative physics of an axion detector. The detector produces the $F \tilde{F}$ source term, and radiates energy away into axions. In the case of the dilaton, we would hope that this provides a relaxation mechanism for the cosmological constant, perhaps even a small relic one if the dilaton is slightly off mass-shell in the universe today.

The idea of using a massless dilaton to eliminate $\Lambda_{\text{cosmological}}$ can only make sense in a world in which we can take $f_D >> M_{\text{Planck}}$. Otherwise, in the case that $f_D/M_{\text{Planck}} \leq 10^{-3}$ the dilaton produces long-range scalar corrections to gravity that can be ruled out by experiment limits. The existence of scales beyond $M_{\text{Planck}}$ removes this hurdle to incorporating a dilaton.

(vii) Softly broken classical scale invariance

Much of this discussion was motivated by a comment a decade ago of Bill Bardeen who pointed out that, in the Standard Model, the Higgs boson mass can be viewed as a “soft classical scale symmetry” breaking parameter [35]. This means that, if the Higgs mass were the only physical scale near to and above the weak scale, then it is technically natural. This is not the case for the usual Standard Model, which has the $U(1)$ Landau pole, but one could imagine a modified Standard Model in which it holds modulo gravity.

Bardeen had in mind scale breaking only by way of logarithmic effects in loop integrals, which lead to the scale anomaly. This was refutation of something we hear often in conference talks about SUSY. Namely, it is often stated that “the Standard Model Higgs mass is subject to additive quadratic divergences in field theory, and is therefore destabilized and unnatural.” Bardeen points out that technically the only source of quantum scale breaking is the scale anomaly, and this is a $d = 4$ operator, and cannot include the $d = 2$ Higgs mass term. The Higgs mass term is a classical input to the standard model, and thus can be viewed as a soft classical scale breaking term, provided we somehow switch off any Landau poles.
Thus, any additive quadratic divergences are an artifact of the choice of a momentum space cut-off, or a Pauli-Villars regulation procedure. One sees this happen in any regulator that violates classical scale invariance. However, if one truly implemented classical scale invariance as a limiting symmetry of the theory, with the Higgs mass as a soft breaking parameter, such effects would be removed by the Ward identities of the theory.

Put another way, if we were to consider massive scalar electrodynamics, and we compute with a regulator that respects classical scale invariance (e.g., dimensional regularization, which may be the only one) then we find that the scalar mass term is multiplicatively renormalized. Essentially, classical scale invariance can provide, in principle, the same degree of technical naturalness that SUSY does for the weak scale.

IV. CONCLUSIONS

We have argued that, for the entire physical world, $T_{\mu}^{\mu} = O(h)$, and has no surviving “classical” term as $h \to 0$. As a principle, this would be a powerful constraint on the world. There is an appealing naturalness to the idea of all mass arising from scale anomalies, a purely quantum mechanical origin, that is satisfying. QCD may be sufficient evidence of this principle. String theory may, through its intrinsic duality, provide an exceptional and nontrivial realization of the principle.

Murray Gell-Mann, to my reckoning, was the first person to close the loop and appreciate how the fundamental scale of the strong interactions is intimately tied to the the scale anomaly, the renormalization group and quantum mechanics. He was also a great teacher and mentor. He made Caltech’s HEP theory group a stimulating place in which to work and study. We are here, in the spectacular surrounds of the Santa Fe Institute, for the celebration of his 75th birthday, and to salute him. To this, may the lights of Babylon burn brightly for many years to come!

I thank Steve Adler, Bill Bardeen and Peter Minkowski for useful discussions.


[6] Gell-Mann and Low in their remarkable paper, M. Gell-Mann and F. E. Low, “Quantum Electrodynamics At Small Distances,” Phys. Rev. 95, 1300 (1954), invented the renormalization group. They defined \( d\ln g/d\ln \mu = \Psi(g) \). In common usage is, \( dg/d\ln(\mu) = \beta(g) \) where \( \beta(g) = g\Psi(g) \). In honor of Murray I will use the original Gell-Mann–Low notation, \( \Psi(g) \).


Typically, under large gauge transformations the Chern-Simons (CS) term in the action shifts by an integer. This integer must be a multiple of $2\pi$ if the path integral is to be invariant. For example, in $1 + 4$ dimensions, the CS term takes the form $\text{Tr}(AdAdA + \ldots)$. One can consider a static independent instanton configuration (an “instantonic soliton” C. T. Hill and P. Ramond, Nucl. Phys. B 596, 243 (2001)) in an $SU(3)$ Yang-Mills theory with $A_0 = 0$ at $t = -\infty$. One then performs a gauge transformation that takes $A_0 = 2\pi$ at $t = \infty$, and the CS term shifts by $96\pi^3$. Thus, the CS term requires a coefficient of $1/48\pi^2$.

W. A. Bardeen, Phys. Rev. 184, 1848 (1969). This anomaly is unambiguous, i.e., it has no counterterm ambiguities. The more common form quoted for the axial anomaly in QED, eq.(49) of this paper, is obtained by the addition of a counterterm (a Chern-Simons term) to make the vector current conserved. This changes the anomaly coefficient. These two forms of the anomaly are often confused in the literature.

S. L. Adler and W. A. Bardeen, “Absence Of Higher Order Corrections In The Anomalous Axial Vector Divergence Equation,” Phys. Rev. 182, 1517 (1969). The axial anomaly can arise in a purely bosonic theory. We can start in $D = 5$ with pure Yang-Mills and a Chern-Simons term. Then, upon orbifold compactification, $A_5$ becomes a “meson” field, represented by the Wilson line $\exp(i \int dx^5 A_5) \sim \exp(i \tilde{\pi})$. The CS term morphs into the Witten-Wess-Zumino term and expresses the anomalies in the chiral currents of the mesons. Thus, anomalies in even $D$ are really controlled by the bosonic topology in $D + 1$ contained in the Chern-Simons term, and no reference need be made to fermions. See, e.g., C. T. Hill and C. K. Zachos, “Dimensional deconstruction and Wess-Zumino-Witten terms,” Phys. Rev. D 71, 046002 (2005).


G. ’t Hooft, “Naturalness, Chiral Symmetry, And Spontaneous Chiral Symmetry Breaking,” PRINT-80-0083 (UTRECHT) Lecture given at Cargese Summer Inst., Cargese, France, Aug 26 - Sep 8, 1979

Lett. **86**, 4757 (2001);


