LUCIEN LECAM

A NOTE ON METASTATISTICS OR 'AN ESSAY TOWARD STATING A PROBLEM IN THE DOCTRINE OF CHANCES'*

1. INTRODUCTION

Discussions about foundations are typically accompanied by much unnecessary proselytism, name calling and personal animosities. Since they rarely contribute to the advancement of the debated discipline one may be strongly tempted to brush them aside in the direction of the appropriate philosophers. However, there is always a ghost of a chance that some new development might be spurred by the arguments. Also the possibly desirable side effects of the squabbles on the teaching and on the standing of the debated disciplines cannot be entirely ignored. This partly explains why the present author reluctantly agreed to add to the extensive literature on the subject.

Another reason for the present essay is the fortuitous circumstance that this author had the benefit of five years of spirited discussions with Etienne Halphen whose views were not entirely out of line with the present neo-Bayesian philosophies. By virtue of these special circumstances, it happened that, contrary to what seems to be the case for most American statisticians, we learned a form of the neo-Bayesian creed *before* being exposed to the classical theory of statistics.

From these long discussions, in which the works of de Morgan, Venn, Keynes, Jeffreys and many others were constantly quoted, we have retained a certain awe for the Bayesian approach itself, but above all for the fascination its attractive simplicity seems to have for the sharpest minds.

However, we could not defend ourselves at that time, and cannot defend ourselves now from the sentiment that this age and times should see an elaboration of a formalized mathematical theory in which the subject could not only be debated but also studied.

Synthese **36** (1977) 133–160.

Since no such formalism is available yet, at least for the purposes which neo-Bayesians seem to have in mind, and since we have failed to build one, it is impossible for us to contribute anything of substance to the study of the problem. We shall therefore content ourselves with an informal description of some of the shortcomings which lurk behind the admittedly attractive façade of the Bayesian approach.

This is not to say that any form of condemnation of the Bayes theory should be suggested, on the contrary. However, even those physicists who are most fascinated by the kinetic theory of gases would hesitate to use it to compute the size of wood beams for their own abode. In the same spirit we would venture to suggest that statisticians restrict their use of the Bayes approach to those cases in which they understand its assumptions and implications.

In the following pages little will be said about statistical inference, for the simple reason that this author does not understand what inference, statistical or otherwise, really means. If inference is what we think it is, the only precept or theory which seems relevant is the following: 'Do the best you can.' This may be taxing for the old noodle, but even the authority of Aristotle is not an acceptable substitute.

The following section describes some of the goals and purposes occasionally attributed to the theory of statistics. The third section gives a brief account of some mathematical structures which enter in the Neyman-Pearson-Wald theory of decision functions. The fourth section gives a short, and hopefully correct, statement of the theorem which underlies L. J. Savage's approach and goes forth to elaborate on the reasons why this approach is too naive for many purposes. Briefly stated the reasons are the following.

(1) The neo-Bayesian theory makes no difference between 'experiences' and 'experiments'.

(2) It confuses 'theories' about nature with 'facts', and makes no provision for the construction of models.

(3) It applies brutally to propositions about theories or models of physical phenomena the same simplified logic which every one of us ordinarily uses for 'events'.

(4) It does not provide a mathematical formalism in which one person can communicate to another the reasons for his opinions or decisions. Neither does it provide an adequate vehicle for transmission

of information. (This, of course, is irrelevant to a purely 'personalistic' approach.)

(5) The theory blends in the same barrel all forms of uncertainty and treats them all alike.

The above shortcomings belong to the neo-Bayesian *theory*. The neo-Bayesian *movement* has additional unattractive facets, the most important of which is its normative interpretation of the role of Statistics. Presumably a statistician who does not abide by its regulations is either irrational, inconsistent or incoherent. We shall attempt to argue on the contrary that the Bayesian viewpoint cannot be held consistently and that even its most recent supporters have had to violate their own rules when expounding it.

In summary, the Bayesian theory is wonderfully attractive but grossly oversimplified. It should be used with the same respect and the same precautions as the kinetic theory of perfect monoatomic gases.

2. THE SCOPE OF STATISTICS

The recent revival of the Bayesian approach is probably correlated with the contemporary trend to define Statistics as the 'Science of decision in the face of uncertainty'. Assuming that the statistician can make appropriate apologies to his colleagues from Economics, Management Science, Operations Research (not to mention true sciences), this may not be a bad definition.

Traditionally, Statistics was a branch of knowledge which appeared helpful in the handling of masses of data, the transmission of information, the design of experiments and the ascertainment of conclusions from experimental evidence. Handling of extensive numerical data is now more and more delegated to electronic machines with the results one would expect from this robotization.

The vigorous young discipline of Information Theory has formalized some of the problems relating to the transmission of information in the specialized context most directly connected with the operation or construction of machinery for the transmission of verbal or written messages between humans. Unfortunately, information theory has little to say about scientific or experimental 'information' in the acceptance of the word often used by statisticians.

The success of the classical theory of statistics in the fields of design of experiments and of ascertainment of behavioral conclusions from experimental evidence is well known. We shall return to this in the next section.

To define Statistics as the 'science of decision in the face of uncertainty' is to embrace a lot of territory. Since it is difficult to dissociate actual decision making from a justification of it, the science of decision is quickly converted to the study of 'confirmation' or 'degree of beliefs' now called pistimetry. An evolution of this type is visible in Savage's own writings. His 1954 book on the Foundations of Statistics is devoted to a personalistic theory of decision, but the 1959 booklet is entitled the Foundations of Statistical Inference. Thus, from 'decision' to 'pistimetry' and formal logic the statistician would be led to assume responsibility for the entirety of the old Leibnizian dream of a Universal Characteristic.

The beginnings of the calculus of probability saw developments of this general nature. The first treatise on the subject (1713) is accordingly called 'Ars Conjectandi'. With the ensuing development of the ideas of mathematical and moral expectations (Daniel Bernoulli among others), the theory of probability seemed closer to fulfill Leibniz' wishes than anything else available at the time. The years following *Ars Conjectandi* saw also the caution of Bayes who attempted to make distinctions between a priori and a posteriori probabilities. Bayes was also unsure of the validity of his arguments. His Essay was only published posthumously (1763).

The major impetus to the theory of probability came only with Laplace's 'Theorie analytique des probabilités' (1812) with its exposition of the method of generating functions and an almost rigorizable proof of the Central Limit Theorem. Laplace also gives examples of tests of hypotheses, a discussion of loss functions and many other things.

It is quite natural that the explosion following the discovery of the theory of probability would lead to excessive claims for its power. It is also natural that it would be tried in all kinds of appropriate and inappropriate domains. However, even the impetus given by Laplace could not delay indefinitely an examination of the abuses to which probability can be subjected.

De Morgan's Formal Logic contains a definitely pistimetric theory of probability, but Venn's Logic of Chance contains an equally devastating review of the deficiencies of such theories.

Even after probabilists turned their attention to the study of the mathematical structure so neatly described by Kolmogorov in 1933 the arguments about pistimetry versus frequency interpretations lingered on, as can be seen for instance from the Proceedings of the Geneva Colloquium of 1937. Probabilists now care little about the foundations of their discipline. They are satisfied with developing its mathematical content. The dispute has been delegated to statisticians.

