A Quantal Hypothesis for Hadrons and the Judging of Physical Numerology

I.J. Good

Introduction

Readers who know little or no physics should not be deterred by the many mentions of physics in this chapter, for most of the required physics is spelt out in Appendix A. The chapter has two parts. Part 1 is an updating of Good (1988b) in which a quantal hypothesis is discussed concerning the rest masses of 'elementary particles'. The discussion depends, perhaps necessarily, on subjective probabilities. The hypothesis is largely *numerological* in the non-occult sense to be described more fully in Part 2. That part deals with the difficult topic of judging more general numerological assertions. Both parts might shed some light on how we choose between scientific theories in general. The chapter is somewhat speculative and this is appropriate in a *Festschrift* for John Hammersley. Moreover some of his early work (Hammersley 1950, 1954) dealt in part with a quantal hypothesis. Quantal hypotheses have occurred in physics, chemistry, genetics, and archaeology.

Warning. This chapter contains subjectively oriented material.

Part 1. Numerology for the Masses of Hadrons

A new edition of the Review of Particle Properties was issued in mid-February, 1989: see Particle Data Group (1988/89), abbreviated here as PDG89. Using this edition, and Cohen and Taylor (1987), together with information from Cohen (1989), who also drew my attention to Kinoshita (1989), I have recomputed some of the numbers, based on PDG86, published in Good (1988b); see also Good (1989c). In the present account, which is self-contained, I report the revised implications. The concept of the *relativistic fine structure constant* might have independent interest.

The first formula was

$$R(p,n) = \frac{m(n) - m(p)}{m(p)} \approx \frac{136\alpha}{720} = \left[\binom{4}{2}^{-1} + \binom{10}{2}^{-1}\right]\alpha \tag{1.1}$$

where $R(p,n)$ may be described as a 'proportional bulge', $m(n)$ and $m(p)$ denote the rest masses of the neutron and proton, and α denotes the fine structure constant (Sommerfeld 1916, p. 91; PDG89, p. 51; Cohen and Taylor 1987, p. 1139; updated by Kinoshita 1989),

$$\alpha = e^2/(\hbar c) = 1/137.0359914(1 \pm 8.1 \times 10^{-9}).$$

(For the sake of a simple and familiar formula for α, I have assumed electrostatic units, in which the permittivity of empty space is unity, but α has the same numerical value, whatever units are used, because it is a dimensionless constant.) Here e denotes the charge on the electron, c denotes the velocity of light, and $\hbar = h/(2\pi)$ where h denotes Planck's constant. (The expressions h and $\hbar$ are equally simple because energy $= h \times$ frequency $= \hbar \times$ angular frequency, while frequency and angular frequency are equally natural concepts. In other words, angles could be measured in *circumfians* instead of radians.) The constant α is generally regarded as measuring the strength of the electromagnetic forces. Kinoshita (1989) argues that α 'may be regarded as the most fundamental parameter of the physical world'.

A very similar conjecture is

$$R(p,n) = \frac{m(n) - m(p)}{m(p)} \approx \frac{136\alpha'}{720} \tag{1.1$'$}$$

where

$$\alpha' = \tanh^{-1}(\alpha) = 1/137.0335589(1 \pm 8.1 \times 10^{-9})$$

which may be regarded as the *relativistic fine-structure constant*. The idea of this minor adjustment to α is that whereas (i) $\alpha = v/c$ where v is the velocity of the electron in the lowest Bohr orbit (for example, Allen 1928, p. 56, or Whittaker 1953, p. 120), and (ii) if we replace the ratio v/c by its *rapidity* in the sense of A.A. Robb (Eddington 1930, p. 22), namely $\tanh^{-1}(v/c)$, then α is replaced by α'. Unlike two velocities, in the same direction, rapidities are strictly additive (in the Special Theory of Relativity). It does not seem too ad hoc to regard α' as a fundamental constant of nature, perhaps as fundamental as α athough the familiar formula for α, mentioned above, is simpler than that for α'.

The expression in (1.1) containing binomial coefficients ('triangulations') is included partly because 4 and 10 are two of the prominent integers in Eddington's *Fundamental Theory*, namely 4, 6, 10, 16, 120, 136, and 256. Moreover 4 is the number of dimensions of ordinary space-time whereas 10 is the number of dimensions in the currently most fashionable form of string theory when the six coiled up dimensions are included (see, for example, Schwarz 1988, p. 72). But the binomial expression in (1.1) will not be mentioned again in this chapter. Observe that the Eddingtonian integers are closely related to one another; for example, 6 and 10 are the lower and upper triangulations of 4, while 120 and 136 bear the same relationships to 16 and the reader will see other even more obvious relationships. It might not be necessary to rely on Eddington's judgement because, for example, a 16-dimensional torus occurs prominently in 'heterotic' string theory (Gross et al. 1985, p. 260). The 16 dimensions are described as 'internal'.

I am going to argue that (1.1) or (1.1′), and some allied assertions, are very probably 'correct'. Of course a convincing physical explanation would be much better. The meaning of *correctness* will be discussed in Part 2.

When trying to estimate the prior probability of (1.1) or (1.1′) it is appropriate to take a little physics into account; because the formulae are not *purely* numerological. The numerator $m(n) - m(p)$ on the left depends only on electromagnetic forces: see, for example, Rowlatt (1966 p. viii). It is therefore natural to have α or α' in the numerator on the right. Since $\alpha^{-1} = 137.0359914(8)$, and since 136 is so prominent in Eddington (1946), and is the closest nice integer to α^{-1} (see the ranking in Part 2), it is natural to introduce the number $\beta' = 1/(136\alpha)$ in preference to Eddington's $\beta = 137/136$ (Bond's factor).

Eddington even had names for 136 and 120. He called 136 'the basal multiplicity' and called 120 'the number of dormant components in the extended energy tensor'. In his theory (Eddington 1946, p. 30) 136 is expressed as $10^2 + 6^2$, and 120 occurs as $2 \times 6 \times 10$.

According to Slater (1957 p. 5), 136 is the number of mechanical degrees of freedom of the hydrogen atom and presumably of any similar two-particle system. We can also think of 136 and 120 as the numbers of real and imaginary components of a 16 by 16 Hermitian matrix or equivalently as the upper and lower triangulations of 16. These ways of expressing 136 and 120 come to much the same thing as Eddington's expressions, from a numerical point of view, because, for all n,

$$\binom{n}{2}^2 + \binom{n+1}{2}^2 = \binom{n^2+1}{2} \quad \text{and} \quad 2\binom{n}{2}\binom{n+1}{2} = \binom{n^2}{2}. \tag{1.2}$$

See also Eddington (1946, pp. 30 and 111) to obtain a further impression of why he liked the number 136. Of course $120 = 5!$ but I don't think Eddington was concerned with this fact.

If a formula contains both 120 and 136 it should not on that account be given independent complexity scores (or independent probabilities), one for 120 and one for 136. (See equation (2.7) below.) There should be an 'interaction term' subtracted from the total complexity (or divided into the product of the probabilities) to allow for the close relationship of these two numbers.

If there is any sense whatever in Eddington (1946), then the number 136 is very probably 'fundamental' (prominent in a good theory: see Part 2), more so than is suggested by its good 'ranking' in Table 2 in Part 2. I think the prior probability that it is physically fundamental is not more than $\frac{1}{4}$, because the testable predictions of Eddington's *Fundamental Theory* have been refuted, but I think the probability is at least $\frac{1}{10}$ because his intuition had been outstanding in other problems of physics, and also because of the current interest in the 16-dimensional torus mentioned above. (My subjective probabilities are my estimates of logical probabilities. There are scientists and statisticians who believe they do not use subjective probabilities. We can ask them for their subjective probabilities of these beliefs.) If 136 deserves to be called fundamental then 136α or $136\alpha'$ very probably deserves to be regarded as a fundamental physical constant because α occurs in Sommerfeld's theory of the hydrogen atom and so does 136 in Eddington's theory.

Again, $[m(n) - m(p)]/m(p)$ (which is of course dimensionless) seems like a reasonable measure of the ratio of the electromagnetic forces to the strong forces although *a priori* a denominator of $m(n)$ or $\frac{1}{2}[m(p) + m(n)]$ would be about as good as $m(p)$. We should therefore pay a factor of 3 (or a little less because the proton is the 'ground state') for 'special selection' of $m(p)$. Conditional on 136α or $136\alpha'$ making sense, we need to decide how impressed we should be by the denominator 720 on the right of (1.1). I think it is the simplest integer in the range of say $[600, 800]$, 625 and 729 being 'runners up'. In accordance with the comment (viii) to Table 2 (in Part 2), I assume that the first stage of information is that the denominator lies in this interval. This 'forces' about $\log_{10}(700/100) = 0.8$ correct significant digits. But the following Bayesian argument makes no use of the number of correct significant digits so it is fair to count 0.8 neither as a penalty nor as a reward.

There are only nine integers other than 720 in the range $[600, 800]$ that are of the simple form $2^a 3^b 5^c$, namely 600, 625, 640, 648, 675, 729, 750, 768, and 800. Moreover $720 = 6!$, it is a 'highly composite number' in the sense of Ramanujan (1915) (that is, it has more factors than any smaller number), and is also the product, 6×120, of two Eddingtonian integers, and one of the two is the 'twin' of 136. Of course 6! is the order of the symmetric group of degree six and the theory of finite groups is already basic to the theory of elementary particles. (See Appendix E for a distinctive property of this group.) So a physical explanation of the number 720 might depend on a theory entirely different from Eddington's. Perhaps it is relevant that

480 bosons occur in the heterotic string theory (Gross et al. 1985, p. 265) and we can think of 720 as sesqui-480. Moreover, 720 occurs prominently in Green et al. (1985, pp. 339, 340, 344) and in Candelas et al. (1985, p. 1123).

We need to judge the prior probability (say between $\frac{1}{20}$ and $\frac{1}{10}$) that there is an unknown *reason* why the denominator on the right of (1.1) or (1.1′) is an integer (or very close) and the prior probability (say between $\frac{1}{20}$ and $\frac{1}{5}$) that it is 720 given that it is an integer between 600 and 800. (Readers should make their own judgements.) With my judgements, the prior probability that (1.1) or (1.1′) is 'correct' lies between $1/(3 \times 10 \times 20 \times 20)$ and $1/(3 \times 4 \times 10 \times 5)$, that is, between $1/12000$ and $1/600$. Estimates should be made, together with some informal reasoning, by several particle physicists, but I have given my estimates to indicate a subjectivistic Bayesian way of thinking about the problem. For this application the approach in Part 2 pays too little attention to the physical background. Perhaps a reader can suggest another approach. Of course, as I said before, it would be better to find a convincing explanation (which by definition must be lucid) instead of just a probability estimate.

The conjectures or hypotheses that (1.1) or (1.1′) is exact, or at least appreciably more accurate than the experiments have proved, will be called H_0 or H_0' respectively. A different but related hypothesis, say H_0'', is that neither is exact but that there is an unknown physical reason why they are very good approximations. Such 'smudging of the null hypothesis' occurs in science more often than not because absolutely precise null hypotheses are rare. To save words, scientists and statisticians often omit explicit mention of this smudging and I shall follow this fashion and usually leave it to the reader to hold H_0'' in mind.

