The Annals of Statistics
1980, Vol. 8, No. 1, 1–24

L. J. SAVAGE—HIS WORK IN PROBABILITY AND STATISTICS

BY D. V. LINDLEY
University College London

Leonard Jimmie Savage was born in Detroit on November 20, 1917 and died in New Haven on November 1, 1971. He spent almost all his working life in the two fields that we today call (for example, in the titles of our journals) probability and statistics. He made important contributions to both subjects and discussed, more than his published works indicate, applications of them to many other branches of knowledge. He was interested in so many things in life, and saw in them the relevance of statistical and probabilistic ideas, so that he came, as near as it is today possible to become, a polymath, in marked contrast to the behaviour of most of us who are content to become experts in a narrow field. He was too a scholar, a man who respected the depository of knowledge of earlier workers and who carefully credited others with what he generously saw as their achievements. (We shall see a striking instance of this below.) But above all he was a revolutionary, in the sense of Kuhn (1970), a man who replaced the accepted paradigm of inference by another, without, at first, realising what he had done. This paper is an attempt by a statistician to describe Savage's technical achievements.

In writing this appreciation of his work I have benefited enormously from some material provided by David Blackwell and Lester Dubins, portions of which, particularly on the more mathematical aspects of Savage's work, have been incorporated into the text. Others who have helped in providing information include W. A. Ericson, J. W. Pratt, R. A. Olshen and I. R. Savage. Unpublished material by W. H. Kruskal, F. Mosteller, W. Allen Wallis, F. J. Anscombe and Paul Feder has been of great assistance. I apologise to the statistical community for the delay in producing the paper caused by my own sense of inadequacy for the task, a sense which increased as the work progressed, and I appreciated the magnitudes of his contributions to our subject.

He was trained as a mathematician and an early paper (1943) testifies to this. It generalizes "a result from an analytic Finsler space and analytic hypersurface to a F-space of class C''' and hypersurface of class C''''". Here we see a dominant feature of modern mathematics, the effort spent in generalization so that a result becomes available under increasingly wider conditions and the essence of the result is more easily appreciated. He was later to exploit this feature in his study of decision-making under uncertainty, so that the results of the probability calculus were seen to hold very generally. This latter generalization was to have important practical consequences. A later paper (1946b) is in a similar mathematical vein. The conditions under which a metric, or more generally, a distance, space M can be embedded isometrically as a subset of a Euclidean space, is a question that had been studied in the 1930's. For instance, Wilson had shown that if M is suitably

Received October 1978.
AMS 1970 subject classifications. Primary 01A70; secondary 00A15.

convex and every set of four or fewer points of M is embeddable in Euclidean space, then so is M. This paper reestablishes and elegantly generalizes this and other related results.

These two papers were preceded by a joint paper with Arthur Smithies (1940). This contains a mathematical treatment of the economic situation of two duopolists facing a common demand function and common costs, and deciding what to do on the basis of what each knows about the other. This is prior to the publication of von Neumann's and Morgenstern's book on game theory (1944). That book also contained their famous axiomatization of utility which showed that preferences amongst outcomes, when allied to assumptions about probabilistic mixtures of outcomes, could be described in terms of utilities for those outcomes in the sense that the optimum course of action was that of maximum expected utility. Savage was very excited by this and it became the starting point for his own wider axiomatization for decision-making embracing both probability and utility, though it should be remembered that Savage had been von Neumann's assistant for a year and was undoubtedly familiar with the seminal paper (1928) so that the book probably only served to reinforce an idea with which he was already familiar.

His first major paper on utility theory, joint with Milton Friedman (1948d), suggested that the utility of many people for money income is increasing, "is convex from above below some income, concave between that income and some larger income, and convex for all higher incomes." This paper has been widely read and quoted and has been twice reprinted. A footnote at the beginning explains that "the fundamental ideas of this paper were worked out jointly by the two authors. The paper was written primarily by the senior author" (MF). It does not adequately distinguish between the normative and descriptive aspects of utility theory but otherwise explores carefully and lucidly some of the implications of the maximum-expected-utility hypothesis and did much to encourage sympathetic consideration of that viewpoint. Savage defended this hypothesis (1948a) and again in a sharp rebuttal to Wold (1952b). He wrote a second paper (1952c) with Friedman in which they again defend the maximization of expected utility. The paper is essentially a reply to one by W. J. Baumol giving an ordinalist view. The paper is especially remarkable for the statement, I think for the first time, of the sure-thing principle, though without that name. The principle says that if action a_1 is preferred to action a_2 if event A obtains, and also if it does not, then a_1 is preferred to a_2 even when one is uncertain about A. The authors describe it as "universally known and recognised; and the Greeks must surely have had a name for it, though current English usage seems not to." The principle is compelling and therefore important as part of an axiom system for decision-making under uncertainty, for if the axioms are compelling so is the whole structure based upon them.

The work of Allais, cited in Savage's book (1954), was the most famous attempt at the time to argue against the notion of maximising expected utility. I think it is fair to say that no counter-example has yet succeeded in demonstrating the unsoundness of the concept, though it is still far from being generally adopted as a criterion for decision. This is not the place to speculate on the reasons for this, but the failure to appreciate the essentially probabilistic character of utility where the probability is that of belief, not frequency, is reasonably the principal cause. Whilst probability is misunderstood, utility will suffer the same fate.

His work in economics was clearly influenced by his contact with Milton Friedman. It is not always recognised how much he owed to Friedman in other respects. For example, Friedman played an important role in developing his writing style. But most of all the two discussed the early ideas of personalistic probability and utility together, and the final form that found expression in his writings owes much to those debates. His respect for de Finetti (see below) was that of a kindred spirit; that for Friedman was of a stimulus, forever questioning and encouraging the new thoughts. His interest in economics persisted and he wrote another paper (1955a) with J. H. Lorie in this field.

Savage's statistical work began with his joining Columbia University's war-time Statistical Research Group. There he interacted with a group of statisticians that included Churchill Eisenhart, Milton Friedman, Abraham Girshick, Harold Hotelling, Frederick Mosteller, Abraham Wald, W. Allen Wallis and Jacob Wolfowitz. In his own words it was "one of the greatest hot-beds statistics has ever had." His first published statistical paper (1946a) (both (1947b and c) were written earlier), with Girshick and Mosteller, on unbiased estimation of a binomial p in the case of a general stopping rule, was immediately influential. It combined the idea that an estimate should be a function of any sufficient statistic with the germ of the concept of completeness. Wolfowitz (1946) immediately recognised that for some purposes bounded completeness was the right concept, and used it in extending their work. The results of the first paper were extended by Savage alone (1947a) who obtained a necessary and sufficient condition for an estimate to be the unique, bounded, unbiased estimate for a closed stopping rule. It is interesting to reflect that this early work was concerned with a notion (unbiased estimation) that he was later to reject. In conversation at the time, though not in print, he, and others, criticised the notion of unbiasedness because of the crazy fact that the unique unbiased estimate of p could be 0 after one had just had a success (following some failures) in the geometric distribution. Two other papers (1947b and c), included as parts of the Statistical Research Group's book, are on the design of experiments. The title of the first, "Abandoning an experiment prior to completion," today immediately suggests the likelihood principle but in fact deals with a modification of a one-sided sequential test to attain required probabilities of error of both kinds. The second contains ideas related to the up-and-down method, presented as "just a suggestion." He was coauthor with several other members of the group of a book (1948) on sampling inspection. His own work is not individually identified though there is a reference to the 1946 paper.

In 1947 he went to the University of Chicago and in 1949, with W. Allen Wallis, founded the Statistics Department. This soon became one of the world's leading statistical groups and remains so today. There he collaborated with Paul Halmos and together they produced a paper (1949a) that was to have considerable influence on the future course of mathematical statistics. The effect on myself is perhaps typical of its effect on all mathematicians who, as a result of the war, had become interested in statistics and wanted to see the subject treated with the same excellent standards of rigour that prevailed in other mathematical subjects. For example, in this paper the discrete and continuous cases were not treated separately but given a unified treatment using the concept of a dominating measure. How aesthetically satisfying this was to us young mathematicians who had been greatly stimulated by Wald's ideas but had found the techniques cumbersome. We were excited by a precise explication of Fisher's brilliant idea of sufficiency and appreciated the idea that a sufficient statistic plus a random number generator give the equivalent of the original data. The paper also gave prominence to a convincing argument for basing inference on sufficient statistics alone, introduced pairwise sufficiency and showed that the likelihood ratios are a minimal pairwise sufficient statistic. Here we have more than a suggestion of the likelihood principle, though he claimed (in the preface to the Dover edition (1972) of his 1954 book) that even in 1954 he was not aware of it, attributing it to Barnard (1947). The recognition of this principle is, for me, one of the great contributions of post-war statistical thinking to the whole problem of inference, and its violation by so many common statistical procedures a source of disquiet. Two quotations from this paper are worth giving.

