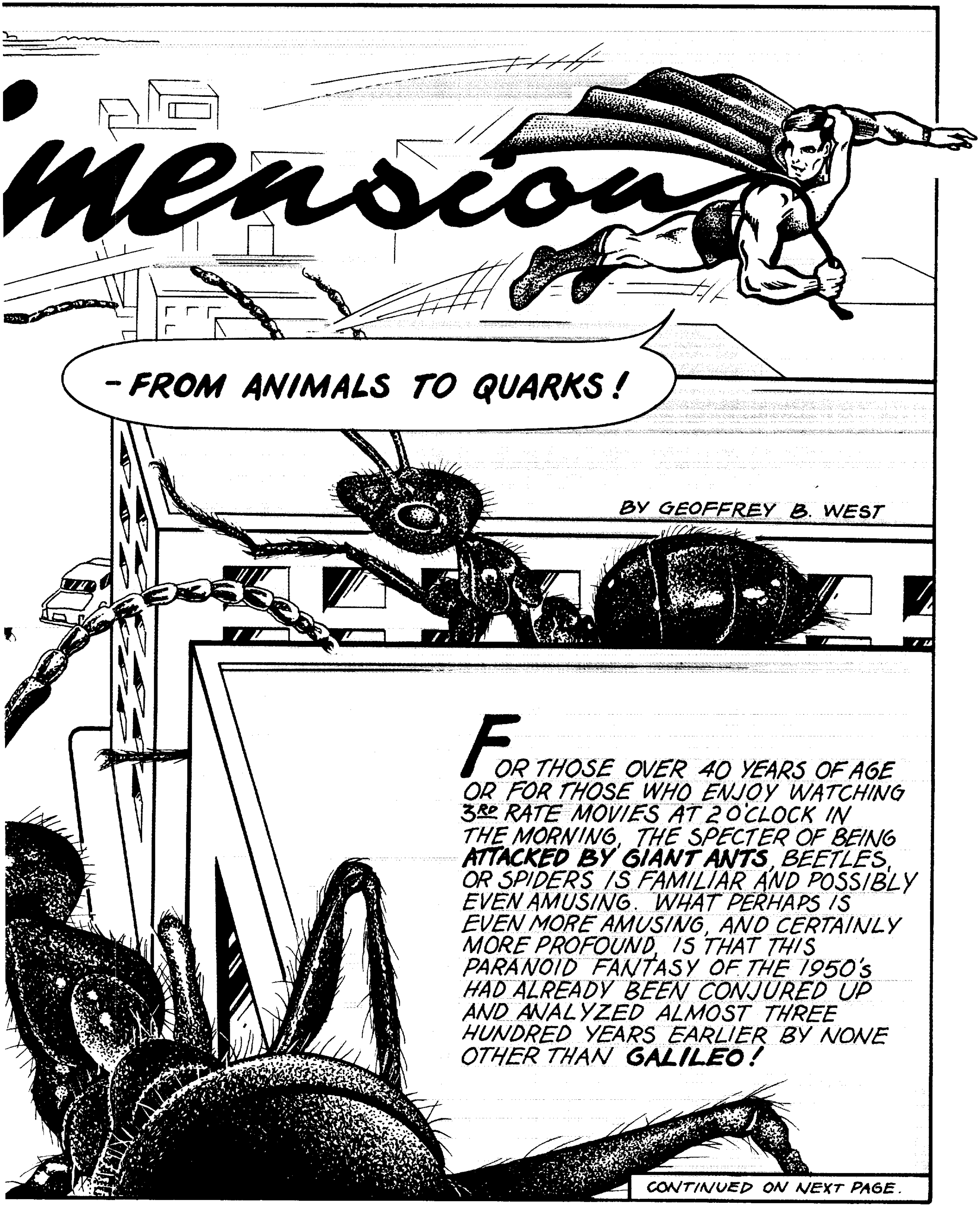

scale & D
J. CRUZ



BY GEOFFREY B. WEST

FOR THOSE OVER 40 YEARS OF AGE OR FOR THOSE WHO ENJOY WATCHING 3RD RATE MOVIES AT 2 O'CLOCK IN THE MORNING, THE SPECTER OF BEING **ATTACKED BY GIANT ANTS**, BEETLES, OR SPIDERS IS FAMILIAR AND POSSIBLY EVEN AMUSING. WHAT PERHAPS IS EVEN MORE AMUSING, AND CERTAINLY MORE PROFOUND, IS THAT THIS PARANOID FANTASY OF THE 1950's HAD ALREADY BEEN CONJURED UP AND ANALYZED ALMOST THREE HUNDRED YEARS EARLIER BY NONE OTHER THAN **GALILEO!**

CONTINUED ON NEXT PAGE.

"I have multiplied visions and used similitudes." — Hosea 7:10

In his marvelous book *Dialogues Concerning Two New Sciences* there is a remarkably clear discussion on the effects of scaling up the dimensions of a physical object. Galileo realized that if one simply scaled up its size, the weight of an animal would increase significantly faster than its strength, causing it ultimately to collapse. As Galileo says (in the words of Salviati during the discorso of the second day), ". . . you can plainly see the impossibility of increasing the size of structures to vast dimensions . . . if his height be increased inordinately, he will fall and be crushed under his own weight." The simple scaling up of an insect to some monstrous size is thus a physical impossibility, and we can rest assured that these old sci-fi images are no more than fiction! Clearly, to create a giant one "must either find a harder and stronger material . . . or admit a diminution of strength," a fact long known to architects.

It is remarkable that so many years before its deep significance could be appreciated, Galileo had investigated one of the most fundamental questions of nature: namely, what happens to a physical system when one changes scale? Nowadays this is the seminal question for quantum field theory, phase transition theory, the dynamics of complex systems, and attempts to unify all forces in nature. Tremendous progress has been made in these areas during the past fifteen years based upon answers to this question, and I shall try in the latter part of this article to give some flavor of what has been accomplished. However, I want first to remind the reader of the power of dimensional analysis in classical physics. Although this is stock-in-trade to all physicists, it is useful (and, more pertinently, fun) to go through several examples that explicate the basic ideas. Be warned, there are some surprises.

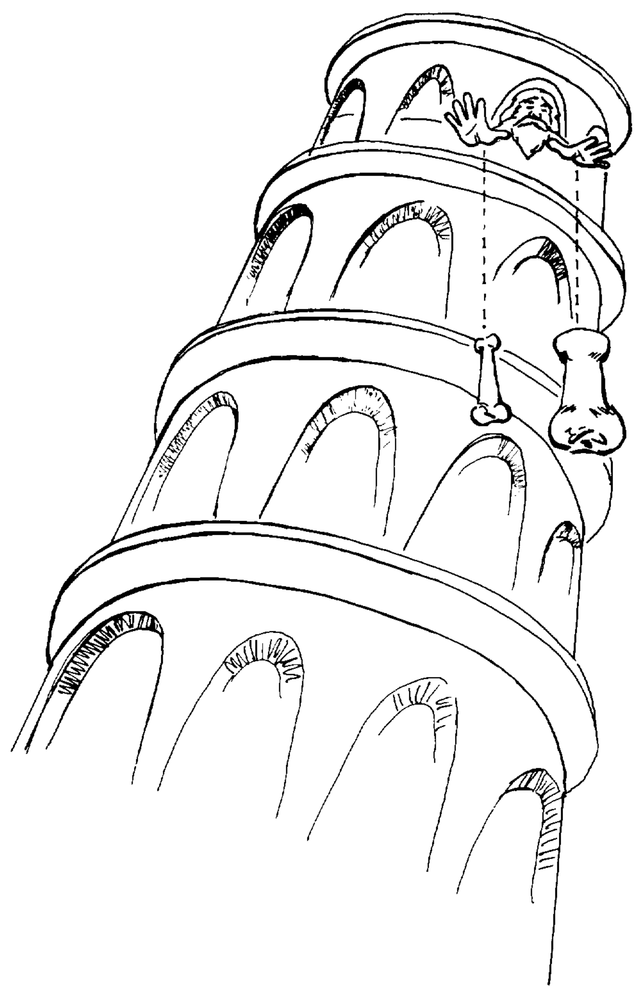
Classical Scaling

Let us first re-examine Galileo's original analysis. For *similar* structures* (that is, structures having the same physical characteristics such as shape, density, or chemical composition) Galileo perceived that weight W increases linearly with volume V, whereas strength increases only like a cross-sectional area A. Since for similar structures $V \propto l^3$ and $A \propto l^2$, where l is some characteristic length (such as the height of the structure), we conclude that

$$\frac{\text{Strength}}{\text{Weight}} \propto \frac{A}{V} \propto \frac{1}{l} \propto \frac{1}{W^{1/3}} . \qquad (1)$$

Thus, as Galileo noted, smaller animals "appear" stronger than larger ones. (It is amusing that Jerome Siegel and Joe Shuster, the creators of Superman, implicitly appealed to such an argument in one of the first issues of their comic.† They rationalized his super strength by drawing a rather dubious analogy with "the lowly ant who can support weights hundreds of times its own" (sic!).) Incidentally, the above discussion can be used to understand why the bones and limbs of larger animals must be proportionately stouter than those of smaller ones, a nice example of which can be seen in Fig. 1.

Arguments of this sort were used extensively during the late 19th century to un-



(a) (b)

Fig. 1. Two extinct mammals: (a) **Neohipparion,** *a small American horse and (b)* **Mastodon,** *a large, elephant-like animal, illustrating that the bones of heavier animals are proportionately stouter and thus proportionately stronger.*

derstand the gross features of the biological world: indeed, the general size and shape of animals and plants can be viewed as nature's way of responding to the constraints of gravity, surface phenomena, viscous flow, and the like. For example, one can understand why man cannot fly under his own muscular power, why small animals leap as high as larger ones, and so on.

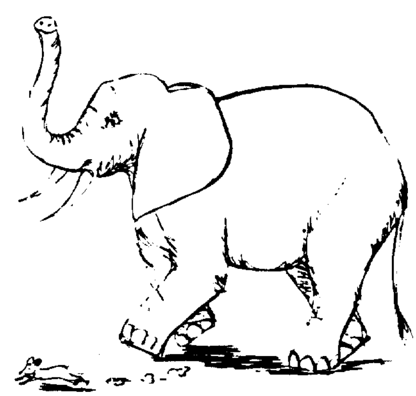
A classic example is the way metabolic rate varies from animal to animal. A measure B of metabolic rate is simply the heat lost by a body in a steady inactive state, which can be expected to be dominated by the surface effects of sweating and radiation. Symbolically, therefore, one expects $B \propto W^{2/3}$. The data (plotted logarithmically in Fig. 2) show that metabolic rate does indeed scale, that is, all animals lie on a single curve in spite of the fact that an elephant is neither a blown-up mouse nor a blown-up chimpanzee. However, the slope of the best-fit curve (the solid line) is closer to 3/4 than to 2/3, indicating that effects other than the pure geometry of surface dependence are at work.‡

It is not my purpose here to discuss why this is so but rather to emphasize the importance of a scaling curve not only for establishing the scaling phenomenon itself but for revealing deviations from some naive prediction (such as the surface law shown as the dashed line in Fig. 2). Typically, deviations from a simple geometrical or kinematical analysis reflect the dynamics of the system and can only be understood by examining it in more detail. Put slightly differently, one can view deviations from naive scaling as a probe of the dynamics.