In this respect, it is particularly disturbing that the neo-Bayesian school does not introduce ideas or techniques essentially different from those of 1763. This school seems to ignore the long and quarrelsome history of the subject and in particular the difficulties pointed out by Venn.

What is perhaps more important is that the neo-Bayesian school has generally added to the confusion by claiming that the personalistic theory of decision ignores pistimetry but is a substitute for or an improvement of the Neyman-Pearson theory.

Since it appears now that the Neyman-Pearson theory, even supplemented by Wald, is essentially a theory of 'experiments' and since the distinction between 'experiences' and 'experiments' is not explicitly recognized by the personalistic decision approach, it is not clear how the latter can be an improvement of the first.

To be more precise we shall now describe a few general features of the Neyman-Pearson theory.

3. SOME CONCEPTS OF CLASSICAL STATISTICS

By classical statistics is meant here the theory usually associated to the names of Neyman, Pearson and Wald. The theory says nothing about inference or pistimetry and precious little about behavior. If you happen to find on the street a page covered with numbers, classical statistics will not tell you what to do with them. In fact, lay opinion to the contrary, classical statistics says practically nothing about numbers.

Whatever may be the historical motivation for the existence of this theory, it appears now that classical statistics is essentially concerned with the study of experiments and of functions defined on or by experiments which are described by a formal mathematical structure as follows.

A single stage experiment consists of a set Θ, a set $\mathscr{X}$, a σ-field $\mathscr{A}$ of subsets of $\mathscr{X}$ and a family $\{P_\theta;\ \theta \in \Theta\}$ of probability measures on $\mathscr{A}$.

Multi-stage, or sequential experiments involve in addition to the sets Θ and $\mathscr{X}$ a partially ordered set S and a family $\{\mathscr{A}_s;\ s \in S\}$ of σ-fields of subsets of $\mathscr{X}$. Corresponding to each $s \in S$ there is a family $\{P_{\theta,s};\ \theta \in \Theta\}$ of probability measures on $\mathscr{A}_s$. It is often assumed that these families are compatible in the sense that if $A \in \mathscr{A}_s$ and $A \in \mathscr{A}_t$ then $P_{\theta,s}(A) = P_{\theta,t}(A)$.

To avoid complications we shall comment mostly on the single stage structures.

Let $\mathscr{E} = \{\Theta, \mathscr{X}, \mathscr{A}, \{P_\theta\}\}$ be an experiment. A test is an $\mathscr{A}$-measurable function φ defined on $\mathscr{X}$ and bounded by zero and unity. The power function of the test φ is the function $\theta \to \beta(\theta) = \int \varphi(x) P_\theta(dx)$.

Confidence sets are defined as follows. Let A be a set in the cartesian product $\Theta \times \mathscr{X}$. Assume that for each $\theta \in \Theta$ the section A_θ of A at θ has P_θ measure at least equal to $\alpha \in [0, 1]$. Let A^x be the section of A at x. Since the relations '$x \in A_\theta$' and '$\theta \in A^x$' are equivalent to the relation '$(\theta, x) \in A$' one can write $P_\theta\{\theta \in A^x\} = P_\theta\{x \in A_\theta\}$.

A sub-σ-field $\mathscr{B}$ of $\mathscr{A}$ is called sufficient if for each bounded $\mathscr{A}$-measurable function φ there is a bounded $\mathscr{B}$-measurable function ψ such that

$$\int \psi(x) u(x) P_\theta(dx) = \int \varphi(x) u(x) P_\theta(dx)$$

identically in θ for every bounded $\mathscr{B}$-measurable function u.

An experiment $\mathscr{E} = \{\Theta, \mathscr{X}, \mathscr{A}, \{P_\theta\}\}$ can be used to define a variety of other mathematical entities, two of which will be used below. Let M_0 be the space of equivalence classes of bounded $\mathscr{A}$-measurable functions. Consider the P_θ as linear functionals on M_0.

The linear functionals on M_0 can be provided with the norm $\|\mu\| = \sup_\varphi \{\int \varphi(x) \mu(dx);\ |\varphi| \leq 1\}$. Let $L(\mathscr{E})$ be the smallest linear space containing all the P_θ, all the positive linear functionals smaller than finite linear combinations $\Sigma a_j P_{\theta_j}$ and all their limits for the norm. Let $M(\mathscr{E})$ be the adjoint of $L(\mathscr{E})$.

Let $\mathscr{F} = \{\Theta, \mathscr{Y}, \mathscr{B}, \{Q_\theta\}\}$ be another experiment having the same index set Θ as $\mathscr{E}$. The *deficiency* $\delta(\mathscr{E}, \mathscr{F})$ of $\mathscr{F}$ with respect to ε is the number

$$\delta(\mathscr{E}, \mathscr{F}) = \inf_\Pi \sup_\theta \|\Pi P_\theta - Q_\theta\|$$

where the infimum is taken over all positive linear maps Π from $L(\mathscr{E})$, to

$L(\mathscr{F})$ which are such that their adjoint (a map from $M(\mathscr{F})$ to $M(\mathscr{E})$) transforms the function identically one on $\mathscr{Y}$ into the function identically one on $\mathscr{X}$.

The 'distance' of $(\mathscr{E}, \mathscr{F})$ is the maximum of $\delta(\mathscr{E}, \mathscr{F})$ and $\delta(\mathscr{F}, \mathscr{E})$. This is only a pseudometric which becomes a metric if two experiments $\mathscr{E}$ and $\mathscr{F}$ such that $\delta(\mathscr{E}, \mathscr{F})+\delta(\mathscr{F}, \mathscr{E})=0$ are identified.

Except for the usual measure theoretic technicalities the equality $\delta(\mathscr{E}, \mathscr{F})=0$ would correspond to the possibility of constructing another experiment $\{\Theta, \mathscr{X}\times\mathscr{Y}, \mathscr{A}\times\mathscr{B}, \{R_\theta\}\}$ where R_θ has P_θ and Q_θ for marginals and where $\mathscr{A}$ is sufficient for $\mathscr{A}\times\mathscr{B}$. Except for further technicalities, this is also equivalent to the possibility of duplicating $\mathscr{F}$ by an appropriate post-experimental randomization of $\mathscr{E}$.

Wald's theory introduces additional elements besides the experiment $\mathscr{E}$. Specifically Wald uses in addition to $\mathscr{E}$ a set T and a function W from $\Theta\times T$ to $(-\infty, +\infty]$. Also we shall assume given a vector lattice C of bounded numerical functions defined on T and denote by $\mathscr{C}$ the smallest σ-field for which they are measurable. Further we shall assume that for each $\theta\in\Theta$ the function $t\to W(\theta, t)$ is bounded below and $\mathscr{C}$-measurable.

A decision function ρ is then a map $x\to p_x$ from $\mathscr{X}$ to the probability measures on $\mathscr{C}$ with the added property that for each $\gamma\in C$ the integral $\int\gamma(t)p_x(dt)$ is $\mathscr{A}$-measurable in $\mathscr{X}$.

For Wald's theory of decision functions, the statistical structure is entirely specified by the experiment $\mathscr{E}$, the sets T and C, function W and a set $\mathscr{D}$ of decision functions.

