



Taylor & Francis
Taylor & Francis Group



The Bayes/Non-Bayes Compromise: A Brief Review

Author(s): I. J. Good

Source: *Journal of the American Statistical Association*, Vol. 87, No. 419 (Sep., 1992), pp. 597-606

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: <https://www.jstor.org/stable/2290192>

Accessed: 14-03-2020 13:29 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Taylor & Francis, Ltd., American Statistical Association are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

The Bayes / Non-Bayes Compromise: A Brief Review

I. J. GOOD*

Various compromises that have occurred between Bayesian and non-Bayesian methods are reviewed. (A citation is provided that discusses the inevitability of compromises *within* the Bayesian approach.) One example deals with the masses of elementary particles, but no knowledge of physics will be assumed.

KEY WORDS: Bayesians, animals as informal; Fine structure constant, relativistic; Hierarchical Bayes; Inexactification; Maximum Likelihood, type II; P values, standardized.

1. HISTORICAL BACKGROUND

Some forms of compromise between Bayesian and non-Bayesian statistics date back perhaps to Laplace, but the concept of such a compromise seems to have been not fully explicit until much more recently. There is an analogy with the explicit way in which Egon Pearson introduced the non-null hypotheses and the earlier and less explicit use by Fisher. Pearson (1939, p. 242) acknowledged that a letter from "Student" had stimulated him to be more explicit about non-null hypotheses than was customary among P value devotees, and that this suggestion "formed the basis of all the later joint researches of Neyman and myself." Student was very familiar with Bayes's theorem, which uses explicit non-null hypotheses. So the influence of Bayesianism on the Neyman-Pearson technique seems to have been fairly direct.

An explicit mention of the compromise was published 35 years ago, hidden in a paper on saddle-point methods (Good 1957, pp. 862-863). The basic idea was that a Bayesian model not necessarily a good one, could be used to compute a Bayes factor F against a (sharp or point) null hypothesis, and that F then could be used as a significance criterion; that is, its distribution under the null hypothesis could be used for computing P values. A subsidiary suggestion was that F should lie in the range

$$\left(\frac{1}{30P}, \frac{3}{10P} \right),$$

and if not we should "think again." This subsidiary suggestion has been improved. Note that computers now are powerful enough to find the distribution of F by Monte Carlo methods, in many circumstances down to tail probabilities as small as $1/1000$ or less.

1.1 Likelihood

An obvious example of the Bayesian influence on non-Bayesian statistics is the importance of the concept of likelihood and of the likelihood principle, both of which are built into Bayes's theorem even if the priors are unknown. Of course Fisher made new uses of the likelihood concept.

1.2 Optional Stopping and P Values

Feller (1950, pp. 140, 190, 197) emphasized that as a consequence of the law of the iterated logarithm, P values

derived in the usual manner are misleading when optional stopping is permitted. But he didn't mention that the weight of evidence against the null hypothesis must depend only on the physical description of the observations, such as the numbers of trials and successes (not allowing for psychokinesis by the experimenter), and not on the experimenter's physically irrelevant thoughts. This fact is obvious and requires no mathematical backing. Therefore, there is something wrong with the naive use of P values. This immediate consequence of the law of the iterated logarithm was made fully explicit by Good (1956a, p. 13) and, with more detail, by Lindley (1957). (Much of his paper didn't lean on the law of the iterated logarithm.) Optional stopping by itself is harmless, but not if it is combined with naive P values. (For some history of this topic, see Good 1982a, p. 322.) For inference problems, the pure Bayesian throws away the use of P values, naive or otherwise. But because clients often want answers having the veneer of objectivity, the use of P values is somewhat justifiable, especially in the planning of experiments (such as when deciding on a sample size). I believe that if something is worth doing, it is at least worth doing badly—the obverse to Tukey's bon mot that if something is not worth doing, it is not worth doing well.

Harold Jeffreys emphatically pointed out the lack of a good logical justification for the use of P values, but as far as I know he never discussed optional stopping.

1.3 The Likelihood Ratio Test as an Implicit Compromise

When Neyman and Egon Pearson (1928, 1933) suggested the likelihood ratio test (ratio of maximum likelihoods), they said it was intuitively appealing. Their suggestion was made practical when Wilks (1938) found the asymptotic distribution given the null hypothesis. The intuitive appeal can be explained on the grounds that the ratio of maximum likelihoods can be regarded as a (very poor) approximation to a Bayes factor in which integrals are replaced by maximum values of integrands (Good 1987/90, p. 449). Thus Neyman and Pearson perhaps were unconscious Bayes/non-Bayes compromisers. Indeed Lindley and Jimmie Savage (Savage et al. 1962, pp. 64-67; Savage 1964, sec. 5) showed that the Neyman-Pearson " (α, β) technique" is implicitly Bayesian. (See also Good 1980, for a relationship of that technique to weight of evidence.)

* I. J. Good is University Distinguished Professor of Statistics and Adjunct Professor of Philosophy, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061. This article was developed from an address given in Atlanta, Georgia, at the joint meetings of the American Statistical Association and the Institute of Mathematical Statistics on August 20, 1991, at the suggestion of former president Arnold Zellner.

1.4 Strength of a Test, a Bayes/ N - P Compromise

When there are more than two parameters, the power functions of tests of hypotheses can be difficult to apprehend intuitively. Good and Crook (1974, p. 711, col. ii) proposed that a strength (a weighted average of powers, the weights being functions of the parameters) might then be used and could constitute a prior density. Crook and Good (1982) applied the method to tests for multinomials and contingency tables. The method is a compromise between Neyman-Pearsonian and Bayesian methods. An ordinary average had been proposed by West and Kempthorne (1972, p. 19) but not pursued. The use of an ordinary average strikes me as a kind of covert Bayesianism.

1.5 The Fiducial Argument

In his fiducial argument, Fisher seemed to obtain posterior distributions for some problems without assuming priors. The fallacy in his argument was pinpointed in Good (1971, p. 139). Harold Jeffreys (1939, p. 311) pointed out which priors, usually improper, would lead to Fisher's fiducial posteriors. This was another relationship between Bayesian and seemingly non-Bayesian ideas.

2. THE HIERARCHICAL BAYES APPROACH TO STATISTICS

For discussing the B/ n B compromise in more detail, a logical place to start is with the hierarchical Bayes (HB) approach to statistical theory. I have been interested in this topic for at least 40 years, but I'll be brief because I have reviewed the topic before for categorical data (Good 1979; Good and Crook 1987). The 1979 review covered much of my work or joint work on the topic but not enough of Tom Leonard's valuable contributions. For continuous data see, for example, Lindley and Smith (1972).

In the HB method, methodology, technique, or philosophy, one has a parameter θ with a prior, as in the usual Bayesian method, but the prior contains a hyperparameter that might have a hyperprior, and so on. In Good (1952, p. 114) I said that "the higher the type, the woollier the probabilities . . . [but] that the higher the type, the less the wooliness matters." Goel and deGroot (1981) showed that this is not always true, but I think it usually is; otherwise, as Stephen Feinberg once remarked in conversation, "science would be impossible." At any type, level, or stage, one can either assume values for the (hyper)"-parameters or else estimate them by type- $(n + 1)$ maximum likelihood (ML). One can terminate the hierarchy by a judgment of diminishing returns (a form of "type II rationality" in which intuitive allowance is made for the "cost" of thinking or calculating). The methods can be denoted by E (empirical; e.g., ML estimation of θ), B (Bayes, a specific prior assumed for θ), EB (pseudo-Bayes or parametric Empirical Bayes and earlier called type II ML estimation of the hyperparameter), BB (Bayes-Bayes, a specific prior assumed for the hyperparameter), EBB (type III ML estimation, for the hyperparameter), and so on (Good 1987, 1991a). These notations are to be interpreted from right to left.