The converse of this is also true: generally, one cannot draw conclusions concerning dynamics from naive scaling. As an illustration of this I now want to discuss some simple aspects of birds' eggs. I will focus on the question of breathing during incubation and how certain physical variables scale from bird to bird. Figure 3, adapted from a *Scientific American* article by Hermann Rahn, Amos Ar, and Charles V. Paganelli entitled "How Bird Eggs Breathe," shows the dependence of oxygen conductance K and pore length l (that is, shell thickness) on egg mass W. The authors, noting the smaller slope for l, conclude that "pore length probably increases slower because the eggshell must be thin enough for the embryo to hatch." This is clearly a dynamical conclusion! However, is it warranted?

From naive geometric scaling one expects that for similar eggs $l \propto W^{1/3}$, which is in reasonable agreement with the data: a best fit (the straight line in the figure) actually gives $l \propto W^{0.4}$. Since these data for pore length agree reasonably well with geometric scaling, no *dynamical* conclusion (such as the shell being thin enough for the egg to hatch) can be drawn. Ironically, rather than showing an anomalously slow growth with egg mass, the data for l actually manifest an anomalously fast growth (0.4 versus 0.33), not so dissimilar from the example of the metabolic rate!

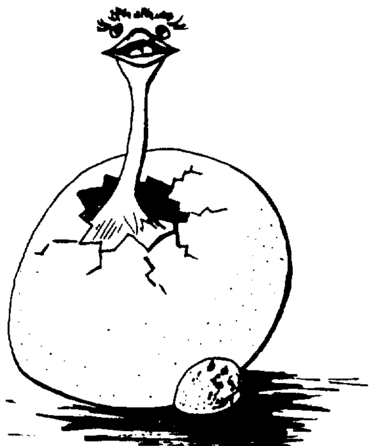
What about the behavior of the conductance, for which $K \propto W^{0.9}$? This relationship can also be understood on geometric grounds. Conductance is proportional to the *total* available pore area and inversely proportional to pore length. However, total pore area is made up of two factors: the total number of pores times the area of individual pores. If one assumes that the number of pores *per unit area* remains constant from bird to bird (a reasonable assumption consistent with other data), then we have two factors that scale like area and one that



**The concept of similitude is usually attributed to Newton, who first spelled it out in the* Principia *when dealing with gravitational attraction. On reading the appropriate section it is clear that this was introduced only as a passing remark and does not have the same profound content as the remarks of Galileo.*

†This amusing observation was brought to my attention by Chris Llewellyn Smith.

‡This relationship with a slope of 3/4 is known as Kleiber's law (M. Kleiber, Hilgardia *6(1932):315), whereas the area law is usually attributed to Rubner (M. Rubner,* Zeitschrift für Biologie (Munich) *19(1883):535).*



scales inversely as length. One thus expects $K \propto (W^{2/3})^2/W^{1/3} = W$, again in reasonable agreement with the data.

Dimensional Analysis. The physical content of scaling is very often formulated in terms of the language of dimensional analysis. The seminal idea seems to be due to Fourier. He is, of course, most famous for the invention of "Fourier analysis," introduced in his great treatise *Theorie Analytique de la Chaleur*, first published in Paris in 1822. However, it is generally not appreciated that this same book contains another great contribution, namely, the use of dimensions for physical quantities. It is the ghost of Fourier that is the scourge of all freshman physics majors, for it was he who first realized that every physical quantity "has one *dimension* proper to itself, and that the terms of one and the same equation could not be compared, if they had not the same *exponent of dimension*." He goes on: "We have introduced this consideration . . . to verify the analysis . . . it is the equivalent of the fundamental lemmas which the Greeks have left us without proof." Indeed it is! Check the dimensions!—the rallying call of all physicists (and, hopefully, all engineers).

However, it was only much later that physicists began to use the "method of dimensions" to *solve* physical problems. In a famous paper on the subject published in

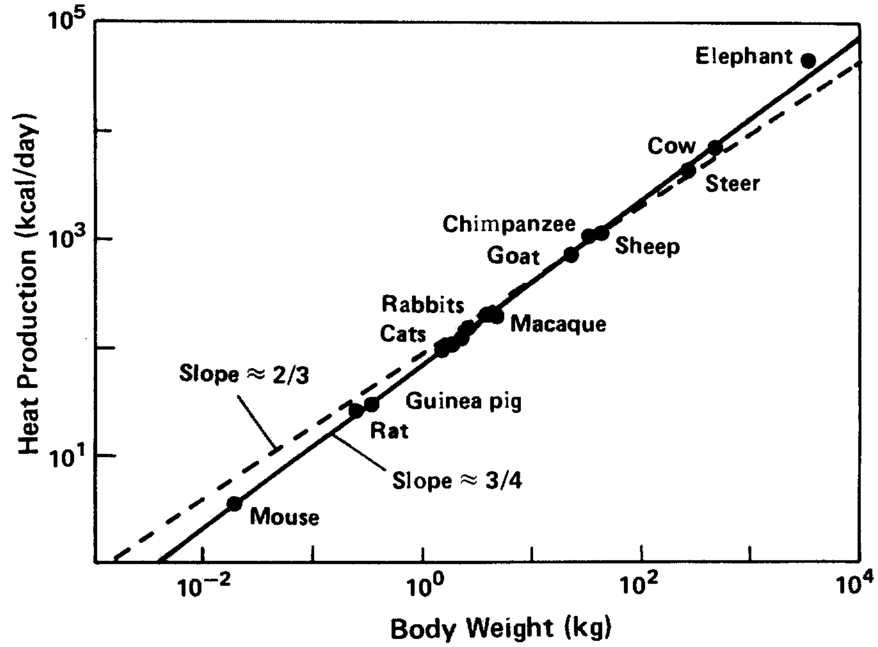


Fig. 2. Metabolic rate, measured as heat produced by the body in a steady state, plotted logarithmically against body weight. An analysis based on a surface dependence for the rate predicts a scaling curve with slope equal to 2/3 (dashed line) whereas the actual scaling curve has a slope equal to 3/4. Such deviation from simple geometrical scaling is indicative of other effects at work. (Figure based on one by Thomas McMahon, Science *179(1973):1201-1204 who, in turn, adapted it from M. Kleiber,* Hilgardia *6(1932):315.)*

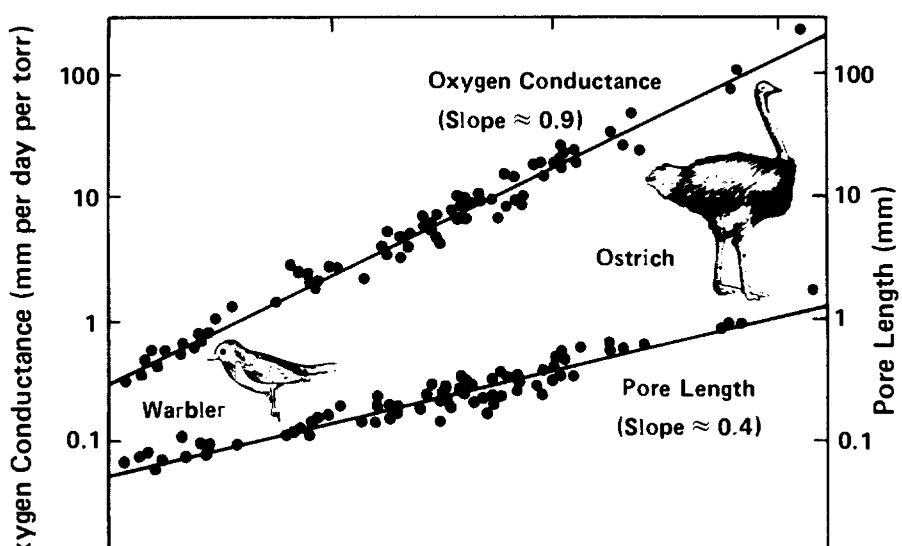
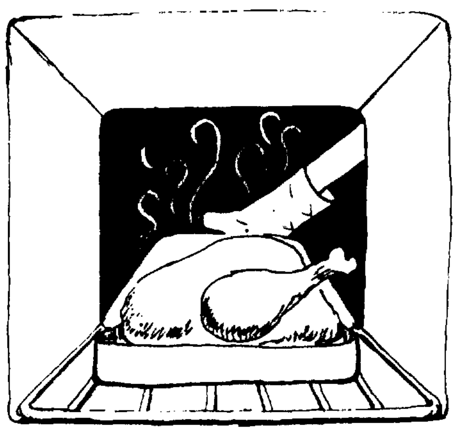


Fig. 3. Logarithmic plot of two parameters relevant to the breathing of birds' eggs during incubation: the conductance of oxygen through the shell and the pore length (or shell thickness) as a function of egg mass. Both plots have slopes close to those predicted by simple geometrical scaling analyses. (Figure adapted from H. Rahn, A. Ar, and C. V. Paganelli, Scientific American *240(February 1979):46-55.)*

Nature in 1915, Rayleigh indignantly begins: "I have often been impressed by the scanty attention paid even by original workers in the field to the great principle of similitude. It happens not infrequently that results in the form of 'laws' are put forward as novelties on the basis of elaborate experiments, which might have been predicted a priori after a few minutes consideration!" He then proceeds to set things right by giving several examples of the power of dimensional analysis. It seems to have been from about this time that the method became standard fare for the physicist. I shall illustrate it with an amusing example.

Most of us are familiar with the traditional Christmas or Thanksgiving problem of how much time to allow for cooking the turkey or goose. Many (inferior) cookbooks simply say something like "20 minutes per pound," implying a linear relationship with weight. However, there exist superior cookbooks, such as the *Better Homes and Gardens Cookbook*, that recognize the nonlinear nature of this relationship.

Figure 4 is based on a chart from this cookbook showing how cooking time t varies with the weight of the bird W. Let us see how



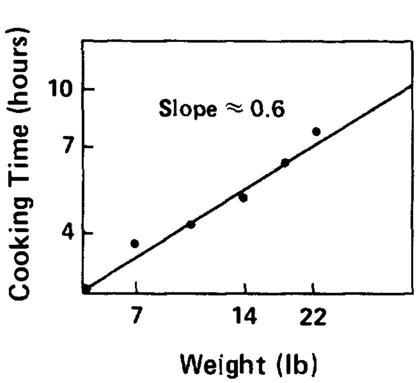


Fig. 4. The cooking time for a turkey or goose as a logarithmic function of its weight. (Based on a table in* Better Homes and Gardens Cookbook, *Des Moines:Meridith Corp., Better Homes and Gardens Books, 1962, p. 272.)

one can understand this variation using "the great principle of similitude." Let T be the temperature distribution inside the turkey and T_0 the oven temperature (both measured relative to the outside air temperature). T satisfies Fourier's heat diffusion equation: $\partial T/\partial t = \kappa\nabla^2 T$, where κ is the diffusion coefficient. Now, in general, for the dimension*al* quantities in this problem, there will be a functional relationship of the form

$$T = f(T_0, W, t, \rho, \kappa) , \qquad (2)$$

where ρ is the bird's density. However, Fourier's basic observation *that the physics be independent of the choice of units*, imposes a constraint on the form of the solution, which can be discerned by writing it in terms of dimension*less* quantities. Only two independent dimensionless quantities can be constructed: T/T_0 and $\rho(\kappa t)^{3/2}/W$. If we use the first of these as the dependent variable, the solution, whatever its form, must be expressible in terms of the other. The relationship must therefore have the structure

$$\frac{T}{T_0} = f\left(\frac{\rho(\kappa t)^{3/2}}{W}\right) . \qquad (3)$$

The important point is that, since the left-hand side is dimensionless, the "arbitrary" function f must be a dimensionless function of a dimensionless variable. Equation 3, unlike the previous one, does not depend upon the choice of units since dimensionless quantities remain invariant to changes in scale.