"We gather from conversations with some able and prominent mathematical statisticians that there is doubt and disagreement about just what a sufficient statistic is sufficient to do, and in particular about in what sense, if any, it contains 'all the information in a sample'." This is a striking illustration of how two of the most elegant mathematical writers of their day interrupted their powerful logical arguments with informative commentary. The second shows an unjustified optimism. "In ordinary statistical parlance one often speaks of a statistic sufficient for some of several parameters. The abstract results mentioned above can undoubtably be extended to treat this concept." This paper, very influential in subsequent work, did not lead to such an extension.

The second Berkeley symposium found him coauthoring again with Girshick (1951a). They studied exponential families and estimation with quadratic loss. Results were obtained that certain estimates were minimax, though much of the work was restricted to translation families. A conjecture regarding admissibility still remains unsettled. We find them writing "the existence of an *a priori* distribution may not be valid, or if such a distribution exists, it may be unknown," a view that he later regarded as incorrect. Another paper (1951b) published the same year is the first substantial indication in print of his interest in the foundations of statistics, though at the Boulder meetings of the Econometric Society (1950) in 1949 he had given a paper on the role of personal probability, commenting on Wald's theory (Wald had also given a paper) that it seems "to lead to insurmountable obstacles" and advocating, instead, de Finetti's views. The paper is an extended review of

Wald's book and in it Savage attempts to fill in the gaps of interpretation that Wald left. He writes enthusiastically about the minimax principle describing it as "the only rule of comparable generality proposed since Bayes' was published in 1763". He is nevertheless not quite happy with it and dislikes the minimax estimate for the binomial p, attempting defences using group decision ideas and by invoking prior opinions about p. Wald had never discussed the interpretation of the concept of a loss function. Savage attempted to do this, introducing what later came to be called the regret function, being, for any state, the difference between the utility for the best action and that for the action actually taken. It is only with this concept of loss that Wald's ideas make coherent sense.

The next few years were mainly taken up with the writing of his book but he takes time off to write a short note with Nunke (1952a) on finitely additive probability measures, showing that they can be nonatomic and yet assign no event a probability, say, between one-third and two-thirds. The point here is that Liapounoff's result to the effect that nonatomic, countably additive, vector-valued probabilities have convex ranges is not true in the finitely additive scenario. The Wald review had shown him to be aware of the work of de Finetti and one is led to suggest that his ideas had led Savage to a study of finite additivity.

Savage's greatest published achievement was undoubtedly his book (1954), *The Foundations of Statistics.* As we have seen he had been impressed by the work of von Neumann on the axiomatization of utility, showing that there was a function, termed utility, whose expected value was a reasonable sole criterion for action (see page 97 of the book). He conceived the idea of generalizing his work to develop, from an axiomatic basis, not only the idea of utility but also that of probability. He succeeded brilliantly. The end papers give the axioms of what he called the personalistic approach, and the first seven chapters develop the results, with the commendable attention to rigour and the pleasant style that we have noted in his earlier work. The formal subject matter of the theory[1] consists in acts, states and consequences, an act being a function from states to consequences, with a preference relation between the acts. This relation is extended to include preferences given an event (a subset of states) and hence to the sure-thing principle. This, in turn, leads to relations between events. Additional restrictions make this last relationship probabilistic. The von Neumann argument then immediately produces utilities, though an additional postulate is required for extension to infinite sets of consequences. Three points are worth noting: first, the essentially economic, action-oriented, basis of the whole work. Second, the simultaneous derivation of both probabilities and utilities, and their combination in the principle of maximization of expected utility. Third, the commendable care with which the whole work is organized, particularly in the quantitative notion of probability. No work of science stands for long for it is the nature of science to improve upon the work of others,

[1]An earlier outline of the ideas exists (1952d).

but these seven chapters will stand as the basic statement of the Bayesian viewpoint for as long as that viewpoint survives.

If this part of the book was a triumph, the last part was a failure, but a failure that led to much clearer understanding. He had been impressed with what he called the British-American school of statistics, with the work of Fisher, Neyman, Pearson, Wald and others based on frequentist ideas of probability. He felt that the considerable success of these ideas, both in theory and practice, must be because they were fundamentally correct and that therefore the foundations developed in the first part of the book would provide them with a sound, logical under-pinning that they then lacked. The last ten chapters are an attempt to justify these established methods. In fact the logical basis shows many of them to be ill-founded. We have his own agreement to this view. In the preface to the 1972 edition he writes: "The second part of the book is indeed devoted to personalistic discussion of frequentist devices, but for one after another it reluctantly admits that justification has not been found. Freud alone could explain how the rash and unfulfilled promise [to show how the frequentist ideas are on the whole consistent with the alternative proposed] went unamended through so many revisions of the manuscript." Illuminating remarks on this point are also to be found in 1962g; including "As I wrote, I became increasingly deaf to my own leitmotif".

I do not think we need to invoke Freud. The views of Kuhn (1970) on the nature of revolutions in science give valuable insight into what had happened. Kuhn argues that normal science (for us, normal statistics) consists of a paradigm within which all work takes place. At intervals scientists studying some aspect of the paradigm come up with a new one, often without realising that they have done so. For a variety of reasons, the new paradigm eventually replaces the old one. The discovers of new paradigms are the true revolutionaries of science and Savage was a revolutionary of the 1950's.

There had been two earlier revolutionaries producing what is essentially the same new paradigm: Ramsey and de Finetti.[2] Ramsey had unfortunately died very young, so that his brief message had not been understood. de Finetti writing mostly in Italian and outside the British-American school (though contending with the Italian paradigm for statistics) had not been appreciated by that school. Savage performed an enormous and scholary service to us by bringing the work of these two writers to our attention and explaining it to us. He was probably the first to understand what Ramsey had said, unlike Lytton Strachey who felt it was wonderful (like Newton, he said) but did not know why. He was also the first to understand the importance of de Finetti's concept of exchangeability in relating the work of the frequentist and personalistic schools. My own guess (confirmed in a letter from W. Allen Wallis) is that he discovered these results for himself and that

[2]Ramsey's work is most easily read in Kyburg and Smokler (1964), which also contains Savage (1961). de Finetti (1974, 1975) has recently published a very complete account of his views.

it was this discovery, and the understanding thereby gained, that enabled him, more than others, to appreciate the earlier work and greatly to extend it. He once wrote to me: "I am reminded when I see expressions like post- and pre-Savage that you are turning too much limelight on me. A reader familiar with Ramsey, Jeffreys, de Finetti and Good has not really so much to learn from Savage. I think, as you seem to, that my main contribution has been to emphasize that the theory of subjective probability finds its place in a natural theory of rational behavior." That was overmodest for 1958: ridiculously so for 1978 when the implications for statistics are better understood.

The book therefore shows him in a transition period. He had accepted the personalistic view of probability but had not abandoned minimax or other ideas of the frequentist school. Minimax ideas play a dominant role in the second part of the book. In 1961 he was to say "those of us who . . . hoped to find in this rule an almost universal answer to the dilemma posed by abstinence from Bayes' theorem have had to accept disappointment." More important, perhaps, he had still not recognized the likelihood principle. He was later to say (1962a) "I came to take . . . Bayesian statistics . . . seriously only through recognition of the likelihood principle." Perhaps because of these flaws, the book did not have an immediate impact and people today still write as if it did not exist. It seems astonishing to me that people, twenty years after the event, can still write learned papers on inference that ignore Savage's work. That they might disagree with it is understandable, but the ignorance is puzzling. It is perhaps idle to speculate on why the Bayesian viewpoint did not catch on as, say, the empirical Bayes attitude did, but part of the difficulty lies in the extremely revolutionary character of the argument and part in the failure to provide workable techniques—a point which is commented on below.

Savage's influence was not confined to his writings. Personal contact with him was a tremendously exciting experience. Not for him the quick, brilliant, sparkling, superficiality of workers wedded to a paradigm. Rather a slow, careful, dissection of the ideas; a rebuilding of them in a clearer form that made them seem so simple that their importance could easily be underestimated. Let me indulge and record my own experiences.