The values, based on PDG89, Cohen and Taylor (1987 p. 1142), and Kinoshita (1989), are

$$\frac{136\alpha}{720} = 0.00137838890 \pm 1 \times 10^{-11}, \tag{1.3}$$

$$\frac{136\alpha'}{720} = 0.00137841336 \pm 1 \times 10^{-11}, \tag{1.3′}$$

and

$$R(p,n) = 0.001378404 \pm 9 \times 10^{-9}, \tag{1.4}$$

but Cohen (1989) updates (1.4) by

$$R(p,n) = 0.001378416 \pm 6 \times 10^{-9}, \tag{1.5}$$

although he believes the uncertainty might well be as large as 8×10^{-9} or as small as 3×10^{-9}. To exaggerate the accuracy of H_0' one could say that the numerological estimate of $m(n)/m(p)$ is 1.0013784134 as compared with the current best experimental value 1.001378416(6). But it is fairer to subtract 1 from both sides when considering the proportional accuracy. (See Appendix C.)

We may infer that

$$\frac{136\alpha}{720R(p,n)} = 0.9999806 \pm \sigma \tag{1.6}$$

and

$$\frac{136\alpha'}{720R(p,n)} = 0.9999981 \pm \sigma, \tag{1.6$'$}$$

where the 'best' estimate of σ is 0.0000044. Thus the fate of the *exactness* of conjecture H_0 depends critically on whether the value of σ is appreciably larger than its nominal value. On the other hand H_0' would be a good fit, even if the uncertainty in (1.5) were, for example, only 3×10^{-9}. If this uncertainty has its nominal value of 6×10^{-9}, so that $\sigma = 0.0000044$ in (1.6′), then H_0' is correct to one part in at least $160,000$ (in accordance with the natural formula (2.5) of Part 2 which allows both for discrepancies and uncertainties), or one part in at least about 10^8 if we 'add 1 to exaggerate'. Another way to present the argument is to start with the simple observation that

$$R(p,n)/\alpha' = 0.1888892 \pm 0.0000008, \tag{1.6$''$}$$

and any schoolgirl would conjecture that $0.1\dot{8} = 17/90$ is exact. But an objective test for 'closeness to rationality' (Good 1969, p. 38, with N there taken as 90 or more) leads to an unimpressive P-value of between 0.078 and 0.098. To be impressed we must write $0.1\dot{8}$ in a more interesting way, for example, as 136/6!. We might want to judge too what fraction of rational numbers, in their lowest terms and with denominators 'subceeding' say 100, and not too distant from $R(p,n)/\alpha'$, can be written in at least as interesting a manner.

Now let X and Y be any pair of hadrons differing in having a u quark in X where there is a d quark in Y or vice versa, and let Y be the heavier. (The 'vice versa' applies only to the pair Λ, Σ^+.) Consider the experimental values of $720R(X,Y)/(136\alpha')$ shown in Table 1, calculated from the data in PDG89 combined with the latest estimate of $m(n)/m(p)$ (see equation (1.5)). As in Good (1989c) the Bayes factors (defined in Appendix B) listed in the last column refer to the hypotheses that each of the ratios is an integer. The method of calculating the factors is described in Appendix D. The product of these factors is the overall Bayes factor in favour of the hypothesis H_1 that all the ratios are integers (at least to an extremely good approximation). If we exclude the pair (B^+, B^0), which is the only pair involving the bottom (or beauty) quarks, the product is 17,500,000. For H_0' alone the factor is 83,000. That H_0' can be extended to H_1 is an example of 'consilience of induction' (Whewell 1847/1967). The overall Bayes factor of 17,500,000 does not allow for the fact that the numbers 48 etc. are all factors of 480 and 720 and are of the simple form $2^a 3^b$. To support my judgement that such numbers are attractively simple, note that Hardy (1940, p. 69) discusses Ramanujan's interest in numbers of this form. The numbers 48 etc. are orders of subgroups of the symmetric group of degree 6, and this fact might be relevant in an explanation.

I estimated the prior probability of H_0 as between 1/12000 and 1/600. For H_0' I am inclined to lean over backwards and to reduce these lower and upper probabilities to 1/36000 and 1/1800. Thus my (subjective) posterior odds of H_0', not allowing for the other evidence in Table 1, are between 2 and 46 (with a geometric mean of about 10). I shall be interested to know the reader's honest estimates.

The initial probability of H_1 is I think not much less than that of H_0', say by a factor of 5. (Don't forget that H_0' is a part of H_1.) The remaining Bayes factor in favour of H_1, from Table 1, is 211 not allowing for (B^+, B^0). Thus, not allowing for (B^+, B^0), my posterior odds that H_1 has a 'physical meaning' would be between 100 and 2000.

Quark compositions	X	Y	$\frac{R(X,Y)}{136\alpha'/720}$	Close integer	Bayes factor
(uud, udd)	p	n	0.9999981 ± 0.0000044	1	83,000
(uds, uus)	Λ	Σ^+	47.95 ± 0.055	48	4.798
(uus, uds)	Σ^+	Σ^0	1.94 ± 0.07	2	3.947
(uds, dds)	Σ^0	Σ^-	2.974 ± 0.048	3	7.177
$(u\bar{s}, d\bar{s})$	K^+	K^0	5.914 ± 0.046	6	1.511
(uss, dss)	Ξ^0	Ξ^-	3.54 ± 0.033	3 or 4	0.775
$(u\bar{c}, d\bar{c})$	$\overline{D}^0$	D^-	1.844 ± 0.11	2	1.327
$(u\bar{b}, d\bar{b})$	B^+	B^0	0.26 ± 0.15	-	-

TABLE 1. Experimental values of $R(X,Y)/(136\alpha'/720)$, based on PDG89 and Cohen (1989).

The result for (B^+, B^0) is somewhat of a setback. The closest integer to 0.26 ± 0.15 is of course zero, but it seems physically unlikely that B^+ and B^0 have the same rest mass (Blecher 1989). (Moreover 0 is of course not a factor of the two numbers 720 and 480 mentioned above.) According to PDG89 (pp. 20 and 218), the mass difference is 1.9 ± 1.1 MeV$/c^2$. This estimate was based on only one experiment (performed by 85 experimenters), so the estimated standard error might not be wholly reliable, but it now seems most unlikely that the pair (B^+, B^0) can be included in our conjecture, H_1.

If we are to retain H_1 we must assume either that

(i) the masses of B^+ and B^0 are equal; or

(ii) the standard error for the (B^+, B^0) mass difference is much larger than 1.1 MeV/c^2 (to allow $R(B^+, B^0)$ to be equal to $R(p,n)$); or

(iii) the hypothesis applies only to particles made of light quarks (u, d, s and c quarks: see PDG89, p. 102, col. i). Then the pair (B^+, B^0) should be excluded. In PDG89 (p. 6, col. ii) the light quarks are defined as the u, d, and s quarks (the c quark not being included), and, with this definition, I would have to exclude the pair $(\overline{D}^0, D^-)$ also, but this would lose a Bayes factor of only 1.327. See Appendix A for information about the masses of the quarks. As in my previous work I have not allowed here for the elegance of the integers 48 etc. although this elegance clearly supports H_1.

I think it will be generally agreed that (iii) is much more reasonable than (i) or (ii). The restriction to the light quarks is not a big restriction because the hypothesis was already restricted to pairs of particles differing only in the interchange of a down quark with an up quark, and these are by far the lightest of the quarks.

The bottom quark is so much heavier than u, d, and s that the removal of the pair (B^+, B^0) is only a small loss of beauty for H_1. As a good analogy, consider the following argument that might have been used against the Copernican system. The moon always shows the same face towards the earth (as if the moon were embodied in a sphere that rotates around the earth) whereas the earth shows a variable face towards the sun. Copernicus could have replied, without much adhockery, that the two pairs (Earth, Moon) and (Sun, Earth) are quantitatively so different that a qualitative difference is not surprising.

We should not forget that the hypothesis H_0' is 'logically' independent of the rest of H_1; but probabilistically (in the epistemic sense) they support one another because they imply that $R(X,Y)$ has a fundamental interpretation, unknown at present but presumably related to the relative strengths of the strong and the electromagnetic forces.

The numerical values of $R(X,Y)/(136\alpha'/720)$ are changed negligibly if the denominator is replaced by $R(p,n)$, so, even if (1.1′) or (1.1) is coincidental, we still have evidence that *the proportional bulges* $R(X,Y)$ *bear a simple rational relationship to one another* when heavy quarks are not involved.

Further Discussion and a Modification of H_1

The restriction to the light quarks suggests that H_1 might be only a good approximation (because if heavy quarks damage the numerology why shouldn't the light ones do a little damage?). Thus the approximations to integers might not be as exact as I hope, but I believe they are at least *close enough to demand an explanation.* For the time being, I define the 'null hypothesis' H_1 in terms of exact integers.

Note that the experimental values of $R(X,Y)/(136\alpha'/720)$ in Table 1 all fall short of the integers 1, 48, etc., though this is ambiguous for the pair (Ξ^0, Ξ^-), and the 'short-fall' is statistically significant only for (B^+, B^0). If we combine the short-falls for the first seven pairs (excluding (B^+, B^0)), each divided by its 'uncertainty' (regarded as a standard deviation), we get 7.42. Since $7.42/\sqrt{7} = 2.80$, the null hypothesis H_1 that all the seven ratios are integers might be rejected with a P-value of 0.0052 (the double-tail) if the non-null hypothesis asserts that the short-falls are all positive or all negative. One would then wish to consider a modified hypothesis H_2 that the true values of the ratios $R(X,Y)/(136\alpha'/720)$ are the integers 1, 48, 2, 3, 6, 4, 2 (and 1?) minus small quantities that are monotonically related to the masses of the corresponding quarks even if the pair (B^+, B^0) is included. Among those pairs of particles for which the strangeness (which is minus the number of s quarks) is 0, ±1, or ±2, the best result is for strangeness 0 (the Bayes factor for H_1 being 83000) and the results are good for the four pairs of particles with strangeness ±1 (Bayes factor 205). For the pair (Ξ^0, Ξ^-), where the strangeness is -2, the hypothesis H_1 loses a little ground (the Bayes factor being 0.775). For giving H_2 a more precise formulation we require improved experimental values for the mass differences, especially for $m(\Xi^-) - m(\Xi^0)$.

I believe that the numerology is good enough to be taken very seriously. How can the numbers 136, 720, 48, etc. be explained by a coherent and intelligible physical theory?

The pair (Λ, Σ^+) stands out in Table 1 in that the value of $R(\Lambda, \Sigma^+)$ is much larger than for the other pairs. This pair is also exceptional in that (i) the 'isospins' I of Λ and Σ^+ are unequal, in fact $I(\Lambda) = 0$ and $I(\Sigma^+) = 1$; and (ii) it is the only 'vice versa' pair, as mentioned a few sentences below equation (1.6″). We could avoid this 'vice versa' property by changing the sign of the numerator of $R(\Sigma^+, \Lambda)$, and then 48 would be replaced by -48 as a 'Pontryagin number' in Green et al. (1985, p. 338).