Let us now consider different but *geometrically similar* birds cooked to the same temperature distribution at the same oven temperature. Clearly, for all such birds there will be a scaling law

$$\frac{\rho(\kappa t)^{3/2}}{W} = \text{constant} . \qquad (4)$$

If the birds have the same physical characteristics (that is, the same ρ and κ), Eq. 4 reduces to

$$t = \text{constant} \times W^{2/3} , \qquad (5)$$

reflecting, not surprisingly, an area law. As can be seen from Fig. 4, this agrees rather well with the "data."

This formal type of analysis could also, of course, have been carried out for the metabolic rate and birds' eggs problems. The advantage of such an analysis is that it delineates the assumptions made in reaching conclusions like $B \propto W^{2/3}$ since, in principle, it focuses upon all the relevant variables. Naturally this is crucial in the discussion of any physics problem. For complicated systems, such as birds' eggs, with a very large number of variables, some prior insight or intuition must be used to decide what the important variables are. The dimensions of these variables are determined by the fundamental laws that they obey (such as the diffusion equation). Once the dimensions are known, the structure of the relationship between the variables is determined by Fourier's principle. There is therefore no magic in dimensional analysis, only the art of choosing the "right" variables, ignoring the irrelevant, and knowing the physical laws they obey.

As a simple example, consider the classic problem of the drag force F on a ship moving through a viscous fluid of density ρ. We shall choose F, ρ, the velocity v, the viscosity of the fluid μ, some length parameter of the ship l, and the acceleration due to gravity g as our

variables. Notice that we exclude other variables, such as the wind velocity and the amplitude of the sea waves because, under calm conditions, these are of secondary importance. Our conclusions may therefore not be valid for sailing ships!

The physics of the problem is governed by the Navier-Stokes equation (which incorporates Newton's law of viscous drag, telling us the dimensions of μ) and the gravitational force law (telling us the dimensions of g). Using these dimensions automatically incorporates the appropriate physics. Since we have limited the variables to a set of six, which must be expressible in terms of three basic units (mass M, length L, and time T), there will only be three independent dimensionless combinations. These are chosen to be $P \equiv F/\rho v^2 l^2$ (the pressure coefficient), $R \equiv vl\rho/\mu$ (Reynold's number), and $N_F \equiv v^2/lg$ (Froude's number). *Although any three similar combinations could have been chosen, these three are special because they delineate the physics.* For example, Reynold's number R relates to the viscous drag on a body moving through a fluid, whereas Froude's number N_F relates to the forces involved with waves and eddies generated on the surface of the fluid by the movement. Thus the rationale for the combinations R and N_F is to separate the role of the viscous forces from that of the gravitational: R does not depend on g, and F does not depend on μ. Furthermore, P does not depend on either!

Dimensional analysis now requires that the solution for the pressure coefficient,

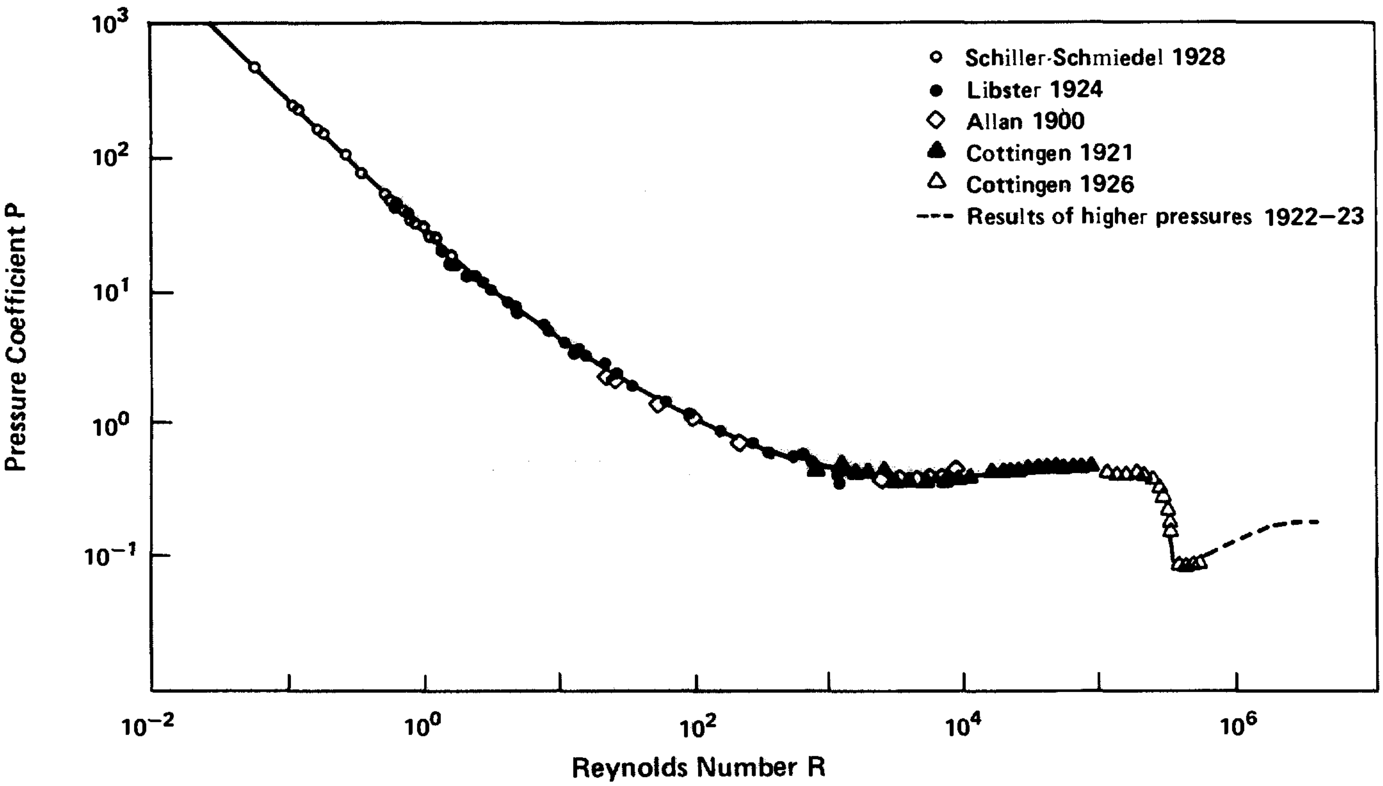


Fig. 5. The scaling curve for the motion of a sphere through a fluid that results when data from a variety of experiments are plotted in terms of two dimensionless variables: the pressure or drag coefficent P *versus Reynolds number* R. *(Figure adapted from* AIP Handbook of Physics, *2nd edition (1963):section II, p. 253.)*

whatever it is, must be expressible in the dimensionless form

$$P = f(R, N_F) . \quad (6)$$

The actual drag force F can easily be obtained from this equation by re-expressing it in terms of the dimensional variables (see Eq. 8 below).

First, however, consider a situation where surface waves generated by the moving object are *un*important (an extreme example is a submarine). In this case g will not enter the solution since it is manifested as the restoring force for surface waves. N_F can then be dropped from the solution, reducing Eq. 6 to the simple form

$$P = f(R) . \quad (7)$$

In terms of the original dimensional variables, this is equivalent to

$$F = \rho v^2 l^2 f(vl\rho/\mu) . \quad (8)$$

Historically, these last equations have been well tested by measuring the speed of different sizes and types of balls moving through different liquids. If the data are plotted using the dimensionless variables, that is, P versus R, then *all* the data should lie on just *one* curve regardless of the size of the ball or the nature of the liquid. Such a curve is called a *scaling curve*, a wonderful example of which is shown in Fig. 5 where one sees a scaling phenomenon that varies over seven orders of magnitude! It is important to recognize that if one had used dimensional variables and plotted F versus l, for example, then, instead of a single curve, there would have been *many* different and apparently unrelated curves for the different liquids. Using carefully chosen dimensionless variables (such as Reynold's number) is not only physically more sound but usually greatly simplifies the task of representing the data.

A remarkable consequence of this analysis is that, for similar bodies, the ratio of drag

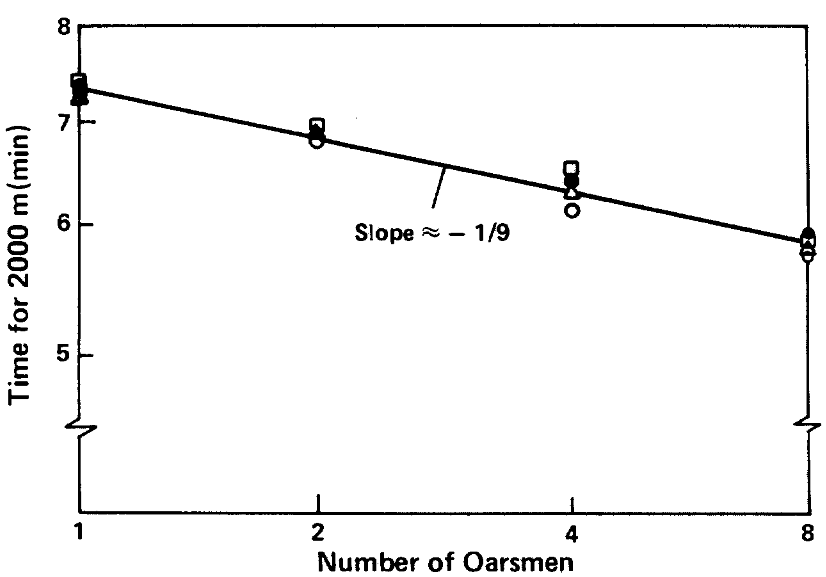
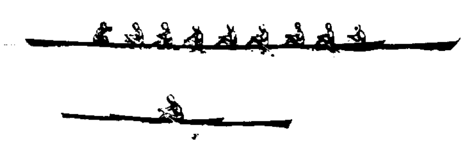


Fig. 6. The time needed for a rowing boat to complete a 2000-meter course in calm conditions as a function of the number of oarsmen. Data were taken from several international rowing championship events and illustrate the surprisingly slow dropoff predicted by modeling theory. (Adapted from T. A. McMahon, Science ***173(1971):349-351.)***



force to weight *decreases* as the size of the structure increases. From Archimedes' principle the volume of water displaced by a ship is proportional to its weight, that is, $W \propto l^3$ (this, incidentally, is why there is no need to include W as an independent variable in deriving these equations). Combined with Eq. 8 this leads to the conclusion that

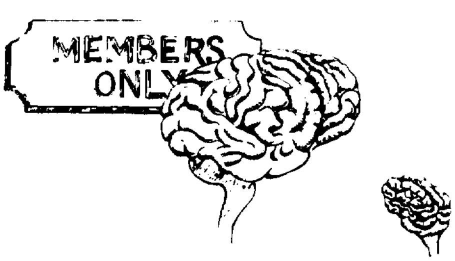
$$\frac{F}{W} \propto \frac{1}{l} . \quad (9)$$

This scaling law was extremely important in the 19th century because it showed that *it was cost effective to build bigger ships*, thereby justifying the use of large iron steamboats!