I had studied under Jeffreys but had never been totally convinced by his ideas for two reasons; firstly, the environment of influential British statistics which was obviously successful and yet unsympathetic to Jeffreys' views; secondly, because he did not provide a convincing reason for thinking that prior probabilities "existed." When W. Allen Wallis took me to tea in a London hotel and persuaded me to go to Chicago for a year, I was impressed by his tale of this extraordinary man in Chicago who had written (this was 1953) a book "justifying" prior probabilities. The year (1954-55) spent in Chicago working with Savage was a revelation. He took my own clumsy ideas, which were leading, within the frequentist framework, to personalistic concepts and patiently showed me how unnecessarily elaborate they were. Eventually we (he?) came up with what Birnbaum has generously called the Lindley-Savage argument. Now the ideas looked so simple that I felt they were obvious and not worth publishing: they were so obviously correct that everybody must know about them.

The importance of the first part of the book is that it provides a formal description of inference and decision-making based on the standard mathematical method of axioms and theorems deduced from them. Savage was the Euclid of statistics. He set out to do this and succeeded, expecting it to provide a firm foundation for the work of the British-American school. The surprise was that the structure set up turned out to be in conflict at many points with standard methods. As we have seen, he was slow to realize this—though faster than the rest of us. Nevertheless he became, in the period between the publication of the book and his early death, the leader of the personalistic school. It is unfortunate that he did not live to see it finally triumph and provide the paradigm for normal statistics as some of us think it will.

One of de Finetti's fruitful ideas was that of exchangeability (or what used to be termed symmetry). Savage, with E. Hewitt, published an important paper generalizing his ideas. A symmetric distribution on an infinite product space is a probability measure that is invariant under the group of permutations of the coordinates that leave all but a finite number of coordinates undisturbed. An example is a power distribution, that is, the distribution of an infinite sequence of independent, identically distributed random variables. That every real symmetric distribution is a probabilistic mixture of power distributions was a discovery that de Finetti made in 1931. Hewitt and Savage's (1955b) paper explores the ramifications of his work, and establishes the now celebrated "Hewitt-Savage Zero-One Law". Their work was not expressed in terms of distributions but rather for the underlying probabilities.

They proved that every element of a convex, weakly compact subset K of a dual space is the barycentre of a countably additive probability measure μ that is supported by the closure of the set E of extreme points of K. Choquet, Bishop and de Leeuw later showed that μ could be taken to live on E itself rather than on its closure.

As abstracted by Hewitt and Savage, de Finetti's work became the study of the convex set K of probability measures invariant under a group G of transformations of the underlying space, with special reference to the extreme points of K. The latter are necessarily ergodic; that is, assume only the values 0 and 1 on the G-invariant sets. They showed that, for the permutation group which is relevant to de Finetti's theorem, the converse also holds and a G-invariant measure is determined by its values on the G-invariant sets.

For K, the set of symmetric distributions, they showed that each element of K is the barycentre of a *unique* distribution on the extreme points of K. This is the first such uniqueness result known to us. If the underlying space is a finite Cartesian product, then each probability invariant under all permutations of the coordinates is the barycentre of hypergeometric distributions. They did not notice, as later did de Finetti, that their result leads to a new and simple proof of de Finetti's original theorem.

This paper had a substantial influence on subsequent mathematical work, but although it lies at the basis of nonparametric statistics, which deals with sequences having the exchangeability property but are otherwise unstructured, it had no effect on that subject because it did not address itself to the operational problem of describing the mixture distributions, a problem that is still unsolved. At least two workers were influenced by Savage's idea here. Dubins and Freedman (1963, 1967) studied random distribution functions. Perhaps nonparametric problems were puzzling him at that time because he wrote another paper with R. R. Bahadur (1956) on that topic. In it they show that if the family of possible distributions of a real, random quantity is sufficiently rich (in particular much richer than one with a finite number of parameters) then there are neither effective tests, point estimates nor confidence intervals for the mean. These conclusions follow since μ is sensitive to the tails of the distribution, about which there is little or no information in a sample.

The next issue of the *Annals* has him (1957a) commenting on some work of Blasbag, who had shown that for the family $dP(x, \theta) = \exp\{r(\theta)A(x) + s(\theta)B(x)\}\, dw(x)$ there exists an infinite set of values $(\theta_0, \theta_1, \alpha, \beta)$ giving the same sequential test of θ_0 against θ_1 with error-rates α and β. He proves that these are (at least practically) the only families thus degenerate.

In 1958 he addressed the International Congress of Mathematicians in Edinburgh (1960a). There he gave one of the many lectures he delivered over the next 14 years in which he talked about the development of his ideas on statistics. We have just seen him, with Hewitt and Bahadur, dealing with points of mathematics. In these papers, in contrast, he was discussing the interpretation of mathematics, how it could be used and understood: and, in particular, how the recent discoveries were affecting one's ideas about current statistical practice. It was this ability to see both sides — the mathematical and the interpretive — that made his understanding of our subject so profound. Rereading these in sequence one has a fascinating glimpse of the new ideas not so much emerging, they were there in the Foundations, but being appreciated as revolutionary. Thus in the Edinburgh paper he is tentatively suggesting that confidence limits may not be of lasting value. As the number of "odd" properties of confidence intervals continues to grow, it looks to me as if he was right in that judgement. By now he has appreciated the likelihood principle, championing Barnard again, and remarking that "it contradicts much that was recently most firmly established in statistical theory and practice." The central thesis of the talk is well summarized by the quotation: "There seems to me nothing at the present time to substitute for the hope that an economic theory of decision in the face of uncertainty will be a valuable guide for the whole problem of inference."

1960 finds him giving a similar paper at the Fourth Berkeley Symposium (1961). After commenting on the various views of probability and describing the behaviouralistic approach, he contrasts objectivity and subjectivity. "(The Bayesian) approach is not to urge the person to ask himself what qualitative properties he likes a procedure to have but to ask himself . . . when he would prefer one procedure to another. A few strongly appealing principles of coherence often succeed in making this task relatively easy." The coherence argument, the sole basis for the personalistic view, is then illustrated with the simple dichotomy. It is interesting how he again attributes so many of the basic ideas to Ramsey and de Finetti in a sincere blend of modesty and scholarship. The paper concludes with a brief discussion of the likelihood principle (including optional stopping) and the principle of precise measurement that says, roughly, if the measurement is precise the prior is (almost) irrelevant. The topic of optional stopping, first recognized by Savage, was much discussed around that time. The idea, derived from the likelihood principle, is that (under wide conditions) the rule that led one to stop taking observations is irrelevant to the final inference. This is nonsense within the sampling-theory framework where the integration over all possible samples is at the heart of the method. A famous example is that of Armitage, in Savage (1962b), where one samples to a foregone conclusion, that is until the result is significant at a preassigned level. The fault, as Savage quickly pointed out, lies in the significance test, not the optional stopping. A good account is contained in Raiffa and Schlaifer (1961). The topic was a good testing ground for Bayesian ideas, and the attempt to overthrow them using it was a failure.

A paper, somewhat similar in scope but directed to an audience less sophisticated statistically, was given in April 1961 in the third symposium on decision and information processes (1962g). There is more awareness of the role of coherence (or consistency) as providing a framework for the assessment of probabilities; but the main topics are the simple dichotomy, the likelihood principle and stable estimation, as before. The style is more informal than at Berkeley and his delightful lecture manner comes through to the printed page as when he says of his lecture "I can only give you a glimpse at what the shouting is about," or when he uses the example of "an experiment to see whether aspirin makes rabbits ears curl." He was fond of such witty illustrations but he has been criticized for not dealing with more serious problems. The difficulty in the latter suggestion is that the example would then have, when treated in a fully Bayesian way, to grow to such an extent that it would dominate the paper. A full personalistic account would not stop at samples from $N(\mu, \sigma^2)$ but would have to consider μ and σ^2, not as Greek letters, or even as means and variances, but as real things in the world about which opinions have to be articulated. Personalistic statistics is so much more practical than any other kind that it becomes hard, and undesirable, to separate it from the science to which it is being applied. Hence the rabbits ears, or later (1968) the speed of neon light in beer.

The important principle of stable estimation was expounded with care in the famous paper that he wrote with W. Edwards and H. Lindman (1963). Conditions were given in the one-parameter case for the posterior distribution to be insensitive to the prior, the conditions being in the form of inequalities on the prior and likelihood. More recently this principle has come in for some criticism in the multiparameter case, though these doubts are already there in the paper (page 232).