The term *parametric empirical Bayes* is well entrenched but is misleading historically and from the point of view of information retrieval. It really is two-stage HB. Empirical Bayes, in its original meaning, assumes hardly more about the prior than its existence and doesn't belong to the Bayesian hierarchy.

2.1 Climbing the Hierarchy

I believe that type II ML estimation usually is much better than ordinary ML. This certainly is true when estimating the parameters of a multinomial. Other techniques of non-Bayesian statistics can be adapted to the various levels. For example, there is a type II likelihood ratio statistic for testing equiprobability of a multinomial and "independence" in a contingency table (Crook and Good, 1980; Good 1976; Good and Crook, 1974). Its asymptotic distribution turns out to be fairly accurate down to extraordinarily small P values such as 10^{-10} , a fact that never has been explained.

Another technique—the simple but useful idea of graphing a likelihood function long advocated by G. A. Barnard—has been adapted to a higher level. (See, for example, the graph of $F(k)$ in Good 1975.) When the likelihood, or type n likelihood, depends on two parameters or (hyper)"-parameters ($n = 1, 2, \dots$), then a table might be better than a graph.

The notions of power and strength also can be promoted up the Bayesian hierarchy, as discussed by Crook and Good (1982, p. 794). Type II minimax was suggested by Good (1951–52) and independently by Hurwicz (1951).

2.2 Testing of Models

HB sheds light on a suggestion by Box (1979–80) that statistical models should be Bayesian but should be tested by significance tests (e.g., P values, sampling theory). The pure Bayesian replies that the model should be embedded in a wider model so that it could be tested Bayesianwise. To this, Box could reply that the wider model needs to be tested by a P value. This alternation between augmenting the model and testing the augmented model can be continued, and it is unclear whether the Bayesian or the non-Bayesian should have the last word (Good 1987–90, p. 452). It is like a game in which the aim is to state a larger number than your opponent.

3. BAYESIANS ALL

All animals act as if they can make decisions. Most must be fairly good at it; otherwise, they wouldn't live as long as they do. They must allow, at least implicitly, for the probable outcomes of their actions and for the utilities of those outcomes. In short all animals, including non-Bayesian statisticians, are at least informal Bayesians. There probably are no perfectly rational people, and conceivably no perfectly rational dogs. General Patton could have called his dog an informal Bayesian instead of calling it, not too informatively, a son of a bitch. But Bayesianism is a matter of degree and of kind, and much depends on how explicit you are about the (epistemic) probabilities of hypotheses or, perhaps more

often, the ratios of such probabilities. Probabilities of hypotheses are not officially used by non-Bayesians, nor by strict followers of de Finetti. In the Neyman–Pearson technique you are supposed to be explicit about what the hypotheses *are*, but not about their probabilities. This matter of explicitness concerning the nonnull hypothesis is relevant to the interpretation of P values or tail area probabilities.

4. P VALUES: THE STATISTICIAN AND HIS CLIENT

Imagine a statistician and a client who would like to know whether some null hypothesis H is true or useful. An experiment is performed and the P value, or tail probability, of some criterion is calculated. There are at least two ways of using the P value.

4.1 The Statistician's Rear End

The P value can be used in a Neymanian manner for making nonprobabilistic statements that are correct in the long run in a certain proportion of cases, thus protecting the statistician's rear end (a card-carrying Neymanian has no posterior) to some extent, but the client's less so. I shall ignore this usage in this article, although it is useful in planning experiments and is convenient in routine applications for quality control. Neyman never claimed that his concept of "inductive behavior" was always appropriate, and Egon Pearson never independently advocated the concept as far as I know.

4.2 P Values and Weights of Evidence

The P value can be used for obtaining some weight of evidence for or against H . This is what the client usually would like and often believes he is getting. I regard this usage as partially Bayesian; it is by no means entirely Bayesian, because a given P value, say .037, conveys very different weights of evidence on different occasions. For the partially Bayesian usage I cite Fisher (1938, p. 83):

If it [a P value] is below .02 it strongly indicates that the hypothesis fails to account for the whole of the facts.

On the next page, Fisher implied that a P value of .001 can be regarded as definite disproof of the hypothesis. In a later book (1956, pp. 98, 100), he actually used the expression "weight of evidence" in an informal manner in connection with P values. From these quotations it seems that his interpretation of this expression would be some decreasing function of P , whatever the application.

To say that an event E is evidence against a hypothesis H can only mean that the event has decreased the epistemic probability (i.e., the logical or subjective probability) of H . Having thus allowed the Bayesian to put one foot in the door, one might as well define the expression "weight of evidence" in its most reasonable formal sense, namely the logarithm of the Bayes factor. (For an easy and convincing non-Bayesian proof of this remark, see Good 1989a,c; for a survey, see Good 1983c.) Of course one might describe the Bayes factor itself as a multiplicative weight of evidence. It is defined as the ratio of the posterior odds (not probability)

of the hypothesis to its prior odds, and in the simplest case (and only then) it is equal to a likelihood ratio.

Fisher (1956, p. 39)—perhaps to discourage anyone from asking "are you some kind of a Bayesian?"—says, in relation to small P values,

Either an exceptionally rare chance has occurred, *or* the theory of random distributions (of stars on the celestial sphere in the specific context) is not true.

This remark is uncontroversial, but it verges on tautology. Consider the following extreme example: Suppose that you superstitiously test hypotheses by tossing a coin ten times and computing a P value according to the number of heads obtained, and that on one occasion you get ten heads. Then Fisher's remark would be valid but unhelpful. The question that this superstitious procedure suggests is how much of a small probability, or its reciprocal, is due to a mere coincidence and how much of it provides evidence against the hypothesis? The amount can even be negative.

In view of what I have said, you might suppose that I'm wholly against the use of P values. Many Bayesians are, but I'm not. That is largely because I don't think epistemic probabilities have sharp values. When they are very vague, you might have to fall back either on P values, with some modification, or on surprise indexes. (See, for example, the indexes of Good 1983a.) Certainly the P value must be based on a sensible criterion, related to the important non-null hypotheses. Further, the sample size N should be taken into account. One way to do this is by using the concept of standardized P values, which I'll now explain.

4.3 The \sqrt{N} Rule and Standardized P Values

Suppose that a random number \tilde{x} has the distribution $\mathcal{N}(\mu, \sigma^2)$ where σ is known, and let H_0 denote the null hypothesis $\mu = 0$ (a "sharp" or "point" hypothesis) and \bar{H}_0 denote the composite hypothesis $\mu \neq 0$. Assume that, given \bar{H}_0 , the mean μ has a continuous prior density $\phi(\mu)$, which might be $\mathcal{N}(0, \tau^2)$. We regard this prior as "existing" even if it is unknown. We wish to test H_0 within (or against) \bar{H}_0 . Take a large number N of observations of \tilde{x} and write \bar{x} for their mean. The standard deviation of \bar{x} , under H_0 , is σ/\sqrt{N} ; the "sigmage" s , here forced to be nonnegative, is

$$s = |\bar{x}|/(\sigma/\sqrt{N}). \quad (1)$$

(The sigmage [which rhymes with "porridge"] is the ratio of the bulge to the standard deviation.)

We shall consider various experiments with various values of N but with an assigned (fixed) double-tail P value, P . Of course,

$$P = \frac{2}{\sqrt{2\pi}} \int_s^\infty e^{-u^2/2} du. \quad (2)$$

This P value is in one-to-one correspondence with s , so s too is regarded as assigned.