The great usefulness of scaling laws is also illustrated by the observation that the behavior of P for large ships ($l \to \infty$) can be derived from the behavior of small ships moving very fast ($v \to \infty$). This is so because both limits are controlled by the same asymptotic behavior of $f(R) = f(vl\rho/\mu)$. Such observations form the basis of *modeling* theory so crucial in the design of aircraft, ships, buildings, and so forth.

Thomas McMahon, in an article in *Science*, has pointed out another, somewhat more amusing, consequence to the drag force equation. He was interested in how the speed of a rowing boat scales with the number of oarsmen n and argued that, at a steady velocity, the power expended by the oarsmen E to overcome the drag force is given by Fv. Thus

$$E = Fv = \rho v^3 l^2 f(R) . \quad (10)$$



Using Archimedes' principle again and the fact that both E and W should be directly proportional to n leads to the remarkable scaling law

$$v \propto n^{1/9}, \tag{11}$$

which shows a *very* slow growth with n. Figure 6 exhibits data collected by McMahon from various rowing events for the time t ($\propto 1/v$) taken to cover a fixed 2000-meter course under calm conditions. One can see quite plainly the verification of his predicted law—a most satisfying result!

There are many other fascinating and exotic examples of the power of dimensional analysis. However, rather than belaboring the point, I would like to mention a slightly different application of scaling before I turn to the mathematical formulation. All the examples considered so far are of a quantitative nature based on well-known laws of physics. There are, however, situations where the qualitative observation of scaling can be used to scientific advantage to reveal phenomenological "laws."

A nice example (Fig. 7), taken from an article by David Pilbeam and Stephen Jay Gould, shows how the endocranial volume V (loosely speaking, the brain size) scales with body weight W for various hominids and pongids. The behavior for modern pongids is typical of most species in that the exponent a, defined by the phenomenological relationship $V \propto W^a$, is approximately 1/3 (for mammals a varies from 0.2 to 0.4). It is very satisfying that a similar behavior is exhibited by australopithecines, extinct cousins of our lineage that died out over a million years ago. However, as Pilbean and Gould point out, our homo sapiens lineage shows a strikingly different behavior, namely: $a \approx 5/3$. Notice that neither this relationship nor the "standard" behavior ($a \approx 1/3$) is close to the naive geometrical scaling prediction of $a = 1$.

These data illustrate dramatically the qualitative evolutionary advance in the brain development of man. Even though the reasons for $a \approx 1/3$ may not be understood, this value can serve as the "standard" for revealing deviations and provoking speculation concerning evolutionary progress: for example, what is the deep significance of a brain size that grows linearly with height versus a brain size that grows like its fifth power? I shall not enter into such questions here, tempting though they be.

Such phenomenological scaling laws (whether for brain volume, tooth area, or some other measurable parameter of the fossil) can also be used as corroborative evidence for assigning a newly found fossil of some large primate to a particular lineage. The fossil's location on such curves can, in principle, be used to distinguish an australopithecine from a homo. Notice, however, that implicit in all this discussion is knowledge of body weight: presumably, anthropologists have developed verifiable techniques for estimating this quantity. Since they necessarily work with fragments only, some further scaling assumptions *must* be involved in their estimates!

Relevant Variables. As already emphasized, the most important and artful aspect of the method of dimensions is the choice of variables relevant to the problem and their grouping into dimensionless combinations that delineate the physics. In spite of the

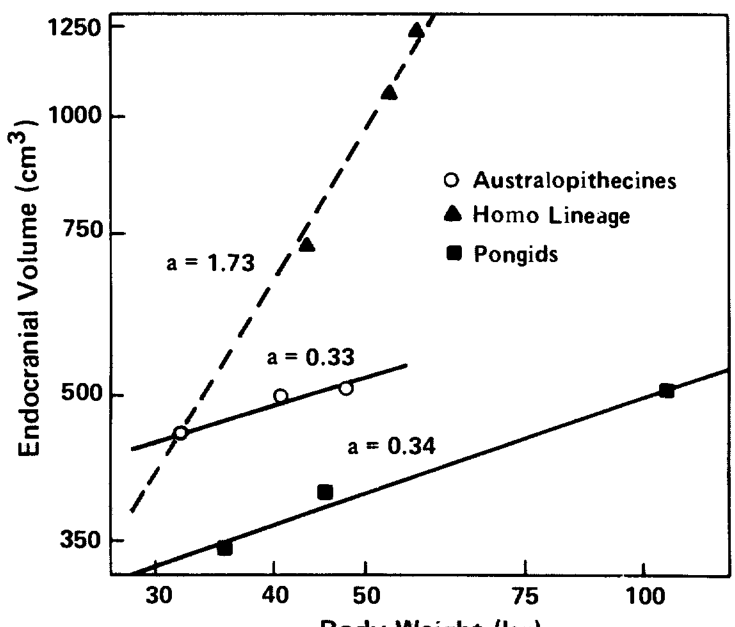


Fig. 7. Scaling curves for endocranial volume (or brain size) as a function of body weight. The slope of the curve for our homo sapiens lineage (dashed line) is markedly different from those for australopithecines, extinct cousins of the homo lineage, and for modern pongids, which include the chimpanzee, gorilla and orangutan. (Adapted from D. Pilbeam and S. J. Gould. Science ***186(1974):892-901.)***

relative simplicity of the method there are inevitably paradoxes and pitfalls, a famous case of which occurs in Rayleigh's 1915 paper mentioned earlier. His last example concerns the rate of heat lost H by a conductor immersed in a stream of inviscid fluid moving past it with velocity v ("Boussinesq's problem"). Rayleigh showed that, if K is the heat conductivity, C the specific heat of the fluid, θ the temperature difference, and l some linear dimension of the conductor, then, in dimensionless form,

$$\frac{H}{kl\theta} = f\left(\frac{lvC}{K}\right). \qquad (12)$$

Approximately four months after Rayleigh's paper appeared, *Nature* published an eight line comment (half column, yet!) by a D. Riabouchinsky pointing out that Rayleigh's result assumed that temperature was a dimension independent from mass, length, and time. However, from the kinetic theory of gases we know that this is not so: temperature can be defined as the mean kinetic energy of the molecules and so is *not* an independent unit! Thus, according to Riabouchinsky, Rayleigh's expression must be replaced by an expression with an additional dimensionless variable:

$$\frac{H}{kl\theta} = f\left(\frac{lvC}{K}, Cl^3\right), \qquad (13)$$

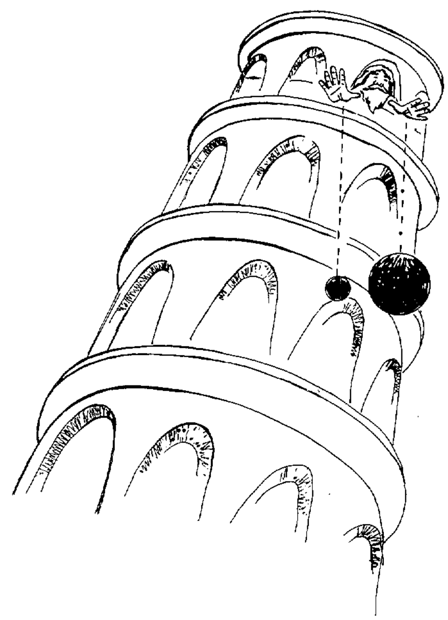
a much *less* restrictive result.

Two weeks later, Rayleigh responded to Riabouchinsky saying that "it would indeed be a paradox if the *further* knowledge of the nature of heat afforded by molecular theory put us in a *worse* position than before in dealing with a particular problem. . . . It would be well worthy of discussion." Indeed it would; its resolution, which no doubt the reader has already discerned, is left as an exercise (for the time being)! Like all paradoxes, this one cautions us that we occasionally make casual assumptions without quite realizing that we have done so (see "Fundamental Constants and the Rayleigh-Riabouchinsky Paradox").

Scale Invariance

Let us now turn our attention to a slightly more abstract mathematical formulation that clarifies the relationship of dimensional analysis to *scale invariance.* By scale invariance we simply mean that the structure of physical laws cannot depend on the choice of units. As already intimated, this is automatically accomplished simply by employing dimensionless variables since these clearly do not change when the system of units changes. However, it may not be immediately obvious that this is equivalent to the *form invariance* of physical equations. Since physical laws are usually expressed in terms of dimensional variables, this is an important point to consider: namely, what are the general constraints that follow from the requirement that the laws of physics look the same regardless of the chosen units. The crucial observation here is that implicit in any equation written in terms of dimensional variables are the "hidden" fundamental scales of mass M, length L, time T, and so forth that are relevant to the problem. Of course, one never actually makes these scale parameters explicit precisely because of form invariance.

Our motivation for investigating this question is to develop a language that can be generalized in a natural way to include the subtleties of quantum field theory. Hopefully classical dimensional analysis and scaling will be sufficiently familiar that its generalization to the more complicated case will be relatively smooth! This generalization has been named the *renormalization group* since its origins lie in the renormalization program used to make sense out of the infinities inherent in quantum field theory. It turns out that renormalization requires the introduction of a new arbitrary "hidden" scale that plays a role similar to the role of the scale parameters implicit in any dimensional equation. Thus any equation derived in quantum field theory that represents a physical quantity must not depend upon this choice of hidden scale. The resulting constraint will simply represent a generalization of ordinary dimensional analysis; the only reason that it is different is that variables in quantum field theory, such as fields, change in a much more complicated fashion with scale than do their classical counterparts.



Nevertheless, just as dimensional analysis allows one to learn much about the behavior of a system without actually solving the dynamical equations, so the analogous constraints of the renormalization group lead to powerful conclusions about the behavior of a quantum field theory without actually being able to solve it. It is for this reason that the renormalization group has played such an important part in the renaissance of quantum field theory during the past decade or so. Before describing how this comes about, I shall discuss the simpler and more familiar case of scale change in ordinary classical systems.

To begin, consider some physical quantity F that has dimensions; it will, of course, be a function of various dimensional variables x_i: $F(x_1,x_2,\ldots,x_n)$. An explicit example is given by Eq. 2 describing the temperature distribution in a cooked turkey or goose.

Fundamental Constants and the

Let us examine Riabouchinsky's paradox a little more carefully and show how its resolution is related to choosing a system of units where the "fundamental constants" (such as Planck's constant h and the speed of light c) can be set equal to unity.

The paradox had to do with whether temperature could be used as an independent dimensional unit even though it can be defined as the mean kinetic energy of the molecular motion. Rayleigh had chosen five physical variables (length l, temperature difference θ, velocity v, specific heat C, and heat conductivity K) to describe Boussinesq's problem and had assumed that there were four independent dimensions (energy E, length L, time T, and temperature Θ). Thus the solution for T/T_0 necessarily is an arbitrary function of *one* dimensionless combination. To see this explicitly, let us examine the dimensions of the five physical variables:

$$[l] = L, [\theta] = \Theta, [v] = LT^{-1}, [C] = EL^{-3}\Theta^{-1},$$

$$\text{and } [K] = EL^{-1}T^{-1}\Theta^{-1}.$$

Clearly the combination chosen by Rayleigh, lvC/K, is dimensionless. Although other dimensionless combinations can be formed, they are not independent of the two combinations (lvC/K and T/T_o) selected by Rayleigh.