This has largely arisen because of the important work of Stein (1956) on the estimation of the multivariate normal mean where the sample mean is inadmissible in dimensions more than two. It is not that the principle is wrong in high dimensions, but rather that the sample size has to be so large before the reasonable conditions for its validity are satisfied. Like so many limiting results it paid inadequate attention to just how fast the limit is approached. The paper also contains some interesting material on Bayesian distribution theory, paying especial attention to the normal case, and a careful account of some of the basic ideas in hypothesis testing, using the Bernoulli case as illustrative material. The latter topic is used to provide an example of a situation where substantial differences arise between the Bayesian and sampling-theory schools, *high* significance for a null hypothesis giving *low* probability that it is false, a result usually known as Jeffreys' paradox. The paper concludes with a discussion of the likelihood principle, by now properly appreciated.

Although this paper is well known, its influence has not, I think, been what the authors had hoped. There are two reasons for this. Firstly, designed as a paper to guide practising psychologists on how to use personalistic ideas in lieu of significance tests and other frequentist methods, it did not pay enough attention to the problem of how scientists should articulate their opinions, opinions that are a necessary ingredient of the new methods. Savage was to study this matter later (1971a) but he never overcame the second, more serious defect. If Bayesian methods are to get off the ground and to be used by scientists, then they have got to be provided in a form that a scientist can use without his being diverted into considerations outside *his* science. The analysis of variance is a popular tool partly because it consists of a series of rules that the scientist can operate to produce a result meaningful to him. It is essential that personalistic statistics have such procedures before it can be used. It is a great shame that he did not devote some of his intellectual effort to the construction of operational techniques. What, for example, would he have done with the results of a factorial experiment with three-factor and higher interactions confounded? I am not here criticizing him in any way; he did enough in all conscience. But perhaps statistics would have benefited more if he had not been so punctilious in reply to correspondents and so helpful with students, and instead developed more operational methods that the writers and graduates could have used. That he was aware of the problems is clear from the papers of his that he helped to write on applications.

The likelihood principle "came of age" with Birnbaum's (1962) important paper. Savage (1962a), commenting on it, said "this is really an historic occasion," and "really momentous in the theory of statistics." His judgement that "it would be hard to point to even a handful of comparable results" is one with which many of us would concur. The principle says that if the observed random quantity, X, is described by a model in which the probability density $p(X|\theta)$ is known, apart from the value of the parameter θ, then when X is observed equal to x, the likelihood function $p(x|\cdot)$ provides the sole source of information about θ from the observation that $X = x$. It is immediate from Bayes theorem, though Birnbaum's approach was different, using principles of sufficiency and conditionality. Almost all the procedures advocated by the British-American school violate the principle. An interesting commentary on Birnbaum's ideas was later provided by Durbin (1970). Savage (1970a), discussing this, persuasively argued in favour of the original approach.

The same year as Birnbaum's paper saw the publication (1962b) of a discussion held in London, opened and closed by Savage with prepared contributions by Bartlett, Barnard, Cox, Pearson and Smith and spontaneous remarks by others. There is substantial overlap with the papers just discussed but the monograph is, as a whole, worth reading, even if some of the ideas have been improved on since. Precise measurement is described again but there is more emphasis on differences between personalistic and frequentist statistics. Savage describes curious confidence intervals and unexpected tests that seem to him to discredit the sampling theory view. (It is curious to me how little attention the discussants pay to these examples. It is the same today when the examples get "curiouser and curiouser." The embarrassments of the advocates of frequentists views is understandable; their choosing to ignore them is hardly in the best scientific tradition. But, according to Kuhn, this attitude to counter-examples is common amongst adherents of a paradigm that is about to disappear, not that it shows much signs of doing so at the moment, despite the growth of data-analysis and Bayesian arguments.) One of his happiest examples, that of Hiero's crown, is included. There is some discussion of randomization, a topic that for many people does not seem to fit into the personalistic framework. But an equally interesting feature of the monograph is the discussion in which the participants show a commendable tolerance towards other viewpoints. Everybody seems to be trying to understand everybody else, and there is little of the sharp writing that sometimes occurs in the discussions at the Royal Statistical Society.[3] All Savage's writing has this tolerant quality even when, as often happens, he is being severely critical. When someone has put forward a formal theory that no one can upset; when that theory goes counter to established practice and provides hosts of counter-examples that flaw that practice; when the practitioners of the establishment continue blithely on their way; when all this happens, it is easy for that someone to rise in wrath and declare the whole pack a lot of fools. Savage never did this in print but tried patiently and persuasively to educate us.

Two later papers are more philosophical in vein. The first (1967a) was based on a talk given to the American Philosophic Association and discusses difficulties in the theory of preference and in the uncertainties we all feel about mathematical statements, such as the truth of Fermat's last theorem. At present the theory is

[3]Savage himself said (1970b): "Some have found this little volume valuable because of the give and take."

inadequate for such questions because its postulates require that in determining your probability for any event you utilize the relevant logical implications of all that you know. The second (1967b), later that year, is concerned with induction. What rational basis is there, he asks, for any of our beliefs about the unobserved? The reply is none; all belief about the unobserved is personal opinion. The paper concludes with a beautiful discussion of possible meanings for and bases for believing in universal sentences and frequency generalizations of great impact in philosophy and science; sentences such as "all men are mortal" and "about 51% of the babies born alive in Boston are boys."

Next year, he is contributing, albeit rather briefly, to the "Future of Statistics" (1968). The talk was not prepared in detail and the discussion has suffered from tape-editing. In his opening remarks on the future he has some wise things to say: "Whoever perceives the present very well will be seeing about as far into the future as there is any hope of seeing." He debates with himself on the possibility of a formal theory of ingenuity, remarking that "most theories of inference tend to stifle thinking about ingenuity"–surprising from one of the originators of an important theory of inference. He then has some kind words to say about descriptive statistics, factor analysis and "fooling around with data." We have already noted his concern with nonparametric statistics, in particular how hard it is to fit it into the Bayesian framework. He returns to the topic here, and also to stepwise regression. His view is much more open and hesitant than my own confident crystal-ball gazing, Lindley (1975). In the discussion his reply to Hartley is amusing and to the point. The 1970 paper on reading suggestions (1970b) is important because it provides an informal, yet thorough, guide to the literature designed for his graduate class at Yale. The astonishing honesty of the note is striking. For example, "This looks important, but I was never able to get into it. A report would be useful to me, but I would not want anyone to suffer too much on my account." is in reference to Harrod's book (1956). Or of Tukey (1962): "Words fail me. Different, important and slippery. We should all tackle it together." I wonder if they did. Two notable omissions from the list are the greatly underrated book by R. T. Cox (1961), which he reviewed in 1962, and that by Braithwaite (1955).

At a symposium in Burcharest in 1971, Savage (1973b) gave a paper repeating the seemingly heretical thesis that the probabilities which are meaningful and important in science are not objective but rather are always personal probabilities. This would appear to contradict the fact of agreement between scientists. Two examples are discussed which show how this agreement arises. The first concerns the final angular position of a spinning wheel where a uniform distribution closely approximates the opinions of most people. The second shows why people with initially different and vague opinions who obtain substantial data concerning past births will all come to assess probabilities close to one in eighty that the next live birth will be that of twins. Extensive use is made of de Finetti's results on exchangeable events.

Apart from this interpretive work on the foundations of statistics, he was, during some of this period, working with Lester Dubins on his second important book[4] with the charming, if slightly misleading title, "How to Gamble if You Must" (1965), though the second part, "Inequalities for Stochastic Processes," does hint of the sterner things to come. The book exists in mimeographed forms dated 1960 and 1963 which do not differ substantially from the published 1965 form. Two papers (1960b, 1965a), which are essentially abstracted from the mimeographed versions, were published earlier. The earliest material is in the report (1957b). An important feature of the book is its use of *finitely* additive measures. They were able to characterize optimal strategies, or what is even more basic, to characterize the optimal yield of a gambling problem, by avoiding the technical measurability difficulties which a *countably* additive theory eventually leads to, and by recognizing that *all* strategies need to be considered. The book grew out of their study in the summer of 1956 of optimal play in red and black with a fixed goal: when your fortune is f, $0 < f < 1$, you may bet any amount s with $0 \leqslant s \leqslant f$, winning s (so your new fortune is $f + s$) with probability $w < \frac{1}{2}$ and losing s (so your new fortune is $f - s$) with probability $\bar{w} = 1 - w$. When your fortune goes outside (0, 1), you stop. What system maximizes your probability of stopping with a fortune of at least one, and what is this maximum probability $U(f)$? Dubins and Savage show that *bold play*: choosing $s = \min(f,\ 1 - f)$, is optimal, and that $U(f)$ is just the distribution function of $X = \Sigma_1^\infty X_n 2^{-n}$, where $X_1, X_2, \cdots$ are independent 0, 1 variables with $P(X_n = 0) = w$. That bold play is optimal will live as long as probability does, and that the probability of success is a singular function is striking.