If N is large enough, we know that \bar{x} is small if H_0 is true. So the probability density of \bar{x} , given \bar{H}_0 , is close to $\phi(0)$; that is,

$$\phi(\bar{x})/\phi(0) \text{ is close to } 1. \quad (2a)$$

(The validity of this condition should be judged separately for each application.) But the probability density of \bar{x} , given H_0 , is

$$\frac{2\sqrt{N}}{\sigma\sqrt{2\pi}} e^{-s^2/2}. \quad (3)$$

Therefore, the Bayes factor against H_0 is approximately

$$\frac{1}{2} \sigma \sqrt{2\pi/N} e^{s^2/2} \phi(0), \quad (4)$$

and this is proportional to $1/\sqrt{N}$ when P is regarded as fixed. Thus when N is large enough, we need to know or to judge the value of ϕ only at or near the "origin."

Formula (4) explains why the factor \sqrt{N} occurs in each table of Bayes factors in Appendix B of Jeffreys (1961). Table 1 exemplifies the near constancy of $FP\sqrt{N}$ for binomial sampling if p has a uniform prior in $(0, 1)$ given \bar{H}_0 . A similar result was found by Good and Crook (1974, p. 715) for multinomial sampling. Thus we have empirical evidence that sensible P values are related to weights of evidence and, therefore, that P values are not entirely without merit. The real objection to P values is not that they usually are utter nonsense, but rather that they can be highly misleading, especially if the value of N is not also taken into account and is large.

Arising from these results is the following rule of thumb for standardizing a P value to sample size 100: Replace P by the standardized value

$$P_{\text{stan}} = \min\left(\frac{1}{2}, P\sqrt{N}/10\right). \quad (5)$$

(For references see Berger and Sellke 1987; Good 1988b, p. 391; and Jefferys 1990.) But this is only a rule of thumb, because it depends on the assumption (2a). The point of introducing P_{stan} is to bring P values into closer relationship with weights of evidence while also preserving the appearance of objectivity.

Incidentally, this standardization prevents the statistician from sampling to a foregone conclusion by optional stopping. Optional stopping makes available a sigmage s close to $\sqrt{2} \log \log N$ for an infinite sequence of values of N (but I don't know how large N has to be for this to be a practical rule). Then the factor $\exp(s^2/2)$ in (4) reduces to $\log N$; this is more than cancelled by the \sqrt{N} in the denominator. (Our logarithms are entirely natural.) As implied earlier, op-

tional stopping has no effect on the Bayes factor, except that knowing the sampler was trying to cheat casts doubt on his honesty. But of course a dishonest sampler can cheat in other ways. (See Good 1991b for a further discussion.)

5. PSYCHOKINESIS

Jahn, Dunne, and Nelson (1987) carried out extensive automated experiments on psychokinesis for over a decade. They performed $N = 104,490,000$ Bernoulli trials, according to W. Jefferys (1990). (One of Jefferys's purposes was to draw attention of parapsychologists to the discrepancies between P values and Bayes factors; compare Good 1982b, which was concerned with neoastrology.) The null hypothesis H_0 is taken as $p = \frac{1}{2}$, where p is the parameter in each trial. We have $\sigma = \frac{1}{2}\sqrt{N} = 5,111$, the standard deviation of the number of successes. The bulge of successes was 18,471 above $N/2$, a sigmage s of 3.614 corresponding to a two-tailed P value of

$$P = \frac{2}{\sqrt{2\pi}} \int_s^\infty e^{-u^2/2} du = .000300. \quad (6)$$

According to Fisher (1938, p. 84), we are entitled to reject the null hypothesis because $P < .001$. (Though if faced with this situation, Fisher probably would have included "no artifact" as part of the definition of H_0 .) One obvious reason why many would regard this P value as not small enough is that we regard the prior probability of the existence of psychokinesis as exceedingly small (perhaps increased a little by the mysteries of quantum mechanics). Another, less obvious reason is that P values are especially misleading when the sample size is very large. If we standardize the P value to sample size 100, by the rule of thumb just given, we get $P_{\text{stan}} = .31$; this is too harsh on psychokinesis, however. In fairness I should add that at a meeting of the Society for Scientific Exploration in Charlottesville, Virginia in 1991, Jahn said the P value was now approaching 10^{-6} . His work is continuing, and an updated evaluation would necessitate a very careful investigation of the experimental conditions to look for artifacts. My purpose here is not to decide whether psychokinesis is possible, but to use the example to illustrate the relationship between P values, sigmages, and Bayes factors.

This Bernoulli sampling is the binomial case of testing a multinomial for equiprobability. A hierarchical Bayesian approach to this problem was considered by Good (1965, 1967, 1975, 1979, 1981-83, 1983d, 1988b) and by Good and Crook (1974). In this work the hyperparameter of a symmetric Dirichlet is assigned a hyperprior. For extensions to contingency tables see Crook and Good (1980), Good (1965, 1976, 1981-83, 1983b), and Good and Crook (1987).

5.1 Max Factor

For the psychokinetic data we make the simplifying assumption that all subjects have the same parameter p and that the prior for $p - \frac{1}{2}$, given \bar{H}_0 , is symmetrical about 0 and is sharply peaked at 0. More precisely, we assume that the variance τ_0^2 of the prior of $p - \frac{1}{2}$ does not exceed $(.05)^2$, because otherwise we would be confident that psychokinesis had been established decades ago. The condition of symmetry

Table 1. The Near Constancy of $FP\sqrt{N}$ for Binomial Sampling

N	r	P	F	$FP\sqrt{N}$
10	8	.058	2.07	.38
20	15	.025	3.22	.37
30	20	.068	1.15	.42
30	25	.00026	243	.35
40	25	.104	.67	.48
40	30	.0016	32	.31
60	35	.20	.36	.55
60	40	.010	4.5	.34
60	45	.00011	350	.30

NOTE: P = Double tail if $p = \frac{1}{2}$ (hypothesis H_0). F = Bayes factor against H_0 (assuming a uniform prior for p under the non-null hypothesis).

Table 2. Relationship Between the Hyperparameter λ and the Bayes Factor F When The Observed Sigmage = 3.614

λ	.01	1	2	3.4	3.5	3.6	4	8	18.3	32	64	686	10,000
F	1.00	18.5	83	115.04	115.08	114.96	113	77	37	21	10.7	1.00	.069

about 0, or evenness, is equivalent to taking the prior probability of “psi-missing” (conditional on \bar{H}_0) as equal to that of positive psi, but this assumption could be dropped at the price of additional complexity of the analysis.

In a Bernoulli sequence of N “trials” the variance of the number r of “successes,” for any given p , is $p(1-p)^N$. This is close to $\sigma^2 = \frac{1}{4}N$ (i.e., $\sigma = 5,111$, as mentioned previously), because $|p - \frac{1}{2}|$ is small (with overwhelming subjective probability).

