Engineering 206: Bayesian Statistics

Details of possible reading course next quarter

There is some interest in organizing a reading course next quarter which complements the ideas of this course. The plan would be to read as much as possible in one quarter of the great unfinished book by Ed Jaynes:


I attach a copy of the preface of the book and a biography of Jaynes, who died in 1998 before his masterwork could be finished.

The idea behind the book is to use the principle of maximum entropy as an overarching concept that can be seen to unify frequentist and Bayesian statistics, and the bottom-line goal is to regard probability theory as the extended logical basis of science (i.e., by using probability to fill in the gray scale between 0 (false) and 1 (true)). You'll have to decide for yourself if the book is a success or a noble failure; what is nearly certain is that you would learn a great deal by studying it.

Herbie Lee <herbie@ams.ucsc.edu>

, an Assistant Professor of Statistics in AMS and an excellent teacher, has said that he's willing to lead this reading course, and

Peter Towbin <ptowbin@soe.ucsc.edu>, <ptowbin@ucsc.edu>

has agreed to be the focal point for organizing the class, which would have a name like *Foundations of Statistics and Scientific Inference* and which would be listed as ENGR 297 (5 credits) (the class actually would appear on your transcript as Independent Study or Research but you could explain in an attachment what its real content was).

Please look over the attached materials and write to Peter Towbin (he asks that you send your email to both of his addresses) if you are interested in enrolling in the reading class, which will not take place unless at least $k$ people take part in it (where $k$ is a number like 6; this number can include auditors if they are serious about participating).
Probability Theory:
The Logic of Science

by
E. T. Jaynes
Wayman Crow Professor of Physics
Washington University
St. Louis, MO 63130, U. S. A.

Dedicated to the Memory of Sir Harold Jeffreys,
who saw the truth and preserved it.
PROBABILITY THEORY – THE LOGIC OF SCIENCE

Short Contents

PART A - PRINCIPLES AND ELEMENTARY APPLICATIONS
Chapter 1  Plausible Reasoning
Chapter 2  Quantitative Rules: The Cox Theorems
Chapter 3  Elementary Sampling Theory
Chapter 4  Elementary Hypothesis Testing
Chapter 5  Queer Uses for Probability Theory
Chapter 6  Elementary Parameter Estimation
Chapter 7  The Central Gaussian, or Normal, Distribution
Chapter 8  Sufficiency, Ancillarity, and All That
Chapter 9  Repetitive Experiments: Probability and Frequency
Chapter 10  Physics of “Random Experiments”
Chapter 11  The Entropy Principle
Chapter 12  Ignorance Priors – Transformation Groups
Chapter 13  Decision Theory: Historical Survey
Chapter 14  Simple Applications of Decision Theory
Chapter 15  Paradoxes of Probability Theory
Chapter 16  Orthodox Statistics: Historical Background
Chapter 17  Principles and Pathology of Orthodox Statistics
Chapter 18  The A_p–Distribution and Rule of Succession

PART B – ADVANCED APPLICATIONS
Chapter 19  Physical Measurements
Chapter 20  Regression and Linear Models
Chapter 21  Estimation with Cauchy and t–Distributions
Chapter 22  Time Series Analysis and Autoregressive Models
Chapter 23  Spectrum / Shape Analysis
Chapter 24  Model Comparison and Robustness
Chapter 25  Image Reconstruction
Chapter 26  Marginalization Theory
Chapter 27  Communication Theory
Chapter 28  Optimal Antenna and Filter Design
Chapter 29  Statistical Mechanics
Chapter 30  Conclusions

APPENDICES
Appendix A  Other Approaches to Probability Theory
Appendix B  Formalities and Mathematical Style
Appendix C  Convolutions and Cumulants
Appendix D  Dirichlet Integrals and Generating Functions
Appendix E  The Binomial – Gaussian Hierarchy of Distributions
Appendix F  Fourier Analysis
Appendix G  Infinite Series
Appendix H  Matrix Analysis and Computation
Appendix I  Computer Programs

REFERENCES
# PART A – PRINCIPLES and ELEMENTARY APPLICATIONS

## Chapter 1  PLAUSIBLE REASONING
- Deductive and Plausible Reasoning  101
- Analogies with Physical Theories  103
- The Thinking Computer  104
- Introducing the Robot  105
- Boolean Algebra  106
- Adequate Sets of Operations  108
- The Basic Desiderata  111

### COMMENTS
- Common Language vs. Formal Logic  114
- Nitpicking  116

## Chapter 2  THE QUANTITATIVE RULES
- The Product Rule  201
- The Sum Rule  206
- Qualitative Properties  210
- Numerical Values  212
- Notation and Finite Sets Policy  217

### COMMENTS
- “Subjective” vs. “Objective”  218
- Gödel’s Theorem  218
- Venn Diagrams  220
- The “Kolmogorov Axioms”  222

## Chapter 3  ELEMENTARY SAMPLING THEORY
- Sampling Without Replacement  301
- Logic Versus Propensity  308
- Reasoning from Less Precise Information  311
- Expectations  313
- Other Forms and Extensions  314
- Probability as a Mathematical Tool  315
- The Binomial Distribution  315
- Sampling With Replacement  318