The pair (Λ, Σ^0) does not qualify for Table 1 because Λ and Σ^0 have the *same* quark composition, *uds*. Moreover the value of $R(\Lambda, \Sigma^0)$ had to be close to $R(\Lambda, \Sigma^+) + R(\Sigma^+, \Sigma^0)$ because the masses of Λ, Σ^+ and Σ^0 don't differ much. In fact $R(\Lambda, \Sigma^0)/(136\alpha'/720) = 50.02 \pm 0.065$.

Philosophical Discussion

Eddington (1946) and Einstein (1949, p. 63) believed that the fundamental constants of physics could be calculated from qualitative assumptions just as π can be calculated from the assumptions of Euclidean geometry. Eddington's main speculations along these lines, though stimulating, seem to have been fairly unsuccessful. There is now a theoretical argument, though an unconvincing one, that it is impossible to attain his Pythagorean goal and I shall mention that argument. First note, however, that if his dream is unattainable then the prior probability that (1.1) or (1.1′) is exact is somewhat reduced. It is not *much* reduced because those equations determine only one computable constraint on the fundamental constants.

The theoretical argument for believing that Eddington's goal cannot be achieved is related to one interpretation of the so-called anthropic principle, better called the *biotic* principle. This interpretation of the principle asserts that the fact that carbon-based life exists implies (but of course does not cause) severe constraints on the fundamental constants; see, for a review, Barrow & Tipler (1986). (The name *anthropic principle* is misleading because it is anthropomorphic to base the deduction on the fact that *humans* exist, and 'astronomomorphic' on the fact that *astronomers* exist: see Barrow & Tipler (1986, p. 15) who say wittily 'certain properties of the universe are necessary if it is to contain carbonaceous astronomers like ourselves.') An early and unconvincing example of the biotic principle can be read into an argument by Boltzmann. He said that the low entropy in the neighbourhood of the earth is *a priori* exceedingly improbable but in an infinite universe everything that is possible occurs somewhere. (See, for example, Porter 1986, pp. 215 and 216.) Boltzmann could have reversed the argument and said that the existence of life on earth is evidence that the universe is infinite or perhaps that there are an infinite number of universes. Another hypothesis is the one of which Laplace 'had no need', the existence of God. A third hypothesis is that Boltzmann's argument is simply wrong. For an extensive discussion of relevant matters see Prigogine and Stengers (1984). To put the matter in general terms: if an explanation requires an amazing coincidence, whether in a legal or a scientific context, then we have probably overlooked something. (Compare the discussion of Sherlock Holmes's law in Good 1950, page 67; and the seeming occurrence of two nearly independent murders in the same house: see, for example, *The Times*, London, 19 September 1970, p. 19.)

As another example, theories of the origin of the solar system based on the close encounter of a second star were proposed because there were difficulties in the theories of a nebular origin. The unlikelihood of a close encounter encouraged astronomers to remove the difficulties in the nebular theories. (See Nieto 1972.) Similarly, the isotropy of the universe would be explained by the 'chaotic cosmological principle' (Barrow & Tipler 1986, §6.11). According to this principle the present isotropic condition of the universe does not require isotropy in the initial conditions, but the principle runs into difficulties pointed out by Collins and Hawking. If, however, the chaotic principle is abandoned the isotropy seems to be an amazing coincidence, and one way to begin to explain the coincidence is to use the biotic principle. But another explanation of the coincidence would be that the chaotic cosmological principle will probably be reinstated by means of a theoretical correction. An example of such a loophole is indeed mentioned by Barrow & Tipler (1986 p. 425).

There is nothing wrong with the biotic principle as such, but a form of it that requires an amazing coincidence needs to be extremely watertight. If such an argument is valid then we are faced with a metaphysical option; very probably either God exists (a 'design argument') or there are myriads of universes (or both). In this case, to believe in Eddington's dream would be analogous to supposing that the first several hundred digits of π would, under some simple encoding into letters, spell out a sonnet by Shakespeare. (To believe that hypothesis you'd have to be a numerologist in the occult sense or perhaps a monkey.) In short, the attaining of Eddington's dream, (1.1′) being a step in that direction, would be effectively incompatible with the 'metaphysical option'.

Part 2. The Judging of Numerological Assertions

The quantal hypothesis in Part 1 was not entirely numerological for it made use of some physical theory though not convincingly. In the present part we consider how one might try to judge numerology with less explicit reference to physical theory.

At one time *numerology* meant divination by numbers, but during the last few decades it has been used in a sense that has nothing to do with the occult and is more fully called *physical numerology.* The expression *numerology* has been applied to one or more proposed formulae of the form (a 'null hypothesis')

$$x = y \text{ or } x \approx y$$

where x and y are numbers that might involve physical constants. The formula is regarded as numerological by a person who thinks it has not been explained. There will also be people who know that an explanation has been proposed, but who have not understood the explanation. These people either accept the judgements of the understanders or they might treat a formula *as if* it were numerological and judge it partly by its simplicity (or elegance) and its accuracy, or they might adopt a compromise position.

Numerological activity can be regarded as the search for patterns in collections of numbers so it is a kind of exploratory data analysis, though not necessarily of a Tukeyesque kind. The ultimate aim is to help in the formulation of scientific theories.

A statement can be partly numerological and partly scientific. For example, in 1815, William Prout suggested that all atomic weights are multiples of that of hydrogen and, as an inference (which contains a little truth), that all elements are composed of hydrogen (see Ihde 1964, p. 154). The evidence at the time was weak but the estimates of 1960 (Ihde, p. 142) are impressive even if we don't take isotopes into account. Of 49 atomic weights given in Ihde's table, 27 are within 0.1 of an integer (apart from oxygen whose atomic weight was taken there as 16 exactly, by definition), whereas only $49/5 = 9.8$ would be expected if Prout's hypothesis were entirely wrong. That Prout's hypothesis is false is obvious from the table, but the discrepancy between 9.8 expected and 27 observed shows that there is enough in the hypothesis to *demand an explanation*. This anachronistic example verifies that a hypothesis can be clearly wrong and yet clearly partly right at the same time. To say it was right would be absurd whereas to say simply that it was wrong would be extremely misleading. Such examples are not exceptional.

Even allowing for isotopes, the atomic weights are not exact integers because of the so-called 'packing fractions' of special relativity, and because the masses of protons and neutrons are not exactly equal. Prout's hypothesis is a good illustration of the need to take approximate laws seriously. Unfortunately such laws cause difficulties for statisticians. For the sake of simplicity, statisticians often assume sharp null hypotheses.

A *sharp hypothesis* is another name for a simple statistical hypothesis. We often test sharp hypotheses although we know they are most unlikely to be exactly true. This activity can often be justified on grounds of simplicity: if the available evidence is not sufficient to reject a sharp hypothesis we might find it convenient to regard it as true to an adequate approximation. But if we *reject* a sharp hypothesis we should be more careful. We ought sometimes to consider smudging, smearing, or desharpening the null hypothesis (Laplace 1774; and independently, but somewhat later, Good 1950, pp. 90–94). If we don't desharpen there'll be a risk of rejecting a hypothesis that points in the general direction of the truth. It is unfortunate that the terminology of 'rejection' is entrenched in statistical jargon, for it causes us, by the 'tyranny of words' (Chase 1938; Good 1969, p. 62), to be too ready to ignore hypotheses that are (probably) 'wrong' but might be suggestive of something better. Examples are Prout's law and the Titius-Bode law. (Good 1969, especially p. 62; Good 1971; Efron 1971; and Nieto 1972.) Perhaps the term *rejected* when applied to hypotheses or theories should often be replaced by *smudged* or *shown to be inexact*.

It is only in the non-occult senses that the word *numerology* is used in this chapter. An example (Lenz 1951) is

$$m(p)/m(e) \approx 6\pi^5 \tag{2.1}$$

where $m(p)$ and $m(e)$ denote the rest masses of the proton and electron. This formula is correct to one part in 50,000 but is now known to be inexact. (See Table 3.) An example from pure mathematics that will seem to some to be numerological, but which has a known explanation (Weber, c. 1908,

§125), is

$$\exp\left(\frac{\pi}{3}\sqrt{67}\right) \approx \text{number of feet in a mile,} \tag{2.2}$$

which is correct to one part in 300,000,000. (Naturally Weber doesn't mention feet or miles.) Of course a mere numerical computation doesn't explain *why* the left side is so close to an integer.

There have been a few examples of numerology that have led to theories that transformed society: see the mention of Kirchhoff and Balmer in Good (1962, p. 316) and in Barrow & Tipler (1986, p. 219ff) and one can well include Kepler on account of his third law. It would be fair enough to say that numerology was the origin of the theories of electromagnetism, quantum mechanics, gravitation, and quantitative chemistry (by Proust's law and Prout's hypothesis). So I intend no disparagement when I describe a formula as numerological.

There is, however, much bad numerology, and we shall discuss methods for attacking the difficult problem of distinguishing the good from the bad when the distinction is not obvious. This project is a special case of the even more difficult and more familiar one of distinguishing between good and bad scientific theories or hypotheses, and the consideration of the special case might provide some insights for the more general project which is of course a main problem in the philosophy of science. I believe that part of the solution should depend on the concept of explicativity (Good 1977) but this approach cannot be carried out without first thinking along somewhat Bayesian lines as in the present work. My purpose here is to contribute ideas that might lead to a more satisfactory solution than has been attained so far. See also Cover (1973).

When a numerological formula is proposed, then we may ask whether it is *correct*. The notion of *exact correctness* has a clear meaning when the formula is purely mathematical, but otherwise some clarification is required. I think an appropriate definition of *correctness* is that the formula has a good explanation, in a Platonic sense, that is, the explanation could be based on a good theory that is not yet known but 'exists' in the universe of possible reasonable ideas. A good but undiscovered theory is like a work of art waiting to be chiselled out of a block of marble.

A formula might be *partly correct* in the sense that some reasonable theory (possibly unknown) might explain why it is a good approximation. Such a theory is *known* for the mathematical formula (1.2) (with 5280 on the right of course). Lenz's formula (2.1) is so simple, and so nearly true, that I would not be surprised if it turned out to be partly correct. It will be discussed again below. Leaving such approximations aside, a correct numerological formula might sometimes be used for predicting new decimal places of observations before an explanation is found. A formula might be partly numerological in the sense that it can already be partly supported by rational arguments, although there is not yet a good explanation. Part 1 provides an example. There is no precise demarcation between numerology and a scientific theory.

We naturally ask the following questions (which are not *demands*) of a numerological formula:

(a) Was it 'consistent' with the experimental observations when it was published? (See below for the meaning of *consistency.*)
(b) Is it consistent with the latest experimental observations?
(c) Has its accuracy improved or deteriorated since it was first suggested?
(d) To how many significant digits (for example, in radix 10) is it correct according to the latest experimental results?
(e) How complex or simple is it? Here, as in (f), subjectivity can hardly be avoided.
(f) What, in some sense, was its prior probability of being correct (without allowing for the experimental results)?
(g) If it is consistent with the latest experimental results, is it likely to remain consistent with future experiments? This question can be attacked by Bayes's theorem, though still with difficulty, if an answer to (f) can be accepted.
(h) Is is part of a set of similar formulae, in other words does it satisfy, in a numerological sense, the desideratum of 'consilience of induction'? (Whewell 1847/1967, Vol. 2, pp. 77–78.)

