kinds of design and artwork are used in illustrating ways of presenting information visually in simplified form. The book provides less basic step-by-step assistance than *Scientific Illustration,* but it does present many imaginative ideas for expressing data pictorially, and certainly many of these are applicable to the presentation of scientific data and experimental situations. Extremely useful also are different sorts of cutaway and sectional representations, both in detail and in simplified diagrams. The book is clearly written, well organized, and wide in scope.

Bowman is obviously a talented and knowledgeable graphic artist, and on the whole the profuse illustrations in his book are excellent in design and elegant in their illustration of points made in the text. In the foreword, Bowman states that "These design solutions are intended to serve as conceptual models, to be interpreted rather than imitated. With this spirit in mind they have been executed informally in pencil, rather than in formal media such as ink and air brush." The pencil renderings are slightly fuzzy and lack contrast, and I hope prospective buyers will take time to discover the worth of the book and not put it back on the shelf because of this shortcoming.

Practical Geometry for Technical Drawing by S. J. Woolven is a delight, at least to anyone who loves geometry or technical drawing. It should be very useful to any illustrator who needs oc- casionally to include accurate geometrical figures in his drawings. The book gives clear, logical, and detailed instructions for the construction of a wide variety of figures and includes both the hows and whys of each problem. The constructions are clearly illustrated, and I think the book should be a boon to students studying technical drawing. The levels of complexity covered by the author run from how simply to bisect an angle and construct accurate scales, to developments of compound objects. As well as being useful, this book is fun.

F. A. McKittrick Watkins
Department of Biological Sciences,
Florida Atlantic University,
Boca Raton

Crosscurrents in Statistics

The Selected Papers of E. S. Pearson. Issued by the *Biometrika* trustees to celebrate his 30 years as editor. University of California Press, Berkeley, 1966. viii + 327 pp., illus. $6.75.

Joint Statistical Papers. J. Neyman and E. S. Pearson. University of California Press, Berkeley, 1967. viii + 299 pp., illus. $7.

A Selection of Early Statistical Papers of J. Neyman. University of California Press, Berkeley, 1967. x + 429 pp., illus. $14.75.

It is 40 years since J. Neyman and E. S. Pearson launched the series of papers which so profoundly influenced the development of theoretical research in statistics and of the teaching and application of statistical theory. At first, their work appeared to be complementing and extending that of R. A. Fisher, who was already established as an innovator of great genius in the theory of statistical inference [see Neyman, *Science* **156**, 1456–60 (1967)]. However, from 1935 until his death in 1962, Fisher consistently and often scathingly attacked the concepts of Neyman and Pearson and railed against the dominance of these concepts in theory and practice. Since the 1950's, moreover, the Bayesian approach to statistical inference has been making significant inroads into the British and American statistical communities, under the influ-

ence of scholars such as D. V. Lindley and L. J. Savage, and thus is regaining much of the important position it held throughout the 19th century. The new Bayesian school has in common with Neyman and Pearson an emphasis on models for decision-making, but the Bayesians find unacceptable the formulations of hypothesis testing and confidence regions which are basic to Neyman and Pearson. In such circumstances, the volumes under review form a welcome platform, amid conflicting intellectual crosscurrents, for the examination and assessment of the approaches and contributions of their distinguished authors.

The Pearson volume covers the longest span of time, with nine papers from the great productive decade 1928 to 1938 and 11 subsequent papers up to 1963. These papers range widely over topics important in applied statistics. They contain early examples of the use of experimental sampling to determine the effects of failures of assumptions on common tests of significance. Throughout his career, Pearson has expounded and defended the view that testing procedures should be viewed as decision rules and evaluated on the basis of long-run frequency properties, always within specific contexts such as the handling of outliers, the analysis of randomized experiments, or the analysis of various kinds of data having the form of a 2×2 contingency table. Pearson strongly emphasizes the role of theory in planning data collection and in treating the planning and the subsequent data analysis as a unit. A recurring metaphor in his writing is that of the statistician as a craftsman whose theoretical conceptions are tools which facilitate statistical design and analysis. Pearson is sensitive and undogmatic, but a persistent advocate of the Neyman-Pearson position.