### Digression: A Sermon on Reality vs. Models
- Correction for Correlations  320
- Simplification  325

### COMMENTS
- A Look Ahead  328
### Chapter 4  ELEMENTARY HYPOTHESIS TESTING
- Prior Probabilities 401
- Testing Binary Hypotheses with Binary Data 404
- Non-Extensibility Beyond the Binary Case 409
- Multiple Hypothesis Testing 411
- Continuous Probability Distributions (pdf’s) 418
- Testing an Infinite Number of Hypotheses 419
- Simple and Compound (or Composite) Hypotheses 424

**COMMENTS**
- Etymology 425
- What Have We Accomplished? 426

### Chapter 5  QUEER USES FOR PROBABILITY THEORY
- Extrasensory Perception 501
- Mrs. Stewart’s Telepathic Powers 502
- Converging and Diverging Views 507
- Visual Perception 511
- The Discovery of Neptune 512
- Digression on Alternative Hypotheses 514
- Horseracing and Weather Forecasting 517
- Paradoxes of Intuition 520
- Bayesian Jurisprudence 521

**COMMENTS** 522

### Chapter 6  ELEMENTARY PARAMETER ESTIMATION
- Inversion of the Urn Distributions 601
- Both N and R Unknown 601
- Uniform Prior 604
- Truncated Uniform Priors 608
- A Concave Prior 609
- The Binomial Monkey Prior 611
- Metamorphosis into Continuous Parameter Estimation 613
- Estimation with a Binomial Sampling Distribution 614
- Digression on Optional Stopping 616
- The Likelihood Principle 617
- Compound Estimation Problems 617
- A Simple Bayesian Estimate: Quantitative Prior Information 618
- From Posterior Distribution to Estimate 621
- Back to the Problem 624
- Effects of Qualitative Prior Information 626
- The Jeffreys Prior 629
- The Point of it All 630
- Interval Estimation 632
- Calculation of Variance 632
- Generalization and Asymptotic Forms 634
- A More Careful Asymptotic Derivation 635

**COMMENTS** 636
Chapter 7  THE CENTRAL GAUSSIAN, OR NORMAL DISTRIBUTION
The Gravitating Phenomenon 701
The Herschel–Maxwell Derivation 702
The Gauss Derivation 703
Historical Importance of Gauss’ Result 704
The Landon Derivation 705
Why the Ubiquitous Use of Gaussian Distributions? 707
Why the Ubiquitous Success? 709
The Near–Irrelevance of Sampling Distributions 711
The Remarkable Efficiency of Information Transfer 712
Nuisance Parameters as Safety Devices 713
More General Properties 714
Convolution of Gaussians 715
Galton’s Discovery 715
Population Dynamics and Darwinian Evolution 717
Resolution of Distributions into Gaussians 719
The Central Limit Theorem 722
Accuracy of Computations 723
COMMENTS 724
Terminology Again 724
The Great Inequality of Jupiter and Saturn 726

Chapter 8  SUFFICIENCY, ANCILLARITY, AND ALL THAT
Sufficiency 801
Fisher Sufficiency 803
Generalized Sufficiency 804
Examples 804
Sufficiency Plus Nuisance Parameters 805
The Pitman–Koopman Theorem 808
The Likelihood Principle 809
Effect of Nuisance Parameters 810
Use of Ancillary Information 811
Relation to the Likelihood Principle 816
Asymptotic Likelihood: Fisher Information 817
Combining Evidence from Different Sources: Meta–Analysis 819
Pooling the Data 822
Fine–Grained Propositions: Sam’s Broken Thermometer 822
COMMENTS 824
The Fallacy of Sample Re–use 824
A Folk–Theorem 825
Effect of Prior Information 825
Clever Tricks and Gamesmanship 826