We could adopt the hierarchical Bayesian model, mentioned earlier, based on a β prior containing a hyperparameter. But instead because $|p - \frac{1}{2}|$ is small, we can adopt a hierarchical Gaussian model in which the number r of “successes” has the distribution $\mathcal{N}(\frac{1}{2}N, \sigma^2)$, given H_0 , whereas the distribution of r given a typical rival hypothesis H_λ is $\mathcal{N}(\frac{1}{2}N, \lambda^2\sigma^2)$, where λ is a hyperparameter. Here,

$$\lambda^2 = 1 + 4\tau_0^2 N, \quad \lambda < 1,024.$$

For this and other details see Good (1991d). A uniform prior for p , in the unit interval, would have corresponded very roughly to taking λ in the region of 10,000, but the normal model would be unsatisfactory if $|p - \frac{1}{2}|$ were not small. Then (again see Good, 1991d) the Bayes factor F against the null hypothesis $p = \frac{1}{2}$, provided by an observed sigmage s , is

$$F = \frac{1}{\sqrt{1 + \lambda^2}} \exp\left[\frac{s^2 \lambda^2}{2(1 + \lambda^2)}\right], \quad (7)$$

if λ is regarded as known for the time being. The value of λ that maximizes F is $(s^2 - 1)^{1/2}$ if $s^2 \geq 1$, and 0 if $s^2 \leq 1$. F_{\max} , the maximum value of F , is given by

$$F_{\max} = s^{-1} \exp \frac{1}{2} (s^2 - 1) \quad (s^2 \geq 1) \\ = 1 \quad (s^2 \leq 1). \quad (8)$$

Formulas equivalent to (7) and (8) appeared in Edwards, Lindman, and Savage (1963, p. 231). The notation F_{\max} corresponds to that used by Good (1967) and Good and Crook (1974) and to L_{normin}^{-1} in Edwards et al. (1963, p. 241). When $s = 3.614$, we have $F_{\max} = 115.08$, which agrees with Table 2, and the corresponding value of λ is 3.47.

In the psychokinetic example we had $s = 3.614$, so (7) gives rise to the relationships between the hyperparameter λ and the Bayes Factor F shown in Table 2. Thus the maximum Bayes factor against the null hypothesis with this model is 115, which is only about $\frac{1}{30}$ of $1/P$. It is the maximum factor of type II (i.e., one level up from the case where no

prior is assumed for p), in which case the “max factor” would be $\exp(\frac{1}{2} s^2) = 686$. If λ is assumed to have a hyperprior density $\psi(\lambda)$, then the Bayes factor against H_0 would be

$$F = \int_0^\infty \frac{\psi(\lambda)}{\sqrt{1 + \lambda^2}} \exp\left[\frac{s^2 \lambda^2}{2(1 + \lambda^2)}\right] d\lambda. \quad (9)$$

For example, with $\psi(\lambda)$ taken as log-uniform from 1 to 1,024 we have $F = 55$. If 1,024 is changed to 512, then F is changed to 59. The distinction between Bayes factors of 55 and 59 is utterly negligible; this exemplifies my 1951 comment about the unimportance of woolliness at the higher levels.

It is of some interest to note that if we had $s = 0$, then we would have obtained a Bayes factor of

$$\sqrt{1 + \lambda^2} \text{ in favor of } H_0. \quad (10)$$

Some non-Bayesians say that you can’t get evidence in favor of a sharp null hypothesis. But it is easily proved from a Bayesian perspective that if an experiment is capable of supplying evidence against a hypothesis, then it also is capable of supplying evidence in favor of that hypothesis (and conversely), provided that all outcomes of the experiment are observable—the “theorem of corroboration and undermining” (Good 1989b).

5.2 The Break-Even Sigmage

Let the value of s that would convey zero weight of evidence, the *evidential break-even sigmage*, be denoted by s^* . This interesting concept occurred in Lindley (1957), but I think that giving it a name will help to focus attention on it. It is obtained by equating the expression (7) to 1, so

$$s^* = \sqrt{[(1 + \lambda^{-2}) \log(1 + \lambda^2)]}, \quad (11)$$

and the corresponding P value (with s^* replacing s in (6)) is called P^* . The relationship between λ , s^* , and P^* is exhibited numerically in Table 3. This table might be helpful for a Bayesian wishing to make coherent judgments for the quantiles of λ , s^* , and P^* .

An elegant but dubious conjecture is that s^* is at least equal to Khintchine’s sigmage (see, for example, Feller 1950–1968) $\sqrt{2 \log \log N} = 2.41$, which would give $F \leq 37$. This conjecture is based on the equally dubious assumption that Khintchine’s sigmage often would be closely attained in 100,000,000 Bernoulli trials when H_0 is true. (George Terrell and I have begun to examine this matter.)

I think that this example requires the use of subjectivistic Bayesianism. Yet objectivistic Bayesianism is a desirable ideal. For a discussion of compromises between these forms of Bayesianism, see Good (1962, 1990b).

Table 3. The Relationship Between λ , s^* , and P^*

λ	.01	1	2	7	18.3	32	90	1,000	10,000
s^*	1.000	1.18	1.42	2.00	2.415	2.63	3.00	3.72	4.29
P^*	.32	.24	.16	.046	.015	.0081	.0027	.00020	.000018

NOTE: $F = 1$.

6. *P* VALUES "IN PARALLEL"

Suppose that two or more *P* values, P_1, P_2, \dots, P_n , are obtained from a single set of observations by various criteria; for example, a parametric and a nonparametric test. The *n* tests were called "tests in parallel" (Good 1958), and a proposed rule of thumb for combining them was to compute their harmonic mean. (This should not be confused with the method of Fisher [1938, pp. 104–106] for combining independent tests, or tests "in series;" but in fact his method also can be regarded as a B/nB compromise.) The informal Bayesian justification was that Bayes factors against a point null hypothesis are very roughly inversely proportional to the reciprocals of the *P* values.

In this example of a B/nB compromise, the standardization mentioned earlier can be applied before the harmonic mean is computed. For various applications see Good (1983e, 1984a, 1984b, 1984c, 1991c) and Good and Gaskins (1980, p. 47).

Sometimes Bayesian and non-Bayesian arguments give similar results. For example, Thatcher (1964) discussed cases in which confidence intervals coincide with Bayesian estimation intervals, and Pratt (1965) emphasized that the *P* value when testing the null hypothesis $\mu \leq 0$ against $\mu > 0$ often is approximately equal to the posterior probability of the null hypothesis (when the prior probability is $\frac{1}{2}$).

7. MAXIMUM ENTROPY

The principle of maximum entropy was used in statistical mechanics by Boltzmann and by Gibbs. Shannon (1948) mentioned that the univariate distribution of given variance and maximum entropy is normal. Jaynes (1957) introduced the principle into Bayesian statistics in the production of "objectivistic priors." Good (1963) returned to the non-Bayesian interpretation as a method for generating hypotheses in continuous and discrete problems. For example, in multidimensional contingency tables the principle of maximum entropy generates loglinear models. These models already had been formulated for intuitive reasons, but those who regard statistics as a science and not just a bag of tricks might find it interesting to see this further relationship between Bayesian and non-Bayesian methods. There also is an earlier pseudo-Bayesian log-linear model, but that's another story (Good 1956b).

In my 1963 paper (p. 931) I made the natural suggestion that maximization of a linear combination of log-likelihood and entropy might be entertained for estimating physical probabilities in the cells of a contingency table. I think that this works best for sparse tables. The point is that ML is sensible if the observed frequencies are large and ME is sensible in the opposite case, where only the marginal totals are known. The method was developed by Pelz (1977), who has not yet written up his program in a form suitable for publication. One can think of this as a Bayesian method with a prior proportional to $\exp(-\lambda \times \text{entropy})$ or else as a natural non-Bayesian method. This "double" point of view also applies to the method of maximum penalized likelihood for estimating probability densities (Good and Gaskins 1971,

1972, 1980; Tapia and Thompson 1978). For some further discussions of the B/nB compromise, see the two indexes of Good (1983a) and the 27 papers listed by Good (1991d, p. 20).