Under the assumptions made above a decision function $\rho: x\to p_x$ possesses a risk function r_ρ defined on Θ by

$$r_\rho(\theta)=\int\int W(\theta, t)p_x(dt)P_\theta(dx).$$

Let $\mathscr{R}$ be the class $\{r_\rho; \rho\in\mathscr{D}\}$. The ordering of real numbers induces on $\mathscr{R}$ a partial ordering. Elements of $\mathscr{R}$ which are minimal for this ordering are called admissible. A subset $\mathscr{R}_0\subset\mathscr{R}$ such that for each $r\in\mathscr{R}$ there is an $r_0\in\mathscr{R}_0$ for which $r_0\leq r$ is called complete. One can also say that $\mathscr{R}_0$ dominates $\mathscr{R}$. An element $r_0\in\mathscr{R}$ which minimizes in $\mathscr{R}$ an integral of the type $\int r(\theta)\mu(d\theta)$ with respect to a finite positive measure μ is called a Bayes risk function for μ. A point r is called low or a Bayes solution in the wide sense if whatever may be $\varepsilon > 0$ there is no element $\mathscr{R}$ smaller than or equal to $r - \varepsilon$.

The general theory of decision functions attempts to find small tractable complete subsets of $\mathscr{R}$. It also attempts to describe decision functions which have admissible risk functions and sets itself other goals of the same general nature.

A particular theorem which is of interest for the present discussion is the following.

THEOREM 1. Let $\mathscr{E} = \{\Theta, \mathscr{X}, \mathscr{A}, \{P_\theta\}\}$ be an experiment and let $\mathscr{D}$ be the set of all decision functions corresponding to the sets T and C. Assume that

(1) there is a finite measure λ on $\mathscr{A}$ such that $\lambda(A) = 0$ implies $P_\theta(A) = 0$,

(2) the set T carries a topology and C is the set of bounded continuous functions on T,

(3) for each α real and each θ the set $\{t \in T;\ W(\theta, t) \leq \alpha\}$ is compact and metrisable.

Further let $\mathscr{D}_1$ be a subset of $\mathscr{D}$ which is closed in $\mathscr{D}$ for the topology of convergence of the integrals $\int u(t) p_x(dt) Q(dx)$ with $u \in C$ and $Q \in L$. Let $\mathscr{R}_1$ be the corresponding set of risk functions. If for each finite positive measure μ carried by a finite subset of Θ the set $\mathscr{R}_1$ contains the Bayes risk functions for μ then $\mathscr{R}_1$ is complete.

Many improvements and many variations of the above theorem are readily available.

Of course the above brief description does not even touch upon some of the most beautiful achievements of the classical theory. It has been given here only for purposes of comparison with the personalistic theory. To this end it will be necessary to elaborate on the motivation and interpretation of the above formalizations.

A convenient interpretation is the following. Each $\theta \in \Theta$ represents a particular theory about the physical phenomena under consideration.

In the particular experiment under study the elements of $\mathscr{A}$ are events which may or may not occur. The class $\mathscr{A}$ includes those events which satisfy two important conditions. First, the experimenter has conceived that they could possibly occur. Second, he has at his disposal instruments which can check whether or not they occur. Events are assumed to follow the usual distributive and associative laws of Boolean algebra. Therefore, there is no danger in taking advantage of their representability by subsets of a certain set $\mathscr{X}$, called the set of possible results of the experiment.

For each θ the measure P_θ is an expression of physical laws governing the machinery of the experiment.

The set T is a set of possible decisions. A decision function is a rule which to each possible result of the experiment associates a particular mode of random selection of an element of T.

The loss function W gives a crude evaluation of the damage which may result from a decision t when the theory θ is applicable.

Some important features of the theory are the following.

(1) The theory deals with *experiments* described by appropriate mathematical structures.

(2) The theory concentrates on the study of properties of decision *functions* not of individual decisions.

(3) Whether a particular physical theory, such as quantum theory, the Newtonian theory of gravitation or Bergeron's theory of meteorological fronts is *correct* or *adequate* remains a vague concept not discussed in formalized language in any work with which the author is acquainted. Similarly the classical theory of statistics takes for granted the possibility of describing the physical regulations governing the machinery of actual experiments by probability measures. However, it does not contain any formalization of the concept of adequacy of a theory θ.

(4) The theory says nothing about the situation in which an experimenter finds himself if he happens to notice the realization of an event which was not included in the list $\mathscr{A}$.

(5) The theory does not have any prescriptions about decisions to be taken if the result of an experiment suggests that a certain theory θ not previously included in Θ is a more adequate description of the physical phenomenon.

In all these cases it is assumed that common sense and the use of our limited brains will be involved.

In particular the theory does not volunteer any statements about the probability that such or such a theory of nature be correct. Above all it does not volunteer any statement about its own correctness or adequacy.

Within the classical theory, one can exercise a liberal amount of freedom in selecting objects of study. For instance, for an experiment $\{\Theta, \mathscr{X}, \mathscr{A}, \{P_\theta\}\}$ it is perfectly feasible to study the properties of a test φ conditionally given a sub-σ-field $\mathscr{B}$ of $\mathscr{A}$. This is done as a matter of course when $\mathscr{B}$ is a sufficient σ-field. It leads to the introduction of 'tests having Neyman structure', that is, tests φ such that $E[\varphi|\mathscr{B}]$ is constant.

If somebody wants to study estimates or tests about the expectation of a normal distribution conditionally given the sample covariance matrix, this remains his privilege. Also if someone wants to avail himself of A. Birnbaum's theorem which states that binary experiments are equivalent (for the distance $\max\{\delta(\mathscr{E}, \mathscr{F}), \delta(\mathscr{F}, \mathscr{E})\}$ defined above) to mixtures of experiments in which the variables can take only two values, this remains his privilege.

In its most restricted form the theory seems to be well adapted to the following type of problem. If two persons disagree about the validity, correctness or adequacy of certain statements about nature they may still be able to agree about conducting an experiment 'to find out'. For this purpose they will have to debate which experiment should be carried out and which rule should be applied to settle the debate. If one of them modifies his requirements after the experiment, if the experiment cannot be carried out, or if another experiment is used instead, or if something occurs that nobody had anticipated, the original contract becomes void.

Since the classical theory is essentially mathematical and clearly not normative it is rather unconcerned about how one interprets the probability measures P_θ. The easiest interpretation is probably that certain experiments such as tossing a coin, drawing a ball out of a bag, spinning a roulette wheel, etc., have in common a number of features which are fairly reasonably described by probability measures. To elaborate a theory or a model of a physical phenomenon in the form of probability measures is then simply to argue by analogy with the properties of the standard 'random' experiments.

The classical statistician will argue about whether a certain mechanism of tossing coins or dice is in fact adequately representable by an 'experiment' in the technical stochastic sense and he will do that in much the same manner and with the same misgivings as a physicist asking whether a particular mechanical system is in fact isolated or not.

Just as the physicist may say 'let us pretend that the system is isolated and apply the laws of conservation of energy', the statistician may say 'let us pretend that the system at hand is representable by an experiment'. Whether the conclusions derived from this 'pretending' game are 'right' or 'wrong' is a question which has received no solution so far, assuming of course that the question has a meaning.

About 'behavior' the only prescription we have been able to find in the classical literature is the one proposed by J. Neyman under the name of 'principle of inductive behavior'. In a soft, informal wording the principle states essentially this. 'If you have no substantial reason to do otherwise it is not altogether silly to use a procedure which has a high probability of success.' Certain theorems underlie this principle, one of which is the following.