Chapter 9  REPETITIVE EXPERIMENTS – PROBABILITY AND FREQUENCY
Physical Experiments 901
The Poorly Informed Robot 902
Induction 905
Partition Function Algorithms 907
Relation to Generating Functions 911
Another Way of Looking At It 912
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability and Frequency</td>
<td>913</td>
</tr>
<tr>
<td>Halley's Mortality Table</td>
<td>915</td>
</tr>
<tr>
<td>COMMENTS: The Irrationalists</td>
<td>918</td>
</tr>
<tr>
<td>Chapter 10 PHYSICS OF &quot;RANDOM EXPERIMENTS&quot;</td>
<td></td>
</tr>
<tr>
<td>An Interesting Correlation</td>
<td>1001</td>
</tr>
<tr>
<td>Historical Background</td>
<td>1002</td>
</tr>
<tr>
<td>How to Cheat at Coin and Die Tossing</td>
<td>1003</td>
</tr>
<tr>
<td>Experimental Evidence</td>
<td>1006</td>
</tr>
<tr>
<td>Bridge Hands</td>
<td>1007</td>
</tr>
<tr>
<td>General Random Experiments</td>
<td>1008</td>
</tr>
<tr>
<td>Induction Revisited</td>
<td>1010</td>
</tr>
<tr>
<td>But What About Quantum Theory?</td>
<td>1011</td>
</tr>
<tr>
<td>Mechanics Under the Clouds</td>
<td>1012</td>
</tr>
<tr>
<td>More on Coins and Symmetry</td>
<td>1013</td>
</tr>
<tr>
<td>Independence of Tosses</td>
<td>1017</td>
</tr>
<tr>
<td>The Arrogance of the Uninformed</td>
<td>1019</td>
</tr>
<tr>
<td>Chapter 11 DISCRETE PRIOR PROBABILITIES – THE ENTROPY PRINCIPLE</td>
<td></td>
</tr>
<tr>
<td>A New Kind of Prior Information</td>
<td>1101</td>
</tr>
<tr>
<td>Minimum $\sum p_i^2$</td>
<td>1103</td>
</tr>
<tr>
<td>Entropy: Shannon's Theorem</td>
<td>1104</td>
</tr>
<tr>
<td>The Wallis Derivation</td>
<td>1108</td>
</tr>
<tr>
<td>An Example</td>
<td>1110</td>
</tr>
<tr>
<td>Generalization: A More Rigorous Proof</td>
<td>1111</td>
</tr>
<tr>
<td>Formal Properties of Maximum Entropy Distributions</td>
<td>1113</td>
</tr>
<tr>
<td>Conceptual Problems: Frequency Correspondence</td>
<td>1120</td>
</tr>
<tr>
<td>COMMENTS</td>
<td>1124</td>
</tr>
<tr>
<td>Chapter 12 UNINFORMATIVE PRIORS – TRANSFORMATION GROUPS</td>
<td></td>
</tr>
<tr>
<td>Chapter 13 DECISION THEORY – HISTORICAL BACKGROUND</td>
<td></td>
</tr>
<tr>
<td>Inference vs. Decision</td>
<td>1301</td>
</tr>
<tr>
<td>Daniel Bernoulli’s Suggestion</td>
<td>1302</td>
</tr>
<tr>
<td>The Rationale of Insurance</td>
<td>1303</td>
</tr>
<tr>
<td>Entropy and Utility</td>
<td>1305</td>
</tr>
<tr>
<td>The Honest Weatherman</td>
<td>1305</td>
</tr>
<tr>
<td>Reactions to Daniel Bernoulli and Laplace</td>
<td>1306</td>
</tr>
<tr>
<td>Wald's Decision Theory</td>
<td>1307</td>
</tr>
<tr>
<td>Parameter Estimation for Minimum Loss</td>
<td>1310</td>
</tr>
<tr>
<td>Reformulation of the Problem</td>
<td>1312</td>
</tr>
<tr>
<td>Effect of Varying Loss Functions</td>
<td>1315</td>
</tr>
<tr>
<td>General Decision Theory</td>
<td>1316</td>
</tr>
<tr>
<td>COMMENTS</td>
<td>1317</td>
</tr>
<tr>
<td>&quot;Objectivity&quot; of Decision Theory</td>
<td>1317</td>
</tr>
<tr>
<td>Loss Functions in Human Society</td>
<td>1319</td>
</tr>
<tr>
<td>A New Look at the Jeffreys Prior</td>
<td>1320</td>
</tr>
<tr>
<td>Decision Theory is not Fundamental</td>
<td>1320</td>
</tr>
<tr>
<td>Another Dimension?</td>
<td>1321</td>
</tr>
<tr>
<td>Chapter 14 SIMPLE APPLICATIONS OF DECISION THEORY</td>
<td></td>
</tr>
<tr>
<td>Definitions and Preliminaries</td>
<td>1401</td>
</tr>
</tbody>
</table>
Sufficiency and Information 1403
Loss Functions and Criteria of Optimal Performance 1404
A Discrete Example 1406
How Would Our Robot Do It? 1410
Historical Remarks 1411
The Widget Problem 1412
Solution for Stage 2 1414
Solution for Stage 3 1416
Solution for Stage 4

Chapter 15  PARADOXES OF PROBABILITY THEORY
How Do Paradoxes Survive and Grow? 1501
Summing a Series the Easy Way 1502
Nonconglomerability 1503
Strong Inconsistency 1505
Finite vs. Countable Additivity 1511
The Borel–Kolmogorov Paradox 1513
The Marginalization Paradox 1516
How to Mass–produce Paradoxes 1517
COMMENTS
Counting Infinite Sets? 1520
The Hausdorff Sphere Paradox 1521

Chapter 16  ORTHODOX STATISTICS – HISTORICAL BACKGROUND
The Early Problems 1601
Sociology of Orthodox Statistics 1602
Ronald Fisher, Harold Jeffreys, and Jerzy Neyman 1603
Pre–data and Post–data Considerations 1608
The Sampling Distribution for an Estimator 1609
Pro–causal and Anti–Causal Bias 1611
What is Real; the Probability or the Phenomenon? 1613
COMMENTS 1613

Chapter 17  PRINCIPLES AND PATHOLOGY OF ORTHODOX STATISTICS
Unbiased Estimators
Confidence Intervals
Nuisance Parameters
Ancillary Statistics
Significance Tests
The Weather in Central Park
More Communication Difficulties
How Can This Be?
Probability Theory is Different
COMMENTS
Gamesmanship
What Does ‘Bayesian’ Mean?