An important result is that optimal strategies in any gambling problem of the type described are just those that are both *thrifty* (never make any mistakes) and *equalizing* (eventually reach positions where you have little to gain by further gambling). This result is often helpful in finding optimal strategies (see Blackwell, 1970, for one example), and it immediately cleared up a previously mysterious phenomenon in dynamic programming, as follows. It had been noticed in many dynamic programming problems that any strategy that satisfied Bellman's optimality principle was optimal, but no general proof could be found. It turned out that satisfying Bellman's optimality principle was essentially thriftiness, and that, in the dynamic programming problems studied, *all* strategies were necessarily equalizing. (In red and black, choosing $s = 0$ is thrifty, but not equalizing.)

Red and black is a special case of a lottery. A lottery is defined by a random variable $\theta > -1$, and not identically zero. When your fortune is f, you may choose any $s \in [0, f]$ and win $s\theta$, so that your new fortune is $f + s\theta$. For what lotteries, θ, is your optimal probability of reaching the goal exactly your initial

[4]An illuminating account of this collaboration is to be found in the preface, by Lester Dubins, to the 1976 reprint of the book.

fortune, f, for all $f \in [0, 1]$? Dubins and Savage show that, for $E(\theta) = 0$,

$$\liminf_{z\to\infty} \frac{-z\int_{-1}^{z} w \, d\theta(w)}{\int_{-1}^{z} w^2 \, d\theta(w)} = 0$$

is necessary and sufficient.

In the course of proving these and other substantial results, many small facts of interest are discovered incidently. Here are two examples. First, the mean of the variable X defined in the above discussion of red and black is less than its median. Second, for each fixed positive integer j, every nonnegative integer n has a unique representation:

$$n = \binom{p(0)}{j} + \binom{p(1)}{j-1} + \cdots + \binom{p(j-1)}{0},$$

where $p(0) > p(1) > \cdots > p(j-1)$ are nonnegative integers.

The main importance of the book may well be the recognition that many probability problems can be formulated as gambling problems and the discovery and application of a powerful and simple method for solving such problems (Theorem 2.12.1).

David Blackwell provides the following simple illustration of the latter. Let $X_1, X_2, \cdots$ be independent with $|X_i| \leqslant 1$, $E(X_i) = 0$, $E(X_i^2) = s > 0$, and let H be a fixed positive number. Put $S_n = X_1 + \cdots + X_n$ and let N be the first n with $|S_n| > H$. We want an upper bound for $E(N)$. Theorem 2.12.1. implies that, if Q is any function such that

(1) $Q(x) \geqslant 0$ for $|x| \leqslant H + 1$ and

(2) $Q(x) \geqslant 1 + EQ(x + X)$ for $|x| \leqslant H$ and all X with $|X| \leqslant 1, E(X) = 0, E(X^2) = s$, then $E(N) \leqslant Q(0)$.

Try $Q(x) = A - Cx^2$. Conditions (2) reduces to $C \geqslant 1/s$. For $C = 1/s$, (1) becomes $A \geqslant (H + 1)^2/s$. So $Q(x) = ((H + 1)^2 - x^2)/s$ works, and $E(N) \leqslant (H + 1)^2/s$.

In fact, when the method works, it often tells us more than we asked. In the above example, Theorem 2.12.1 tells us that, with $S_n = x + X_1 + \cdots X_n$, $|x| \leqslant H$, that is, starting at x, $Q(x)$ is an upper bound for $E(N)$, not only for independent X_i, but for any X_i satisfying $|X_i| \leqslant 1$, $E(X_i|X_1, \cdots, X_{i-1}) = 0$, $E(X_i^2|X_1, \cdots, X_{i-1}) = s$. Also, studying the variables X where equality is achieved in (2) (only these variables can possibly make $E(N) = Q$, a fact related to thriftiness) can be informative. In our example equality is achieved only for X concentrating on -1, 0, and 1, and for those X, $E(N)$ is precisely Q provided x and H are integers.

An early use of the method by Dubins and Savage themselves gave the following nice result (1965a).

THEOREM. *For any sequence* $\{X_n\}$ *for which* $\mu_n = E(X_n|X_1, \cdots, X_{n-1})$ *is finite* a.s. *and for any positive numbers* α *and* β, $P\{X_1 + \cdots + X_n \geqslant \beta + \mu_1 + \cdots + \mu_n) + \alpha(v_1 + \cdots + v_n)$ *for some* $n\} < 1/(1 + \alpha\beta)$, *where* $v_i = E(X_i^2|X_1, \cdots, X_{i-1})$. *This bound is sharp.*

Their proof is a simple application of Theorem 2.12.1. It reduces to checking $Q(x) \geqslant EQ(x + X)$ for all X with $E(X) \leqslant -\alpha\sigma^2(X)$, where $Q(x) = 1/(1 - \alpha x)$ for $x \leqslant 0$, and 1 for $x > 0$.

The method seems to be not yet widely known, let alone used, but one can be confident that it will be.

During the period between the book with Dubins and his death, he wrote several papers on topics in mathematical statistics. One that he wrote with his brother (1965b) is concerned with the extensions of the sequential, probability-ratio test to general sequences of random variables. Conditions for certain termination and for finite expected sample size are obtained and applied to a problem in ranking. The appreciation of the importance of de Finetti's work on exchangeable (or symmetric) distributions had led him to an interest in symmetry and his 1969 paper is a result. This beautiful short paper offers new insight into the stability of the Cauchy distribution and of the distribution of the reciprocal of the square of a normal random variable. These known results are here seen as immediate corollaries to certain elementary but deep observations of his about circularly symmetric distributions. For instance, he shows that if the Euclidean plane is endowed with a distribution that is invariant under the group of rotations about the origin, and X and Y are the usual orthogonal coordinate variables, then

$$\frac{p}{X^2} + \frac{q}{Y^2} \sim \frac{\left(p^{\frac{1}{2}} + q^{\frac{1}{2}}\right)^2}{X^2},$$

where p and q are any nonnegative numbers and $U \sim V$ means U is distributed as is V. Applied to the case in which X and Y are independent and normally distributed with mean 0 and variance 1, the distribution of $1/X^2$ is immediately seen to be stable of order $1/2$. A not dissimilar geometric argument exhibits the Cauchy distribution, that is, the distribution of the ratio of X to Y, as stable of order 2.

The paper with Richard Olshen (1970c) introduces, and makes a thorough study of, a one-parameter family of intensities of unimodality, one intensity for each positive number α. If, for a probability μ on an n-dimensional real vector space V, $\mu(tS)$ is at least $t^\alpha\mu(S)$ for all Borel subsets S of V, and all t in the unit interval then μ is *α-unimodal* (about the origin). In the one-dimensional case, 1-unimodality is equivalent to the ordinary notion of unimodality at 0.

A principal result is this. Necessary and sufficient for μ to be α-unimodal is that μ be the distribution of $U^{1/\alpha}X$ where U and X are independent, U is uniform on (0, 1). (The special case in which α and n are both 1 is a celebrated result of Khinchine.) The authors also show that μ is α-unimodal if, and only if, μ satisfies certain interesting inequalities which imply that, for any convex neighborhood K of the origin, $\mu(t^{1/\alpha}K)$ is then concave in t.

Some years ago, K. L. Chung made the surprising discovery that the convolution of two unimodal distributions on the real line need not be unimodal. Olshen and Savage salvage and improve upon this discovery by showing that the convolution of an α-unimodal μ with an α'-unimodal μ' is always $(\alpha + \alpha')$-unimodal, but need not be β-unimodal for any β less than $\alpha + \alpha'$.