For each $j = 1, 2, \cdots, n$, let $\mathscr{E}_j = \{\Theta_j, \mathscr{X}_j, \mathscr{A}_j, \{P_{j,\theta}\}\}$ be an experiment. Let $\{\Theta, \mathscr{X}, \mathscr{A}\}$ be the cartesian product of the $\{\Theta_j, \mathscr{X}_j, \mathscr{A}_j\}$. For $\theta\omega = \{\theta_1, \theta_2, \cdots, \theta_n\} \in \Theta$ let P_ω be the cartesian product of the P_j, θ_j. Further, for each j let φ_j be a test function having β_j for power function.

Finally, for each $x_j \in \mathscr{X}_j$ let Z_j be a random variable which, conditionally given $x = x_1, x_2, \cdots, x_n$ takes value unity with probability $\varphi_j(x_j)$ and value zero with probability $1 - \varphi_j(x_j)$. Let $\beta(\omega) = (1/n)\Sigma\beta_j(\theta_j)$. Then

$$P_\omega\left\{\left|\frac{1}{n}\sum_j^n Z_j - \beta\right| > \varepsilon\right\} \leqslant \frac{1}{4n\varepsilon^2}.$$

According to this theorem, if one applies in n different and unrelated experiments procedures with a small average probability of error, then the average number of mistakes will be small with overwhelming probability as soon as n is sufficiently large.

In other words, the actual batting average will be close to the average probability of being correct.

It has been shown recently by H. Robbins and others that for certain sequences $\mathscr{E}_j$; $j = 1, 2, \cdots, n$; it is possible to improve the batting average by treating all the problems together (or sequentially) instead of individually but this is a refinement which does not endanger the principle itself.

Since we are concerned here with a comparison of the classical and neo-Bayesian approaches it seems necessary to comment on the importance of Theorem 1. Suppose to avoid infinite difficulties that Θ, $\mathscr{X}$, $\mathscr{A}$ and everything else in sight is finite. Then Theorem 1 says that Bayes solutions form a complete class if the set $\mathscr{D}$ is the set of all decision procedures. This is an extension of classical results concerning Lagrange multipliers. One can understand its philosophical implications better by comparing it with the following remark.

Let $\mathscr{R}_0$ be the subset of $\mathscr{R}$ consisting of those $r \in \mathscr{R}$ which minimize expressions of the type $\sup\{|r(\theta)-f(\theta)|;\ \theta \in \Theta\}$ for functions f which are neither in $\mathscr{R}$ nor larger than elements of $\mathscr{R}$. The class $\mathscr{R}_0$ is complete.

In other words, admissible solutions are minimax solutions for suitable functions f, and in fact this statement remains valid under the conditions of Theorem 1 assuming only in addition that the elements of $\mathscr{R}$ are real functions.

Apparently nobody claimed that this result gives the minimax principle an advantage over other principles. Nobody even seems to claim that the minimax principle should be used.

It is also clear that neither the completeness of the class of Bayes solutions nor the completeness of the class of modified minimax solutions will necessarily hold for cases where $\mathscr{D}$ is a proper nonconvex subset of the class of all decision functions. In such cases there may be admissible functions which are not Bayes solutions and the classical statistician may then feel that it is not necessary for him to behave as if he had an a priori distribution.

Another point on which there is some disagreement is the following. Suppose that θ actually was a random variable with a distribution μ known to the statistician. Should he use the Bayes procedure relative to μ? The classical theory does not say. He may prefer to look at risk functions conditionally given θ (and possibly other things) and take various precautions instead of shooting for the best average with respect to μ. To be sure if the same problem presented itself over and over again there may be some advantage in using μ. However, if θ has been selected, it is not random even if it is unknown and two persons having an argument over its actual value may well agree to use a procedure which does not use μ explicitly.

4. THE PERSONALISTIC THEORY OF DECISION

Any given person indulges in beliefs, preferences and prejudices. As far as beliefs are concerned ordinary language makes a small and obviously inadequate provision for their expressions by means of words such as probable, very probable, likely, certain, possible, impossible. It is very tempting to try to construct a mathematical framework describing these states of mind more accurately as well as their possible modifications through the intervention of fresh observational evidence or other information. Such an attempt should be considered as reasonable as the early attempt to formalize the concepts of temperature and quantity of caloric. Unfortunately the attempts of a theory of pistimetry which have come to our attention are not very convincing and not very precise.

Several authors have noted that introspection on degrees of belief is difficult to formalize but that one may be able to pin down expressions of opinion by offering suitable bets. Unless the bets are actually performed this is only an assisted form of introspection. However, with the introduction of such considerations one can build a formal theory of a 'homo economicus' which deals at the same time with beliefs and preferences.

The most widely known theory of this type is perhaps the one expounded in 1954 by L. J. Savage. In certain respects it is also the simplest and the most beautiful. Thus we shall restrict our comments to Savage's theory, ignoring in particular the more recent, but less convincing, papers.

It is characteristic of the pistimetric and preferential theories available at the present time that they do not attempt a formalization of the concept of experiment and tend to treat experiments and fortuitous observations alike. In fact, the main reason for their periodic return to fashion seems to be that they claim to hold the magic which permits to draw conclusions from whatever data and whatever features one happens to notice.

It seems also that their recurrent fading is very closely linked to the fact that they are imprecise, naive and unable to allow substantific discussion involving several persons. To exemplify we shall now restate in our own language the very first result of Savage's theory.

Savage uses nine axioms or postulates couched in almost formal mathematical language. They are restated below in six parts. We hope that the informality of the language has not led us to gross misinterpretations.

The axioms are relative to two sets S and C and to the set $\mathscr{F}(S, C)$ of all functions from S to C. If $f \in \mathscr{F}(S, C)$ and $A \subset S$ the restriction of f to A will be denoted $f|A$. The complement of A will be denoted A^c. An element f of $\mathscr{F}(S, C)$ which coincides with g on A and h on A^c will be denoted $f = gI_A + hI_{A^c}$. The set of functions from $A \subset S$ to C will be denoted $\mathscr{F}(A, C)$.

Axiom 1. Each one of the spaces $\mathscr{F}(A, C)$, $A \subset S$ is totally preordered by a relation denoted '$\leq$', and the preorder of $\mathscr{F}(S, C)$ is not the trivial one where $f \leq g$ for every pair (f, g).

For a particular $A \subset S$ let f be a function identically equal to α on A and let g be a function identically equal to β. The preorder of $\mathscr{F}(A, C)$ induces a preorder on C by considering that $\alpha \leq \beta$ is equivalent to $f \leq g$.

Axiom 2. For every $A \subset S$ the preorder induced by $\mathscr{F}(A, C)$ on C is either trivial or identical to the order induced by $\mathscr{F}(S, C)$.

Axiom 3. The relation '$(f|A) \leq (g|A)$' is equivalent to the relation:

$$fI_A + hI_{A^c} \leq gI_A + hI_{A^c} \quad \text{for every } h \in \mathscr{F}(S, C)\,.$$

Axiom 4. Suppose that $g < h$ in $\mathscr{F}(S, C)$. Then for every $\alpha \in C$ there is a partition $\{A_j; j = 1, 2, \ldots, k\}$ of S such that if $g_j = gI_{A_j^c} + \alpha I_{A_j}$ and $h_j = hI_{A_j^c} + \alpha I_{A_j}$ then either $g_j < h$ or $g < h_j$.