We now elaborate on aspects (a) to (g).

(a) and (b). *Consistency with experiment.* Let x and y have estimated standard errors σ and τ, and estimated correlation coefficient ρ. (For the sake of simple notation I am not writing $\hat{\sigma}$, $\hat{\tau}$, and $\hat{\rho}$.) Let

$$s = \frac{|x-y|}{(\sigma^2+\tau^2-2\rho\sigma\tau)^{1/2}} \tag{2.3}$$

and call s the *sigmage* (to rhyme with *porridge*) of $|x-y|$. (In the special case of a purely mathematical piece of numerology we have $\sigma = \tau = 0$, and the sigmage is infinite unless the piece of numerology is exact.) We should hold in mind that, historically, standard errors of physical constants have tended to be too small (Henrion & Fischhoff 1986). Also ρ is seldom presented, and we might have to assume that $\rho = 0$. We might say that the equation $x = y$ is *inconsistent* with present experiments if $s > 3$ (compare Jeffreys 1937, p. 83, or 1957, p. 72) and *consistent* if $s \leq 2$. Like most things, consistency is a matter of degree, for example $s = 1$ is appreciably more consistent with experiment than is $s = 2$. In fact if $s \leq 1$ it would hardly be misleading to say that the agreement with experiment is *perfect* and in many Bayesian models such good agreement would give a little support to the null hypothesis, the *Bayes factor* (see Appendix B) being roughly proportional to $\exp(-s^2/2)$. A rough rule of thumb for the weight of evidence (logarithm of a Bayes factor), against the null hypothesis, would be $\frac{1}{2}(s^2-1)$ 'natural bans' or $0.217(s^2-1)$ bans (Turing's name for decimal units). This takes the 'cut even' sigmage as unity. As recalled by Good (1989a), if it is possible for the outcome of an experiment to undermine a hypothesis, then it is also possible for the outcome to support the hypothesis provided that all outcomes are observable. A proof of this little theorem is obvious, once the theorem is stated, but the result might surprise some Popperians and non-Bayesian statisticians. An example of a hypothesis tested by an experiment for which not all possible outcomes are observable is that there is life after death.

(c) and (d). *Number of correct significant digits (n.c.s.d.).* People sometimes use an integer to measure the number of correct significant digits of an approximation, but this constraint is of course unnecessary and ties one too much to radix 10. It is more informative to measure the number of correct significant digits by means of a real positive number and to take the standard errors into account. In Good (1988a) I suggested two related definitions but made an error (plus some verbal ones). Of the two, I think most statisticians will prefer *the smaller of the two numbers*

$$\left|\log_{10}\left|\left\{\left|\log_e\left(\frac{x}{y}\right)\right| \pm \left(\frac{\sigma^2}{x^2}+\frac{\tau^2}{y^2}-\frac{2\rho\sigma\tau}{xy}\right)^{1/2}\right\}\right|\right|, \tag{2.4}$$

where σ, τ, and ρ denote respectively the 'uncertainties' or standard errors of x and y and the estimated correlation between the measurements. The proportional accuracy is then defined as

$$1 \text{ in } 10^{\text{n.c.s.d.}}, \tag{2.5}$$

where one may well add the words 'or better' or 'at least' when $|x-y| \leq 2(\sigma^2+\tau^2-2\rho\sigma\tau)^{1/2}$. The proportional accuracy (2.5) does not depend on the radix 10.

To allow both for the n.c.s.d. and the sigmage s, the rule of thumb mentioned above for the weight of evidence in bans in favour of the null hypothesis (to be added to the prior log-odds to get the posterior log-odds) is

$$\text{n.c.s.d.} - 0.217(s^2-1), \tag{2.6}$$

the corresponding Bayes factor being

$$10^{\text{n.c.s.d.}} e^{-(s^2-1)/2}. \tag{2.6A}$$

But the most difficult and controversial problem is the estimation of the initial odds of the null hypothesis. The initial odds, the n.c.s.d., and the sigmage are all relevant for an evaluation.

(e) *How complex or simple is the numerology?* Although no one has found an entirely satisfactory definition of simplicity or complexity, it is obviously easier to *force* several significant digits of accuracy by inventing complex formulae. Certainly simplicity has something to do with brevity as hardly anyone would deny apart from some Humpty Dumpty philosophers who change the meanings of words completely. In Good (1968) I suggested that the complexity $\kappa(H)$ of a proposition H could be defined by a measure of information,

$$\kappa(H) = -\log P(H), \tag{2.7}$$

where $P(H)$ denotes the prior probability of H. This definition, which links headings (e) and (f), leads to some difficulties when applied to arbitrary propositions (for example, $H \vee K$ is less simple but more probable than H), and a modification was discussed by Good (1975, pp. 46–48 = 1983, pp. 154–156) where negations and logical disjunctions are avoided. For the computable numbers that I shall discuss I believe formula (2.7) is reliable enough. The base of the logarithms merely determines the unit of complexity and I shall assume base 10 for convenience. My article on surprise, Good (1984/88), contains some further discussion of this matter.

A natural axiom for complexity is

$$\kappa(H\&H') = \kappa(H) + \kappa(H' \mid H) \tag{2.8}$$

in which the second term is a *conditional complexity*. This axiom is an immediate deduction from (2.7) when (2.7) is applicable. Also the axiom forces (2.7) if $\kappa(H)$ is assumed to depend only on $P(H)$, just as in information theory (Good 1950, p. 75).

When considering the complexity of an *arbitrary-looking* real positive number based on measurement (and hence not 'computable') such as 4357.073 we should hold in mind that this almost certainly is just an abbreviated way of denoting an estimate with a standard error which is *sometimes* roughly equal to 0.0005 (or this divided by $\sqrt{3}$). The definition of the complexity of the number ought to depend on how large it is as compared with its standard error, say σ (Rissanen 1983, p. 419). An approximate measure of the complexity is $\log_{10}(4357.073/\sigma)$ in decimal units. This definition is consistent with (2.7) if we assume the Jeffreys-Haldane improper prior density $1/u$ for a random positive number u. The definition does not depend on the use of the decimal system any more than the theories of information and weight of evidence depend on the units 'bits', 'decibans', etc.

(f) and (g). *What was the prior (epistemic) probability?* We shall be discussing, in two different senses, the prior probability of a computable number x. Sometimes the prior probability refers to the probability that a computable number chosen at random from some context will be x, and sometimes it refers to the prior probability that $x = y$ where y is some physical quantity. These two senses are very different and to distinguish between them I shall call the prior probability in the former sense the *preprior* probability. It is the probability of coming across x in some context and is not at all the same as the prior probability that $x = y$. But sometimes the context is influenced by that in which y arises. The complexity of x can be defined as minus the logarithm of its preprior probability. Thus the complexity of x might be relative to a context. This comment is in the spirit of the more general concept of complexity discussed in Good (1975), as referenced here just below equation (2.7). The complexity of the equation $x = y$ is equal to that of y plus that of x in the contexts in which y appears.

The preprior probability of x is especially pertinent when a piece of numerology is consistent with experimental results. It is more fundamental than the concept of simplicity (or complexity) but the concept of simplicity helps one to judge the preprior probability. The preprior probability can be expressed as the product of that of the functional form or algorithm assumed in the piece of numerology, and that of the choice of computable numbers that are the arguments of this functional form. (More precisely, this probability would be computed for the simplest algorithm that gives the same answer. Still more precisely, the probability should be estimated for all the algorithms, the maximum of these probabilities being selected.) In this article I have mainly in mind the functional form of $a_1 a_2 \ldots a_m/(b_1 b_2 \ldots b_n)$ (a *pseudorational number* so to speak), where the a's and b's are positive integers or π or e or one of a few operations such as square-rooting, but what I say about the probabilities of numbers of this form might well apply to most functional forms.

We have mentioned the Jeffreys-Haldane prior for measured quantities but for computable numbers it is much more difficult to suggest reasonable prior probabilities. In principle, this can be done consistently because computable numbers (like statable hypotheses) form only a countable set. This problem of assigning probabilities to computable numbers was propounded but not attacked by Good (1950 p. 55n). Here I shall try to make some small steps towards a solution. See also Rissanen (1983). In Part 1 I gave an example in which the computable number is 136/720. This example shows that in a special case one might need to invoke special arguments.

When judging any prior probability, such as the occurrence of an integer, it is necessary to take the context into account. Consider, for example, M.H.A. Newman's bus problem (Jeffreys 1938, p. 186): We know that a town has N buses, numbered 1 to N, but we don't know N, and we have little idea of the population of the town. We see a bus numbered k. What is the posterior probability distribution of N? Jeffreys obtained a reasonable solution by assuming an (improper) prior roughly proportional to $1/N$ by analogy with the Jeffreys-Haldane prior density $1/x$ for a positive random real number x (which has the property that powers of x have the same distribution as x has). This can also be called suggestively the log-uniform prior. Some proper distributions that 'approximate' the Jeffreys-Haldane distribution are mentioned by Good (1989b).

When the density $1/x$ is 'approximated' by a proper density $f(x)$ the definition of the complexity $\kappa(x)$ of a real number x with standard deviation σ needs to be modified. A natural definition is $\kappa(x) = -\log_{10} f(x/\sigma)$.

For general physical numerology one needs to put the positive integers and a few operations in some rank order, and a distribution somewhat resembling $1/N$ should be applied to the *ranks* and not to the original integers. For example, we would tend to favour composite integers over primes of about the same size because many formulae in physics, and in mathematics, consist of the products of various quantities. It would be possible to sample computable numbers in many texts, but so far my sample is too small. It will be described below. When sampling we must of course avoid including numbers, like 100, when they occur as approximations merely because of the use of radix 10. Also angles should be measured in radians or 'circumfians' and not in degrees.

Apart from the large amount of work required to obtain an adequate sample, there are further sampling difficulties such as

(i) A sample of individual numbers ignores relationships between pairs or larger groups of numbers; for example, the two 'triangulations' of n, namely $\frac{1}{2}n(n-1)$ and $\frac{1}{2}n(n+1)$, are logically related so it would be unfair to multiply their preprior probabilities when they both occur in the same piece of numerology. This case occurred in Part 1.

(ii) The field of theoretical physics is not homogeneous, and which population is appropriate for a given piece of numerology might be difficult to judge.

For these and other reasons a large reliance on subjective judgement is necessary. This necessity was exemplified in Part 1 and will be further exemplified below. Science seems objective only when it is compared with non-science. Fortunately, an exact ordering is unimportant.