Chapter 18  THE A_P–DISTRIBUTION AND RULE OF SUCCESSION
Memory Storage for Old Robots 1801
Relevance 1803
A Surprising Consequence 1804
An Application 1806
PART B - ADVANCED APPLICATIONS

Chapter 19 PHYSICAL MEASUREMENTS
  Reduction of Equations of Condition 1901
  Reformulation as a Decision Problem 1903
  Sermon on Gaussian Error Distributions 1904
  The Underdetermined Case: K is Singular 1906
  The Overdetermined Case: K Can be Made Nonsingular 1906
  Numerical Evaluation of the Result 1907
  Accuracy of the Estimates 1909
  COMMENTS: a Paradox 1910

Chapter 20 REGRESSION AND LINEAR MODELS

Chapter 21 ESTIMATION WITH CAUCHY AND t-DISTRIBUTIONS

Chapter 22 TIME SERIES ANALYSIS AND AUTOREGRESSIVE MODELS

Chapter 23 SPECTRUM / SHAPE ANALYSIS

Chapter 24 MODEL COMPARISON AND ROBUSTNESS
  The Bayesian Basis of it All 2401
  The Occam Factors 2402

Chapter 25 MARGINALIZATION THEORY

Chapter 26 IMAGE RECONSTRUCTION

Chapter 27 COMMUNICATION THEORY
  Origins of the Theory 2701
  The Noiseless Channel 2702
  The Information Source 2706
  Does the English Language Have Statistical Properties? 2708
  Optimum Encoding: Letter Frequencies Known 2709
  Better Encoding from Knowledge of Digram Frequencies 2712
  Relation to a Stochastic Model 2715
  The Noisy Channel 2718
  Fixing a Noisy Channel: the Checksum Algorithm 2718

Chapter 28 OPTIMAL ANTENNA AND FILTER DESIGN

Chapter 29 STATISTICAL MECHANICS

Chapter 30 CONCLUSIONS

APPENDICES

Appendix A Other Approaches to Probability Theory
  The Kolmogorov System of Probability A 1
  The de Finetti System of Probability A 5
  Comparative Probability A 6
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holdouts Against Comparability</td>
<td>A 7</td>
</tr>
<tr>
<td>Speculations About Lattice Theories</td>
<td>A 8</td>
</tr>
<tr>
<td>Appendix B Formalities and Mathematical Style</td>
<td></td>
</tr>
<tr>
<td>Notation and Logical Hierarchy</td>
<td>B 1</td>
</tr>
<tr>
<td>Our &quot;Cautious Approach&quot; Policy</td>
<td>B 3</td>
</tr>
<tr>
<td>Willy Feller on Measure Theory</td>
<td>B 3</td>
</tr>
<tr>
<td>Kronecker vs. Weierstrasz</td>
<td>B 5</td>
</tr>
<tr>
<td>What is a Legitimate Mathematical Function?</td>
<td>B 6</td>
</tr>
<tr>
<td>Nondifferentiable Functions</td>
<td>B 8</td>
</tr>
<tr>
<td>What am I Supposed to Publish?</td>
<td>B 10</td>
</tr>
<tr>
<td>Mathematical Courtesy</td>
<td>B 11</td>
</tr>
<tr>
<td>Appendix C Convolutions and Cumulants</td>
<td></td>
</tr>
<tr>
<td>Relation of Cumulants and Moments</td>
<td>C 4</td>
</tr>
<tr>
<td>Examples</td>
<td></td>
</tr>
<tr>
<td>Appendix D Dirichlet Integrals and Generating Functions</td>
<td>C 5</td>
</tr>
<tr>
<td>Appendix E The Binomial – Gaussian Hierarchy of Distributions</td>
<td></td>
</tr>
<tr>
<td>Appendix F Fourier Theory</td>
<td></td>
</tr>
<tr>
<td>Appendix G Infinite Series</td>
<td></td>
</tr>
<tr>
<td>Appendix H Matrix Analysis and Computation</td>
<td></td>
</tr>
<tr>
<td>Appendix I Computer Programs</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
</tr>
<tr>
<td>NAME INDEX</td>
<td></td>
</tr>
<tr>
<td>SUBJECT INDEX</td>
<td></td>
</tr>
</tbody>
</table>
PREFACE

The following material is addressed to readers who are already familiar with applied mathematics at the advanced undergraduate level or preferably higher; and with some field, such as physics, chemistry, biology, geology, medicine, economics, sociology, engineering, operations research, etc., where inference is needed.\(^1\) A previous acquaintance with probability and statistics is not necessary; indeed, a certain amount of innocence in this area may be desirable, because there will be less to unlearn.

We are concerned with probability theory and all of its conventional mathematics, but now viewed in a wider context than that of the standard textbooks. Every Chapter after the first has "new" (i.e., not previously published) results that we think will be found interesting and useful. Many of our applications lie outside the scope of conventional probability theory as currently taught. But we think that the results will speak for themselves, and that something like the theory expounded here will become the conventional probability theory of the future.