Now suppose, along with Riabouchinsky, we use our knowledge of the kinetic theory to define temperature "as the mean kinetic energy of the molecules" so that Θ is no longer an independent dimension. This means there are now only *three* independent dimensions and the solution will depend on an arbitrary function of *two* dimensionless combinations. With $\Theta \propto E$, the dimensions of the physical variables become:

$$[l] = L, [\theta] = E, [v] = LT^{-1}, [C] = L^{-3}, \text{and } [K] = L^{-1}T^{-1}.$$

It is clear that, in addition to Rayleigh's dimensionless variable, there is now a new *independent* combination, Cl^3 for example, that is dimensionless. To reiterate Rayleigh: "it would indeed be a paradox if the *further* knowledge of the nature of heat . . . put us in a *worse* position than before . . . it would be well worthy of discussion."

Like almost all paradoxes, there is a bogus aspect to the argument. It is certainly true that the kinetic theory allows one to express an energy as a temperature. However, this is only useful and *appropriate* for situations where the physics is dominated by molecular considerations. For macroscopic situations such as Boussinesq's problem, the molecular nature of the system is irrelevant; the microscopic variables have been replaced by macroscopic averages embodied in phenomenological properties such as the specific heat and conductivity. To make Riabouchinsky's identification of energy with temperature is to introduce *irrelevant* physics into the problem.

Exploring this further, we recall that such an energy-temperature identification implicitly involves the introduction of Boltzmann's factor k. By its very nature, k will only play an explicit role in a physical problem that directly involves the molecular nature of the system; otherwise it will not enter. Thus one could describe the system from the molecular viewpoint (so that k is involved) and then take a macroscopic limit. Taking the limit is equivalent to setting $k = 0$; the presence of a finite k indicates that explicit effects due to the kinetic theory are important.

With this in mind, we can return to Boussinesq's problem and derive Riabouchinsky's result in a somewhat more illuminating fashion. Let us follow Rayleigh and keep E, L, T, and Θ as the

Each of these variables, including F itself, is always expressible in terms of some standard set of independent units, which can be chosen to be mass M, length L, and time T. These are the hidden scale parameters. Obviously, other combinations could be used. There could even be other independent units, such as temperature (but remember Riabouchinsky!), or more than one independent length (say, transverse and longitudinal). In this discussion, we shall simply use the conventional M, L, and T. Any generalization is straightforward.

In terms of this standard set of units, the magnitude of each x_i is given by

$$x_i = M^{\alpha_i} L^{\beta_i} T^{\gamma_i} \tag{15}$$

The numbers α_i, β_i, and γ_i will be recognized as "the dimensions" of x_i. Now suppose we change the system of units by some scale transformation of the form

$$M \to M' = \lambda_M M,$$

$$L \to L' = \lambda_L L,$$

and

$$T \to T' = \lambda_T T. \tag{16}$$

Each variable then responds as follows:

$$x_i \to x_i' = Z_i(\lambda) x_i, \tag{17}$$

where

$$Z_i(\lambda) = \lambda_M^{\alpha_i} \lambda_L^{\beta_i} \lambda_T^{\gamma_i}. \tag{18}$$

and λ is shorthand for λ_M, λ_L, and λ_T. Since F is itself a dimensional physical quantity, it transforms in an identical fashion under this scale change:

$$F \to F' = Z(\lambda) F(x_1, x_2, \ldots, x_n), \tag{19}$$

where

$$Z(\lambda) = \lambda_M^{\alpha} \lambda_L^{\beta} \lambda_T^{\gamma}. \tag{20}$$

Here α, β, and γ are the dimensions of F.

There is, however, an alternate but equivalent way to transform from F to F', namely, by transforming each of the variables x_i separately. Explicitly we therefore also have

Rayleigh-Riabouchinsky Paradox

independent dimensions but add k (with dimensions $E\Theta^{-1}$) as a new physical variable. The solution will now be an arbitrary function of *two* independent dimensionless variables: lvC/K and kCl^3. When Riabouchinsky chose to make Cl^3 his other dimensionless variable, he, in effect, chose a system of units where $k = 1$. But that was a terrible thing to do here since the physics dictates that $k = 0$! Indeed, if $k = 0$ we regain Rayleigh's original result, that is, we have only *one* dimensionless variable. It is somewhat ironic that Rayleigh's remarks miss the point: "further knowledge of the nature of heat afforded by molecular theory" does not put one in a better position for solving the problem—rather, it leads to a microscopic description of K and C. The important point pertinent to the problem set up by Rayleigh is that knowledge of the molecular theory is irrelevant and k must not enter.

The lesson here is an important one because it illustrates the role played by the fundamental constants. Consider Planck's constant $\hbar \equiv h/2\pi$: it would be completely inappropriate to introduce it into a problem of classical dynamics. For example, any solution of the scattering of two billiard balls will depend on macroscopic variables such as the masses, velocities, friction coefficients, and so on. Since billiard balls are made of protons, it might be tempting to the purist to include as a dependent variable the proton-proton total cross section, which, of course, involves $\hbar$. This would clearly be totally inappropriate but is analogous to what Riabouchinsky did in Boussinesq's problem.

Obviously, if the scattering is between two microscopic "atomic billiard balls" then $\hbar$ *must be* included. In this case it is not only quite legitimate but often convenient to choose a system of units where $\hbar = 1$. However, having done so one cannot directly recover the classical limit corresponding to $\hbar = 0$. With $\hbar = 1$, one is stuck in quantum mechanics just as, with $k = 1$, one is stuck in kinetic theory.

A similar situation obviously occurs in relativity: the velocity of light c must not occur in the classical Newtonian limit. However, in a relativistic situation one is quite at liberty to choose units where $c = 1$. Making that choice, though, presumes the physics involves relativity.

The core of particle physics, relativistic quantum field theory, is a synthesis of quantum mechanics and relativity. For this reason, particle physicists find that a system of units in which $\hbar = c = 1$ is not only convenient but is a manifesto that quantum mechanics and relativity are the basic physical laws governing their area of physics. In quantum mechanics, momentum p and wavelength λ are related by the de Broglie relation: $p = 2\pi\hbar/\lambda$; similarly, energy E and frequency ω are related by Planck's formula: $E = \hbar\omega$. In relativity we have the famous Einstein relation: $E = mc^2$. Obviously if we choose $\hbar = c = 1$, all energies, masses, and momenta have the same units (for example, electron volts (eV)), and these are the same as *inverse* lengths and times. Thus larger energies and momenta *inevitably* correspond to shorter times and lengths.

Using this choice of units automatically incorporates the profound physics of the uncertainty principle: to probe short space-time intervals one needs large energies. A useful number to remember is that 10^{-13} centimeter, or 1 fermi (fm), equals the reciprocal of 200 MeV. We then find that the electron mass ($\approx$ 1/2 MeV) corresponds to a length of $\approx$ 400 fm—its Compton wavelength. Or the 20 TeV (2×10^7 MeV) typically proposed for a possible future facility corresponds to a length of 10^{-18} centimeter. This is the scale distance that such a machine will probe! ■

$$F \rightarrow F' = F(Z_1(\lambda)x_1, Z_2(\lambda)x_2, \ldots, Z_n(\lambda)x_n)\,. \qquad (21)$$

Equating these two different ways of effecting a scale change leads to the identity

$$F(Z_1(\lambda)x_1, Z_2(\lambda)x_2, \ldots, Z_n(\lambda)x_n) = Z(\lambda)\, F(x_1, x_2, \ldots, x_n)\,. \qquad (22)$$

As a concrete example, consider the equation $E = mc^2$. To change scale one can either transform E directly or transform m and c separately and multiply the results appropriately—obviously the final result must be the same.

We now want to ensure that the resulting form of the equation does not depend on λ. This is best accomplished using Euler's trick of taking $\partial/\partial\lambda$ and then setting $\lambda = 1$. For example, if we were to consider changes in the mass scale, we would use $\partial/\partial\lambda_M$ and the chain rule for partial differentiation to arrive at

$$\sum_{i=1}^{n} x_i \frac{\partial Z_i}{\partial \lambda_M} \frac{\partial F}{\partial x_i'} = \frac{\partial Z}{\partial \lambda_M} F\,. \qquad (23)$$

When we set $\lambda_M = 1$, differentiation of Eqs. 18 and 20 yields

$$\left(\frac{\partial Z_i}{\partial \lambda_M}\right)_{\lambda_M = 1} = \alpha_i\,,$$

$$\left(\frac{\partial Z}{\partial \lambda_M}\right)_{\lambda_M = 1} = \alpha\,. \qquad (24)$$

and $x_i' = x_i$, so that Eq. 23 reduces to

$$\alpha_1 x_1 \frac{\partial F}{\partial x_1} + \alpha_2 x_2 \frac{\partial F}{\partial x_2} + \alpha_3 x_3 \frac{\partial F}{\partial x_3} + \ldots + \alpha_n x_n \frac{\partial F}{\partial x_n} = \alpha F\,. \qquad (25)$$

Obviously this can be repeated with λ_L and λ_T to obtain a set of three coupled partial differential equations expressing the fundamental *scale invariance of physical laws (that is, the invariance of the physics to the choice of units)* implicit in Fourier's original work. These equations can be solved without too much difficulty; their solution is, in fact, a special case of the solution to the renormalization group equation (given explicitly as Eq. 35 below). Not too surprisingly, one finds that the solution is precisely equivalent to the constraints of dimensional analysis. Thus there is never any explicit need to use these rather cumbersome equations: ordinary dimensional analysis takes care of it for you!

Quantum Field Theory

We have gone through this little mathematical exercise to illustrate the well-known relationship of dimensional analysis to scale and form invariance. I now want to discuss how the formalism must be amended when applied to quantum field theory and give a sense of the profound consequences that follow. Using the above chain of reasoning as a guide, I shall examine the response of a quantum field theoretic system to a change in scale and derive a partial differential equation analogous to Eq. 25. This equation is known as the *renormalization group equation* since its origins lay in the somewhat arcane area of the renormalization procedure used to tame the infinities of quantum field theory. I shall therefore have to digress momentarily to give a brief résumé of this subject before returning to the question of scale change.

Renormalization. Perhaps the most unnerving characteristic of quantum field theory for the beginning student (and possibly also for the wise old men) is that almost all calculations of its physical consequences naively lead to infinite answers. These infinities stem from divergences at high momenta associated with virtual processes that are always present in any transition amplitude. The renormalization scheme, developed by Richard P. Feynman, Julian S. Schwinger, Sin-Itiro Tomonaga, and Freeman Dyson, was invented to make sense out of this for quantum electrodynamics (QED).

To get a feel for how this works I shall focus on the photon, which carries the force associated with the electromagnetic field. At the classical limit the *propagator** for the photon represents the usual static $1/r$ Coulomb potential. The corresponding Fourier transform (that is, the propagator's representation in momentum space) in this limit is $1/q^2$, where q is the momentum carried by the photon. Now consider the "classical" scattering of two charged particles (represented by the Feynman diagram in Fig. 8 (a)). For this event the exchange of a single photon gives a transition amplitude proportional to e_0^2/q^2, where e_0 is the charge (or coupling constant) occurring in the Lagrangian. A standard calculation results in the classical Rutherford formula, which can be extended relativistically to the spin-1/2 case embodied in the diagram.