8. INEXACTIFICATION OF HYPOTHESES

Scientific hypotheses and theories often are shown to be inexact rather than refuted. The statistician's term *rejected* often should be replaced by a more precise, albeit unpoetic, word such as *inexactified*. Instead of saying that the Newtonian theory has been refuted, it would be better to say that special relativity explains why the Newtonian theory is so good! We need methods for estimating the probability that a hypothesis contains some truth, or is "causal," or that it is accurate enough to be "more than just a coincidence." I'll illustrate this matter by considering a case study on "physical numerology."

9. PHYSICAL NUMEROLOGY

A piece of nonoccult numerology is an unexplained numerical statement related to physics or to some other natural science or to mathematics. Like quality, numerology can be good, bad, or indifferent.

Let us consider a piece of good numerology, discovered by hand, concerning "elementary particles." (This discussion is condensed from that in Good 1988c and 1990a, but contains several corrections and new points.) These remarks should be fairly intelligible even to those who know nothing about such matters; the words in quotes have meaning at least for particle physicists. I am discussing this topic to illustrate the use of both Bayesian and non-Bayesian thinking in the same scientific context. I believe that the piece of numerology presented is probably not just a coincidence, but this opinion is controversial.

There are three kinds of light (not heavy) "quarks" known as *u*, *d*, and *s*. Each "ordinary baryon" contains just three of these light quarks; for example, a proton "is" a *uud* (Particle Data Group 1990, p. III 65). Two particles can have the same quark composition and "spin" but can have different "isospins." A "meson" consists of a quark and an "antiquark," for example, a K^+ particle "is" a $u\bar{s}$, where \bar{s} denotes the antiquark corresponding to the quark called *s*, but $u\bar{u}$ is not a particle. Now consider a pair of particles (*X*, *Y*), both with the same "spin" and such that if one *d* quark in *Y* is replaced by a *u* quark, then we obtain *X*. (See also Note c to Table 4.) Then form the ratio $R(X, Y)$ defined by

$$R(X, Y) = \frac{m(Y) - m(X)}{\min[m(Y), m(X)]}, \quad (12)$$

where $m(X)$ and $m(Y)$ denote the rest masses of *X* and *Y*. (This is a slight modification of the definition in Good 1990a.) If not for the electromagnetic forces, *X* and *Y* would exhibit similar behavior (cf. Sudbury 1986, p. 228), so the numerator probably depends only on the electromagnetic forces (Rowlatt 1966, p. viii). Thus $R(X, Y)$ can be regarded as a measure of the ratio of electromagnetic forces to "strong" forces (those that bind quarks together).

Next let α denote the “fine structure constant” defined, in electrostatic units, by

$$\alpha = \frac{e^2}{\hbar c}, \quad (13)$$

where e denotes the charge on an electron; $\hbar = h/(2\pi)$, where h denotes Planck’s constant; and c denotes the velocity of light. (These are all the standard notations.) This is a measure of the electromagnetic force. It is dimensionless and its measured value, independent of the units used, is

$$\alpha = 1/137.0359895(1 \pm 4.5 \times 10^{-8}). \quad (14)$$

Note that e^2/\hbar is a simply defined velocity and, therefore, has a reasonable chance of occurring in a fundamental theory. The corresponding “rapidity” is

$$\alpha' = \tanh^{-1}(\alpha) = 1/137.0335570(1 \pm 4.5 \times 10^{-8}), \quad (15)$$

which may be regarded as the “relativistic fine-structure constant” (not yet a standard definition). Rapidities are dimensionless. They were introduced into the special theory of relativity partly so as to obtain additivity, a property not shared by ordinary velocities in the same direction. (For references to rapidities see, for example, Eddington 1930, p. 22 and the index of Particle Data Group 1990.)

In Eddington’s *Fundamental Theory* (1946), the integer 136 was absolutely basic and called the “basal multiplicity.” (See Note i to Table 4 and see also McCrea [1991] for a eulogy for Eddington.) Moreover, $1/\alpha$ was experimentally indistinguishable from 137, so it was natural to guess that the closeness of these two integers was not coincidental. Eddington formulated the hypothesis that associated with the proton is a bare particle, called a “hydrocule,” of mass $1/\beta$ times that of the “fully dressed” proton complete with its own energy field, where $\beta = 137/136$; he termed this the *Bond factor*. The concept of a bare particle is current in modern quantum electrodynamics (see, for example, Quigg 1985, p. 88), but I don’t know whether anyone relates the concept to Eddington’s hydrocules. I write $\beta' = 1/(136\alpha)$ for the *corrected* Bond factor and $\gamma = 1/(136\alpha')$ for the relativistic Bond factor. Although $1/\alpha$ is not an integer, I believe (almost following Eddington’s lead) that it is reasonable to infer that one of these two ratios— β' or γ —has a fundamental significance; otherwise the closeness of $1/\alpha$ to 136 would have to be considered a coincidence. (See also Note m to Table 4.) We may regard $m(p)/\gamma$ as a relativistically revised mass of the hydrocule of the proton; it then is

natural enough to replace $R(p, n)$ by $\gamma R(p, n)$. Note now that

$$6!\gamma R(p, n) = 1.0000019 \pm .0000044. \quad (16)$$

This strongly suggests the hypothesis, which the non-Bayesian certainly cannot reject, that

$$6!\gamma R(p, n) = 1, \quad (17)$$

or at least that there is a reason or explanation (unknown) of why the left side is so close to 1. What’s special about 6! or 720? It has more factors than any smaller number; that is, it is a highly composite number in the sense of Ramanujan (1915) and so has many opportunities of an explanation. It is the largest highly composite number having only three prime factors and also is a multiple of every smaller highly composite number. (The highly composite numbers below 1,000 are 2, 4, 6, 12, 24, 36, 48, 60, 120, 180, 240, 360, 720, and 840; probably the only highly composite numbers that are multiples of all smaller highly composite numbers are 4, 12, 24, 720, and 5,040.) Also, 720 is the order of the symmetric group of degree 6 and is 6 times 120, where 6 and 120 are two of the numbers (i.e., 4, 6, 10, 16, 120, 136, and 256) that are highly conspicuous in Eddington’s fundamental theory. Finite groups are basic to current theories of the elementary particles. (See also Good 1990a, app. E.)

Because 720 is highly composite, that it has geometrical interpretations is not surprising. Indeed, of the 16 regular polytopes in four dimensions, 10 have $N_1 = 720$ or $N_2 = 720$ or both, and seven have $N_3 = 120$, where N_1 , N_2 , and N_3 denote the numbers of edges, two-dimensional faces, and three-dimensional faces. (See Coxeter 1963–1973, pp. 292–294, and Good 1990a, pp. 132–133.) Furthermore, 120 is the order, g , of the “extended polyhedral group” (including reflections as well as rotations) of six of the nine regular polyhedra in ordinary space. (The other three polyhedra have $g = 24$ or 48.) These facts, combined with Note h to Table 4, suggest (but of course do not prove) that a geometrical explanation of our numerology might be found.