History: The present form of this work is the result of an evolutionary growth over many years. My interest in probability theory was stimulated first by reading the work of Harold Jeffreys (1939) and realizing that his viewpoint makes all the problems of theoretical physics appear in a very different light. But then in quick succession discovery of the work of R. T. Cox (1946), C. E. Shannon (1948) and G. Pólya (1954) opened up new worlds of thought, whose exploration has occupied my mind for some forty years. In this much larger and permanent world of rational thinking in general, the current problems of theoretical physics appeared as only details of temporary interest.

The actual writing started as notes for a series of five lectures given at Stanford University in 1956, expounding the then new and exciting work of George Pólya on "Mathematics and Plausible Reasoning". He dissected our intuitive "common sense" into a set of elementary qualitative desiderata and showed that mathematicians had been using them all along to guide the early stages of discovery, which necessarily precede the finding of a rigorous proof. The results were much like those of James Bernoulli's "Art of Conjecture" (1713), developed analytically by Laplace in the late 18th Century; but Pólya thought the resemblance to be only qualitative.

However, Pólya demonstrated this qualitative agreement in such complete, exhaustive detail as to suggest that there must be more to it. Fortunately, the consistency theorems of R. T. Cox were enough to clinch matters; when one added Pólya's qualitative conditions to them the result was a proof that, if degrees of plausibility are represented by real numbers, then there is a uniquely determined set of quantitative rules for conducting inference. That is, any other rules which conflict with them will necessarily violate an elementary desideratum of rationality or consistency.

But the final result was just the standard rules of probability theory, given already by Bernoulli and Laplace; so why all the fuss? The important new feature was that these rules were now seen as uniquely valid principles of logic in general, making no reference to "chance" or "random variables"; so their range of application is vastly greater than had been supposed in the conventional probability theory that was developed in the early twentieth Century. As a result, the imaginary distinction between "probability theory" and "statistical inference" disappears, and the field achieves not only logical unity and simplicity, but far greater technical power and flexibility in applications.

In the writer's lectures, the emphasis was therefore on the quantitative formulation of Pólya's viewpoint, so it could be used for general problems of scientific inference, almost all of which arise out of incomplete information rather than "randomness". Some personal reminiscences about Pólya and this start of the work are in Chapter 5.

\(^1\) By "inference" we mean simply: deductive reasoning whenever enough information is at hand to permit it; inductive or plausible reasoning when — as is almost invariably the case in real problems — the necessary information is not available. Thus our topic is the optimal processing of incomplete information.
But once the development of applications started, the work of Harold Jeffreys, who had seen so much of it intuitively and seemed to anticipate every problem I would encounter, became again the central focus of attention. My debt to him is only partially indicated by the dedication of this book to his memory. Further comments about his work and its influence on mine are scattered about in several Chapters.

In the years 1957–1970 the lectures were repeated, with steadily increasing content, at many other Universities and research laboratories.¹ In this growth it became clear gradually that the outstanding difficulties of conventional “statistical inference” are easily understood and overcome. But the rules which now took their place were quite subtle conceptually, and it required some deep thinking to see how to apply them correctly. Past difficulties which had led to rejection of Laplace’s work, were seen finally as only misapplications, arising usually from failure to define the problem unambiguously or to appreciate the cogency of seemingly trivial side information, and easy to correct once this is recognized. The various relations between our “extended logic” approach and the usual “random variable” one appear in almost every Chapter, in many different forms.

Eventually, the material grew to far more than could be presented in a short series of lectures, and the work evolved out of the pedagogical phase; with the clearing up of old difficulties accomplished, we found ourselves in possession of a powerful tool for dealing with new problems. Since about 1970 the accretion has continued at the same pace, but fed instead by the research activity of the writer and his colleagues. We hope that the final result has retained enough of its hybrid origins to be usable either as a textbook or as a reference work.

In view of the above, we repeat the sentence that Charles Darwin wrote in the Introduction to his Origin of Species: “I hope that I may be excused for entering on these personal details, as I give them to show that I have not been hasty in coming to a decision.” But it might be thought that work done thirty years ago would be obsolete today. Fortunately, the work of Jeffreys, Pólya and Cox was of a fundamental, timeless character whose truth does not change and whose importance grows with time. Their perception about the nature of inference, which was merely curious thirty years ago, is very important in a half–dozen different areas of science today; and it will be crucially important in all areas 100 years hence.

**Foundations:** From thirty years of experience with its applications in hundreds of real problems, our views on the foundations of probability theory have evolved into something quite complex, which cannot be described in any such simplistic terms as “pro–this” or “anti–that.” For example our system of probability could hardly, in style, philosophy, and purpose, be more different from that of Kolmogorov. What we consider to be fully half of probability theory as it is needed in current applications – the principles for assigning probabilities by logical analysis of incomplete information – is not present at all in the Kolmogorov system.