A typical quantum-mechanical correction to the scattering formula is illustrated in Fig. 8 (b). The exchanged photon can, by virtue of the uncertainty principle, create for a very short time a virtual electron-positron pair, which is represented in the diagram by the loop. We shall use k to denote the momentum carried around the loop by the two particles.

There are, of course, many such corrections that serve to modify the $1/q^2$ single-photon behavior, and this is represented schematically in part (c). It is convenient to include all these corrections in a single multiplicative factor D_0 that represents deviations from the single-photon term. The "full" photon propagator including all possible radiative corrections is therefore D_0/q^2. The reason for doing this is that D_0 is a *dimensionless* function that gives a measure of the polarization of the vacuum caused by the production of virtual particles. (The origin of the Lamb shift is vacuum polarization.)

We now come to the central problem: upon evaluation it is found that contributions from diagrams like (b) are infinite because there is no restriction on the magnitude of the momentum k flowing *in* the loop! Thus, typical calculations lead to integrals of the form

$$\int_0^\infty \frac{dk^2}{k^2 + aq^2}, \qquad (26)$$

which diverge logarithmically. Several prescriptions have been invented for making such integrals finite; they all involve "reg-

Roughly speaking, the photon propagator can be thought of as the Green's function for the electromagnetic field. In the relativistically covariant Lorentz gauge, the classical Maxwell's equations read

$$\Box^2 A(x) = j(x),$$

where A(x) is the vector potential and j(x) is the current source term derived in QED from the motion of the electrons. (To keep things simple I am suppressing all space-time indices, thereby ignoring spin.) This equation can be solved in the standard way using a Green's function:

$$A(x) = \int d^4x'\, G(x'-x)\, j(x'),$$

with

$$\Box^2 G(x) = \delta(x)\ .$$

Now a transition amplitude is proportional to the interaction energy, and this is given by

$$H_I = \int d^4x\, j(x)\, A(x) =$$

$$\int d^4x \int d^4x'\, j(x)\, G(x-x')\, j(x'),$$

illustrating how G "mediates" the force between two currents separated by a space-time interval (x-x'). It is usually more convenient to work with Fourier transforms of these quantities (that is, in momentum space). For example, the momentum space solution for G is $\tilde{G}(q) = 1/q^2$, and this is usually called the free photon propagator since it is essentially classical. The corresponding "classical" transition amplitude in momentum space is just $\tilde{j}(q)(1/q^2)\tilde{j}(q)$, which is represented by the Feynman graph in Fig. 8 (a).

In quantum field theory, life gets much more complicated because of radiative corrections as discussed in the text and illustrated in (b) and (c) of Fig. 8. The definition of the propagator is generally in terms of a correlation function in which a photon is created at point x out of the vacuum for a period x-x' and then returns to the vacuum at point x'. Symbolically, this is represented by

$$G(x-x') \sim \langle vac|A(x')\, A(x)|vac\rangle\ .$$

During propagation, anything allowed by the uncertainty principle can happen—these are the radiative corrections that make an exact calculation of G almost impossible.

ularizing" the integrals by introducing some large mass parameter Λ. A standard technique is the so-called Pauli-Villars scheme in which a factor $\Lambda^2/(k^2 + \Lambda^2)$ is introduced into the integrand with the understanding that Λ is to be taken to infinity at the end of the calculation (notice that in this limit the regulating factor approaches one). With this prescription, the above integral is therefore replaced by

$$\lim_{\Lambda^2 \to \infty} \int_0^\infty \frac{dk^2 \Lambda^2}{(k^2 + aq^2)(k^2 + \Lambda^2)} = \ln \frac{\Lambda^2}{aq^2} . \qquad (27)$$

The integral can now be evaluated and its divergence expressed in terms of the (infinite) mass parameter Λ. All the infinities arising from quantum fluctuations can be dealt with in a similar fashion with the result that the following series is generated:

$$D_0(q,e_0) \approx 1 + a_1 e_0^2 \left(\ln \frac{\Lambda^2}{q^2} + \ldots \right) + e_0^4 \left[a_2 \left(\ln \frac{\Lambda^2}{q^2} \right)^2 + b_2 \ln \frac{\Lambda^2}{q^2} + \ldots \right] + \ldots . \qquad (28)$$

In this way the structure of the infinite divergences in the theory are parameterized in terms of Λ, which can serve as a finite *cutoff* in the integrals over virtual momenta.*

The remarkable triumph of the renormalization program is that, rather than imposing such an arbitrary cutoff, all these divergences can be swallowed up by an *infinite rescaling* of the fields and coupling con-

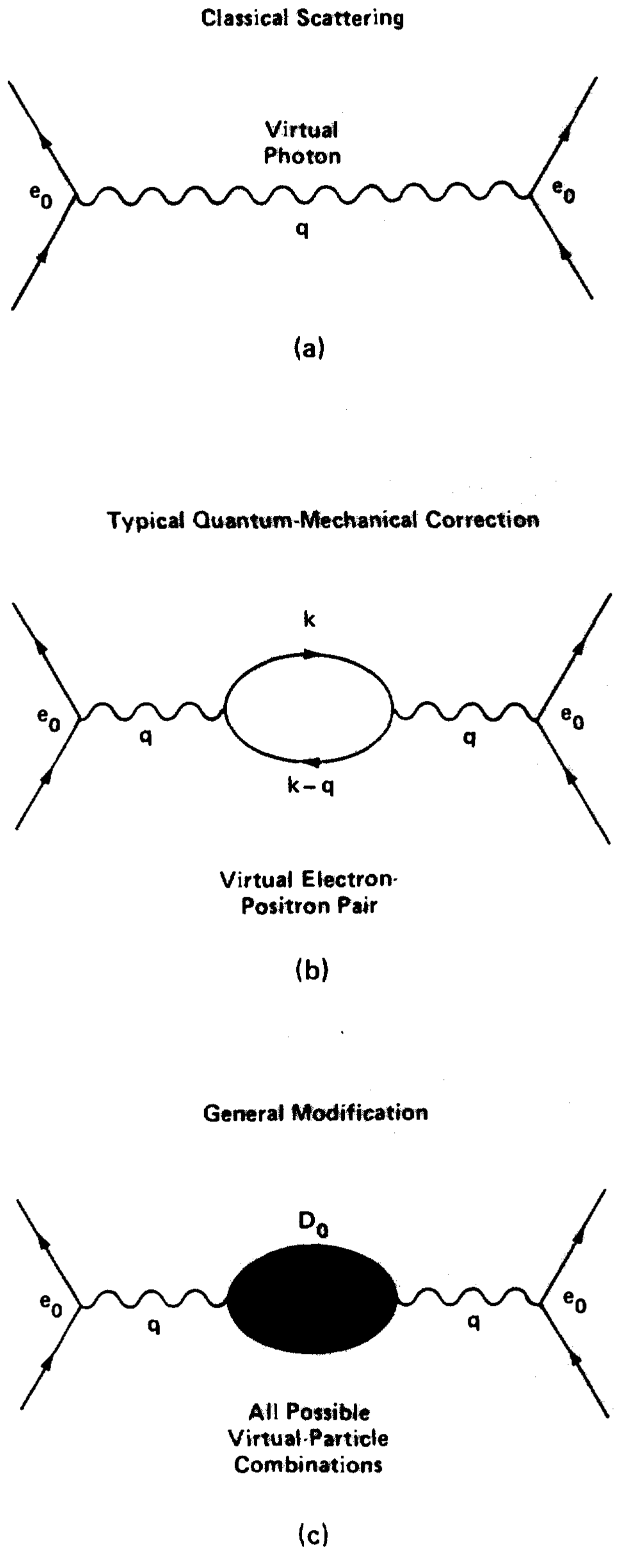


Fig. 8. Feynman diagrams for (a) the classical scattering of two particles of charge e_0, (b) a typical correction that must be made to that scattering—here because of the creation of a virtual electron-positron pair—and (c) a diagram representing all such possible corrections. The matrix element is proportional for (a) to e_0^2/q^2 and for (c) to D_0/q^2 where D_0 includes all corrections.

****In this discussion I assumed, for simplicity, that the original Lagrangian was massless; that is, it contained no explicit mass parameter. The addition of such a mass term would only complicate the discussion unnecessarily without giving any new insights.***

stants. Thus, a *finite* propagator D, that does not depend on Λ, can be derived from D_0 by rescaling if, at the same time, one rescales the charge similarly. These rescalings take the form

$$D = Z_D D_0 \text{ and } e = Z_e e_0 . \qquad (29)$$

The crucial property of these scaling factors is that they are independent of the physical momenta (such as q) but depend on Λ in such a way that when the cutoff is removed, D and e remain finite. In other words, when $\Lambda \to \infty$, Z_D *and* Z_e must develop infinities of their own that precisely compensate for the infinities of D_0 and e_0. The original so-called *bare* parameters in the theory calculated from the Lagrangian (D_0 and e_0) therefore have no physical meaning—only the renormalized parameters (D and e) do.

Now let us apply some ordinary dimensional analysis to these remarks. Because they are simply scale factors, the Z's must be dimensionless. However, the Z's are functions of Λ but not of q. But that is very peculiar: a dimensionless function cannot depend on a *single* mass parameter! Thus, in order to express the Z's in dimensionless form, *a new finite mass scale* μ *must be introduced* so that one can write $Z = Z(\Lambda^2/\mu^2, e_0)$. *An immediate consequence of renormalization is therefore to induce a mass scale not manifest in the Lagrangian.* This is extremely interesting because it provides a possible mechanism for generating mass even though no mass parameter appears in the Lagrangian. We therefore have the exciting possibility of being able to calculate the masses of *all* the elementary particles in terms of just *one* of them. Similar considerations for the dimensionless D's clearly require that they be expressible as $D_0 = D_0(q^2/\Lambda^2, e_0)$, as in Eq. 28, and $D = D(q^2/\mu^2, e)$. (The dream of particle theorists is to write down a Lagrangian with *no* mass parameter that describes all the interations in terms of just *one* coupling constant. The mass spectrum and scattering amplitudes for all the elementary particles would then be calculable in terms of the value of this single coupling at some given scale! A wonderful fantasy.)

To recapitulate, the physical finite renormalized propagator D is related to its bare and divergent counterpart D_0 (calculated from the Lagrangian using a cutoff mass) by an infinite rescaling:

$$D\left(\frac{q^2}{\mu^2}, e\right) = \lim_{\Lambda \to \infty} Z_D\left(\frac{\mu^2}{\Lambda^2}, e_0\right) D_0\left(\frac{q^2}{\Lambda^2}, e_0\right) . \qquad (30)$$

Similarly, the physical finite charge e is given by an infinite rescaling of the bare charge e_0 that occurs in the Lagrangian

$$e = \lim_{\Lambda \to \infty} Z_e\left(\frac{\mu^2}{\Lambda^2}, e_0\right) e_0 . \qquad (31)$$

Notice that the physical coupling e now depends implicitly on the renormalization scale parameter μ. Thus, in QED, for example, it is not strictly sufficient to state that the fine structure constant $\alpha \approx 1/137$; rather, one must also specify the corresponding scale. From this point of view there is nothing magic about the particular number 137 since a change of scale would produce a different value.

At this stage, some words of consolation to a possibly bewildered reader are in order. It is not intended to be obvious how such infinite rescalings of infinite complex objects lead to consistent finite results! An obvious question is what happens with more complicated processes such as scattering amplitudes and particle production? These are surely even more divergent than the relatively simple photon propagator. How does one know that a similar rescaling procedure can be carried through in the general case?