My sample consisted of all the integers and π, 2π, and e occurring in Ramond (1985) and Goldman & Haber (1985), although e had zero frequency. I chose these two articles because my examples relate to elementary particles and because these two articles contain a lot of integers. I included also the operations $\times$, $\div$, $+$, $-$, squaring, powering other than squaring, ! (factorial), and triangulating up or down. Let us think first of numerological formulae of the form $x = y$ where x is of the form $a_1 a_2 \ldots a_m/(b_1 b_2 \ldots b_n)$ in which the a's and b's all belong to the set consisting of positive integers, etc., and their square roots, squares and other integer powers. I assume further that ζ and $1/\zeta$ have equal complexities and preprior probabilities, where ζ is any computable number, so at first we need to consider only products $a_1 a_2 \dots a_n$. For this limited project the preprior probabilities of $\times$, $\div$, $+$, $-$, and 1 are irrelevant and are mentioned only in the *heading* of Table 2 and not in the body of that table. The table consists of the integers etc. in a ranking order that I have chosen mainly subjectively because my samples are too small to give very much information about the ranking.

A principle of this ranking is to begin by thinking of the (unique) prime factorization of integers, where small primes are preferred to larger primes. Note that unique factorization also applies to rational (or pseudorational) numbers expressed in their lowest terms, even if π and e are regarded as pseudoprimes, for presumably π is not a rational or pseudorational power of e. (That looks like a difficult problem in the theory of numbers.) A simple measure of complexity of a rational number

$$2^{n_1} 3^{n_2} 5^{n_3} 7^{n_4} \cdots , \tag{2.9}$$

where the n_i are integers, positive, negative or zero, would be

$$\lambda_1 |n_1| + \lambda_2 |n_2| + \lambda_3 |n_3| + \cdots , \tag{2.10}$$

where λ_r is $-\log$(relative frequency of the rth prime). The frequencies of 2 and 3, as prime factors, in the two samples combined, were respectively 1237 and 269 (out of 1665), so the ratio of the frequencies of 2 and 3, regarded as *prime factors*, differs substantially from the ratio when they are regarded just as integers. This discrepancy suggests that the simple rule (2.10) can be only a crude approximation. If used, the λ's could instead be estimated by means of monotonic regression, where the 'dependent variable' is someone's subjective measure of complexity.

The samples suffer from the same disadvantage as the use of continuous patches of prose for sampling the frequencies of English words — that a specific number can occur many times in a single context and yet can be rare in a more general population. I have listed the two samples separately to show how much they differ. This difference shows the desirability of a very large sample of samples. Having only two samples is like estimating the time by consulting two bad clocks, and averaging the result. But I shall regard the sum of the two samples as adequate, in this early discussion, for estimating a smooth fit to the relative frequencies of the *ranks*.

A fairly good smoothing is given by

$$p_r = \frac{2}{(r+1)(r+2)} \quad (r = 1, 2, 3, \dots), \tag{2.11}$$

where p_r denotes the probability of the *rank* r. (For example, the rank of 2 is 1.) Compare (2.11) to H_9 of Good (1953, p. 249) where, for greater flexibility, a factor of ξ^r occurs in the numerator, ξ being close to and less than 1. I have found that H_9 often fits biological data. As an example note that the frequencies of 2, 3, and *squaring* according to (2.11), would be 342, 171, and 103, while the observations were 327, 138, and 104. The fit p_r is good for the ranges $1 \leq r \leq 39$ and $r \geq 60$ but not when $40 \leq r \leq 59$. The fit would be improved by giving the number 56 a much smaller rank or else by assuming that the relative frequency of 56 will be greatly reduced when the sample of samples is much larger. Because p_r gives a better fit than the other distributions mentioned in Good (1989b), namely the log-Cauchy and the distribution of Rissanen (1983), let us accept p_r for the present. The fit leads to results that are harsher on numerology than the less formal method used by Good (1984/88, p. 107).

TABLE 2. *Frequency counts of positive integers etc. in two articles on physics, and a subjective ranking.* Because I was aiming only to evaluate products (and quotients), the following pairs of counts were excluded from the table: $1 : (273, 43)$; $\times : (145, 126)$; $\div : (60, 84)$; $+ : (133, 41)$; $- : (40, 34)$. Although the ranking is mainly subjective it is influenced somewhat by the two samples.

Rank, r	Number	Sample 1	Sample 2	Total
1	2	286	41	327
2	3	93	45	138
3	squaring	7	97	104
4	4	56	14	70
5	8	58	26	84
6	6	26	3	29
7	powering	1	20	21
8	$\surd$	-	17	17
9	$\pi, 2\pi$	-	14	14
10	16	8	4	12
11	24	1	-	1
12	e	-	-	-
13	5	11	14	25
14	10	13	4	17
15	!	1	-	1
16	32	2	4	6
17	9	11	4	15
18	12	4	1	5
19	21	-	-	-
20	7	16	-	16
21	48	-	2	2

22	720	-	-	-
23	120	-	-	-
24	28	6	-	6
25	64	1	-	1
26	triang.	3	-	3
27	27	1	9	10
28	36	-	-	-
29	60	-	-	-
30	72	-	-	-
31	96	-	-	-
32	15	3	-	3
33	20	7	1	8
34	136	-	-	-
35	144	-	-	-
36	216	3	1	4
37	81	-	-	-
38	105	-	-	-
39	7!	-	-	-
40	25	1	-	1
41	50	1	-	1
42	18	-	1	1
43	56	10	10	20
44	45	-	-	-
45	55	1	-	1
46	256	3	1	4
47	1024	-	-	-
48	512	-	-	-
49	128	8	1	9
50	576	-	1	1
51	125	-	-	-
52	49	-	-	-
53	480	2	-	2
54	35	6	-	6
55	42	1	-	1
56	65536	1	-	1
57	32768	3	2	5
58	40	3	-	3
59	30	3	-	3

The remaining numbers occurred but are not *ranked* here.
(The multiples of 10 are exact integers, not mere roundings.)

22	1	-	1
26	1	-	1
31	1	-	1

	44	1	-	1
	82	2	-	2
	84	2	-	2
	110	2	-	2
	126	1	-	1
	133	2	-	2
	560	1	-	1
	672	1	-	1
	1120	2	-	2
	1160	1	-	1
	1520	3	-	3
	2040	1	-	1
	2640	2	-	2
	3200	1	-	1
	3696	1	-	1
	8800	1	-	1
Totals		688	337	1025

Table 2 should be considered together with the following notes. (i) The numbers beyond rank 59 in this table have not been ranked but occurred in the first sample. (ii) Composite numbers can be represented in more than one way, for example $576 = 24^2$. In this example we might well regard the prior probability of 576 as the larger of $\frac{1}{2}p_{50}$ and $\frac{1}{4}p_{11}^2 p_3$ where 50, 11, and 3 are the ranks of 576, 24 and *squaring* as listed and where the quotients 2 and 4 are explained under note (vii). (iii) Similarly, the probability of $v = a_1 a_2 \ldots a_m/(b_1 b_2 \ldots b_n)$ should strictly be computed as the maximum of the probabilities of all the different ways of expressing v. This leads to a difficult unsolved mathematical problem and will not be explicitly taken into account. When v is a rational number it can be expressed *uniquely* as a product of prime powers (or, more generally, one could include π and e as 'pseudoprimes') where the powers can be negative. The primes and pseudoprimes could be ranked in order, such as $2, 3, \pi$ or $2\pi, e, 5, 7, 11, \ldots,$ and their probabilities estimated from samples. (Allowance for square roots, factorials etc. is still necessary.) Some of this idea is of course implicit in the ranking in Table 2. (iv) The subjective ranking would be modified if the number of samples of samples were greatly increased but p_r might still be adequately estimated by formula (2.11). (v) The number 120 deserves a good ranking because it is both a factorial and a triangulation of a nice number, 16. See also Part 1. (vi) When estimating the probability of a product (or quotient) it is necessary to take into consideration whether the factors are in some manner logically related to each other. (vii) For each parameter one should pay a Bayes factor of 2 because the parameter could have occurred as its reciprocal. I shall call this the *binary factor*. Instead, we could pay a factor of 2 for each multiplication or division. (viii) It helps the judgement to imagine that the information about the experimental observations arrives in two stages. In the first stage we are told only that the unknown number lies in some wide (but not excessively wide) interval (y_1, y_2), with a uniform prior, or slightly more accurately a log-uniform prior, for the non-null hypothesis, within that interval. For example, the interval might be $(y/\sqrt{2}, y\sqrt{2})$. The full information (in the form $y \pm \tau$) arrives at the second stage. The first stage gives so little information that it is reasonable to 'condition' on it. This device might help you to replace the preprior by a prior. The conditioning on $(y_1 < x < y_2)$ provides a method for multiplying by a *ballpark factor* to obtain the prior probability. The ballpark factor is a 'reward' for x's being in the right ballpark, and, when the ballpark interval is $(y/\sqrt{2}, y\sqrt{2})$ I take this factor as $2y$ in accordance with the following argument.

Let us condition on $x \neq 1$, assume a probability of $\frac{1}{2}$ that a positive computable number exceeds 1 (since a rational number and its reciprocal are assumed to be equally probable), and that the probabilities have the geometric distribution

$$P(2^n < x \leq 2^{n+1}) = 2^{-(n+2)} \quad (n = 0, 1, 2, \ldots). \tag{2.12}$$

(This is more consistent with (2.11) than it looks.) Then, if $y > 1$, we have

$$P(y/\sqrt{2} < x \leq y\sqrt{2}) \approx \frac{1}{2^{3/2}y}. \tag{2.12A}$$

By conditioning on x's lying in this interval we force a proportional accuracy of $1/\sqrt{2}$ so the n.c.s.d. of the observed x should strictly be adjusted to allow for this. Instead, I shall absorb this small adjustment into the ballpark factor which is therefore taken as $2y$. (ix) $65536 = 2^{2^{2^2}}$ is the number of possible functionals of two binary variables (Good 1985) whereas 16 is the number of functions. (x) From a geometrical point of view π and 2π are equally simple for an obvious reason. (xi) I hope to carry out a small survey to see to what extent judgement of the ranking varies from one judge to another.

I now apply my ranking of computable numbers, such as it is, to several examples related to $m(p)/m(e)$.

The Mass Ratio of Proton to Electron

I have assembled eight pieces of numerology for $m(p)/m(e)$, the ratio of the rest masses of the proton and the electron. (All but one were previously assembled in Good, 1987.) They are not all of the form $a_1a_2 \ldots a_m/(b_1b_2 \ldots b_n)$ which was mentioned above. They are shown in Table 3 together with the observational value and I here use them as examples for a method of evaluation. Although it is obvious, I emphasize that these evaluations are largely based on my personal judgement, but the kind of reasoning might help others to make their own judgements about numerological assertions. The sources were

(i) Cohen and Taylor (1987, pp. 1126 and 1139), the observational value.
(ii)* Eddington (1946, pp. 38 and 58).
(iii)* Lenz (1951).
(iv) Worrall (1960, p. 602).
(v)* Good (1960, 1962).
(vi) Good (1962, p. 318).
(vii) Sirag (1977).
(viii)* Parker-Rhodes (1981, p. 185).
(ix) A modification of (iii) proposed here.