Yet when all is said and done we find ourselves, to our own surprise, in agreement with Kolmogorov and in disagreement with his critics, on nearly all technical issues. As noted in Appendix A, each of his axioms turns out to be, for all practical purposes, derivable from the Pólya–Cox desiderata of rationality and consistency. In short, we regard our system of probability as not contradicting Kolmogorov’s; but rather seeking a deeper logical foundation that permits its extension in the directions that are needed for modern applications. In this endeavor, many problems have been solved, and those still unsolved appear where we should naturally expect them: in breaking into new ground.

As another example, it appears at first glance to everyone that we are in very close agreement with the de Finetti system of probability. Indeed, the writer believed this for some time. Yet when all is said and done we find, to our own surprise, that little more than a loose philosophical

¹ Some of the material in the early Chapters was issued in 1958 by the Socony–Mobil Oil Company as Number 4 in their series “Colloquium Lectures in Pure and Applied Science”.
agreement remains; on many technical issues we disagree strongly with de Finetti. It appears to us that his way of treating infinite sets has opened up a Pandora’s box of useless and unnecessary paradoxes; nonconglomerability and finite additivity are examples discussed in Chapter 15.

Infinite set paradoxology has become a morbid infection that is today spreading in a way that threatens the very life of probability theory, and requires immediate surgical removal. In our system, after this surgery, such paradoxes are avoided automatically; they cannot arise from correct application of our basic rules, because those rules admit only finite sets and infinite sets that arise as well-defined and well-behaved limits of finite sets. The paradoxing was caused by (1) jumping directly into an infinite set without specifying any limiting process to define its properties; and then (2) asking questions whose answers depend on how the limit was approached.

For example, the question: “What is the probability that an integer is even?” can have any answer we please in (0, 1), depending on what limiting process is to define the “set of all integers” (just as a conditionally convergent series can be made to converge to any number we please, depending on the order in which we arrange the terms).

In our view, an infinite set cannot be said to possess any “existence” and mathematical properties at all – at least, in probability theory – until we have specified the limiting process that is to generate it from a finite set. In other words, we sail under the banner of Gauss, Kronecker, and Poincaré rather than Cantor, Hilbert, and Bourbaki. We hope that readers who are shocked by this will study the indictment of Bourbakism by the mathematician Morris Kline (1980), and then bear with us long enough to see the advantages of our approach. Examples appear in almost every Chapter.

Comparisons: For many years there has been controversy over “frequentist” versus “Bayesian” methods of inference, in which the writer has been an outspoken partisan on the Bayesian side. The record of this up to 1981 is given in an earlier book (Jaynes, 1983). In these old works there was a strong tendency, on both sides, to argue on the level of philosophy or ideology. We can now hold ourselves somewhat aloof from this because, thanks to recent work, there is no longer any need to appeal to such arguments. We are now in possession of proven theorems and masses of worked-out numerical examples demonstrating the facts of actual performance. As a result, the superiority of Bayesian methods is now a thoroughly demonstrated fact in a hundred different areas. We point this out in some detail whenever it makes a substantial difference in the final results. Thus we continue to argue vigorously for the Bayesian methods; but we ask the reader to note that our arguments now proceed by citing facts rather than proclaiming a philosophical or ideological position.

However, neither the Bayesian nor the frequentist approach is universally applicable, so in the present more general work we take a broader view of things. Our theme is simply: Probability Theory as Extended Logic. The “new” perception amounts to the recognition that the mathematical rules of probability theory are not merely rules for calculating frequencies of “random variables”; they are also the unique consistent rules for conducting inference (i.e. plausible reasoning) of any kind, and we shall apply them in full generality to that end.

It is true that all “Bayesian” calculations are included automatically as particular cases of our rules; but so are all “frequentist” calculations. Nevertheless, our basic rules are broader than either of these, and in many applications our calculations do not fit into either category.

To explain the situation as we see it presently: The traditional “frequentist” methods which use only sampling distributions are usable and useful in many particularly simple, idealized problems; but they represent the most proscribed special cases of probability theory, because they presuppose conditions (independent repetitions of a “random experiment” but no relevant prior information) that are hardly ever met in real problems. This approach is quite inadequate for the current needs of science.
In addition, frequentist methods provide no technical means to eliminate nuisance parameters or to take prior information into account, no way even to use all the information in the data when sufficient or ancillary statistics do not exist. Lacking the necessary theoretical principles, they force one to "choose a statistic" on intuitive grounds rather than from probability theory, and then to invent arbitrary \textit{ad hoc} devices (such as unbiased estimators, confidence intervals, power functions, tail-area significance tests) not contained in the rules of probability theory. But, as Cox's theorems guarantee, such devices always generate inconsistencies or absurd results when applied far enough; we shall see dozens of examples.

Bayesian methods represent a great generalization, and are adequate to deal with what we might call "well–developed" problems of inference. As Harold Jeffreys demonstrated, they have a superb analytical apparatus, able to deal effortlessly with the technical problems on which frequentist methods fail. Therefore they enable us to solve problems of far greater complexity than can be discussed at all in frequentist terms. One of our main purposes is to show how all this capability was contained already in the simple product and sum rules of probability theory interpreted as extended logic, with no need for -- indeed, no room for -- any \textit{ad hoc} devices.