The proof that such a procedure does indeed work consistently for *any* transition amplitude in the theory was a real tour de force. A crucial aspect of this proof was the remarkable discovery that in QED only a *finite* number (three) of such rescalings was necessary to render the theory finite. This is terribly important because it means that once we have renormalized a few basic entities, such as e_0, all further rescalings of more complicated quantities are completely determined. Thus, the theory retains predictive power—in marked contrast to the highly unsuitable scenario in which each transition amplitude would require its own infinite rescaling to render it finite. Such theories, termed nonrenormalizable, would apparently have no predictive power. High energy physicists have, by and large, restricted their attention to renormalizable theories just because all their consequences can, in principle, be calculated and predicted in terms of just a few parameters (such as the physical charge and some masses).

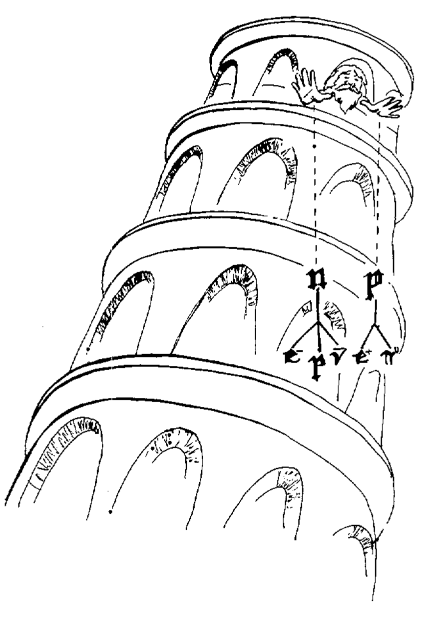
I should emphasize the phrase "in principle" since in practice there are very few techniques available for actually carrying out honest calculations. The most prominent of these is perturbation theory in the guise of Feynman graphs. Most recently a great deal of effort, spurred by the work of K. G. Wilson, has gone into trying to adapt quantum field theory to the computer using lattice gauge theories.* In spite of this it remains sadly true that perturbation theory is our only "global" calculational technique. Certainly its success in QED has been nothing less than phenomenal.

Actually only a very small class of renormalizable theories exist and these are characterized by dimensionless coupling constants. Within this class are gauge theories like QED and its non-Abelian extension in which the photon interacts with itself. All modern particle physics is based upon such theories. One of the main reasons for their popularity, besides the fact they are renormalizable, is that they possess the property of being *asymptotically free*. In such theories one finds that the renormalization group constraint, to be discussed shortly, requires that the large momentum behavior

In recent years there has been some effort to come to grips analytically with the nonperturbative aspects of gauge theories.

be equivalent to the small coupling limit; thus for large momenta the renormalized coupling effectively vanishes thereby allowing the use of perturbation theory to calculate physical processes.

This idea was of paramount importance in substantiating the existence of quarks from deep inelastic electron scattering experiments. In these experiments quarks behaved as if they were quasi-free even though they must be bound with very strong forces (since they are never observed as free particles). Asymptotic freedom gives a perfect explanation for this: the effective coupling, though strong at low energies, gets vanishingly small as q^2 becomes large (or equivalently, as distance becomes small).



In seeing how this comes about we will be led back to the question of *how the field theory responds to scale change.* We shall follow the exact same procedure used in the classical case: first we scale the hidden parameter (μ, in this case) and see how a typical transition amplitude, such as a propagator, responds. A partial differential equation, analogous to Eq. 25, is then derived using Euler's trick. This is solved to yield the general constraints due to renormalization analogous to the constraints of dimensional analysis. I will then show how these constraints can be exploited, using asymptotic freedom as an example.

The Renormalization Group Equation. As already mentioned, renormalization makes the bare parameters occurring in the Lagrangian effectively irrelevant; the theory has been transformed into one that is now specified by the value of its physical coupling constants at some mass scale μ. In this sense μ plays the role of the hidden scale parameter M in ordinary dimensional analysis by setting the scale of units by which all quantities are measured.

This analogy can be made almost exact by considering a scale change for the arbitrary parameter μ in which $\mu \rightarrow \lambda^{1/2}\mu$. This change allows us to rewrite Eq. 30 in a form that expresses the response of D to a scale change:

$$D\left(\frac{q^2}{\lambda\mu^2}, g(\lambda\mu^2)\right) = Z(\lambda)\, D\left(\frac{q^2}{\mu^2}, g(\mu^2)\right). \tag{32}$$

(From now on I will use g to denote the coupling rather than e because e is usually reserved for the electric charge in QED.)

The scale factor $Z(\lambda)$, which is independent of q^2 and g, must, unlike the Z's of Eqs. 30 and 31, be *finite* since it relates two finite quantities. Notice that all explicit reference to the bare quantities has now been eliminated. The structure of this equation is *identical* to Eq. 22, the scaling equation derived for the classical case; *the crucial difference is that $Z(\lambda)$ no longer has the simple power law behavior expressed in Eq. 18.* In fact, the general structure of $Z(\lambda)$ and $g(\mu)$ are not known in field theories of interest. Nevertheless we can still learn much by converting this equation to the differential form analogous to Eq. 25 that expresses scale invariance. As before we simply take $\partial/\partial\lambda$ and set $\lambda = 1$, thereby deriving the so-called *renormalization group equation:*

$$-q^2\frac{\partial D}{\partial q^2} + \beta(g)\frac{\partial D}{\partial g} = \gamma(g)\, D\,, \tag{33}$$

where

$$\beta(g) = \mu^2\frac{\partial g}{\partial \mu^2} \tag{34}$$

and

$$\gamma(g) = \frac{\partial \ln Z(\lambda)}{\partial\lambda}\Big|_{\lambda=1}\,. \tag{35}$$

Comparing Eq. 33 with the scaling equation of classical dimensional analysis (Eq. 25), we see that the role of the dimension is played by γ. For this reason, and to distinguish it from ordinary dimensions, γ is usually called the *anomalous dimension* of D, a phrase originally coined by Wilson. (We say anomalous because, in terms of ordinary dimensions and again by analogy with Eq. 25, D is actually dimensionless!) It would similarly have been natural to call $\beta(g)/g$ the anomalous dimension of g; however, conventionally, one simply refers to $\beta(g)$ as the β-function. Notice that $\beta(g)$ characterizes the theory as a whole (as does g itself since it represents the coupling) whereas $\gamma(g)$ is a property of the particular object or field one is examining.

The general solution of the renormalization group equation (Eq. 33) is given by

$$D\left(\frac{q^2}{\mu^2}, g\right) = e^{A(g)} f\left(\frac{q^2}{\mu^2}e^{K(g)}\right), \tag{36}$$

where

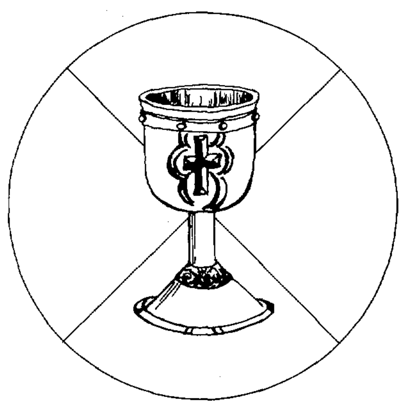
$$A(g) = \int^g dg\,\frac{\gamma(g)}{\beta(g)} \tag{37}$$

and

$$K(g) = \int^g \frac{dg}{\beta(g)}\,. \tag{38}$$

The arbitrary function f is, in principle, fixed by imposing suitable boundary conditions. (Equation 25 can be viewed as a special and rather simple case of Eq. 33. If this is done, the analogues of $\gamma(g)$ and $\beta(g)/g$ are constants, resulting in trivial integrals for A and K. One can then straightforwardly use this general solution (Eq. 36) to verify the claim that the scaling equation (Eq. 22) is indeed exactly equivalent to using ordinary dimensional analysis.) The general solution reveals what is perhaps the most profound consequence of the renormalization group, namely, that in quantum field theory the momentum variables and the coupling constant are inextricably linked. The photon propagator (D/q^2), for instance, appears at first sight to depend separately on the momentum q^2 and the coupling constant g. Actually, however, the renormalizability of the theory constrains it to depend effectively, as shown in Eq. 36, on only *one* variable ($q^2e^{K(g)}/\mu^2$). This, of course, is exactly what happens in ordinary dimensional analysis. For example, recall the turkey cooking problem. The temperature distribution at first sight depended on several different variables: however, scale invariance, in the guise of dimensional analysis, quickly showed that there was in fact only a single relevant variable.

The observation that renormalization introduces an arbitrary mass scale upon which no physical consequences must depend was first made in 1953 by E. Stueckelberg and A. Peterman. Shortly thereafter Murray Gell-Mann and F. Low attempted to exploit this idea to understand the high-energy structure of QED and, in so doing, exposed the intimate connection between g and q^2. Not much use was made of these general ideas until the pioneering work of Wilson in the late 1960s. I shall not review here his seminal work on phase transitions but simply remark that the scaling constraint implicit in the renormalization group can be applied to correlation functions to learn about critical exponents.* Instead I shall concentrate on the particle physics successes, including Wilson's, that led to the discovery that non-Abelian gauge theories were asymptotically free. Although the foci of particle and condensed matter physics are quite different, they become unified in a spectacular way through the language of field theory and the renormalization group. The analogy with dimensional analysis is a good one, for, as we saw in the first part of this article, its constraints can be applied to completely diverse problems to give powerful and insightful results. In a similar fashion, the renormalization group can be applied to *any* problem that can be expressed as a field theory (such as particle physics or statistical physics).



Often in physics, progress is made by examining the system in some asymptotic regime where the underlying dynamics simplifies sufficiently for the general structure to become transparent. With luck, having understood the system in some extreme region, one can work backwards into the murky regions of the problem to understand its more complex structures. This is essentially the philosophy behind bigger and bigger accelerators: keep pushing to higher energies in the hope that the problem will crack, revealing itself in all its beauty and simplicity. 'Tis indeed a faithful quest for the holy grail. As I shall now demonstrate, the paradigm of looking first for simplicity in asymptotic regimes is strongly supported by the methodology of the renormalization group.

In essence, we use the same modeling-theory scaling technique used by ship designers. Going back to Eq. 36, one can see immediately that the high-energy or short-distance limit ($q^2 \rightarrow \infty$ with g fixed) is identical to keeping q^2 fixed while taking $K \rightarrow \infty$. However, from its definition (Eq. 38), K diverges whenever $\beta(g)$ has a zero. Similarly, the low-energy or long-distance limit ($q^2 \rightarrow 0$ while g is fixed) is equivalent to $K \rightarrow -\infty$, which also occurs when $\beta \rightarrow 0$. Thus *knowledge of the zeros of* β, *the so-called fixed points of the equation, determines the high- and low-energy behaviors of the theory.*

If one assumes that for small coupling quantum field theory is governed by ordinary perturbation theory, then the β-function has a zero at zero coupling ($g \rightarrow 0$). In this limit one typically finds $\beta(g) \approx -bg^3$ where b is a calculable coefficient. Of course, β might have other zeroes, but, in general, this is unknown. In any case, for small g we find (using Eq. 38) that $K(g) \approx (2bg^2)^{-1}$, which diverges to either $+\infty$ or $-\infty$ depending on the sign of b. In QED, the case originally studied by Gell-Mann and Low, $b < 0$ so that $K \rightarrow -\infty$, which is equivalent to the low-energy limit. One can think of this as an explanation of why perturbation theory works so well in the low-energy regime of QED: the smaller the energy, the smaller the effective coupling constant.