Axiom 5. Let α, α', β and β' be constant elements of $\mathscr{F}(S, C)$ such that $\alpha' < \alpha$ and $\beta' < \beta$ in the ordering of $\mathscr{F}(S, C)$. If

$$\alpha I_A + \alpha' I_{A^c} \leq \alpha I_B + \alpha' I_{B^c}$$

then

$$\beta I_A + \beta I_{A^c} \leq \alpha I_B + \alpha' I_{B^c}\,.$$

Let g be an element of $\mathscr{F}(A, C)$. If $s \in A$ let g_s be the element of $\mathscr{F}(A, C)$ which is identically equal on A to the value of g at s.

Axiom 6. Let f and g be two elements of $\mathscr{F}(A, C)$. If for every $s \in A$ the relation $f \leq g_s$ holds then $f \leq g$. Similarly $f \geq g_s$ for all $s \in A$ implies $f \geq g$.

From these postulates Savage derives a result, which, if our interpretation and translation is correct, can be stated as follows.

THEOREM 2. If the axioms 1 to 6 are satisfied there is a numerical function U defined on C and a finitely additive probability measure P defined on the σ-field of all subsets of S such that $f \leqslant g$ is equivalent to

$$\int U[f(s)]P(ds) \leqslant \int U[g(s)]P(ds)$$

for every pair (f, g) of elements of $\mathscr{F}(S, C)$.

Furthermore, P is uniquely defined and U is unique up to a linear transformation.

Although we have not checked the proof of this theorem in detail and may have misinterpreted Savage's assertions it is fairly clear that the theorem as stated or a small modification of it must be correct. Assuming this, it is a beautiful theorem, deriving from purely qualitative assumptions a precise quantitative conclusion.

In this respect previous theorems by von Neumann, de Finetti and others, as well as the more recent work of Pratt, Raiffa and Schlaifer are not nearly as appealing.

Of course axiom 4 implies that S is an infinite set, so that one may be reluctant to use all the functions from S to C and look at measures defined for all the subsets of S. However, it is fairly clear that one could assume that S carries a σ-field $\mathscr{A}$ and restrict oneself to sets which belong to $\mathscr{A}$ and to functions which are $\mathscr{A}$-measurable. The measure P would then be defined only on $\mathscr{A}$. Savage's proof seems to be left unchanged by this modification, although, here again, we have not checked the details.

Axiom 5, which is an anti-bribery law, is not relevant if C has two elements. Similarly Axiom 6 loses in this case much of its strength. Since for the purpose of proving the existence of P one can choose two elements of C and work with a set having only two elements, the main axioms leading to the existence of P are essentially the first four.

The first four axioms are closely related to those of B. O. Koopman, but they are more attractive, even though they are subject to the same kind of criticism that Koopman himself raised against his own axioms.

It is of course possible to reject the theory altogether because the axioms are intuitively too strong and unacceptable. However, the situation is unfortunately not that simple as the following elaboration will show.

Let $\mathscr{A}$ be an arbitrary Boolean algebra. There is, of course, no loss of generality in assuming that $\mathscr{A}$ is a Boolean algebra of subsets of certain set S. If $A \in \mathscr{A}$, let I_A be unity on A and zero otherwise. Let $\mathscr{V}$ be the set of all such indicators. Let $\mathscr{S}$ be the vector space of step functions on $\mathscr{A}$. Explicitly $f \in \mathscr{S}$ if f is a finite sum $f = \Sigma \alpha_j I_{A_j}$ where the α_j are real and where $\{A_j\}$ is a partition of S by elements of $\mathscr{A}$. Instead of real numbers, one could use rational numbers without essential changes.

Let R be a subset of $\mathscr{V} \times \mathscr{V}$. It will be assumed throughout that $(0, v) \in R$ for every $v \in \mathscr{V}$.

Following the method of Kraft, Pratt, and Seidenberg one can use the relation R to induce an order, or more precisely a preorder on $\mathscr{S}$.

Let K be the set of elements f of $\mathscr{S}$ which can be written in the form of finite sums

$$f = \Sigma \alpha_j (v_j - u_j)$$

with $\alpha_j \geq 0$ and $(u_j, v_j) \in R$. The requirement $(0, v) \in R$ for every v implies that K contains all the positive elements of $\mathscr{S}$. The set K is a convex cone in $\mathscr{S}$. Thus, it induces a preorder on $\mathscr{S}$. This preorder is obviously the smallest one which is compatible with the linear structure of $\mathscr{S}$ and the relation R.

Since the existence of a positive linear functional μ such that $(u, v) \in R$ implies $\mu(u) \leq \mu(v)$ is a very weak property it is convenient to require somewhat more, as follows.

Let R and R_1 be two subsets of $\mathscr{V} \times \mathscr{V}$ such that

(1) $R_1 \subset R$.

(2) $(0, v) \in R$ for every $v \in \mathscr{V}$.

(3) $(0, I) \in R_1$.

A positive linear functional μ on $\mathscr{S}$ will be called compatible with the pair (R, R_1) if $(u, v) \in R$ implies $\mu(u) \leq \mu(v)$ and $(u, v) \in R_1$ implies $\mu(u) < \mu(v)$. The case where only R is under consideration reduces to this by putting for R_1 the set consisting only of the point $(0, I)$.

THEOREM 3. Let (R, R_1) be a pair of subsets of $\mathscr{V} \times \mathscr{V}$ such that $R_1 \subset R$ and $(0, v) \in R$ for every $v \in \mathscr{V}$ and $(0, I) \in R_1$. Let K be the cone of functions $f \in \mathscr{S}$ having the form

$$f = \Sigma \alpha_j (v_j - u_j)$$

with $\alpha_j \geq 0$ and $(u_j, v_j) \in R$.

There exists a positive linear functional μ which is compatible with the pair (R, R_1) if and only if there is a function ε from R_1 to the open half line $(0, \infty)$ such that no function g of the type

$$g = f + h, \quad \text{with } f \in K \quad \text{and}$$

$$h = \Sigma\alpha_j[(v_j - u_j) - \varepsilon(u_j, v_j)], \qquad \alpha_j > 0,$$

$(u_j, v_j) \in R_1$ vanishes identically.

Proof. Let C be the convex cone formed by functions f of the type $f = \Sigma\alpha_j(v_j - u_j)$ with $\alpha_j > 0$ and $(u_j, v_j) \in R_1$. The existence of a μ compatible with (R, R_1) is equivalent to the existence of an open convex cone G such that $0 \notin G$ and $C + K \subset G$.

Indeed, if such a cone exists, the Hahn-Banach theorem insures the existence of a linear functional μ such that $\mu(f) > 0$ for $f \in G$. This implies $\mu(f + \alpha) > 0$ for every $f \in K$ and every $\alpha > 0$, hence $\mu(f) \geqslant 0$ for $f \in K$. Since K contains the cone of positive elements of $\mathscr{S}$, the functional μ is necessarily positive. The converse implication is trivial.