The four items marked with an asterisk were 'within experimental error' when they were announced but, in common with the other four items, they are now 'contradicted' (shown not to be exact) by the observed value given in row (i). The sigmages are given in the third column of the table. The numbers of correct significant digits (n.c.s.d.) are shown in the fourth column. In the fifth column I give a rough subjective estimate of the complexity κ_0 of the numerology measured in decimal digits and adjusted by the binary factor and ballpark factor of notes (vii) and (viii) to Table 2. These estimates are based largely on formulae (2.7) and (2.11) where r is the *rank* of an integer or symbol as shown in Table 2. The adjusted complexities are denoted by κ_0 to distinguish them from the κ used above and in Good (1988c) where no allowance was made for the binary and ballpark factors. Because I am ignoring note (iii) to Table 2, apart from allowing for the obvious permutations, the measures of adjusted complexity might be somewhat too harsh. The last column is discussed at the end of the article. The approach leaves much to be desired, but I don't know a better one, and the results seem to me to make overall approximate sense.

Elaborate arguments were provided by the authors of items (ii) and (viii) but I have not yet understood their arguments so I here treat these items *as if* they were purely numerological. This treatment might not be excessively unfair because the formulae seem to have been empirically refuted. Note, however, that item (viii) is stated only as an approximation on page 475 of Bastin et al. (1979). Those who have understood the arguments of the authors won't need to adopt the numerological approach for items (ii) and (viii) but they provide examples for my discussion.

The following notes describe how I arrived at the rough estimates of the adjusted complexity measures κ_0. In each case the ballpark factor is taken as $2 \times 1836 = 3672$. The paragraph numbers (ii) to (viii) correspond to those in Table 3.

		'sigmage'	n.c.s.d.	κ_0 (adjusted complexity)	n.c.s.d.–κ_0 (score)
(i)	1836.152701 ± 0.000037	0	7.6	-	-
(ii)*	1836.34, the ratio of the roots of $10x^2 - 136x + \beta^{5/6} = 0$ where $\beta = 137/136$	5000	4.0	7.5	-3.5
(iii)*	$1836.118 = 6\pi^5$	900	4.7	3.8	0.9
(iv)	$1839.39 = 10^4/(2e)$	90000	2.8	4.1	-1.3
(v)*	$1836.10 = 137^2/(10\gamma)$ where $\gamma = 46/45$	1400	4.5	8.9	-4.4
(vi)	$1836 = 1728 + 108$	4000	4.1	4.7	-0.6
(vii)	$1836 = 136 \times 135/10$	4000	4.1	4.4	-0.3
(viii)*	$1836.15150 = \frac{137\pi\times(5/4)}{\frac{3}{14}\left(1+\frac{2}{7}+\frac{4}{49}\right)}$	32	6.2	11.0	-4.8
(ix)	$1836.15070 = 6\pi^5\alpha'/\alpha$	54	6.0	-	-

TABLE 3. Numerology for $m(p)/m(e)$.

(i) This is the best available experimental value.

(ii) I judge that the prior probability that $m(p)/m(e)$ is equal to the ratio of the roots of a quadratic, with interesting and ultimately explicable coefficients, is less than 1/200, but I'll use this value to be generous. (The use of this judgement shows that my prior probability in this example is not entirely 'preprior'.) Conditioning on that assumption we have to measure the complexities of 10, -136, 137/136, powering, 5 and 6. The numbers 6, 10, and 136 are somewhat related to one another, as said in Part 1, so it would not be fair to multiply all of the values of p_r $(= 2/[(r+1)(r+2)])$ for the ranks. To allow for the relationship just mentioned, and out of respect for Eddington, Bastin, and others, I shall generously allow no penalty for 10 and 137. We recover a factor of 2 because the quadratic could have been written in reverse order. The adjusted complexity (in decimal units) then comes to at least

$$\kappa_0 = -\log_{10}\left(\tfrac{2}{200}\cdot\tfrac{2}{35\cdot 36}\cdot\tfrac{2}{14\cdot 15}\cdot\tfrac{2}{7\cdot 8}\cdot\tfrac{2}{8\cdot 9}\cdot\tfrac{3672}{16}\right) = 7.5.$$

(iii) Lenz's paper is probably the shortest physics paper on record. He did not try, at least not there, to dress up his formula with a theory, whether lucid or obscure. The values of r for 6, π, powering and 5, are 6, 9, 7, and 13. We gain a factor of 2 because the product can be written as $\pi^5 6$. This gives $\kappa_0 = 3.8$. This is slightly less than the number of correct significant digits, so I judge that Lenz's formula was better than an evens bet at the time it was first proposed. It is interesting to write the formula as

$$m(p)/m(e) \approx 6!\, V_{10}/C_{10} \tag{2.13}$$

where V_{10} is the volume of a ten-dimensional ball of radius say k and C_{10} is that of a ten-dimensional cube of side k (compare Good 1970). The occurrence of 6! again is somewhat striking.

(iv) The ranks for 10, powering, 4, 2, and e are 14, 7, 4, 1, and 12, and we recover a factor of $3! = 6$ for permuting the factors, so we get $\kappa_0 = 4.1$. This item should not have been published because it was extremely inaccurate when first suggested. Moreover the author tried to explain the factor e^{-1} as 'exponential decay'. The correct factor representing exponential decay is of the form $e^{-\lambda t}$ and it is hardly conceivable that there can be any reason for λt to be equal to 1.

(v) The constant $\gamma = 46/45$ was introduced by analogy with 137/136 which Eddington denoted by β. Whereas $136 = 10^2 + 6^2$ (or the number of independent elements in a 16×16 symmetric matrix), we have $45 = 6^2 + 3^2$ (as for a 9×9 matrix) where 3, 6, and 10 are three consecutive triangular numbers. To allow for the relationships I count 10 only once, having regarded 45 as a triangulation of 10 (while losing a factor of 2 because there are always two ways to triangulate a number). I think of 46 as $45 + 1$ and 137 as $136 + 1$. Take the ranks of 'squaring', 10, and triangulation $\left(\binom{10}{2} = 45\right)$ (ranks 3, 14, and 26), and judge a rank of 12 for $+1$. We recover a factor 4! for permutations and get

$$\kappa_0 = -\log_{10}\left(\tfrac{2}{35\cdot 36}\cdot\tfrac{2}{13\cdot 14}\cdot\tfrac{2}{4\cdot 5}\cdot\tfrac{2}{15\cdot 16}\cdot\tfrac{2}{27\cdot 28}\cdot\tfrac{2}{13\cdot 14}\cdot\tfrac{24\cdot 3672}{16\cdot 2}\right) = 8.9.$$

(vi) The numbers 1728 and 108 were originally given credit for arising in Klein's theories of groups connected with the icosahedron and dodecahedron. If (on grounds of self-criticism) we harshly credit nothing for these reasons we can write the formula as $12^3(1 + \frac{1}{16})$, and use ranks for 12, powering, 3, 16, and addition of 1, namely 18, 7, 2, 10 and say 12. The expression is unchanged under four permutations so $\kappa_0 = 4.7$.

(vii) The formula can be written as $\binom{136}{2}/5$. The ranks for 136, triangulation and 5 are 34, 26, and 13. We should pay a factor of 2 for the choice of the lower triangulation and recover a factor of 2 for 'permutation'. We get $\kappa_0 = 4.4$.

(viii) Write the formula as

$$\frac{137\pi(5/4)}{\frac{3}{2\times 7}\left(1 + \frac{2}{7} + \frac{4}{7^2}\right)}$$

and take ranks for 136, +1, π, 5, 4, 3, 2, 7, +1, 2, and squaring, namely 34, 12, 9, 13, 4, 2, 1, 20, 12, 1, and 3. It seems roughly right to count the 7 only once and allow for only one plus sign in the denominator. We recover $8!\,2!$ for 'permutations' of the factors 137, π, 5, 4, 3, 2, 7 and $1+\frac{2}{7}+\frac{4}{49}$ and for reversal of the terms in the last factor. The binary factor is 512. The outcome is $\kappa_0 = 11.0$. It is possible that Parker-Rhodes constructed his obscure explanation of (viii) after noticing that, according to the best value of $m(p)/m(e)$ known at the time,

$$\frac{1}{137\pi}\cdot\frac{m(p)}{m(e)} \approx \frac{1715}{402} = \frac{5\times 7^3}{6\times 67}$$

which would have been readily obtained by using continued fractions. He then might have rewritten this in various ways while constructing his explanation. (For example, $67 = 7^2 + 2\times 7 + 2^2$.) If only he were still alive he could confirm or deny this conjecture. The fundamental correctness of his explanation would have been far more convincing (to those who do not understand it) if he had produced it without first knowing the observed value. The same statement applies to Eddington's explanations which I surmise were largely numerological though nominally based on a theory.

(ix) This modification of Lenz's formula is correct to one part in a million. I have found it too difficult to estimate the adjusted complexity, for I cannot decide how much to 'pay' for the factor α'/α. This factor seems to require a self-contradictory explanation. It might be better to replace it by $1+\frac{1}{3}\alpha^2$ or $1+\frac{1}{3}{\alpha'}^2$ or $\exp(\frac{1}{3}\alpha^2)$ etc.

The last column of the table gives the difference n.c.s.d. $-\kappa_0$ and is a rough measure of how good each piece of numerology is when the sigmage s is ignored or equivalently is assumed to be equal to 1. The difference is suggestive of the posterior log-odds of the corresponding piece of numerology, at the time it was proposed, if it was consistent with the observations at that time as were items (ii), (iii), (v), and (viii). I say 'suggestive of' rather than 'roughly equal to' because my arguments are not rigorous enough to justify the latter expression. It is safe enough to describe the expression n.c.s.d. $-\,\kappa_0$ as a *score* in its ordinary English sense. It gives an indication of whether the numerology (even if only an approximation) might point towards the truth when its sigmage is not taken into account by means of the factor $\exp\left[-\frac{1}{2}(s^2-1)\right]$ of formula (2.6A).

Lenz's formula is the only one, among formulae (ii) to (viii), having a positive score.

Concluding Comments

The methods used in Parts 1 and 2 differ considerably. This is because the examples in Part 2 are treated as almost purely numerological apart from my judgement of the rankings of numbers and operations in Table 2. But the two parts share the 'two stages of information' as described in note (viii) to Table 2. There are precedents for two stages of information. For example, in statistical consulting a client might suggest a null hypothesis, and then the statistician might take this hypothesis seriously out of respect for the client's scientific judgement. Similarly a medical or legal investigation often begins with 'presenting symptoms' or a 'prima facie case'.

Part 1 supports the hypothesis of *the relative rationality of proportional bulges (of hadron masses)*, at least to a good approximation when the heavy quarks are not involved. The numbers 720, 48 etc. suggest that an explanation based on finite groups might be found, the symmetric group of degree 6 being a candidate, or the 'heterotic string theory' might be relevant. Part 2 argues that the judgement of physical numerology does not need to be made only in a gestalt manner, but can be largely analyzed in terms of judgements concerning the complexities of integers, familiar constants, and familiar mathematical operations. At present these judgements are subjective and depend on what mathematical language, or calculator, one believes to be appropriate for a specific application. (Compare, for example, Good 1977, pp. 326–327.)

The methods of Part 2 are exemplified by various numerological expressions for $m(p)/m(e)$. One conclusion was that Lenz's formula $6\pi^5$ was seemingly 'odds on' when it was suggested though it is now known not to be accurate. Equation (2.13) expresses it in a form that might lead to a geometrical interpretation in ten dimensions.