Quantum Chromodynamics. It appears that some non-Abelian gauge theories and, in particular, QCD (see "Particle Physics and the Standard Model") possess the unique property of having a *positive* b. This marvelous observation was first made by H. D. Politzer and independently by D. J. Gross and F. A. Wilczek in 1973 and was crucial in understanding the behavior of quarks in the famous deep inelastic scattering experiments at the Stanford Linear Accelerator Center. As a result, it promoted QCD to the star position of being a member of "the standard model." With $b > 0$ the high-energy limit is

Since the photon propagator is defined as the correlation function of two electromagnetic fields in the vacuum it is not difficult to imagine that the formalism discussed here can be directly applied to the correlation functions of statistical physics.

related to perturbation theory and is therefore calculable and understandable. I shall now give an explicit example of how this comes about.

First we note that no boundary conditions have yet been imposed on the general solution (Eq. 36). The one boundary condition that *must* be imposed is the known free field theory limit ($g = 0$). For the photon in QED, or the gluon in QCD, the propagator G ($= D/q^2$) in this limit is just $1/q^2$. Thus $D(q^2/\mu^2, 0) = 1$. Imposing this on Eq. 36 gives

$$D\left(\frac{q^2}{\mu^2}, 0\right) = \lim_{g\to 0} e^{A(g)} f\left(\frac{q^2}{\mu^2} e^{K(g)}\right) = 1 \,. \tag{39}$$

Now when $g \to 0$, $\gamma(g) \approx -ag^2$, where a is a calculable coefficient. Combining this with the fact that $\beta(g) \approx -bg^3$ leads, by way of Eq. 37, to $A(g) \approx (a/b) \ln g$. Since $K(g) \approx (2bg^2)^{-1}$, the boundary condition (Eq. 39) gives

$$\lim_{q\to 0} f\left(\frac{q^2}{\mu^2} e^{1/(2bg^2)}\right) = g^{-a/b} \,. \tag{40}$$

Defining the dimensionless variable in the function f as

$$x \equiv \left(\frac{q^2}{\mu^2}\right) e^{1/(2bg^2)} \,, \tag{41}$$

it can be shown that with $b > 0$ Eq. 40 is equivalent to

$$\lim_{x\to\infty} f(x) = (2b \ln x)^{a/2b} \,. \tag{42}$$

An important point here is that the $x \to \infty$ limit can be reached either by letting $g \to 0$ or by taking $q^2 \to \infty$. Since the $g \to 0$ limit is calculable, so is the $q^2 \to \infty$ limit. The free field ($g \to 0$) boundary condition therefore determines the large x behavior of $f(x)$, and, once again, the "modeling technique" can be used—here to determine the large q^2 behavior of the propagator G.

In fact, combining Eq. 36 with Eq. 42 leads to the conclusion that

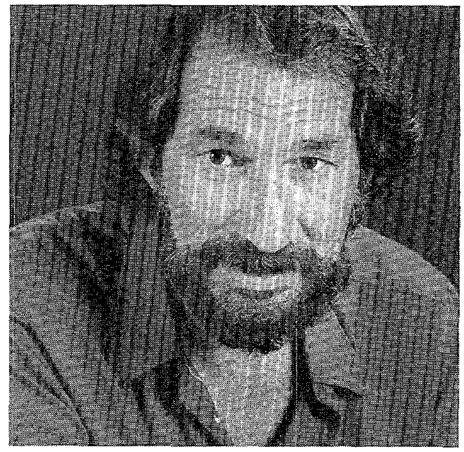
$$\lim_{q^2\to\infty} D\left(\frac{q^2}{\mu^2}, g\right) = e^{A(g)}\left(2b \ln \frac{q^2}{\mu^2}\right)^{a/2b} \,. \tag{43}$$

This is the generic structure that finally emerges: the high-energy or large-q^2 behavior of the propagator $G = D/q^2$ is given by free field theory ($1/q^2$) modulated by calculable powers of logarithms. The wonderful miracle that has happened is that all the powers of $\ln(\Lambda^2/q^2)$ originally generated from the divergences in the "bare" theory (as illustrated by the series in Eq. 28) have been summed by the renormalization group to give the simple expression of Eq. 43. The amazing thing about this "exact" result is that is is far easier to calculate than having to sum an infinite number of individual terms in a series. Not only does the methodology do the summing, but, more important, it justifies it!

I have already mentioned that asymptotic freedom (that is, the equivalence of vanishingly small coupling with increasing momentum) provides a natural explanation of the apparent paradox that quarks could appear free in high-energy experiments even though they could not be isolated in the laboratory. Furthermore, with lepton probes, where the theoretical analysis is least ambiguous, the predicted logarithmic modulation of free-field theory expressed in Eq. 43 has, in fact, been brilliantly verified. Indeed, this was the main reason that QCD was accepted as the standard model for the strong interactions.

There is, however, an even more profound consequence of the application of the renormalization group to the standard model that leads to interesting speculations concerning unified field theories. As discussed in "Particle Physics and the Standard Model," QED and the weak interactions are partially unified into the electroweak theory. Both of these have a negative b and so are *not* asymptotically free; their effective couplings grow with energy rather than decrease. By the same token, the QCD coupling should grow as the energy *decreases*, ultimately leading to the confinement of quarks. Thus as energy increases, the two small electroweak couplings grow and the relatively large QCD coupling decreases. In 1974, Georgi, Quinn, and Weinberg made the remarkable observation that all *three* couplings eventually became equal at an energy scale of about 10^{14} GeV! The reason that this energy turns out to be so large is simply due to the very slow logarithmic variation of the couplings. This is a very suggestive result because it is extremely tempting to conjecture that beyond 10^{14} GeV (that is, at distances below 10^{-27} cm) all three interactions become unified and are governed by the same *single* coupling. Thus, the strong, weak, and electromagnetic forces, which at low energies appear quite disparate, may actually be manifestations of the same field theory. The search for such a unified field theory (and its possible extension to gravity) is certainly one of the central themes of present-day particle physics. It has proven to be a very exciting but frustrating quest that has sparked the imagination of many physicists. Such ideas are, of course, the legacy of Einstein, who devoted the last twenty years of his life to the search for a unified field theory. May his dreams become reality! On this note of fantasy and hope we end our brief discourse about the role of scale and dimension in understanding the world—or even the universe—around us. The seemingly innocuous investigations into the size and scale of animals, ships, and buildings that started with Galileo have led us, via some minor diversions, into baked turkey, incubating eggs, old bones, and the obscure infinities of Feynman diagrams to the ultimate question of unified field theories. Indeed, similitudes have been used and visions multiplied. ■

Geoffrey B. West was born in the county town of Taunton in Somerset, England. He received his B.A. from Cambridge University in 1961 and his Ph.D. from Stanford in 1966. His thesis, under the aegis of Leonard Schiff, dealt mostly with the electromagnetic interaction, an interest he has sustained throughout his career. He was a postdoctoral fellow at Cornell and Harvard before returning to Stanford in 1970 as a faculty member. He came to Los Alamos in 1974 as Leader of what was then called the High-Energy Physics Group in the Theoretical Division, a position he held until 1981 when he was made a Laboratory Fellow. His present interests revolve around the structure and consistency of quantum field theory and, in particular, its relevance to quantum chromodynamics and unified field theories. He has served on several advisory panels and as a member of the executive committee of the Division of Particles and Fields of the American Physical Society.



Further Reading

The following are books on the classical application of dimensional analysis:

Percy Williams Bridgman. *Dimensional Analysis.* New Haven: Yale University Press, 1963.

Leonid Ivanovich Sedov. *Similarity and Dimensional Methods in Mechanics.* New York: Academic Press, 1959.

Garrett Birkhoff. *Hydrodynamics: A Study in Logic, Fact and Similitude.* Princeton: Princeton University Press, 1960.

D'Arcy Wentworth Thompson. *On Growth and Form.* Cambridge: Cambridge University Press, 1917. This book is, in some respects, comparable to Galileo's and should be required reading for all budding young scientists.

Benoit B. Mandelbrot. *The Fractal Geometry of Nature.* New York: W. H. Freeman, 1983. This recent, very interesting book represents a modern evolution of the subject into the area of fractals; in principle, the book deals with related problems, though I find it somewhat obscure in spite of its very appealing format.

Examples of classical scaling were drawn from the following:

Thomas McMahon. "Size and Shape in Biology." *Science* 179(1973):1201-1204.

Hermann Rahn, Amos Ar, and Charles V. Paganelli. "How Bird Eggs Breathe." *Scientific American* 240(February 1979):46-55.

Thomas A. McMahon. "Rowing: a Similarity Analysis." *Science* 173(1971):349-351.

David Pilbeam and Stephen Jay Gould. "Size and Scaling in Human Evolution." *Science* 186(1974): 892-901.

The Rayleigh-Riabouchinsky exchange is to be found in:

Rayleigh. "The Principle of Similitude." *Nature* 95(1915):66-68.

D. Riabouchinsky. "Letters to Editor." *Nature* 95(1915):591.

Rayleigh. "Letters to Editor." *Nature* 95(1915):644.

Books on quantum electrodynamics (QED) include:

Julian Schwinger, editor. *Selected Papers on Quantum Electrodynamics.* New York: Dover, 1958. This book gives a historical perspective and general review.

James D. Bjorken and Sidney D. Drell. *Relativistic Quantum Mechanics.* New York: McGraw-Hill, 1964.

N. N. Bogoliubov and D. V. Shirkov. *Introduction to the Theory of Quantized Fields.* New York: Interscience, 1959.

H. David Politzer. "Asymptotic Freedom: An Approach to Strong Interactions." *Physics Reports* 14(1974):129-180. This and the previous reference include a technical review of the renormalization group.

Claudio Rebbi. "The Lattice Theory of Quark Confinement." *Scientific American* 248(February 1983):54-65. This reference is also a nontechnical review of lattice gauge theories.

For a review of the deep inelastic electron scattering experiments see:

Henry W. Kendall and Wolfgang K. H. Panofsky. "The Structure of the Proton and the Neutron." *Scientific American* 224(June 1971):60-76.

Geoffrey B. West. "Electron Scattering from Atoms, Nuclei and Nucleons." *Physics Reports* 18(1975):263-323.

References dealing with detailed aspects of renormalization and its consequences are:

Kenneth G. Wilson. "Non-Lagrangian Models of Current Algebra." *Physical Review* 179(1969):1499-1512.

Geoffrey B. West. "Asymptotic Freedom and the Infrared Problem: A Novel Solution to the Renormalization-Group Equations." *Physical Review D* 27(1983):1402-1405.

E. C. G. Stueckelberg and A. Petermann. "La Normalisation des Constantes dans la Theorie des Quanta." *Helvetica Physica Acta* 26(1953):499-520.

M. Gell-Mann and F. E. Low. "Quantum Electrodynamics at Small Distances." *Physical Review* 95(1954):1300-1312.

H. David Politzer. "Reliable Perturbative Results for Strong Interactions?" *Physical Review Letters* 30(1973):1346-1349.

David J. Gross and Frank Wilczck. "Ultraviolet Behavior of Non-Abelian Gauge Theories." *Physical Review Letters* 30(1973):1343-1346.