If an open convex cone G exists, for every pair $(u, v) \in R_1$ there is an $\varepsilon > 0$ such that $(v - u) - \varepsilon \in G$ hence the necessity of the condition expressed in the theorem. Conversely if the function ε exists, then the set C_1 of functions g such that $g \geqslant \Sigma\alpha_j[(v_j - u_j) - \eta(u_j, v_j)]$ with $\alpha_j > 0$ and $(u_j, v_j) \in R_1$ and $\eta(u, v) < \varepsilon(u, v)$ is an open convex set containing C. The cone $C_1 + K$ is an open cone which contains $C + K$ and the origin does not belong to $C_1 + K$. This completes the proof of the theorem.

In the particular case where R_1 is reduced to the single point $(0, I)$ of $\mathscr{V} \times \mathscr{V}$ the condition can be rephrased as follows.

COROLLARY. If $R_1 = \{(0, I)\}$ there is a positive linear μ compatible with (R, R_1) if and only if the function identically equal to -1 does not belong to K.

This can be checked directly or can be deduced from Theorem 3 by taking $\varepsilon(0, I) = \frac{1}{2}$.

This corollary can be given the following interpretation.

Suppose that a statistician is challenged to express his opinions about the location of point $s \in S$. For every pair (A, B) of elements of $\mathscr{A}$ he may either decline to answer or indicate that if he were forced to choose he would choose B in preference to A.

It is assumed that the statistician will prefer any nonempty set $A \in \mathscr{A}$ to the empty set and that he does not prefer the empty set to the whole space.

The challenger decides to test the preferences expressed by the statistician as follows. For a pair (A, B) for which the statistician expresses preference of B over A the challenger selects a positive number α. If B occurs but not A, the statistician receives α. If A occurs but not B, the statistician pays α. This is done for a finite number of pairs (A_j, B_j) with the effect that the total amount paid by the statistician is $f(s) = \Sigma_j \alpha_j [I_{A_j}(s) - I_{B_j}(s)]$.

Unless the preferences expressed by the statistician are compatible with some finitely additive probability measure on $\mathscr{A}$ the challenger can select pairs (A_j, B_j) and numbers α_j such that $f(s)$ is always larger than unity, for all values of $s \in S$.

When the algebra $\mathscr{A}$ is finite the condition given in the theorem can be simplified. The function ε is not relevant. It is necessary and sufficient that no function $f+g$, $f \in K$, $g = \Sigma\alpha_j(v_j - u_j)$ with $\alpha_j > 0$ and $(u_j, v_j) \in R_1$ be identically zero. This is the result of Kraft, Pratt and Seidenberg. The finiteness of $\mathscr{A}$ implies that $\mathscr{S}$ is finite dimensional. Hence the simplification.

In general, the function ε defined on R_1 can be interpreted as follows: there are prizes $\varepsilon(u, v)$ so small that even if they are offered a bet on v against u still seems preferable to the sure prize $\varepsilon(u, v)$.

Combining Theorem 3 with the theorem of de Finetti which states that preferences expressed by odds must follow the rules of the calculus of probabilities, and with the various theorems of Koopman, Savage and others it does seem that one would be forced to admit that opinions or preferences must be probabilistic to be coherent.

At least it is impossible to deny that a pair of relations (R, R_1) satisfying the general conditions of Theorem 3 but incompatible with every finitely additive probability measure, must have rather disagreeable properties.

Fortunately, or unfortunately depending on one's personal views, the implications of this for statistical purposes are rather meager as we shall now try to show.

First, under most circumstances, statisticians are not called upon to express a whole spectrum of opinions about the location of a point s or the value of a parameter. Clearly, any expression of opinion of the type 'A is better than A^c and B is better than B^c' will be compatible with a measure provided only that $A \cap B$ be not empty.

Second, whether or not one adopts arguments of the pistimetric type, the validity of the statements of the classical theory of statistics remains entirely unaffected.

Third, pairs of relations (R, R_1) which are compatible with a *unique* probability measure do not arise very commonly. It is true that the total order and the partitioning possibilities invoked by Savage (in Axiom 4 of our list) will enforce uniqueness of the compatible measure. However, the practical cases where such an extensive relation can effectively be spelled out are very few and very special. To claim that an ideal person could in principle specify such a relation is to beg the question. To claim that since an ideal person could do it, a real person should do it is to introduce a dogma for which we have little justification.

It appears therefore more acceptable to resign oneself to the fact that in many if not most cases the pair (R, R_1) does not specify μ entirely.

Fourth, some people have complicated minds. If they did attempt to express their opinions on the location of a point $s \in S$ the opinions would include statements such as these: A_1 is less likely than B_1, but A_2 is much less likely than B_2 and A_3 is incomparably less likely than B_3. To formalize this one would have to introduce not a pair of relations (R, R_1) but a whole string of relations $\{R_0, R_1, R_2, \dots, R_n\}$.

Let us suppose for the sake of argument that we have such a string of relations with $R_j \supset R_{j+1}$. If there is a μ compatible with (R_0, R_1) and even if such a μ is unique, it does not follow in any way that the relative magnitudes of the numbers $\mu(A)$ and $\mu(B)$ are any reflection or indication of the validity of the relations R_2 or R_n.

To put it differently, suppose that we have three relations $(A, B) \in R_0$ expressing that A is not preferable to B, $(A, B) \in R_1$ expressing that B is strictly preferred to A and $(A, B) \in R_2$ expressing that B is vastly preferred to A.

Suppose that the pair (R_0, R_1) is compatible with one and only one measure μ. If it turns out that $c\mu(A) \leq \mu(B)$ for $c = (10)^{137}$, this does not in any way imply $(A, B) \in R_2$. In fact, it might turn out that for all pairs $(A, B) \in R_2$ with $A \neq \varnothing$ one has $(10)\mu(A) \geqslant \mu(B)$ but that for some pairs (A, B) which are in R_1 and not in R_2 one has $c\mu(A) \leqslant \mu(B)$.

Fifth, from a technical point of view the existence of measures compatible with relations (R, R_1) is closely connected with the fact that the relations were supposed to be relations on a Boolean algebra. If the only sets entering in the relations R and R_1 were reduced to sets consisting of one single point the whole analysis would be quite spurious. In this case similar arguments would lead to the existence of a function defined on S not to the existence of a measure on the subsets of S. From 'states of nature' which are described in quantum mechanical terms it is often impossible to construct Boolean algebras unless one deliberately ignores the laws of physics. In this case again theorems such as Theorem 3 lose their relevance.

Finally, some people have complicated minds and some people do have most complicated minds to the extent that even a string of relations $\{R_0, R_1, \ldots, R_n\}$ would not allow them to express themselves. True, one may be forced to satisfy oneself with unsatisfactory choices, and one does so every day, however a theory of decision which claims that one ought to be satisfied goes too far. The situations in which one would be forced to admit that $\mu(A) = \mu(B)$ because he has no relevant information and the situations in which one definitely states $\mu(A) = \mu(B)$ because the system selecting s has been programmed to give A and B equal chances (as in roulette wheels) are vastly different. One should perhaps investigate relations not only in $\mathscr{A} \times \mathscr{A}$ but on strings of relations. The firmness of a belief that $\mu(A) = \mu(B)$ does not appear to be expressible in terms of μ.

With this in mind, let us pass to some considerations on the possibility of applying Savage's theory and for this purpose let us return to the framework proposed by Savage. In this framework there were two sets S and C and the set $\mathscr{F}(S, C)$ of functions from S to C. For purported applications the set S is supposed to represent the set of possible 'states of nature'. The set C is a set of 'consequences' and the elements of $\mathscr{F}(S, C)$ are 'acts' which associate consequences to states of nature.

