Book Review from Am. Jour. Phys. vol. 31, p. 66 (1963).

The Algebra of Probable Inference. RICHARD T. COX. Pp. 114, The Johns Hopkins Press, Baltimore 18, Maryland, 1961. Price $5.00.

It is a particular pleasure to write a review of this book for the American Journal of Physics, because the book had its beginnings in an article by Professor Cox which appeared in this journal in 1946, and for several years I have been giving lectures on probability theory which draw their inspiration from that article. It is, in my opinion, one of the most important ever written on the foundations of probability theory, and the greatest advance in the conceptual, as opposed to the purely mathematical, formulation of the theory since Laplace.

In the meantime discussion among statisticians and others, concerned with the practical use of probability theory has continued, and intensified, on a very basic question: Is it or is it not legitimate to use the mathematical rules of probability theory in problems of inductive inference, where the probabilities cannot be regarded as frequencies in any random experiment, but denote only a "degree of reasonable belief?" It is clear from the recent literature that protagonists on both sides are still unaware of the fundamental contribution made sixteen years ago by Cox, which all but settled this issue. Professor Cox has performed another valuable service by enlarging the original article, with several improvements in the derivations, and making it available to a wider audience in this small, but very important, book.

Professor Cox undertook to state in mathematical terms the most general rules for a "calculus of inductive reasoning" which would agree with certain elementary commonsense requirements and would represent degrees of plausibility by real numbers. He not only succeeded, but he showed us an entirely new technique for the construction of mathematical theories.

Let us recall the usual axiomatic technique in mathematics, of which perhaps the purest and most familiar example is the geometry of Euclid. We start with a few axioms, the rules of the game, which are basically arbitrary. Then we deduce their consequences, carrying the process as far as possible. The game continues until we discover a contradiction. The we know that our axioms were in some way inconsistent. We go back and modify them, and repeat the game. In this way, one hopes, we arrive eventually at a set of rules which can be applied as far as we please without generating contradictions.

Many authors have tried to develop probability theory as an extension of logic to the case of inductive inference in just this way; by arbitrarily stating the rules for associating degrees of plausibility with real numbers, and the rules for combining these numbers. This was done, for example, by Laplace (1812), de Morgan (1847), Keynes (1921), Jeffreys (1938) and Koopman (1940). But none of these efforts was taken seriously by twentieth-century mathematicians and statisticians, for a very simple reason; if the rules of combination are merely stated as arbitrary axioms, how do we know that they are unique? What makes these rules any better than a hundred other arbitrary ones we could invent?

Professor Cox's great contribution was to notice that there is another way of developing mathematical theories. Instead of stating the rules arbitrarily and hoping that they are free of inconsistencies, the requirement that the rules be consistent can be taken as one of the basic conditions imposed on the theory from the start. As in all really fundamental advances, the key to the situation lay simply in learning how to ask the right question. Cox found that the conditions for consistency could be written in the form of three functional equations, whose general solutions he proceeded to find. These conditions restricted the possibilities so greatly that for all practical purposes, there is only one way of writing the basic rules. Mathematical transformations can alter their form, but not their content.

In terms of the standard notation for conditional probabilities, $p(A|B)$="probability of A, given B," the basic rules derived by Cox are simply, $p(AB|C) = p(A|BC)\,p(B|C)$, and $p(A|B)+p(\bar{A}|B)=1$, where AB stands for the proposition, "both A and B are true," and $\bar{A}$ is the proposition, "A is false." These are, of course, the fundamental equations of probability theory; all others follow from their repeated application. This result established for the first time that the mathematical rules of probability theory given by Laplace not only constitue a valid calculus of inductive reasoning; they are in fact unique in the sense that any set of rules in which we represent degrees of plausibility by real numbers is either equivalent to Laplace's, or inconsistent.

Two years later, Professor Cox's method was used again, with equal success and infinitely greater fanfare, by C. E. Shannon. He wrote down conditions for a reasonable "information measure." Again the conditions of consistency took the form of functional equations, and with their solution the notion of entropy became a fundamental new concept in probability theory.

I want to emphasize the importance of Cox's result, because Cox himself has not seen fit to do so. While his great modesty and gift for understatement are charming and make every page of this book a delight to read, a less happy consequence is that the average reader would never realize the implications for practical problems. There are no applications, and not even a hint that applications exist.

In order to understand the full importance of Cox's work, we have to look at the recent history of statistics. Because of supposed difficulties with Laplace's rules, a belief arose that no valid "calculus of inductive reasoning" had been produced, and that the term "probability" should be used only in the restricted sense of "physical" or "statistical" probability, which refers not to anybody's judgments but only to limiting frequencies in the outcome of some random experiment. In other words, probability statements can be made only about random variables; it is meaningless to speak of the probability that an hypothesis is true, or that an unknown constant parameter lies in a certain interval. This viewpoint has been expounded by R. von Mises, R. A. Fisher, W. Feller, and many others, and it was adopted by almost all statisticians until very recently.

Unfortunately, a theory of probability restricted in this way is totally inadequate to meet the needs of practice; for as soon as we depart from the most artificial textbook-type problems, almost every useful application of probability theory concerns some problem of inductive infer-

ence, and not primarily any calculation of frequencies. Thus, whenever we decide between alternative hypotheses, estimate an unknown parameter, predict a future trend on the basis of present data, we are doing inductive, or plausible, reasoning *about a quantity which is not "random."* Adoption of the strict frequency interpretation of probability thus forced statisticians to relegate such problems to a new field, statistical inference, which was considered to be distinct from probability theory. Its aim was to avoid the supposed mistakes of Laplace by developing entirely new approaches.

Now a very strange thing happened. Some of our finest mathematicians labored for some fifty years, developing this new field. The culmination came in the 1950's with the decision theory of Abraham Wald, where for the first time general rules of procedure were uniquely derived from certain very elementary requirements on a reasonable theory. But by 1954, several workers in this field had realized that if we simply ignore Wald's entirely different vocabulary and diametrically opposed philosophy, and look only at the specific mathematical steps that were now to be used in solving specific problems, *they were identical witht he original rules given by Laplace in the 18th century, which two generations of statisticians had held to be metaphysical nonsense!* The real road to progress lay not in rejecting Laplace's methods, but in learning how to use them properly.

All the while this dramatic development (which statisticians refer to as the "Bayesian revolution", since Laplace's use of Bayes' theorem was the main bone of contention) was taking place, Cox's article, which contained the real key to understanding this situation, lay there ignored. The problems of statistical inference are all problems of inductive, or plausible, reasoning; and Cox's argument shows clearly why it does not make any difference which philosophy of interpretation you have when you approach these problems. By the time you have made your methods fully consistent, you will be forced to rediscover Laplace's principles, just as Wald did. But Cox's argument is much simpler and more direct than Wald's.

The emancipating effect of this development, which makes the methods of probability theory once more available for problems of inductive inference, cannot be adequately conveyed in a short discussion. Equally important is the pedagogical advantage; a student who has approached probability theory via Cox's theorems has at his fingertips the basis of all modern statistical practice. In the writer's Statistical Mechanics course, by the end of the second week the students are familiar with statistical principles such as maximum likelihood, confidence intervals, sequential analysis, etc., and the Gibbs canonical formalism has been derived as a general method of inductive reasoning, applicable to problems in or out of physics. We are thus able to devote much more time to the important applications of the theory than was formerly possible, while giving the students a glimpse of other statistical methods useful in physics.

E. T. JAYNES
Washington University