Now look at Table 4, where only the light quarks are used. Also note the following remarks about the table:

- $I =$ Isospin, $J =$ spin. The masses of the Δ particles are not yet known accurately enough to be used in this table.
- In Good (1990a) .33 was misprinted as .033, which, if correct, would have forced the omission of the pair (Ξ^0, Ξ^-) . Also, the reciprocal of 1.0000019 was entered in error, but that affects only Note f in this list. (An-

Table 4. Experimental Values of $6!\gamma R(X, Y)$

Quark compositions	J	I	X	Y	$6!\gamma R(X, Y)$	Close integer	Bayes factor
(uud, udd)	$\frac{1}{2}$	$(\frac{1}{2}, \frac{1}{2})$	p	n	$1.0000019 \pm .0000044$	1	83,000
(uus, uds)	$\frac{1}{2}$	(1, 0)	Σ^+	Λ	$-47.95 \pm .085$	–48	3.947
(uus, uds)	$\frac{1}{2}$	(1, 1)	Σ^+	Σ^0	$1.94 \pm .07$	2	3.950
(uds, dds)	$\frac{1}{2}$	(1, 1)	Σ^0	Σ^-	$2.974 \pm .048$	3	7.177
(u \bar{s} , d \bar{s})	0	$(\frac{1}{2}, \frac{1}{2})$	K^+	K^0	$5.914 \pm .046$	6	1.511
(uss, dss)	$\frac{1}{2}$	$(\frac{1}{2}, \frac{1}{2})$	Ξ^0	Ξ^-	$3.53 \pm .33$	3 or 4	0.771

NOTE: Values are based on those in Cohen (1989) and Particle Data Group (1990).

other misprint occurred on page 137 line 11, where “-48 as a ” should read “-48. Note the occurrence of -48 as a .”)

- c. A pair (X, Y) appears in the table if X and Y have the same spin and X is obtained from Y by replacing one d by a u . Also, apart from the maverick (Σ^+, Λ) , X and Y have the same isospin. When the isospins are different, a constraint is assumed; namely, that the X particle is the one with the larger isospin. The effect of this slight adhocery is to select the pair (Σ^+, Λ) but not (Λ, Σ^-) . I believe that this piece of adhocery is more than compensated for by the niceness of the number 48, but some readers might prefer the cleaner numerology with the maverick deleted.
- d. I presented an earlier form of the result for $R(p, n)$ in Good (1970). When later observations gave improved numerical results, I computed the values of $R(X, Y)$ for the other specified pairs (X, Y) . The hypothesis that $6! \gamma R(X, Y)$ is close to an integer was formulated after the calculations were done, violating a principle sometimes stated as dogma in elementary non-Bayesian textbooks whose authors worship at the shrine of objectivity. Nevertheless, many scientific theories violate that dogma, as does exploratory data analysis. The reason for the dogma is, of course, that its violation enables one to achieve high “significance” by inventing complicated hypotheses. But the intelligent scientist informally balances complexity against goodness of fit, or adds the prior log-odds to the weight of evidence. If, in our example, the numbers 136 and 720 had not been special, the numerology could be confidently rejected as being too complex and having too small a prior probability. In fact, both of these numbers are very special indeed.
- e. Each Bayes factor in the last column of the table is a factor in favor of the corresponding number being an integer or close to an integer. (For the theory see Good 1990a, p. 159.) The product of the six Bayes factors is about 11,000,000. This doesn’t allow for the “niceness” of those integers.
- f. The values of $6! \gamma R(X, Y)$ don’t need to be exact integers for the numerology to be regarded as probably “causal.” Indeed the absolute values of these numbers, taken as a group, seem to have a tendency to fall short of the “close integers,” though none of the shortfalls is “statistically significant at the 5% level,” (in the usual jargon). But if we combine the shortfalls for all six pairs, each divided by its standard error, we get $-\frac{19}{44} + \frac{50}{55} + \frac{6}{7} + \frac{26}{48} + \frac{86}{46} + \frac{47}{33} = 5.17$. Because $5.17/\sqrt{6} = 2.11$, the null hypothesis that all six ratios are integers might be weakly “inexactified” with a P value of .035 (the double tail) if the non-null hypothesis asserts that the true shortfalls are all positive or all negative. That part of my argument is non-Bayesian; it could be made Bayesian, but I don’t think the effort would be justified.
- g. If the very heavy b quark were introduced, the numerology would be inexactified by the pair (B^+, B^0) (see Good 1990a, p. 135), again in a non-Bayesian manner.
- h. The set of numbers 1, 2, 3, 4, and 6 is familiar in crystallography. These numbers are the only possible orders for the symmetry rotation axes of a simple crystal. In this context the number 1 indicates no rotational symmetry. Because the number 48 is again “highly composite” (in fact the largest highly composite number having only two prime factors), that it occurs in several geometrical contexts is not surprising; for example, it is the order of the automorphism group of the simple three-dimensional cubic lattice (Boisen and Gibbs 1990, p. 123; Conway and Sloane 1988, p. 91; Lovett 1989, p. 10; Good 1990c). The number -48 is mentioned as a “Pontryagin number” in a paper on superstring theory by Green, Schwarz, and West (1985, p. 338). I mention this for the benefit of those few people who understand the theory of superstrings, of whom I am not one.
- i. I believe that the main part of the evaluation of this numerology is necessarily subjective, but I also have used a few P values. Here I’ll mention only how I started the argument in Good (1990a). I said that allowing for Eddington’s reputation as a physicist, the probability that his *Fundamental Theory* contains a little sense is at least .1, and if so the number 136 has an essential part to play in the foundation of physics. For example, according to Slater (1957, p. 5) 136 is the number of mechanical degrees of freedom of a two-particle system. (For other properties of 136, see Eddington 1946 and Good 1990.) Kilminster (1966, p. 271) implied that such numbers as $1^2 + 3^2$ and $6^2 + 10^2$ must occur in any theory that separates space-time into space and time. Thus an explanation of our numerology might emerge from a theory that overlaps only slightly with Eddington’s speculations.
When making a disinterested interested judgment, I hope the reader will take into account Good (1990a). I’d be grateful to receive any new arguments, pro or con, together with overall judgments. I estimated the prior probability that the numerology is causal as between 1/36,000 and 1/1,800 and thus posited that the posterior probability is substantial (not allowing for competition from other sources).
- j. If we had not replaced α by α' , the entry 1.0000019 would have been 1.0000194 (with only four 0s following the 1), which still is strikingly close to 1 although statistically significantly above 1. The numerology in that form cannot be exact, but might very well be causal even if the introduction of the relativistic fine structure constant turned out to be a bad move. One can invert the argument and say that the numerology strengthens the case for regarding the relativistic fine structure constant as fundamental.
- k. The following simple rule seems to give the “close integers” for the pairs of ordinary baryons with equal spins and equal isospins: Consider the X particle; score 0 for each u , 1 for each d , and 2 for each s , then add

the scores. This formula predicts the close integer 4 for the pair (Ξ^0, Ξ^-) .

- I. The methods I have used to try to evaluate whether my numerology is “causal” don’t allow for any competitive theory or numerology. But according to Cheng and O’Neill (1979, p. 316), “. . . neither their values $[m(n)]$ and $m(p)$ nor their ratio can be predicted by SU_3 .” I will be grateful to any reader who can supply a reference to any fairly accurate prediction that has been made, based on a fairly widely accepted and intelligible theory or else on *good* numerology.
- m. As a pure speculation, an explanation of our numerology might be found in terms of the “law of equipartition of energy in statistical equilibrium” among degrees of freedom; see Kilmister (1966, p. 225), in which a 1935 paper of Eddington’s is quoted.

ACKNOWLEDGMENTS

I am indebted to Oxford University Press for permission to use quotes from Fisher (1938 and 1956) that were reprinted in *Statistical Methods, Experimental Design and Scientific Inference* (1991), J. H. Bennett, ed., and from Pearson et al. (1990). I also thank Addison-Wesley for permission to quote from Cheng and O’Neil (1979).

[Received October 1991. Revised November 1991.]