In the volume of papers in honour of Pompilj (1971b), Savage, not content with complex number magic, casts new light on why it is natural to study the expected value of e^{itX}, the characteristic function of a random quantity X. He defines a function f to be *manageable* with respect to a class D of distributions if there is a pair of functions g and h with values in finite dimensional vector spaces such that, for all independent X and Y with distributions in D, the expectation of $f(X + Y)$ is determined by the expectations of $g(X)$ and $h(Y)$; that is, for some function T,

$$Ef(X + Y) = T(Eg(X), Eh(Y)).$$

He then goes on to determine the class of manageable f for various classes D. In particular, if D is the class of all distributions on the real line, then every manageable f is a trigonometric polynomial, that is, a finite linear combination of trigonometric monomials $\sin tx$ and $\cos t'x$ where t is real and positive and t' is nonnegative, and hence $Ef(X)$ is a finite linear combination of values of the characteristic function of X.

The sixth Berkeley symposium (1973a) sees him coauthoring a joint paper with five others. The topic is related to the 1969 paper being concerned with the more general situation of elliptically-contoured distributions having densities of the form $|\Sigma|^{-\frac{1}{2}}f(x^T\Sigma^{-1}x)$ in p dimensions: necessarily $\int_0^\infty r^{p-1}f(r^2)dr$ is finite. The authors obtain inequalities for probabilities such as $p\{|x_i| < h_i, \text{ all } i\}$ and $p\{x_i < l_i, \text{ all } i\}$.

A person who had a most important intellectual influence on him was Bruno de Finetti. We have seen how around 1950 he became aware of de Finetti's work and how he appreciated its enormous importance for probability and statistics. In his Purdue paper (1962g), he says "In my opinion . . . it is now wrong to repeat over and over again that the personal view [of probability] must be wrong without reading something that de Finetti and others have said about it." There are at least four important facets to de Finetti's work: (a) exchangeability and its role in explaining, within the personalistic viewpoint, certain results previously given a frequency interpretation; (b) his criticism of σ-additivity and its replacement by finite additivity; (c) the assessment, particularly through the use of proper scoring rules, of personal properties; and (d) an aspect that it is hard to pinpoint but essentially amounts to a "new look" for the concept of probability—a new look that could possibly change the whole of statistics as we know it. De Finetti has (1974) summarized it in the aphorism, "Probability does not exist." Part of the new look involves coherence; the idea that the only thing that matters about opinions is how they fit together.

Savage learned Italian and spent some time in Rome with de Finetti, though they appear only to have written one paper together (1962c). This "rather long"

(Savage's own words) paper in Italian concerns the subjective evaluation of "initial probabilities" and shows how remote it is from being arbitrary or aprioristic. The subjective view sees all probabilities as the opinions of some person. In particular these opinions should cohere and no one opinion differs in kind from another; "It is only chronologically, not logically, that one has priority over another." Attention is paid to the concept of vagueness in connection with personal probabilities and the role of second-order probabilities discussed. Stable estimation is again considered and relationship with the important work of Jeffreys emphasized. Finally there is a discussion of the thorny problem of randomization. There is one surprise in the paper, namely the views expressed about the Neyman-Pearson theory. They remark that "Bayesian statistics can be viewed as a continuation, rather than a contradiction, of the Neyman-Pearson theory. That theory leads to subjective choices that amount to the adoption by the user of the theory of some system of prior probabilities." This is incomplete because it fails to recognise that the prior probabilities so used are incoherent in that they are affected by irrelevant features, such as the sample size. The failure by Neyman and Pearson, and later Wald, to consider coherence is both surprising and unfortunate for the development of statistics. Other papers in Italian, of his own (1962d, e, f) explore similar topics.

One important outcome of the work with de Finetti was the paper on the elicitation of personal probabilities and expectations (1971a). With typical modesty he makes clear de Finetti's important contribution to the paper which is, in my view, one of the most significant of his papers. A central problem for the personalistic viewpoint is how the probabilities are to be measured. One idea is to use a scoring rule. Suppose a subject is required to assess his probability for an event E. Let him be motivated by rewarding him with a prize $f(x)$ if he declares $p(E)$ to be x and E occurs, and a prize $g(x)$ if E does not occur. (Assuming he would declare $1 - x$ to be $p(\bar{E})$ we can take $g(x)$ to be $f(1 - x)$.) If p is his probability for E, his expected prize is $pf(x) + (1 - p)g(x)$ and he will naturally choose x to maximize this expression. We would like this maximum to occur when x is equal to p and an assignment of prizes that achieves this is said to be *proper;* we speak of proper scoring rules. Familiar examples are the quadratic $f(x) = x^2$ and the logarithmic $f(x) = \log x$. In the paper he carefully develops the conditions for a rule to be proper. He does this in the more general framework of elicitation of rates of substitution, for which probability is a special case. The argument is extended to several rates (or events) when, under a mild condition, the logarithmic rule appears in a special light and leads to Shannon's concept of information. The paper concludes with extensive practical comments on the ideas, again illustrating his ability to blend the pure and applied aspects of the subject. There is a useful, and interestingly wide-ranging, bibliography.

That paper appeared shortly after his death. At the end of 1970 he had given the Fisher memorial lecture, in which he had held an audience spellbound for well over the period on the subject "Rereading Fisher." It has been prepared for publication by J. W. Pratt (1976). Savage relates how, like others of his generation, he had been brought up to believe that Fisher's *Statistical Methods for Research Workers* was "the serious man's introduction to statistics," and so he had felt that on the occasion of the lecture he would go back and look at this and the other writings of this great man. He obviously enjoyed the task, did it with great care, and ended up by producing a most complete account of Fisher's statistical work.

Although in many ways Fisher's viewpoint was the antithesis of the Bayesian, in others it comes close to it, and Savage's treatment is extraordinarily sympathetic. You come away with the strong feeling that he was surprised and delighted by what he discovered in the rereading. At the least, the essay is a glorious lesson in how to appreciate the views of someone not completely sharing your own. It seems reasonable to me that the Bayesian viewpoint should encourage this sympathy, because of its insistence on open statement (through probabilities and utilities) of personal opinions. But even if I see cause and effect where none exists, we can all appreciate the results. He emphasizes how good a mathematician Fisher was, in spite of Fisher's occasional criticisms of that class of men, and how energetic a calculator. He shows that the development of Fisher's ideas followed a steady pattern and was not grossly distorted by the controversies in which he engaged, controversies that were marred by the failure of either side seriously to understand what the other was saying.

After referring briefly to some branches of statistics that Fisher contributed to substantially or even created, such as sampling distributions, experimental design, analysis of variance, k-statistics and exact tests, and describing certain technical concepts that Fisher introduced, such as 'statistic', 'estimate', 'sufficient', 'consistent' and 'information', Savage spends much of the essay discussing topics where Fisher had a lot to say but where his ideas are imperfectly understood. He has some pertinent remarks to make about small-sample information, about approximating situations by a multinomial distribution and consistency. Maximum likelihood is carefully discussed and an ingenious example provided of an m.l. estimate that is not sufficient. On the subject of probability, Fisher was very slippery and no one view seems to be acceptable. Savage relates how he "was somewhat taken aback in my reading to find how vehemently he denies that probability is a limiting frequency in an indefinite sequence of repeated trials." One thing Fisher was clear about was the need for the nonexistence of recognizable subsets, though quite what this meant is unclear. Exchangeability is relevant here and it is surprising to find no mention of it by Savage in this context. There is a substantial section on inference, inductive behaviour and decision-making, and on that slippery subject of fiducial probability. Finally there is an illuminating account of the conflict there seemed to be within Fisher between his advocacy of the likelihood principle and his denial of it in discussing tail-area tests of hypotheses.

The essay is remarkable for the preparation that must have gone into it (and for Pratt's continuation of the work) with the copious references to Fisher's own writings, and for the way in which Savage keeps to the topic and only rarely diverges to discuss developments that have taken place since Fisher wrote. It is interesting that one great statistician should have provided in the last year of his life so fine a memorial to another when, on first glance, their views seem so dissimilar. Perhaps someday someone with equal skill will give us a lecture on rereading Savage. The discussants add much to the printed version.