But before Bayesian methods can be used, a problem must be developed beyond John Tukey's "exploratory phase" to the point where it has enough structure to determine all the needed apparatus (a model, sample space, hypothesis space, prior probabilities, sampling distribution). Almost all scientific problems pass through an initial exploratory phase in which we have need for inference, but the frequentist assumptions are invalid and the Bayesian apparatus is not yet available. Indeed, some of them never evolve out of the exploratory phase. Problems at this level call for more primitive means of assigning probabilities directly out of our incomplete information.

For this purpose, the Principle of Maximum Entropy has at present the clearest theoretical justification and is the most highly developed computationally, with an analytical apparatus as powerful and versatile as the Bayesian one. To apply it we must define a sample space, but do not need any model or sampling distribution. In effect, entropy maximization creates a model for us out of our data, which proves to be optimal by so many different criteria\footnote{These concern efficient information handling; for example, (1) The model created is the simplest one that captures all the information in the constraints (Chapter 11); (2) It is the unique model for which the constraints would have been sufficient statistics (Chapter 8); (3) If viewed as constructing a sampling distribution for subsequent Bayesian inference from new data \(D\), the only property of the measurement errors in \(D\) that are used in that subsequent inference are the ones about which that sampling distribution contained some definite prior information (Chapter 7). Thus the formalism automatically takes into account all the information we have, but avoids assuming information that we do not have. This contrasts sharply with orthodox methods, where one does not think in terms of information at all, and in general violates both of these desiderata.} that it is hard to imagine circumstances where one would not want to use it in a problem where we have a sample space but no model.

Bayesian and maximum entropy methods differ in another respect. Both procedures yield the optimal inferences from the information that went into them, but we may choose a model for Bayesian analysis; this amounts to expressing some prior knowledge – or some working hypothesis – about the phenomenon being observed. Usually such hypotheses extend beyond what is directly observable in the data, and in that sense we might say that Bayesian methods are – or at least may be – speculative. If the extra hypotheses are true, then we expect that they will improve the Bayesian results; if they are false, the inferences will likely be worse.

On the other hand, maximum entropy is a nonspeculative procedure, in the sense that it invokes no hypotheses beyond the sample space and the evidence that is in the available data. Thus it predicts only observable facts (functions of future or past observations) rather than values of parameters which may exist only in our imagination. It is just for that reason that maximum
entropy is the appropriate (safest) tool when we have very little knowledge beyond the raw data; it protects us against drawing conclusions not warranted by the data. But when the information is extremely vague it may be difficult to define any appropriate sample space, and one may wonder whether more primitive principles than Maximum Entropy can be found. There is room for much new creative thought here.

For the present, there are many important and highly nontrivial applications where Maximum Entropy is the only tool we need. The planned second volume of this work is to consider them in detail; usually, they require more technical knowledge of the subject–matter area than do the more general applications studied in this volume. All of presently known statistical mechanics, for example, is included in this, as are the highly successful maximum entropy spectrum analysis and image reconstruction algorithms in current use. However, we think that in the future the latter two applications will evolve on into the Bayesian phase, as we become more aware of the appropriate models and hypothesis spaces.

Mental Activity: As one would expect already from Pólya’s examples, probability theory as extended logic reproduces many aspects of human mental activity, sometimes in surprising and even disturbing detail. In Chapter 5 we find our equations exhibiting the phenomenon of a person who tells the truth and is not believed, even though the disbelievers are reasoning consistently. The theory explains why and under what circumstances this will happen.

The equations also reproduce a more complicated phenomenon, divergence of opinions. One might expect that open discussion of public issues would tend to bring about a general consensus. On the contrary, we observe repeatedly that when some controversial issue has been discussed vigorously for a few years, society becomes polarized into two opposite extreme camps; it is almost impossible to find anyone who retains a moderate view. Probability theory as logic shows how two persons, given the same information, may have their opinions driven in opposite directions by it, and what must be done to avoid this.

In such respects, probability theory is undoubtedly telling us something about the way our own minds operate when we form intuitive judgments, of which we may not have been consciously aware. Some may feel uncomfortable at these revelations; others may see in them useful tools for psychological, sociological, or legal research.

What is ‘safe’? We are not concerned here only with abstract issues of mathematics and logic. One of the main practical messages of this work is the great effect of prior information on the conclusions that one should draw from a given data set. Currently much discussed issues such as environmental hazards or the toxicity of a food additive, cannot be judged rationally if one looks only at the current data and ignores our prior information about the phenomenon. As we demonstrate, this can lead us to greatly overestimate or underestimate the danger.

A common error is to assume a linear response model without threshold when judging the effects of radioactivity or the toxicity of some substance. Presumably there is no threshold effect for cumulative poisons like heavy metal ions (mercury, lead), which are eliminated only very slowly if at all. But for virtually every organic substance (such as saccharin or cyclamates), the existence of a finite metabolic rate means that there must exist a finite threshold dose rate, below which the substance is decomposed or eliminated so rapidly that it has no ill effects. If this were not true, the human race could never have survived to the present time, in view of all the things we have been eating.