REFERENCES

- Berger, J. O., and Sellke, T. (1987), “Testing of a Point Null Hypothesis: The Irreconcilability of Significance Levels and Evidence,” *Journal of the American Statistical Association*, 82, 112–122.
- Boisen, M. B. Jr., and Gibbs, G. V. (1990), *Mathematical Crystallography* (2nd ed.), Washington, D.C.: Mineralogical Society of America.
- Box, G. E. P. (1981), “Sampling Inference, Bayes’ Inference, and Robustness in the Advancement of Learning,” in *Bayesian Statistics: Proceedings of the First International Meeting Held in Valencia, Spain*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Valencia, Spain: University of Valencia, pp. 366–381.
- Cheng, D., and O’Neil, G. K. (1979), *Elementary Particle Physics: An Introduction*, Reading, MA.: Addison-Wesley.
- Cohen, E. R. (1989), Letter dated March 7 based on recent information obtained from Robert S. Van Dyck Jr., G. Audi, and A. H. Wapstra.
- Conway, J. H., and Sloane, N. J. A. (1988), *Sphere Packings, Lattices and Groups*, New York: Springer-Verlag.
- Coxeter, H. M. S. (1963/1973), *Regular Polytopes* (2nd ed.), London: Constable. (Reprinted by Dover Publications, New York.)
- Crook, J. F., and Good, I. J. (1980), “On the Application of Symmetric Dirichlet Distributions and Their Mixtures to Contingency Tables, Part II,” *The Annals of Statistics*, 8, 1198–1218.
- (1982), “The Powers and Strengths of Tests for Multinomials and Contingency Tables,” *Journal of the American Statistical Associations*, 77, 793–802.
- Eddington, A. S. (1930), *The Mathematical Theory of Relativity*, Cambridge, U.K.: Cambridge University Press.
- (1946), *Fundamental Theory*, Cambridge, U.K.: Cambridge University Press.
- Edwards, W., Lindman, H., and Savage, L. J. (1963), “Bayesian Statistical Inference for Psychological Research,” *Psychological Review*, 70, 193–242.
- Feller, W. (1950/1968), *An Introduction to Probability Theory and its Applications*, Vol. 1 (1st and 3rd eds.), New York: John Wiley.
- Fisher, R. A. (1983), *Statistical Methods for Research Workers*, Edinburgh, U.K.: Oliver & Boyd.
- (1956), *Statistical Methods and Scientific Inference*, Edinburgh, U.K.: Oliver & Boyd.
- Goel, P. K., and deGroot, M. H. (1981), “Information About Hyperparameters in Hierarchical Models,” *Journal of the American Statistical Association*, 76, 140–147.
- Good, I. J. (1952), “Rational Decisions,” *Journal of the Royal Statistical Society*, Ser. B, 14, 107–114.
- (1956a), Contributions to the discussion in *Information Theory, Third London Symposium 1955*, ed. C. Cherry, London: Butterworth pp. 13–14, 33, 36, 44–45, 95, 109, 229–230, 298, 360, 371.
- (1956b), “On the Estimation of Small Frequencies in Contingency Tables,” *Journal of the Royal Statistical Society*, Ser. B, 18, 113–124.
- (1957), “Saddle-Point Methods for the Multinomial Distribution,” *The Annals of Mathematical Statistics*, 28, 861–881.
- (1958), “Significance Tests in Parallel and in Series,” *Journal of the American Statistical Association*, 53, 799–813.
- (1962), “A Compromise Between Credibility and Subjective Probability,” in *International Congress of Mathematicians Abstracts of Short Communications (Stockholm)*, Uppsala, Sweden: Almqvist & Wiksells, p. 160.
- (1963), “Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables,” *The Annals of Mathematical Statistics*, 34, 911–934.
- (1965), *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*, Cambridge, MA: MIT Press.
- (1967), “A Bayesian Significance Test for Multinomial Distributions” (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 29, 399–431. Corrigendum (1974), 36, 109.
- (1970), “The Proton and Neutron Masses and a Conjecture for the Gravitational Constant,” *Physics Letters A*, 23, 383–384.
- (1971), “The Probabilistic Explication of Information, Evidence, Surprise, Causality, Explanation, and Utility” (with discussion), in *Foundations of Statistical Inference: Proceedings of the Symposium at the University of Waterloo, Ontario, Canada, 1970*, eds. V. P. Godambe and D. A. Sprott, Toronto: Holt, Rinehart and Winston, pp. 108–141.
- (1975), “The Bayes Factor Against Equiprobability of a Multinomial Population Assuming a Symmetric Dirichlet Prior,” *The Annals of Statistics*, 3, 246–250.
- (1976), “On the Application of Symmetric Dirichlet Distributions and Their Mixtures to Contingency Tables,” *The Annals of Statistics*, 4, 1159–1189.
- (1979), “Some History of the Hierarchical Bayesian Methodology” (with discussion), in *Bayesian Statistics: Proceedings of the First International Meeting in Valencia, Spain 1979*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Valencia, Spain: University of Valencia, pp. 489–510, 512–519.
- (1980), “Another Relationship Between Weight of Evidence and Errors of the First and Second Kinds,” *Journal of Statistical Computation and Simulation* 10, 315–316.
- (1982a), Comments by G. Shafer, “Lindley’s Paradox,” *Journal of the American Statistical Association*, 77, 342–344.
- (1982b), “Is the Mars Effect an Artifact?,” *Zetetic Scholar*, 9, 65–69.
- (1981/1983), “The Robustness of a Hierarchical Model for Multinomials and Contingency Tables,” in *Scientific Interference, Data Analysis, and Robustness*, eds. G. E. P. Box, T. Leonard, and C.-F. Wu, New York: Academic Press, pp. 191–211.
- (1983a), *Good Thinking: The Foundations of Probability and its Applications*, Minneapolis: University of Minnesota Press, pp. xviii, 332.
- (1983b), “Probability Estimation by Maximum Penalized Likelihood for Large Contingency Tables and for Other Categorical Data,” *Journal of Statistical Computation and Simulation*, 17, 66–67.
- (1983c), “Weight of Evidence: A Brief Survey” (with discussion), in *Bayesian Statistics: Proceedings of the Second International Meeting in Valencia, Spain*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, New York: North-Holland, pp. 249–269.
- (1983d), “Hierarchical Bayes, Mixed Dirichlet Priors, Penalized Likelihood, Etc.,” *Journal of Statistical Computation and Simulation*, 18, 231–234.
- (1983e), “The Bayes/Non-Bayes Compromise or Synthesis and Box’s Current Philosophy,” *Journal of Statistical Computation and Simulation*, 18, 234–236.
- (1984a), “One Tail Versus Two Tails, and the Harmonic-Mean Rule of Thumb,” *Journal of Statistical Computation and Simulation*, 19, 174–176.
- (1984b), “The Harmonic-Mean Rule of Thumb: Some Classes of Applications,” *Journal of Statistical Computation and Simulation*, 20, 176–179.
- (1984c), “A Sharpening of the Harmonic-Mean Rule of Thumb

