Scaling Mount Planck I: A View from the Bottom

Frank Wilczek

Gravity dominates the large-scale structure of the universe, but only by default, so to speak. Matter arranges itself to cancel electromagnetism, and the strong and weak forces are intrinsically short range. At a more fundamental level, gravity is extravagantly feeble. Acting between protons, gravitational attraction is about 10^{-36} times weaker than electrical repulsion. Where does this outlandish disparity come from? What does it mean?

These questions greatly disturbed Richard Feynman. His famous paper on quantizing general relativity,[1] in which he first described his discovery of the "ghost particles" that eventually played a crucial role in understanding modern gauge field theories, begins with a discussion of the smallness of gravitational effects on subatomic scales, after which he concludes,

> There's a certain irrationality to any work on [quantum] gravitation, so it's hard to explain why you do any of it. . . . It is therefore clear that the problem we [are] working on is not the correct problem; the correct problem is: What determines the size of gravitation?

The same question drove Paul Dirac[2] to consider the radical idea that the fundamental "constants" of nature are time dependent, so that the weakness of gravity could be related to the great age of the universe, through the following numerology: The observed expansion rate of the universe suggests that it began with a bang roughly 10^{17} seconds ago. On the other hand, the time it takes light to traverse the diameter of a proton is roughly 10^{-24} seconds. Squinting through rose-colored glasses, we can see that the ratio, 10^{-41}, is not so far from our mysterious 10^{-36}. (For what it's worth, the numbers agree better if we compare gravitational attraction versus electrical repulsion for electrons, instead of protons.) But the age of the universe, of course, changes with time. So if the numerological coincidence is to abide, something else—the relative strength of gravity, or the size of protons—will have to change in proportion. There are powerful experimental constraints on such effects, and Dirac's idea is not easy to reconcile with our standard modern theories of cosmology and fundamental interactions, which are tremendously successful.

In this column, I show that today it is natural to see the problem of why gravity is extravagantly feeble in a new way—upside down and through a distorting lens compared to its superficial appearance. When viewed this way, the feebleness of gravity comes to seem much less enigmatic. In a sequel, I'll make a case that we're getting close to understanding it.

First let's quantify the problem. The mass of ordinary matter is dominated by protons (and neutrons), and the force of gravity is proportional to mass squared. Using Newton's constant, the proton mass, and fundamental constants, we can form the pure dimensionless number

$$N = G_{\mathrm{N}} m_{\mathrm{p}}^2 / \hbar c,$$

where G_{N} is Newton's constant, m_{p} is the proton mass, $\hbar$ is Planck's constant, and c is the speed of light. Substituting the measured values, we obtain

$$N \approx 3 \times 10^{-39}.$$

This is what we mean, quantitatively, when we say that gravity is extravagantly feeble.

We can interpret N directly in physical terms, too. Since the proton's geometrical size R is roughly the same as its Compton radius, $\hbar/m_{\mathrm{p}}c$, the gravitational binding energy of a proton is roughly $G_{\mathrm{N}} m_{\mathrm{p}}^2/R \approx N m_{\mathrm{p}} c^2$. So N is the fractional contribution of gravitational binding energy to the proton's rest mass!

Soon after Max Planck introduced his constant $\hbar$ in the course of a phenomenological fit to the blackbody radiation spectrum, he pointed out the possibility[3] of building a system of units based on the three fundamental constants $\hbar$, c, and G_{N}. Indeed, from these three we can define a unit of mass $(\hbar c/G_{\mathrm{N}})^{1/2}$, a unit of length $(\hbar G_{\mathrm{N}}/c^3)^{1/2}$, and a unit of time $(\hbar G_{\mathrm{N}}/c^5)^{1/2}$—what we now call the Planck mass, length, and time, respectively. Planck's proposal for a system of units based on fundamental physical constants was, when it was made, formally correct but rather thinly rooted in fundamental physics. Over the course of the 20th century, however, his proposal became compelling. Now there are profound reasons to regard c as the fundamental unit of velocity and $\hbar$ as the fundamental unit of action. In the special theory of relativity, there are symmetries relating space and time—and c serves as a conversion factor between the units in which space intervals and time intervals are measured. In quantum theory, the energy of a state is proportional to the frequency of its oscillations—and $\hbar$ is the conversion factor. Thus c and $\hbar$ appear directly as primary units of measurement in the basic laws of these two great theories. Finally, in general relativity theory, spacetime curvature is proportional to the density of energy—and G_{N} (actually $1/G_{\mathrm{N}} c^4$) is the conversion factor.