The volume of joint papers consists of ten papers from the years 1928 to 1938 which set forth the main features of the Neyman-Pearson theory of hypothesis testing, whose theoretical development continues today over a wide spectrum, extending from very general mathematical theories to specific situations arising in day-by-day statistical practice. It is interesting that the initial paper sets the Bayesian approach on a par with the approach via sampling distributions, but gradually loses interest in the former owing to the apparent arbitrariness of (that is, the absence of a frequency basis for) prior distributions. Much of the early work involves the introduction and application of the likelihood ratio criterion for tests. But the idea of controlling long-run frequencies of errors is present from the start, along with the appealing concept of errors of two kinds, namely the error of accepting a false hypothesis and the error of rejecting a true hypothesis, which grafted a new dimension on the old puzzling (and still puzzling) concept of what Fisher called a significance test. An important breakthrough came in the

1933 paper published in the *Philosophical Transactions* of the Royal Society which established a connection between likelihood ratios and the optimal control of frequencies of errors. The simple lemma (pp. 149–52 in the joint papers) now often called the fundamental lemma provided a basic technical device in much of the theory developed by Neyman, Pearson, and many others, not only in hypothesis testing but in estimation theory and in other branches of frequency-oriented decision theory. After 1933, Neyman and Pearson moved rapidly into exciting areas of new theoretical concepts such as uniformly most powerful tests, unbiasedness, and similar regions based on sufficient statistics.

Neyman's earliest papers create the picture of a young Continental mathematician wavering between a career in mathematical research and a bent for involvement in real-world problems of applying mathematics. The latter won out. The volume under review shows that, during the period of active collaboration with Pearson, Neyman was involved with statistical problems in bacteriology, sample surveys, and agricultural experimentation. A later volume will no doubt record his contributions to stochastic models and statistical methods, up to the present, in such diverse fields as cosmology, weather control, carcinogenesis, and accident-proneness. In statistical inference per se, Neyman's distinctive contribution was the technique of confidence regions, introduced in 1934 as a variant or explication of Fisher's fiducial argument, and later defended against Fisher's attacks as a separate method of "interval estimation" deriving justification from pre-assigned long-run frequencies of covering unknown parameter points. Another series of papers, not represented in the volume of early papers under review, is concerned to build a philosophical viewpoint toward statistical inference, which Neyman called "inductive behavior" and which encompassed the decision-theoretic school that grew up with the Neyman-Pearson work. The mathematical content of Neyman's papers is always very clear, and he was an early exponent of the trend toward removing inessential mysteries from applied mathematics by separating out and exposing the purely mathematical aspects of models, along the lines of modern abstract mathematical theory. Also, as he mentions in his introductory note, a feature of his long and successful research career has been "a tendency of concentrating on some 'big' problem"— a tendency made abundantly clear in his early papers. A valuable feature of the Neyman volume is the inclusion of the discussion of his papers read to meetings of the Royal Statistical Society in 1934 and 1935. These show the reactions to new ideas from the then-current establishment. In particular, they show Neyman acquitting himself with dignity and reason under the first of a long series of rather intemperate attacks from R. A. Fisher, thus establishing a precedent that has served Neyman's cause well over subsequent decades.