Acknowledgements

This work was supported in part by a grant from the U.S. National Institute of Health. I am also much indebted to E. Richard Cohen for a correction and for some important information concerning the latest observational values of physical constants. My thanks are also due to Leslie Pendleton Miller who typed the original draft of this long paper.

Appendix A. Physics

Many of the 'elementary particles', namely the *hadrons*, are believed to be mainly made up of *quarks* of which there are various kinds. The quarks are often described as up, down, strange, charmed, bottom (or beauty) and top (or truth). These can be regarded as nicknames or mnemonics for the official names, u, d, s, c, b, and t (Cohen and Giacomo 1987, p. 12). Corresponding antiquarks are denoted by $\overline{u}$, $\overline{d}$, etc. The charges on u, d, s, c, b, and t are respectively $-\frac{1}{3}$, $\frac{2}{3}$, $-\frac{1}{3}$, $\frac{2}{3}$, $-\frac{1}{3}$ and $\frac{2}{3}$ where the unit is the charge on the electron, while the antiparticles have the signs of the charges reversed. The compositions, in terms of quarks, of the particles mentioned in Table 1, are shown in the first column of that table. The particles containing three quarks are baryons while those containing one quark and one antiquark are mesons.

The standard notation for the mass of the proton, for example, is m_p, but I have used the notation $m(p)$ to make the production of the document a little easier. The 'mass' $m(X)$ of a particle X means its rest mass. The relativistic mass of X moving with velocity v, relative to a specified frame of reference, is equal to $m(X)(1 - v^2/c^2)^{-1/2}$.

The masses of the quarks. PDG89 (p. 102, col. i) gives the masses of the quarks, in units of MeV$/c^2$, as $m(u) = 5.6 \pm 1.1$, $m(d) = 9.9 \pm 1.1$, $m(s) = 199 \pm 33$, $m(c) = 1350 \pm 50$, $m(b) \approx 5000$, and $m(t) > 50,000$. These are described as 'running masses evaluated at 1 GeV'. Perhaps the charmed quark should be regarded as of intermediate mass, neither light nor heavy. Much of the mass of a quark is converted into 'packing energy' so a particle can be lighter than the sum of the masses of the quarks that lie within it.

From Balmer to Bohr. The well-respected textbook Messiah (1961, p. 38n) is historically somewhat inaccurate when it says 'The quantization of circular orbits led Bohr to find the Balmer formula ...'. For Bohr postulated this quantization to *explain*, not to 'predict', Balmer's formula when that formula was shown to him by Hans Marius Hansen (Barrow & Tipler 1986, p. 222).

Appendix B. Odds and Bayes Factors

If an event or proposition has (possibly conditional) probability p, then its odds are defined as $p/(1-p)$. (Odds of, for example, 3.5 are also expressed as 7 to 2 on.) If the result of an experiment is denoted by E, then the prior odds $O(H)$ of a hypothesis H are multiplied by $B(H : E)$ to obtain the posterior odds of H, where $B(H : E)$ is called the *Bayes factor* in favour of H provided by E, and is given by

$$B(H : E) = \frac{O(H|E)}{O(H)} = \frac{P(E|H)}{P(E|\overline{H})} \tag{B1}$$

where $\overline{H}$ denotes the negation of H. This odds form of Bayes's theorem was stated by Wrinch and Jeffreys (1921), although they did not use the terminology of odds. One can think of $O(H|E)/O(H)$ as the definition of the Bayes factor, and the right side of (B1) as the method usually used for its calculation or estimation. When H and $\overline{H}$ are simple statistical hypotheses the right side is an uncontroversial simple likelihood ratio, otherwise it is undefined in non-Bayesian statistics. In Bayesian statistics, in this latter case, some judgement is needed to estimate the right side. Sometimes much depends on how the negation of H is interpreted.

Some writers use the term 'odds-ratio' which can mean, ambiguously, odds or the ratio of odds. Hence the expression *Bayes factor* is linguistically better as well as being historically earlier.

Appendix C. The One-Plus Exaggeration

The one-plus exaggeration, which was mentioned in the text, has occurred in relation to quantum electrodynamics (QED). While mentioning this I have no wish to question that highly successful theory.

The experimental value of the magnetic moment of the electron, in units of $e\hbar/(2cm(e))$ (where e in the numerator denotes the charge of an electron), according to PDG89 (p. 24) or Cohen and Taylor (1987, p. 1141), was 1.001 159 652 193(10), while the value given by QED was 1.001 159 652 46(20) (see Feynman 1985, pp. 6 and 7). In an interesting book, Watkins (1986, p. 46), said that the accuracy was better than one part in a million million. He confirmed in correspondence that he was referring to the magnetic moment of the electron. If we take the observed and theoretical values as the x and y of our formula, we find that n.c.s.d. was only (at least) 8.3 which is an accuracy of one part in at least two hundred million. But, according to Dirac's previous theory, which did not allow for the interaction of electrons with light, the theoretical value would be 1. If we are evaluating the *further* advance of QED, for this observation, it seems to me that we should consider that

$$x/y = (115965246 \pm 20)/(115965219 \pm 1)$$

and this reduces the n.c.s.d. to at least 6.6 or one part in at least 4,000,000 (instead of one part in a million million). Thus this crowning achievement of QED was 24 times as accurate as my piece of numerology H_0' (and both had small sigmages and therefore had 'room for improvement'). Of course a numerical success based on an attractive and otherwise successful theory is very much more convincing than even an equally accurate largely numerological result. This is because the result based on a successful theory has the higher prior probability as judged by most of the people who are paid to do physics.

The latest experimental and theoretical values for the magnetic moment of the electron are (Kinoshita 1989) 1.001159652164(7) and 1.001159652188(4) with an accuracy of 1 in 30 billion, or 1 in 40 million if we don't 'add one to exaggerate'. The 'official' sigmage is now 3.0 *so the theory has possible reached the limit of its accuracy.*

Feynman (1985, p. 9) says that nobody understands QED and then proceeds to explain it brilliantly! Similarly Bohr said that any one who is not shocked by quantum mechanics hasn't understood it. This can be reworded, paradoxically: Anyone who understands quantum mechanics knows that she has not understood it. And this is at least as true for QED. The 'instrumentalism' of modern physical theories detracts somewhat from their stati as explanations and in this respect they have a soupçon of numerology.

Appendix D. Quantal Hypotheses

Suppose that a hypothesis or theory H states that a certain physical constant is exactly equal to an unspecified integer n, while the experimental estimate is $N(x, \sigma^2)$. For the sake of elegance I allow n to be positive, negative, or zero, and I assume that σ is known precisely.

Suppose that, before the measurement was made, we had little idea about what value x would have. Then I claim that the Bayes factor in favour of H is approximately

$$\frac{1}{\sigma\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \exp\left(-\frac{(n-x)^2}{2\sigma^2}\right) \tag{D1}$$

$$= 1 + 2\sum_{n=1}^{\infty} e^{-2\pi^2\sigma^2 n^2} \cos(2\pi n x) \tag{D2}$$

$$= \vartheta_3(\pi x \mid 2\pi i \sigma^2). \tag{D3}$$

The equivalence of (D1) and (D2) is a special case of Poisson's summation formula given as (5.13) of Good (1986) where further details and applications of formula (D2), as well as historical comments, can be found.

Formula (D1) is based on the idea that, given H, the prior distribution of the relevant integer is nearly uniform over a wide range of integers, while, given the negation of H, the corresponding real number has a prior that is nearly uniform, as a real number, over much the same range. In fact I am regarding this as the *definition* of the negation of H. It is important not to forget that this assumption has been made because, for example, the result would be very different if the negation of H stated that the real number is equal to half an odd integer. The Bayes factor would be expressible as ϑ_3/ϑ_4.

If $\sigma < \frac{1}{2}$, formula (D1) can be well approximated by just a few terms of the series, and if σ is small a single term is adequate. If $\sigma > \frac{1}{2}$, the Bayes factor is close to 1, as can be readily seen both intuitively from the meaning of the statistical problem and also from formula (D2). It is also interesting to note the check that, if x is an integer, expression (D2) exceeds unity, as it should, while if x is half an odd integer the expression 'subceeds' unity, and this again makes perfect intuitive sense. This last fact follows at once from Jacobi's infinite product for ϑ_4.

If σ is regarded as having a prior distribution we could multiply formula (D1), (D2), or (D3) by that distribution and integrate to get an improved value for the overall Bayes factor. But I shall not try to carry out this refinement.

Example. In Part 1 the hypothesis H_1 states that seven independent physical constants are all equal to integers. The corresponding observations are 0.9999981 ± 0.0000044, 47.95 ± 0.055, etc. The seven corresponding Bayes factors are therefore approximately

$$\frac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{0.0000019^2}{2\sigma_1^2}\right) \quad (\sigma_1 = 0.0000044),\ \text{etc.},$$

$$\frac{1}{\sigma_6\sqrt{2\pi}}\left[\exp\left(-\frac{0.46^2}{2\sigma_6^2}\right)+\exp\left(-\frac{0.54^2}{2\sigma_6^2}\right)\right] \quad (\sigma_6 = 0.33),$$

and

$$\frac{1}{\sigma_7\sqrt{2\pi}}\exp\left(-\frac{0.156^2}{2\sigma_7^2}\right) \quad (\sigma_7 = 0.11).$$

The seven Bayes factors are respectively as shown in Table 1.

Discussion. The topic of this appendix is closely related to that of 'Quantum hunting' which is surveyed by Kendall (1986). In quantum hunting one searches for a quantity q such that all observations are multiples of q 'within experimental error', where the experimental error (standard deviation σ) is assumed to be the same for all observations. Our problem is the case where q has a specified value, but where σ varies from one observation to another and has an approximately known value for each observation. As far as I know, the published work on quantum hunting has all been non-Bayesian but it could be tackled by a Bayesian approach. Even without assuming a prior distribution $F(q)$ for q it would be of interest to draw a graph of $B(q)$ where $B(q)$ is the Bayes factor (or *a* Bayes factor) in favour of the quantum hypothesis (say H_q) that q has a specified value, the rival hypothesis being that *no* value of q exists. Clearly $B(q) \to 1$ as $q \to 0$. By definition H_q is supposed to denote the hypothesis that q is the largest quantum. This definition makes the various hypotheses H_q mutually exclusive. Without this constraint, H_q would *imply* $H_{q/2}$, $H_{q/3}$, etc.

The integral of $B(q)dF(q)$ would be the overall Bayes factor in favor of the quantal hypothesis without specifying a value for q.

Another quantal problem was treated by Hammersley (1950), that of estimating a parameter when it is known in advance of sampling that the parameter certainly belongs to a specified set of numbers, such as the set of integers, whereas in this appendix we have been concerned with *testing* this hypothesis. In the example, the problem of estimation is not entirely absent because the favoured integers seem all to be of the form $2^a 3^b$, but I have not taken this nice-looking feature into account when calculating the overall Bayes factor.

Appendix E. The Symmetric Group of Degree 6

Burnside (1911 or 1955, p. 209) states the following theorem which gives a distinctive property of the symmetric group of degree 6:

> The symmetric group of degree n ($n \neq 6$) contains n and only n sub-groups of order $(n-1)!$ .... The symmetric group of degree 6 contains 12 sub-groups of order 5!, which are simply isomorphic with one another and form two conjugate sets of 6 each.