The present author finds himself quite unable to understand what is meant by 'states of nature'. Savage's own explanations are inadequate and careless. In the famous six egg omelette, Savage lists two states of nature and six consequences but only three acts instead of the more usual

6^2 possible acts. This would not be too serious if it did not go to the heart of the applicability of the theory. One can argue that in many cases it is impossible for certain consequences to be associated with certain events. Examples where 'consequences' are inextricably mixed with the 'events', or states of nature are easily constructed. For instance, when debating the possibility of buying some life insurance, or contributing to a retirement fund, can one compare the values independently of whether one will be dead or alive? Can one invent acts which will permit one equally to enjoy retirement in the countryside whether dead or alive?

In the example of the six egg omelette, was there an act which permitted us to have a six egg omelette even if the sixth egg was rotten?

Typically the answer to such questions is no, and this is probably why in all the examples given by Savage to each pair (s, f) consisting of a state of nature s and an act f corresponded a consequence $c(s, f)$ not all conceivable assignments being possible.

In fact, since generally, the preference ordering of 'consequences' may be strongly affected or even reversed according to whichever state of nature happens to be true one could argue with Drèze that Savage's theory applies only to those cases where our actions modify the probabilities of the various states of nature.

Since, according to all appearances, even Theorem 3 would require us to specify what is meant by 'state of nature' one cannot evade the difficulty by slight modifications of the axioms.

In some cases it is easy to specify what the 'states of nature' are since only questions of fact are involved. This happens for instance if one asks whether a certain passage in the Federalist Papers was written by Madison or Hamilton. However, what are the 'states of nature' in the question 'is Krebiozen effective against cancer?' If this appears too clear, what are the states of nature in the problem of finding out whether cancer is caused by a virus or by one mutation or by two successive mutations?

In many problems to which statistical methods are now routinely considered applicable there are no recognizable 'states of nature' only various 'theories' or 'models'.

Whether some specified theory is an adequate or 'correct' representation of the physical world is not satisfactorily formalizable at present. It is possible for instance for two contradictory theories to be representations of the same phenomenon with different and overlapping domains of adequacy. Thus one should not expect 'sets of theories' to behave according to the laws of Boolean algebra.

To go further, most models or theories of nature which are encountered in statistical practice are probabilistic or stochastic. The probability measures entering in these models are only vaguely related to opinions and preferences. On the contrary, they are used to indicate a certain structure which can, in final analysis, be reduced to this 'Everything is as if one were drawing balls from a well-mixed bag.'

One of the peculiarities which is bound to occur in the application of Savage's system to such situations is the following. Suppose that one is faced with the problem of testing some hypothesis about the value of a parameter $\theta \in \{0, 1\}$. Suppose also that the experiment to be conducted is a Binomial one where

$$\Pr[X = k|\theta] = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Presumably a neo-Bayesian statistician can look at the subsets of the set $S = \Theta \times J$ where J is the set of integers $J = \{0, 1, 2, \ldots, n\}$ and induce a pair (R, R_1) of relations on the algebra $\mathscr{A}$ of subsets of S. Presumably there exists at least one measure μ compatible with (R, R_1) on $\mathscr{A}$ and presumably this measure μ is also compatible with the Binomial probabilities listed above.

If μ is not uniquely defined, the theory does not tell us what to do. If μ is unique then Bayes theorems will give the required answers after X is observed.

In this form the neo-Bayesian statistician is in just about the same kind of situation as the classical one except that he may have somewhat more trouble convincing his customer.

However, one of the advantages of the neo-Bayesian theory is that it pretends to answer questions about θ even when X has already been drawn and made available so that for the classical statistician nothing stochastic would be left. In this case there are two possibilities, one is to ignore the value of X and proceed to the same introspection as above *without letting oneself be influenced by the known value* $X = k$. The other is to assign a measure in the set Θ directly without regard to the machinery by which k was obtained. Both possibilities appear to us quite unrealistic. But if one has to have recourse to the arguments which take place before the experiment is conducted, the neo-Bayesian approach has no particular advantage over the classical one.

This same binomial example can be used to illustrate the fact that the theory does not allow sufficient formalism for communication of information to others. Nor does it allow any debate on the reasons why certain decisions are taken or why certain opinions are held.

Consider the same binomial experiment with $\theta \in \Theta = [0, 1]$. It seems to be a consequence of the theory that after the observation has taken place and X has been found equal to k the entirety of the relevant evidence or opinion of the neo-Bayesian statistician is expressed by the measure which possesses a density

$$p(\theta) = \frac{\theta^k (1-\theta)^{n-k}}{\int t^k (1-t)^{n-k} \lambda(dt)}$$

with respect to the initial measure λ on Θ. This measure λ could for instance be any one of a family having densities $\theta^\alpha (1-\theta)^\beta$ with respect to Lebesgue measure. The density $p(\theta)$ is then proportional to $\theta^{k+\alpha}(1-\alpha)^{n-k+\beta}$. Thus if we follow the theory and communicate to another person a density $C\theta^{100}(1-\theta)^{100}$ this person has no way of knowing whether (1) an experiment with 200 trials has taken place or (2) no experiment took place and this is simply an a priori expression of opinion.

Since some of us would argue that the case with 200 trials is more 'reliable' than the other, something is missing in the transmission of information.

If the neo-Bayesian has to give not only his final measure, but also his initial measure, the description of the experiment and the result obtained there, the simplicity of the Bayes approach is lost.

What is more, an admission that such a complete description should be given is an admission of the fact that a probability measure is unable to convey the whole of the relevant evidence. Why should one then argue that the theory is still adequate for the behavior of one single person, since this same person will forget from one day to the next and will have to refresh his memory by asking what experiments have been performed?

To comment briefly on another point, the neo-Bayesian theory seems to treat alike all forms of uncertainty. However, it is clear that we can be uncertain for many reasons. For instance, we may be uncertain because (1) we lack definite information, (2) the events involved will occur or not according to the results of the spin of a roulette wheel, (3) we could find out by pure logic but it is too hard. The first type of uncertainty occurs in practically every question. The second assumes a well-defined mechanism. However, the neo-Bayesian theory seems to make no real distinction between probabilities attached to the three types. It answers in the same manner the following questions.

(1) What is the probability that Eudoxus had bigger feet than Euclid?

(2) What is the probability that a toss of a 'fair' coin will result in tails?

(3) What is the probability that the $10^{137}+1$ digit of π is a 7?

Even Savage and de Finetti admit that, especially in cases involving the third kind of uncertainty, our personal probabilities are fleeting, more or less rapidly in that the very act of cogitating to evaluate precisely the probabilities is enough or can be enough to modify or totally overcome the uncertainty situation which one wanted to express.

Thus, presumably, when neo-Bayesians state that a certain event A has probability one-half, this may mean either that he did not bother to think about it, or that he has no information on the subject, or that whether A occurs or not will be decided by the toss of a fair coin. The number $\frac{1}{2}$ itself does not contain any information about the process by which it was obtained, fleetingly or not.

As a final comment, it seemes necessary to mention that in certain respects the theory of personal probability is very similar to a theory of personal mass, which exhibits the same shortcomings.