The fundamental contributions of Neyman and Pearson are those yielding new general concepts and theories for statistical inference. These should be viewed against the tension which has existed from the time of Laplace between the approaches of Bernoulli and Bayes, or between the interpretation of sampling distributions yielding probabilities relevant before samples are drawn and the interpretation of posterior distributions yielding probabilities relevant after samples have been observed, or, in the 1928 words of Neyman and Pearson, between "two distinct methods of approach, one to start from the population Π and to ask what is the probability that a sample such as Σ should have been drawn from it and the other the inverse method of starting from Σ and seeking the probability that Π is the population sampled." Among statisticians there is, of course, no division into camps of those who always choose one approach and those who always choose the other, since both modes come naturally to probabilistic thinkers and the appropriateness of one or both is partly a matter of taste and good judgment. Through his long career from 1780 onward, Laplace rather confused the picture by prescribing the Bayesian approach as a matter of principle, while often using sampling distributions apparently under the impression that they necessarily would yield the same results. Gauss published a remarkable paper in 1816 in which the use of a sample mean square to estimate a variance was justified as being the most probable value under the posterior distribution, and then to buttress his position he consciously switched to sampling distributions and computed asymptotic relative efficiencies for a range of moment estimators plus the sample median for good measure. In his 1837 book Poisson reacted to La- place by elucidating a Bayesian approach to significance testing, whereas Cournot in his 1843 book reacted by criticizing Laplace's arbitrary prior distributions and advocated the use of a Bernoullian approach not unlike that of many present-day statisticians. Cournot clearly recognized that in large samples the choice of a prior density has little effect and that sampling distribution considerations also converge to the common Bayesian answer. These ideas were also present in the writings of Edgeworth from 1880 to 1920. Edgeworth made clear his understanding of both approaches and their difficulties, while his sympathies lay in the Bayesian direction. Edgeworth's contemporary Karl Pearson, the father of E. S. Pearson, also wrote in both modes, and although he came to deal largely with sampling distributions he can sometimes be seen looking over his shoulder, perhaps with a nod towards Edgeworth. Fisher came on the scene about 1912 with a strong, overt anti-Bayes bias, but gradually came to feel that his methods of likelihood and fiducial probability, while based on sampling distributions, provided in limited circumstances answers possessing Bayesian merits but avoiding the opprobrium attached to prior distributions.

Such are a few of the highlights in the history of statistical inference from Laplace to Neyman and Pearson. Neyman and Pearson rode roughshod over the elaborate but shaky logical structure of Fisher, and started a movement which pushed the Bernoullian approach to a high-water mark from which, I believe, it is now returning to a more normal equilibrium with the Bayesian view.

Two characteristic features of the Neyman-Pearson outlook are (i) the interpretation of statistical procedures as rules of behavior and (ii) the use of sampling distributions to compute expectations interpretable as operating characteristics of the rules of behavior. In a less formalized state, this approach is clearly visible in the Gauss paper mentioned above and in the more famous work of Gauss on linear models where the operating characteristic of an estimation procedure is expected squared error. Neyman and Pearson developed a similar approach for testing rules where power is the operating characteristic, and for interval estimates where a suitable operating characteristic can be expected length of interval. Later, Wald greatly generalized and unified the ap-

proach based on (i) and (ii) above and tied it to the von Neumann and Morgenstern theory of games. A third theme of Neyman and Pearson was to stress the long-run frequency interpretation of operating characteristics, at the expense of, almost to the point of denial of, the interpretation of individual bodies of statistical data.

To accept the relevance of operating characteristics is to accept the theory on its own terms. At this level, criticism of the theory becomes a technical subject largely beyond the scope of this review. Perhaps the chief weakness of the theory is that of inconclusiveness, for, if such conditions as unbiasedness of estimators or fixed size of tests are set aside as artificial, then Wald's theory of admissible decision functions reaches the conclusion that some Bayes rule must be used, thus leading full circle to the central Bayesian difficulty of choosing a prior distribution. On the other hand, if fixed-size tests are to be evaluated in terms of power functions, inconclusiveness again appears because optimum tests in general vary greatly with the choice of alternative hypotheses, and it may be no more plausible to pick out a specific alternative hypothesis than to pick out an intuitively desirable test statistic. Of course, any scientific theory must contend with internal difficulties, so that judgment and good sense play important parts in its use. It seems probable that many statisticians will long continue to accept the Neyman-Pearson theories in such terms, and the future role of these theories in teaching and practice is thus assured.