- for Combining Tests 'In Parallel',” *Journal of Statistical Computation and Simulation*, 20, 173–176.
- (1987), “Hierarchical Bayesian and Empirical Bayesian Methods” (letter), *American Statistician*, 41, 92.
- (1987/1990), “Speculations Concerning the Future of Statistics,” in *Journal of Statistical Planning and Inference: Special Issue on the Foundations of Statistics and Probability* (proceedings of a conference in honor of I. J. Good), pp. 441–466.
- (1988a), “Scientific Method and Statistics,” in *Encyclopedia of Statistical Sciences* Vol. 8, eds. N. L. Johnson and S. Kotz, New York: John Wiley. 291–304.
- (1988b), “The Interface Between Statistics and Philosophy of Science” (with discussion), *Statistical Science*, 3, 386–412.
- (1988c), “What are the Masses of Elementary Particles?,” *Nature*, 332, 495–496.
- (1989a), “Yet Another Argument for the Explication of Weight of Evidence,” *Journal of Statistical Computation and Simulation*, 31, 58–59.
- (1989b), “The Theorem of Corroboration and Undermining, and Popper’s Demarcation Rule,” *Journal of Statistical Computation and Simulation*, 31, 119–120.
- (1989c), “Weight of Evidence and a Compelling Metaprinciple,” *Journal of Statistical Computation and Simulation*, 31, 121–123.
- (1990a), “A Quantal Hypothesis for Hadrons and the Judging of Physical Numerology,” in *Disorder in Physical Systems: Essays in Honour of John M. Hammersley*, eds. G. Grimmet and D. J. A. Welsh, London: Oxford University Press, pp. 129–165.
- (1990b), “A Compromise Between Credibility and Subjective Probability,” *Journal of Statistical Computation and Simulation*, 36, 186–193.
- (1990c), “Comments Concerning the Hadron Quantal Hypothesis,” *Journal of Statistical Computation and Simulation*, 37, 245–247.
- (1991a), “Historical Introduction to Herbert Robbins,” in *Breakthroughs in Statistics* Volume 1, eds. N. L. Johnson and S. Kotz, eds. S. Kotz and N. L. Johnson, New York: Springer-Verlag, pp. 379–387.
- (1991b), “A Comment Concerning Optional Stopping,” *Journal of Statistical Computation and Simulation*, 39, 191–192.
- (1991c), “Ridge Regression and the Harmonic-Mean Rule of Thumb,” *Journal of Statistical Computation and Simulation*, in press.
- (1991d), “The Bayes/Non-Bayes Compromise: a Brief Review (the Longer Version),” Technical Report 91-16B, Virginia Polytechnic Institute and State University, Dept. of Statistics.
- Good, I. J., and Crook, J. F. (1974), “The Bayes/Non-Bayes Compromise and the Multinomial Distribution,” *Journal of the American Statistical Association*, 69, 711–720.
- (1987), “The Robustness and Sensitivity of the Mixed Dirichlet Bayesian Test for ‘Independence’ on Contingency Tables,” *The Annals of Statistics*, 15, 670–693.
- Good, I. J., and Gaskins, R. A. (1971), “Non-Parametric Roughness Penalties for Probability Densities,” *Biometrika*, 58, 255–277.
- (1972), “Global Nonparametric Estimation of Probability Densities,” *Virginia Journal of Science*, 23, 171–193.
- (1980), “Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data” (with discussion), *Journal of the American Statistical Association*, 75, 42–73.
- Green, M. B., Schwarz, J. H., and West, P. C. (1985), “Anomaly-Free Chiral Theories in Six Dimensions,” *Nuclear Physics*, B254, 327–345.
- Hurwicz, L. (1951), “Some Specification Problems and Applications to Econometric Models,” *Econometrics*, 19, 343–344.
- Jahn, R. G., Dunne, B. J., and Nelson, R. D. (1987), “Engineering Anomalies Research,” *Journal of Scientific Exploration*, 1, 21–50.
- Jaynes, E. T. (1957), “Information Theory and Statistical Mechanics,” *Physical Review*, 106, 620–630; 108, 171–190.
- Jeffreys, H. (1938–1961), *Theory of Probability* (1st and 3rd eds.), Oxford, U.K.: Clarendon Press.
- Jefferys, W. H. (1990), “Bayesian Analysis of Random Event Generation Data,” *Journal of Scientific Exploration*, 4, 153–169.
- Kilmister, C. W. (1966), *Men of Physics: Sir Arthur Eddington*, Oxford, U.K.: Pergamon Press.
- Lindley, D. V. (1957), “A Statistical Paradox,” *Biometrika*, 44, 187–192.
- Lindley, D. V., and Smith, A. F. M. (1972), “Bayes Estimates for the Linear Model” (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 34, 1–41.
- Lovett, D. R. (1989), *Tensor Properties of Crystals*, Philadelphia: Adam Hilger.
- McCrea, W. (1991), “Arthur Stanley Eddington,” *Scientific American*, 265, 92–97.
- Neyman, J., and Pearson, E. S. (1928), “On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference,” *Biometrika*, 20A, 175–240, 263–294.
- (1933), “On the Problem of the Most Efficient Tests of Statistical Hypotheses,” *Transactions of the Royal Society London A* 231, 289–337.
- Particle Data Group (1989), “Review of Particle Properties,” *Physics Letters B*, 204, 1–486.
- Particle Data Group (1990), “Review of Particle Properties,” *Physics Letters B*, 239, 1–516.
- Pearson, E. S. (1939), “William Sealy Gosset, 1876–1937 (2) ‘Student’ as Statistician,” *Biometrika*, 30, 210–250.
- Pearson, E. S., Plackett, R. L., and Barnard, G. A. (1990), “Student”: A Statistical Biography of William Sealy Gosset, Oxford, U.K.: Clarendon Press.
- Pelz, W. (1977), “Topics on the Estimation of Small Probabilities,” unpublished Ph.D. dissertation, Virginia Polytechnic Institute and State University, Dept. of Statistics.
- Pratt, J. W. (1965), “Bayesian Interpretation of Standard Inference Statements” (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 27, 169–203.
- Quigg, C. (1985), “Elementary Particles and Forces,” *Scientific American* 252(4), 84–95.
- Ramanujan, S. (1915), “Highly Composite Numbers,” in *Proceedings of the London Mathematical Society* 2, 14, 347–409. Republished in *Collected Papers of Srinivasa Ramanujan*, eds. G. H. Hardy, P. V. S. Aiyar, and B. M. Wilson, Cambridge, UK: Cambridge University Press, 1927. Reprint, New York: Chelsea Publishing, 1962.
- Rowlatt, P. A. (1966), *Group Theory and Elementary Particle*, London: Longmans.
- Savage, L. J. (1964), “The Foundations of Statistics Reconsidered,” in *Studies in Subjective Probability* (1st ed.), eds. H. E. Kyburg and H. E. Smokler, New York: John Wiley, pp. 173–188.
- Savage, L. J., Bartlett, M. S., Barnard, G. A., Cox, D. R., Pearson, E. S., Smith, C. A. B., Armitage, P., Good, I. J., Jenkins, G. M., Lindley, D. V., Pearson, E. S., Ruben, H., Syski, R., van Rest, E. D., and Winsten, C. B. (1962), *The Foundations of Statistical Inference*, London: Methuen; New York: John Wiley.
- Schwarz, J. H. (ed.) (1985), *Superstrings* (Vol. 2), Singapore: World Scientific.
- Shannon, C. E. (1948), “A Mathematical Theory of Communication,” *Bell System Technical Journal*, 27, 379–423, 623–656.
- Slater, N. B. (1957), *The Development and Meaning of Eddington’s “Fundamental Theory”*, Cambridge, U.K.: Cambridge University Press.
- Sudbury, A. (1986), *Quantum Mechanics and the Particles of Nature*, Cambridge, U.K.: Cambridge University Press.
- Tapia, R. A., and Thompson, J. R. (1978), *Nonparametric Probability Density Estimation*, Baltimore: The John Hopkins University Press.
- Thatcher, A. R. (1964), “Relationships Between Bayesian and Confidence Limits for Predictions” (with discussion) *Journal of the Royal Statistical Society, Ser. B*, 26, 176–210.
- West, E. N., and Kempthorne, O. (1972), “A Comparison of the χ^2 and Likelihood Ratio Tests for Composite Alternatives,” *Journal of Statistical Computation and Simulation*, 1, 1–33.
- Wilks, S. S. (1938), “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses,” *The Annals of Mathematical Statistics*, 9, 60–62.