There remains one group of papers that has not been discussed (1947d, 1948c, 1951c, 1952e, 1972a). These are all joint papers which Savage coauthored on applications of mathematics, statistics and probability to other fields of science. These correspond to the visible part of an iceberg, representing only a small fraction of the time and energy he spent on working with other scientists on *their* problems. It is difficult to identify Savage's contribution in these five papers; it is easier to understand from the testimony of those with whom he worked. He had a most astonishing ability to see through a problem, to isolate the important ideas and to subject these to analysis. This could be exasperating to someone who wanted a quick answer such as "take logs and do an analysis of variance . . . " and there are recorded cases of misunderstandings. But for the scientist (or even, in one case, a musician!) who judged the analysis of the data to be as important as its collection, the experience could be extremely rewarding. By all accounts he was excellent as a statistical consultant, and yet there does not seem to be a single statistical procedure due to him that might be used in a practical problem. The last of these papers, on glottochronology, is typical and consists essentially of a critique of some work by Chrétien that suffers a confusion between the expectation of the number of words, and the actual number. Here the critique goes to the heart of the problem and the paper is surely of importance to linguists and anthropologists. His remarks on medical diagnosis (1972b) are brief but include one remark of considerable practical import. He points out that in several situations a physician should report the likelihood function, *not* his personal probability. Thus when the doctor in charge of the case calls for an elaborate test, what he wants from the physican conducting the test is the likelihood, for the doctor can use this in *his* probability assessment more easily than he can the physician's personal probability.

There is another valuable and important aspect of his work that only appears in one paper (1970b) already discussed, that is his work with students. I have touched on this when describing my own experiences in Chicago. He had firstly enormous patience with them. Secondly, he took great pains to lead them through an argument in such a way that they felt were creating the result for themselves, so that their confidence in themselves was enhanced, rather than deflated, by the criticism. He also went to a lot of trouble in the preparation of material for his graduate courses, as the paper just mentioned testifies. I have in my possession some notes of his (dated 1960) on Thomas Bayes' original paper. These seven typewritten pages carefully take the reader through this important paper and help him substantially to appreciate Bayes' argument. Again there is this tendency to get to the heart of the matter, to throw away, for a moment, the necessary impedimenta that prevent a clear view.

David Blackwell said in 1950, "of all things going on in statistics today, the only work sure to be significant fifty years from now is Savage's." In my view that is still true a quarter of a century later. His name will live if only because of the first seven chapters of the Foundations. In it he built a new paradigm for statistics that will endure.

L. J. SAVAGE: PUBLICATIONS

Articles

(1940) A dynamic problem in duopoly (with Arthur Smithies). *Econometrica* **8** 130–143.

(1943) On the crossing of extremals at focal points. *Bull. Amer. Math. Soc.* **49** 467–469.

(1946a) Unbiased estimates for certain binomial sampling problems with applications (with M. A. Girshick and F. Mosteller). *Ann. Math. Statist.* **17** 13–23.

(1946b) The application of vectorial methods to metric geometry. *Duke Math. J.* **13** 521–528.

(1947a) A uniqueness theorem for unbiased sequential binomial estimation. *Ann. Math. Statist.* **18** 295–297.

(1947b) Abandoning an experiment prior to completion (with K. J. Arnold and M. A. Girshick). In *Techniques of Statistical Analysis* (C. Eisenhart, M. W. Hastay and W. A. Wallis, eds.) Ch. 12, 353-362. McGraw-Hill, New York.

(1947c) Planning experiments seeking maxima (with M. Friedman). In *Techniques of Statistical Analysis* (C. Eisenhart, M. W. Hastay and W. A. Wallis, eds.) Ch. 13, 363–372. McGraw-Hill, New York.

(1947d) Appendix to "Evidence for the formation of cell aggregates by chemotaxis in the development of the slime mold dictyostelium discoideum" by J. T. Bonner (*J. Exp. Zool.* **106** 1–24) *J. Exp. Zool.* **106** 24–25.

(1948a) Samuelson's foundations: its mathematics. *J. Political Econ.* **56** 200–202.

(1948b) Games with circular symmetry. In *Rand Corporation Research Memo* RM597G.

(1948c) Diet studies in pregnant patients (with W. J. Dieckmann, D. F. Turner, E. J. Meiller, M. T. Straube, K. B. Grossnickle, R. E. Pottinger, A. J. Hill, J. B. Forman, H. D. Priddle, E. S. Beckette and E. M. Schumacher). *Obstetrical and Gynecological Survey* **3** 731–745.

(1948d) The utility analysis of choices involving risk (with M. Friedman). *J. Political Econ.* **56** 279–304. (Reprinted in *Landmarks in Political Economy* (E. J. Hamilton, A. Rees and H. G. Johnson, eds.) 297–336. University of Chicago Press, Chicago, (1962); and in *Theory of Business Finance: A Book of Readings.* (S. H. Archer and C. A. D'Ambrosio, eds.) 36–66. Macmillan, New York, (1967).

(1948e) A game theoretic study of tactics of area defense (with M. M. Flood). In *Rand Corporation Research Memo,* RM51.

(1949a) Application of the Radon-Nikodym theorem to the theory of sufficient statistics (with P. R. Halmos). *Ann. Math. Statist.* **20** 225–241.

(1949b) An apparent ambiguity in the interpretation of minimum risk. In *Rand Corporation Research Memo,* RM184.

(1950) The role of personal probability in statistics. *Econometrica* **18** 183–184.

(1951a) Bayes and minimax estimates for quadratic loss functions (with M. A. Girshick). In *Proc. Second Berkeley Symp. Math. Statist. Probability,* 53–73. Univ. California Press, Berkeley.

(1951b) The theory of statistical decision. *J. Amer. Statist. Assoc.* **46** 55–67.

(1951c) Epidemiological aspects of histoplasmin, tuberculin and coccidioidin sensitivity (with C. G. Loosli, W. G. Beadenkopf and F. A. Rice). *Amer. J. Hygiene* **53** 33–57.

(1952a) On the set of values of a nonatomic, finitely additive, finite measure (with R. J. Nunke). *Proc. Amer. Math. Soc.* **3** 217–218.

(1952b) Discussion of "Ordinal preferences or cardinal utility?" by H. O. Wold (*Econometrica* **20** 661–663) *Econometrica* **20** 663–664.

(1952c) The expected-utility hypothesis and the measurability of utility (with M. Friedman). *J. Political Econ.* **60** 463–474.

(1952d) Une axiomatisation de comportement raisonnable face a l'incertitude (with discussion). In *Colloques Internationaux du Centre National de la Recherche Scientifique XL, Econométrie* 29–40.

(1952e) Histoplasmin and tuberculin sensitivity among Illinois residents (with W. G. Beadenkopf, J. T. Grayson, J. M. Ward, C. G. Loosli and C. Hall). *Amer. J. Hygiene* **57** 328–343.

(1955a) Three problems in rationing capital (with J. H. Lorie). *J. of Business* (University of Chicago) **28** 229–239.

(1955b) Symmetric measures on cartesian products (with E. Hewitt). *Trans. Amer. Math. Soc.* **80** 470–501.

(1956) The non-existence of certain statistical procedures in non-parametric problems (with R. R. Bahadur). *Ann. Math. Statist.* **27** 1115–1122.

(1957a) When different pairs of hypotheses have the same family of likelihood-ratio test regions. *Ann. Math. Statist.* **28** 1028–1032.

(1957b) The casino that takes a percentage and what you can do about it. In *Rand Corporation Paper,* P1132.

(1959) La probabilità soggettiva nei problemi practici della statistica. In *Induzione e Statistica,* (CIME Lectures), Instituto Mathematico dell'Università, Roma.

(1960a) Recent tendencies in the foundations of statistics. *Proc. Eighth International Congress of Mathematicians* 540–544. Cambridge Univ. Press.

(1960b) Optimal gambling systems (with L. E. Dubins). *Proc. Nat. Acad. Sci. U.S.A.* **46** 1597–1598.

(1961) The foundations of statistics reconsidered. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 575–586. Univ. California Press, Berkeley. (Reprinted in *Studies in Subjective Probability* (H. E. Kyburg, Jr. and H. E. Smokler, eds.) 1964, 175–188. Wiley, New York.)

(1962a) Discussion of "On the foundations of statistical inference" by Allan Birnbaum. (*J. Amer. Statist. Assoc.* **57** 269–306) *J. Amer. Statist. Assoc.* **57** 307–308.

(1962b) Subjective probability and statistical practice. In *The Foundations of Statistical Inference: A Discussion.* 9–35 + discussion. Methuen, London and Wiley, New York.

(1962c) Sul modo di scegliere le probabilità inizali (with B. de Finetti). In *Biblioteca del "Metron", Series C: Note E Commenti.* **1** 81–154. (Includes an English summary reprinted in *Probability, Induction and Statistics* by B. de Finetti, 1972, Chapter 8. Wiley, New York.)

(1962d) Campi di applicazione e techniche della statistica. In *La Mathematica Negli·Istituti Tecnici Commerciali.* 165–178. Ministero della Pubblica Istruzione, Aziende Tipografiche Eredi Dott. G. Bardi, Roma.