If we accept that G_{N} is a primary quantity, together with $\hbar$ and c, then the enigma of N's smallness looks quite different. We see that the question it poses is not, "Why is gravity so feeble?" but rather, "Why is the proton's mass so small?" For in natural (Planck) units, the strength of gravity simply is what it is, a primary quantity, while the proton's mass is the tiny number $\sqrt{N}$.

That's a provocative and fruitful way to invert the question, because we've attained quite a deep understanding of the origin of the proton's mass, as I discussed in an earlier col-

Frank Wilczek *is the Herman Feshbach Professor of Physics at the Massachusetts Institute of Technology in Cambridge, Massachusetts.*

umn (PHYSICS TODAY, November 1999, page 11). The lion's share of the proton's mass can be accounted for in an approximation to quantum chromodynamics (QCD), where all the relevant particles—gluons, and up and down quarks—are taken to be massless. In that earlier column, I discussed this in conceptual terms; now let's look under the hood.

The key dynamical phenomenon is the running of the coupling (see my article, "QCD Made Simple," PHYSICS TODAY, August 2000, page 22). Looking at the classical equations of QCD, one would expect an attractive force between quarks that varies with the distance as g^2/r^2, where g is the coupling constant. This result is modified, however, by the effects of quantum fluctuations. The omnipresent evanescence of virtual particles renders empty space into a dynamical medium, whose response alters the force law.

In QCD, the antiscreening effect of virtual color gluons (asymptotic freedom) enhances the strength of the attraction by a factor that grows with the distance. This effect can be captured by defining an effective coupling, $g(r)$, that grows with distance. The attractive interaction among quarks wants to bind them together, but the potential energy to be gained by bringing quarks together must be weighed against its cost in kinetic energy. In a more familiar application, just this sort of competition between Coulomb attraction and localization energy is responsible for the stability and finite size of atoms.[4] Here, quantum-mechanical uncertainty implies that quark wavefunctions localized in space must contain a substantial admixture of high momentum, which translates directly, for a relativistic particle, into energy. If the attraction followed Coulomb's law, with a small coupling, the energetic price for staying localized would always outweigh the profit from attraction, and the quarks would not form a bound state. But the running coupling of QCD grows with distance, and that tips the balance. The quarks finally get reined in at distances where $g(r)$ becomes large.

The mechanism leading to this binding dynamics explains the "coincidence," noted above, that the geometric size of the proton is close to its Compton radius. That is because a substantial portion of its formation energy, from which its mass arises as $m_p = E/c^2$, is associated with the momenta of order $p \approx \hbar/R$, required for the quarks' localization, through $E = pc$. Simple algebra then yields $R \approx \hbar/m_p c$. But really no detailed calculation was required to reach this conclusion. Since only the broadest principles of special relativity and quantum mechanics come into the dynamics, the relationship follows by dimensional analysis.

Thus the proton mass is determined by the distance at which the running QCD coupling becomes strong. Let's call this the QCD distance. Our question, "Why is the proton mass so small?" has been transformed into the question, "Why is the QCD distance much larger than the Planck length?" To close our circle of ideas, we need to explain how, if only the Planck length is truly fundamental, this vastly different length can arise naturally.

This last elucidation, profound and beautiful, is worthy of the problem. It has to do with how the coupling runs. When the QCD coupling is weak, "running" is a bit of a misnomer. Actually the coupling creeps along, like a wounded snail. To be precise (and we can in fact calculate the behavior precisely, following the rules of quantum field theory, and even check it out experimentally[5]), the inverse coupling varies logarithmically with distance. As a result, the distance will need to change by many orders of magnitude for a moderately weak coupling to evolve into a strong one. So, finally, all we require to generate our large QCD distance dynamically is that, at the Planck length, the QCD coupling is moderately small (between a third and a half of what it is observed to be at 10^{-15} cm). From this modest and innocuous starting point, by following our logical flow upstream, we arrive at the tiny value of N, which at first sight seemed so absurd.

I've explained how the ridiculously feeble appearance of gravity is consistent with the idea that this force sets the scale for a fundamental theory of nature. But does it? Stay tuned.

References

1. R. P. Feynman, *Acta Physica Polonica* **24**, 697 (1963).
2. P. A. M. Dirac, *Proc. R. Soc. London* **A165**, 199 (1939).
3. M. Planck, *Sitzungsber. Dtsch. Akad. Wiss. Berlin, Math-Phys. Tech. Kl.*, 440 (1899).
4. R. P. Feynman, R. B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. 1, Addison-Wesley, Reading, Mass. (1963), p. 38–4.
5. Summarized in *The Review of Particle Physics*, D. Groom et al. (Particle Data Group), European Physical Journal **C15**, 1 (2000), "Quantum Chromodynamics" entry. ■