It is tempting to conjecture that the two conjugate sets correspond to the six quarks and six antiquarks.

References

Allen, H.S. (1928). *The Quantum and its Interpretation*. Methuen, London.

Barrow, J.D. and Tipler, F.J. (1986). *The Anthropic Cosmological Principle*. Clarendon Press, Oxford.

Bastin, Ted, Noyes, H.P., Amson, J., and Kilmister, C.W. (1979). On the physical interpretation and the mathematical structure of the combinatorial hierarchy. *International Journal of Theoretical Physics* 18, 445–488.

Blecher, Marvin (1989). Private communication.

Burnside, W. (1911). *Theory of Groups of Finite Order*. 2nd edition. Cambridge University Press, Cambridge. Reprinted by Dover Publications, New York, 1955.

Candelas, P., Horowitz, G.T., Strominger, A., and Witten, E. (1985). Vacuum configurations for superstrings. In Schwarz 1985, 1107–1133.

Chase, Stuart (1938). *The Tyranny of Words*. Harcourt, Brace, and Co., New York.

Cohen, E. Richard (1989). A letter dated March 7 based on recent information obtained from Robert S. Van Dyck Jr., G. Audi, and A.H. Wapstra.

Cohen, E. Richard and Giacomo, Pierre (1987). Symbols, units, nomenclature and fundamental constants in physics. *Physica* 146A, 1–68.

Cohen, E. Richard and Taylor, Barry N. (1987). The 1986 adjustment of the fundamental physical constants. *Reviews of Modern Physics* 59, 1121–1148.

Cover, T. (1973). On determining the irrationality of the mean of a random variable. *Annals of Statistics* 1, 862–871.

Davies, P.C.W. and Brown, Julian (1988). *Superstrings: A Theory of Everything?* Editors, Cambridge University Press, Cambridge.

Eddington, A.S. (1930). *The Mathematical Theory of Relativity*. Cambridge University Press, Cambridge.

——— (1946). *Fundamental Theory*. Cambridge University Press, Cambridge.

Efron, B. (1971). Does an observed sequence of numbers follow a simple rule? (Another look at Bode's law.) (with discussion). *Journal of the American Statistical Association* 66, 552–568.

Einstein, A. (1949). Autobiographical notes. Translated from the German by P.A. Schilpp, written in 1946. In *Albert Einstein: Philosopher-Scientist*, ed. P.A. Schilpp, Tudor, New York, 3–95.

Feynman, R.P. (1985). *QED: The Strange Theory of Light and Matter*. Princeton University Press, Princeton, N. J.

Goldman, T. and Haber, H.E. (1985). Gluinonium: the hydrogen atom of supersymmetry. *Physica* 15D, 181–196. Reprinted in Kostelecký and Campbell (1985).

Good, I.J. (1950). *Probability and the Weighing of Evidence*. Charles Griffin, London, and Hafners, New York.

——— (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* 40, 237–264.

——— (1960). Some numerology concerning the elementary particles or things. *Journal of the Royal Naval Scientific Service* 15, 213.

——— (1962). Physical numerology. In *The Scientist Speculates*, ed. I.J. Good, A.J. Mayne, and J. Maynard Smith, Heinemann, London, 315–319.

——— (1968). Corroboration, explanation, evolving probability, simplicity, and a sharpened razor. *British Journal of Philosophical Science* 19, 123–143.

——— (1969). A subjective evaluation of Bode's law and an 'objective' test for approximate numerical rationality (with discussion). *Journal of the American Statistical Association* 64, 23–66. Reprinted in part in Good 1983.

——— (1970). The proton and neutron masses and a conjecture for the gravitational constant. *Physics Letters* 33A, 383–384.

——— (1971). The evolving explanation of a 'numerological' law, being an invited 'rebuttal' to Efron (1971). *Journal of the American Statistical Association* 66, 559–562.

——— (1975). Explicativity, corroboration, and the relative odds of hypotheses. *Synthèse* 30, 39–73. Reprinted in Good 1983.

——— (1977). Explicativity: a mathematical theory of explanation with statistical applications. *Proceedings of the Royal Society (London) A* 354, 303–330, and 377 (1981), 504. Reprinted in part in Good 1983.

——— (1983). *Good Thinking: The Foundations of Probability and its Applications*. University of Minnesota Press.

——— (1984/88). Surprise index. *Encylopedia of Statistical Sciences* 9, 104–109.

——— (1985). The number of functionals: a combinatorial curiosity. C228 in *Journal of Statistical Computing and Simulation* 21, 90.

——— (1986). Some statistical applications of Poisson's work. *Statistical Science* 1, 157–170.

——— (1987). Good and bad scientific speculations. An invited lecture at the meeting of the British Association for the Advancement of Science, Belfast.

——— (1988a). The number of correct significant digits in a piece of physical numerology. C305 in *Journal of Statistical Computing and Simulation* 29, 358–359.

——— (1988b). What are the masses of the elementary particles? *Nature* 332, 495–496 .

——— (1988c). Physical numerology. Technical Report No. 88–26, Department of Statistics, Virginia Polytechnic Institute and State University (December 30), 44 pp.

——— (1989a). The theorem of corroboration and undermining, and Popper's demarcation rule. C317 in *Journal of Statistical Computing and Simulation* 31, 119–120.

——— (1989b). On the judgement of "numerological" assertions. C333 in *Journal of Statistical Computing and Simulation*, in press. [Or see Good 1988c.]

——— (1989c). On the neutron and proton masses: a numerological case study. C334 in *Journal of Statistical Computing and Simulation*, in press. [Or see Good 1988c.]

Green, M.B., Schwarz, J.H., and West, P.C. (1985). Anomaly-free chiral theories in six dimensions. *Nuclear Physics* B254, 327–348. Reprinted in Schwarz 1985.

Gross, D.J., Harvey, J.A., Martinec, E., and Rohm, R. (The 'Princeton String Quartet') (1985). Heterotic string theory, I. *Nuclear Physics* B256, 253–284. Reprinted in Schwarz (1985).

Hammersley, J.M. (1950). On estimating restricted parameters (with discussion). *Journal of the Royal Statistical Society, Series B* 12, 192–240.

Hammersley, J.M. and Morton, K.W. (1954). Poor man's Monte Carlo. *Journal of the Royal Statistical Society, Series B* 16, 23–38, with discussion on pp. 61–75.

Hardy, G.H. (1940). *Ramanujan*. Cambridge University Press, Cambridge.

Henrion, M. and Fischhoff, B. (1986). Assessing uncertainty in physical constants. *American Journal of Physics* 54 (9), 791–798.

Ihde, A.J. (1964). *The Development of Modern Chemistry*. Harper & Row, New York.

Jeffreys, H. (1937/57). *Scientific Inference*. 1st and 2nd editions. Cambridge University Press, Cambridge.

——— (1938). *Theory of Probability*. Clarendon Press, Oxford.

Kendall, D.G. (1986). Quantum hunting. In *Encyclopedia of Statistical Sciences*, Volume 7, ed. S. Kotz, N.L. Johnson, and G.B. Read, John Wiley & Sons, New York, 435–439.

Kinoshita, T. (1989). Accuracy of the fine structure constant. *I.E.E.E. I.M. [Instrumentation and Measurement]* 38, 172–174.

Kosteleckỳ, V.A. and Campbell, D.K. (1985). *Supersymmetry in Physics*. Proceedings of a conference on supersymmetry in physics held at the Center for Nonlinear Studies, Los Alamos, NM 87545, USA, December 15–20, 1983.

Laplace, P.S. (1774/1986). Mémoire sur les probabilités des causes par les événements. *Mémoires de Mathématique et de Physique, Académie Royale des Sciences* 6, 621–656. Reprinted in Laplace's *Oeuvres Complètes* 8, 27–65.

English translation by S.M. Stigler, *Statistical Science* 1 (1986), 359–378, with an introduction by Stigler.

Lenz, P. (1951). The ratio of proton and electron masses. *Physical Review* 82, 554.

Messiah, A. (1961). *Quantum Mechanics*, Volume 1. John Wiley& Sons, New York.

Nieto, M.M. (1972). *The Titius-Bode Law of Planetary Distances: Its History and Theory*. Pergamon Press, New York.

Parker-Rhodes, A.F. (1981). *The Theory of Indistinguishables*. Reidel, Dordrecht.

Particle Data Group (1986) ['PDG86']. Review of particle properties. *Physics Letters* 170B, 1–350.

——— (1988/89) ['PDG89']. Review of Particle Properties, *Physics Letters* 204B, 1–486. (Imprinted April 1988, but distributed in February 1989. Covers data up to December 1987.)

Porter, T.M. (1986). *The Rise of Statistical Thinking, 1820–1900*. Princeton University Press, Princeton, N.J.

Prigogine, I. and Stengers, I. (1984). *Order Out of Chaos*. Bantam, New York.

Ramanujan, S. (1915). Highly composite numbers. *Proceedings of the London Mathematical Society* 14, 347–409. Reprinted in *Collected Papers of Srinivasa Ramanujan*, ed. G.H. Hardy, P.V.S. Aiyar, and B.M. Wilson; Cambridge University Press, Cambridge, 1927, and Dover Publications, 1962.

Ramond, P. (1985). Supersymmetry in physics: an algebraic overview. *Physica* 15D, 25–41. Reprinted in Kostelecký and Campbell (1985).

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description-length. *Annals of Statistics* 11, 416–431.

Rowlatt, P.A. (1966). *Group Theory and Elementary Particles*. Longmans, London. (Elementary particles are hardly mentioned except in the Introduction.)

Schwarz, J.H. (1985). *Superstrings*, Volume 2. Editor, World Scientific, Singapore.

——— (1988). Chapter 2 of Davies and Brown (1988).

Sirag, S.P. (1977). A combination [combinatorial] derivation of the proton-electron mass ratio. *Nature* 268, 294.

Slater, N.B. (1957). *Eddington's Fundamental Theory*. Cambridge University Press, Cambridge.

Sommerfeld, A. (1916). Zur Quantentheorie der Spektrallinien. *Annalen der Physik* 51, 1–94.

Watkins, P. (1986). *Story of the W and Z*. Cambridge University Press, Cambridge.

Weber, H. (c. 1908). *Lehrbuch der Algebra*, Volume 3. Chelsea reprint, New York, undated.

Whewell, W. (1847/1967). *The Philosophy of the Inductive Sciences Founded Upon Their History*, two volumes (with an introduction by John Herival). Johnston Reprint Corporation, New York and London.

Whittaker, E.T. (1953). *A History of the Theories of Aether and Electricity: The Modern Theories, 1906–1926.*

Worrall, R.L. (1960). Ratio of neutron and electron mass. *Nature* 185 (Feb. 27), 602.

Wrinch, D. and Jeffreys, H. (1921). On certain fundamental principles of scientific inquiry. *Philosophical Magazine, Series 6* 42, 369–390.

Department of Statistics
Virginia Polytechnic Institute and State University
Blacksburg
Virginia 24061.