The deeper difficulties of the theory appear when it is recognized that the interpretation of individual sets of data is after all a necessity. A theory which deals only in long-run frequencies does not possess in itself the resources to deal with such interpretation. Consider for example the comments of Bowley on Neyman's initial and lucid exposition of the theory of confidence intervals in 1934. Regarding such limits for an unknown population proportion, Neyman quotes Bowley's essential comments as:

a) Does it really lead us towards what we need—the chance that, in the universe which we are sampling, the proportion is within these certain limits? I think it does not;

b) I think that we are in the position of knowing that *either* an improbable event has occurred, *or* the proportion in the population is within the limits.

Neyman essentially agreed with Bowley. [In a paper included in his *Selected Papers* Pearson had made a point similar to (a) some five years before.] But Neyman went on to identify the confidence method with long-run frequency interpretations. Perhaps this is too great a jump, for if (b) holds then it is logically necessary and not just a convenience to restrict confidence coefficients to fractions close to unity. Moreover, if (a) and (b) hold, then it need trouble no one that a 95-percent confidence interval might, conditional on some observable aspect of the data, have confidence only 80 percent or even 100 percent, where 100 percent obtains if a confidence procedure should produce an interval containing all possible values of the parameter. But most users of confidence regions are quite disturbed by such properties, their uneasiness being a sign of the difficulty of not reading a posterior probability interpretation into confidence coefficients. A similar point may be made regarding hypothesis testing, conveniently in terms of Pearson's hypothetical example of King Hiero's crown (*Selected Papers*, pp. 279–83). The king is represented as being happy with a decision procedure which offers chance .01 of hanging an innocent man and chance .01 of failing to hang a guilty man, under plausible assumptions. Would the king feel happy about applying the rule to hang a man about whom he had some extra, perhaps intangible, information pointing towards innocence? The existence of such information need in no way alter the long-run objective operating characteristic .01, but it could certainly affect one's feeling about the relevance of the operating characteristic to a specific real situation.

The criticisms just made are very much in the spirit of Fisher, who did believe in the necessity of being concerned with individual circumstances and did believe that his fiducial argument, suitably hedged, was protected in a way that the confidence argument is not. Moreover, the Bayesian view is very close to Fisher in this regard. I believe that Neyman and Pearson have never met the criticisms head on, nor indeed is it possible to do so satisfactorily. The best defense is perhaps an appeal to pragmatism and common sense; after all, when confidence coefficients or significance levels are obviously not relevant this fact will usually be recognized and a new,

more suitable procedure constructed. In view of the counterbalancing difficulties in applying reasonable Bayesian methods, the pragmatic defense is probably acceptable for most widely used procedures; but constant care is required.

We are in little danger, I think, of rejecting any true theory of statistical inference in the near future, and must be satisfied with partially accepting each of two false, or at least oversimplified, theories. An uneasy truce between the posthumous contributions of Bernoulli and Bayes seems likely to go on for some time. But we have Neyman and Pearson to thank for their important role in lifting the debate to new levels of theoretical sophistication where the issues can be seen with great clarity.

A. P. DEMPSTER

Department of Statistics,
Harvard University,
Cambridge, Massachusetts

Quaternions and Other Topics

The Mathematical Papers of Sir William Rowan Hamilton. Vol. 3, Algebra. H. HALBERSTAM and R. E. INGRAM, Eds. Cambridge University Press, New York, 1967. xxiv + 672 pp., illus. $37.50. Cunningham Memoir No. 15.

The first two volumes of *The Mathematical Papers of Sir William Rowan Hamilton*, edited by A. W. Conway and J. L. Synge under the auspices of the Royal Irish Academy, appeared in 1931 and 1940. Now the work is to be brought to completion by H. Halberstam and R. E. Ingram. Volumes 1 and 2 dealt with geometrical optics and dynamics, respectively. The fourth and final volume, promised in a few years, will be on analysis and geometry. The main topic of volume 3 is the algebra of quaternions. This volume also includes early work on complex numbers and papers on the icosian calculus.

Almost all the articles in volume 3 are published papers. The editors have included the preface to *Lectures on Quaternions* because it "lends cohesion and perspective" to other material in the volume. They reprint the (previously published) manuscript pages written on the day when quaternions were discovered. Several other manuscripts are also included; the major document on the icosian calculus is perhaps the most