(1962e) Uno squardo sulla statistica di oggi. In *Mathematica ed Economia* (Convegno di Bressanone, 1961 74–85). Università di Padova, Italy.

(1962f) Il problema delle strategie ottime di giòco. In *Mathematica ed Economia* (Convegno di Bressanone, 1961 86–93). Università di Padova. Italy.

(1962g) Bayesian statistics. In *Recent Developments in Decision and Information Processes* (R. E. Machol and P. Gray, eds.) 161–194. Macmillan, New York.

(1963) Bayesian statistical inference for psychological research (with W. Edwards and H. Lindman). *Psych. Rev.* **70** 193–242. (Reprinted in *Readings in Mathematical Psychology* **2** 519–568 (R. D. Luce, R. R. Bush and E. Galanter, eds.) Wiley, New York (1965).

(1965a) A Tchebycheff-like inequality for stochastic processes (with L. E. Dubins). *Proc. Nat. Acad. Sci. U.S.A.* **53** 274–275.

(1965b) Finite stopping time and finite expected stopping time (with I. R. Savage). *J. Roy. Statist. Soc. Ser. B* **27** 284–289.

(1967a) Difficulties in the theory of personal probability. *Philos. Sci.* **34** 305–310. (Reprinted in translation in *Mathematiques et Sciences Humaines* **21** 5–9 (1968).

(1967b) Implications of personal probability for induction. *J. Philos.* **64** 593–607.

(1968) Discussion with J. Keifer and L. Le Cam on "Statistical inference." In *The Future of Statistics* (D. A. Watts, ed.) 139–160. Academic Press, New York.

(1969) A geometrical approach to the special stable distributions. *Zastos. Mat.* **10** 43–46.

(1970a) Comments on a weakened principle of conditionality. *J. Amer. Statist. Assoc.* **65** 399 –401.

(1970b) Reading suggestions for the foundations of statistics. *Amer. Statist.* **24** 23–27.

(1970c) A generalized unimodality (with R. Olshen). *J. Appl. Prob.* **7** 21–34.

(1970d) Die Bayessche Entwicklungsstufe der Statistichen Schlussweise. In *Vorträge der II. Ungarischen Biometrischen Konferenz* 23–26. Akademiai Kiadó, Budapest.

(1971a) Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801.

(1971b) The characteristic function characterized and the momentousness of moments. In *Studi di probabilità, statistica e ricerca operativa in onore di Guiseppe Pompilj,* 131–141. Tipografia Oderisi Editrice, Gubbio.

(1972a) The mathematics of glottochronology revisited (with A. J. Dobson, D. Sankoff and J. B. Kruskal). *Anthropological Linguistics* **14** 205–212.

(1972b) Diagnostic and the Bayesian viewpoint. In *Computer Diagnosis and Diagnostic Methods* (J. A. Jacquez, ed.) 131–138. Charles C. Thomas, Springfield, Ill.

(1973a) Inequalities on the probability content of convex regions for elliptically contoured distributions (with S. Das Gupta, I. Olkin, M. L. Eaton, M. Perlman and M. Sobel). In *Proc. Sixth Berkeley Symp. Math. Statist. Probability.* **2** 241–265. Univ. California Press, Berkeley.

(1973b) Probability in science: a personalistic account. *Proc. Fourth International Symp. on Logic, Methodology and Philosophy of Science* (P. Suppes, ed.). Bucharest, Romania.

(1976) On rereading R. A. Fisher. *Ann Statist.* **4** 441–500.

Books

(1948) *Sampling Inspection.* (with H. A. Freeman, M. Friedman, F. Mosteller, D. H. Schwartz and W. A. Wallis) (H. A. Freeman, M. Friedman, F. Mosteller and W. A. Wallis, eds.) McGraw-Hill, New York.

(1954) *The Foundations of Statistics.* Wiley, New York.

(1965) *How to Gamble if You Must: Inequalities for Stochastic Processes.* McGraw-Hill, New York.

(1972) *The Foundations of Statistics.* (Second Revised Edition) Dover Publications, New York.

(1976) *Inequalities for Stochastic Processes: How to Gamble if You Must.* Dover Publications, New York.

Reviews

(1951a) Good, I. J. *Probability and the Weighing of Evidence. J. Amer. Statist. Assoc.* **46** 383–384.

(1951b) Carnap, R. J. *The Nature and Application of Inductive Logic. J. Amer. Statist. Assoc.* **46** 534.

(1952) Carnap, R. *Logical Foundations of Probability. Econometrica* **20** 688–690.

(1955a) Gregg, John R. *The Language of Taxonomy—An Application of Symbolic Logic to the Study of Classificatory Systems. Bull. Amer. Math. Soc.* **61** 460–462.

(1955b) Polya, George *Mathematics and Plausible Reasoning,* **1**: *Induction and Analogy in Mathematics,* **2**: *Patterns of Plausible Inference. J. Amer. Statist. Assoc.* **50** 1352–1354.

(1956a) Bartlett, M. S. *An Introduction to Stochastic Processes with Special Reference to Methods and Applications. J. Amer. Statist. Assoc.* **51** 383–385.

(1956b) Bochner, Salomon *Harmonic Analysis and the Theory of Probability. Science* **123** 511.

(1957) Fréchet, Maurice *Les Mathematiques et le Concret. Econometrica* **25** 498.

(1958) Grenander, U. and Szegö, G. *Toeplitz Forms and Their Applications. J. Amer. Statist. Assoc.* **53** 763.

(1959) Cantelli, F. P. *Alcune Memorie Mathematiche. J. Amer. Statist. Assoc.* **54** 827.

(1961) Weibull, Christer *Some Aspects of Statistical Inference with Applications to Sample Survey Theory. J. Amer. Statist. Assoc.* **56** 746–747.

(1962) Cox, Richard T. *The Algebra of Probable Inference. J. Amer. Statist. Assoc.* **57** 921–922.

(1964) Raiffa, H. and Schlaifer, R. O. *Applied Statistical Decision Theory. Psychometrika* **29** 213.

(1971) Fishburn, Peter C. *Utility Theory for Decision Making. The American Scientist* **59** 771–772.

REFERENCES

BARNARD, G. A. (1947). Review of *Sequential Analysis* by Abraham Wald. *J. Amer. Statist. Assoc.* **42** 658–664.

BIRNBAUM, ALLAN (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.* **57** 269–306.

BLACKWELL, DAVID (1970). On stationary policies. *J. Royal Statist. Soc. Ser. A.* **133** 33–37.

BRAITHWAITE, R. B. (1955). *Scientific Explanation.* Cambridge Univ. Press.

COX, RICHARD T. (1961). *The Algebra of Probable Inference.* John Hopkins Press, Baltimore.

DE FINETTI, BRUNO (1974, 1975). *Theory of Probability.* Wiley, New York. 2 volumes. (Translation by A. Machi and A. Smith from the 1970 Italian edition.)

DUBINS, L. E. and FREEDMAN, D. A. (1963). Random distribution functions. *Bull. Amer. Math. Soc.* **69** 548–551.

DUBINS, L. E. and FREEDMAN, D. A. (1967). Random distribution functions. *Proc. Fifth Berkeley Symp. Math. Statist. Probability* **2** 183–214.

DURBIN, J. (1970). On Birnbaum's theorem on the relation between sufficiency, conditionality and likelihood. *J. Amer. Statist. Assoc.* **65** 395–398.

HARROD, ROY (1956). *Foundations of Inductive Logic.* Macmillan, London.

KUHN, THOMAS S. (1970). *The Structure of Scientific Revolutions.* Univ. Chicago Press.

KYBURG, H. E. and SMOKLER, H. E. (1964). *Studies in Subjective Probability.* Wiley, New York.

LINDLEY, D. V. (1975). The future of statistics—a Bayesian 21st century. *Adv. Appl. Prob. (Supplement)* 106–115.

VON NEUMANN, J. (1928). Zur theorie der Gesellschaftsspiele. *Math. Annalen.* **100** 295–320.

VON NEUMANN, J. and MORGENSTERN, O. (1944). *Theory of Games and Economic Behavior.* Princeton Univ. Press.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory.* Business School, Boston.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probability.* **1** 197–206.

TUKEY, JOHN W. (1962). The future of data analysis. *Ann. Math. Statist.* **33** 1–67.

WOLFOWITZ, J. (1946). On sequential binomial estimation. *Ann. Math. Statist.* **17** 489–493.

2, PERITON LANE
MINEHEAD TA24 8AQ
ENGLAND