Suppose that a store owner is asked to assign weights to the items in his store. For this purpose he can group items in sets and compare them by hand. If a set A appears to him lighter than a set B we shall say that $(A, B) \in R$. It is fairly easy to see, in the spirit of Theorem 3, that if the relation R is not compatible with an assignment of individual masses to the items and with the additivity of masses, the system is not very coherent. It is also possible to show that if there are enough items which could be indefinitely divided into 'equally weighty parts' the assignment of masses will be unique up to a multiplicative constant.

Nobody would be particularly surprised however if it turned out that ten thousand peas which were judged all alike when compared pairwise turn out to be quite different when parted into two sets of 5000.

In theory one would try to reconcile these contradictory feelings but it is not clear that it could be achieved nor that it would be worth the trouble, since similar difficulties may then crop up elsewhere.

In spite of the theoretical possibility of assigning masses by hand comparison in this manner, nobody seems to claim that this is just what should be done in stores. Nobody even claims that since masses are masses there is no point in specifying whether they were obtained by hand comparison, or by using a spring scale or by using a balance. In addition the hand comparison system would lead to classify people in categories according to their ability to guess weights and according to their ability to avoid self-deceptions due to size of containers or density of the material.

The parallelism between this and the proposals of the neo-Bayesian school is quite evident. The proposal to classify people according to the sharpness of their ability for statistical guessing has already been made. For instance, Halphen could state that there is no good and bad statistics, there are good and bad statisticians.

If the process of measuring something as definite as masses by hand comparison seems rather unreliable, can one really expect a similar theory of measurement of ethereal opinions to inspire much confidence? If an indication of the process of measurement is helpful in the masses problem, it also appears necessary in the opinion problem.

Finally, an assignment of masses may conceivably be checked by experimenting with a scale, but the neo-Bayesian theory does not even pretend to make statements which could be checked by an impartial observer.

5. CONCLUSION

In the foregoing pages we have attempted to indicate that the classical theory is essentially a formalization and study of experiments and functions defined on experiments. We also have attempted to show that the backbone of the neo-Bayesian theory is a naive assumption that since by forcing an individual to make decisions one can perhaps force him to behave as if there was a probability measure on certain undefined states of nature, there is no point in elaborating further what kind of prejudices and what more or less unreliable evidence entered in the decisions.

Classical statistics achieves some kind of interpersonal communicability by restricting its domain to well-defined experiments. Whether the neo-Bayesian theory can reach this goal without losing its flavor appears doubtful. However, it seems clear that any theory which pretends to permit discussion between two persons of the quality and nature of the evidence available will have to be more complicated than the Bayesian theory. It also will have to imbed some indication of the possibility of verification of the validity of the statements made.

For the practical purposes of teaching statistics, the neo-Bayesian approach does not seem to have any definite advantage over the classical one but it has already led to some strange claims, which, if taken seriously, would tend to undo a generation of patient and often painful public education.

One of the claims is that the experiment matters little, what matters is the likelihood function after experimentation. Whether this is true, false, unacceptable or inspiring, it tends to undo what classical statisticians have been preaching for many years: think about your experiment, design it as best you can to answer specific questions, take all sorts of precautions against selection bias and your subconscious prejudices. It is only at the design stage that the statistician can help you.

Another claim is the very curious one that if one follows the neo-Bayesian theory strictly one would not randomize experiments. The advocates of the neo-Bayesian creed admit that the theory is not so perfect that one should follow its dictates in this instance. This author would say that no theory should be followed, that a theory can only suggest certain paths. However, in this particular case the injunction against randomization is a typical product of a theory which ignores differences between experiments and experiences and refuses to admit that there is a difference between events which are *made* equiprobable by appropriate mechanisms and events which are equiprobable by virtue of ignorance. Furthermore, the theory would turn against itself if the neo-Bayesian statistician was asked to bet on the results of a poll carried out by picking the 100 persons most likely to give the right answers.

In spite of this the neo-Bayesian theory places randomization on some kind of limbo, and thus attempts to distract from the classical preaching that double blind randomized experiments are the only ones really convincing.

There are many other curious statements concerning confidence intervals, levels of significance, power, and so forth. These statements are only confusing to an otherwise abused public. If the neo-Bayesian would advocate certain specific methods, one could presumably investigate what the consequences of these methods would be in classical terms, but this is an evaluation they refuse. It must then be very confusing to the layman to be told that the neo-Bayesian theory is vastly superior, but that its claims must be taken on faith.

In view of this, we can only conclude that the neo-Bayesian theory is a premature and confusing return to 1763.

Department of Statistics,
University of California, Berkeley

NOTE

* This paper was prepared with the partial support of the National Science Foundation, Grant GP-2593.

REFERENCES

Bernoulli, D., 'Specium Theoriae Novae De Mensura Sortis', *Commentarii Academiae Scientarum Imperialis Petropolitanae* **5** (1738), 175–192.
Birnbaum, A., 'On the Foundations of Statistical Inference', *J. Amer. Statist. Assoc.* **57** (1962), 269–326.
Bodiou, G., *Théorie dialectique des probabilités, englobant leurs calculs classique et quantique*, Gauthier-Villars, Paris, 1964.
Ellsberg, D., 'Risk, Ambiguity and the Savage Axioms', *Quarterly J. of Economics* **75** (1961), 644–661.
de Finetti, B. and Savage, L. J., 'Sul modo di scegliere le probabilita iniziali', special issue 'Sui Fondamenti della Statistica' *Biblioteca del Metron, Series C* **1** (1962), 81–147.
de Finetti, B., *Probability, Induction and Statistics*, J. Wiley & Sons, N.Y., 1972.
de Morgan, A., *Formal Logic*, Taylor and Walton, London, 1847.
Drèze, J., 'Les probabilités "subjectives" ont-elles une signification objective?', *Economie appliquée* **13** (1960), 55–70.
Halphen, E., 'La notion de vraisemblance – Essai sur les fondements du calcul des probabilités et de la statistique mathématique', *Publications de l'Institut de Statistique de l'Université de Paris* **4** (1955), 41–92.
Kraft, C., Pratt, J. and Seidenberg, A., 'Intuitive Probabilities on Finite Sets', *Ann. Math. Statist.* **30** (1959), 408–419.
Neyman, J. and Pearson, E. S., *Joint Statistical Papers of J. Neyman and E. S. Pearson*, University of California Press, Berkeley, 1967.
Savage, L. J., *The Foundations of Statistics*, J. Wiley & Sons, N.Y., 1954.

Savage, L. J., *The Foundations of Statistical Inference*, Methuens Monographs, London, 1962.
Suppes, P., 'The Measurement of Belief', *J. Roy. Statist. Soc.* (Ser. B) **36** (1974).
van Dantzig, D., 'Statistical Priesthood', *Statistica Neerlandica* **11** N.1 (1957), 1–16.
Venn, J., *The Logic of Chance* (1866), reprinted, Chelsea, N.Y., 1962.
Wald, A., *Statistical Decision Functions*, J. Wiley & Sons, N.Y., 1950.
Wolfowitz, J., 'Bayesian Inference and Axioms of Consistent Decision', *Econometrica* **30** (1962), 470–479.
Wolfowitz, J., 'Reflections on the Future of Mathematical Statistics', Chapter 39 of *Essays in Probability and Statistics in Honor of S. N. Roy*, University of North Carolina Press, 1969.