Indeed, every mouthful of food you and I have ever taken contained many billions of kinds of complex molecules whose structure and physiological effects have never been determined – and many millions of which would be toxic or fatal in large doses. We cannot doubt that we are daily ingesting thousands of substances that are far more dangerous than saccharin – but in amounts that are safe, because they are far below the various thresholds. But at the present time there is hardly any substance except common drugs, for which we actually know the threshold.
Therefore, the goal of inference in this field should be to estimate not only the slope of the response curve, but far more importantly, to decide whether there is evidence for a threshold; and if so, to estimate its magnitude (the “maximum safe dose”). For example, to tell us that a sugar substitute is dangerous in doses a thousand times greater than would ever be encountered in practice, is hardly an argument against using the substitute; indeed, the fact that it is necessary to go to kilodoses in order to detect any ill effects at all, is rather conclusive evidence, not of the danger, but of the safety, of a tested substance, and probability theory confirms this whenever it is allowed to do so (that is, whenever we use a model that is flexible enough to admit the possibility of a threshold). A similar overdose of sugar would be far more dangerous, leading not to barely detectable harmful effects, but to sure, immediate death by diabetic coma; yet nobody has proposed to ban the use of sugar in food.

Kilodore effects are irrelevant because we do not take kilodoses; in the case of a sugar substitute the important question is: What are the threshold doses for toxicity of a sugar substitute and for sugar, compared to the normal doses? If that of a sugar substitute is higher, then the rational conclusion would be that the substitute is actually safer than sugar, as a food ingredient. To analyze one’s data in terms of a model which does not allow even the possibility of a threshold effect, is to prejudice the issue in a way that can lead to false conclusions from any amount of data.

We emphasize this in the Preface because false conclusions of just this kind are now not only causing major economic waste, but also creating unnecessary dangers to public health. Society has only finite resources to deal with such problems, so any effort expended on imaginary dangers means that real dangers are going unattended. Use of models which correctly represent the prior information that scientists have about the mechanism at work (such as metabolic rate, chemical reactivity) can prevent such folly in the future.

**Style of Presentation**: In part A, expounding principles and elementary applications, most Chapters start with several pages of verbal discussion of the nature of the problem. Here we try to explain the constructive ways of looking at it, and the logical pitfalls responsible for past errors. Only then do we turn to the mathematics, solving a few of the problems of the genre. In part B, expounding more advanced applications, we can concentrate more on the mathematical details.

The writer has learned from much experience that this primary emphasis on the logic of the problem, rather than the mathematics, is necessary in the early stages. For modern students, the mathematics is the easy part; once a problem has been reduced to a definite mathematical exercise, most students can solve it effortlessly and extend it endlessly, without further help from any book or teacher. It is in the conceptual matters (how to make the initial connection between the real-world problem and the abstract mathematics) that they are perplexed and unsure how to proceed.

Recent history demonstrates that anyone foolhardy enough to describe his own work as “rigorous” is headed for a fall. Therefore, we shall claim only that we do not knowingly give erroneous arguments. We are conscious also of writing for a large and varied audience, for most of whom clarity of meaning is more important than “rigor” in the narrow mathematical sense.

There are two more, even stronger reasons for placing our primary emphasis on logic and clarity. Firstly, no argument is stronger than the premises that go into it, and as Harold Jeffreys noted, those who lay the greatest stress on mathematical rigor are just the ones who, lacking a sure sense of the real world, tie their arguments to unrealistic premises and thus destroy their relevance. Jeffreys likened this to trying to strengthen a building by anchoring steel beams into plaster. An argument which makes it clear intuitively why a result is correct, is actually more trustworthy and more likely of a permanent place in science, than is one that makes a great overt show of mathematical rigor unaccompanied by understanding.

Secondly, we have to recognize that there are no really trustworthy standards of rigor in a mathematics that has embraced the theory of infinite sets. Morris Kline (1980, p. 351) came close to the Jeffreys simile: “Should one design a bridge using theory involving infinite sets or the axiom
of choice? Might not the bridge collapse?” The only real rigor we have today is in the operations of elementary arithmetic on finite sets of finite integers, and our own bridge will be safest from collapse if we keep this in mind.

Of course, it is essential that we follow this “finite sets” policy whenever it matters for our results; but we do not propose to become fanatical about it. In particular, the arts of computation and approximation are on a different level than that of basic principle; and so once a result is derived from strict application of the rules, we allow ourselves to use any convenient analytical methods for evaluation or approximation (such as replacing a sum by an integral) without feeling obliged to show how to generate an uncountable set as the limit of a finite one.

But we impose on ourselves a far stricter adherence to the mathematical rules of probability theory than was ever exhibited in the “orthodox” statistical literature, in which authors repeatedly invoke the aforementioned intuitive ad hoc devices to do, arbitrarily and imperfectly, what the rules of probability theory as logic would have done for them uniquely and optimally. It is just this strict adherence that enables us to avoid the artificial paradoxes and contradictions of orthodox statistics, as described in Chapter 17.

Equally important, this policy often simplifies the computations in two ways: (A) The problem of determining the sampling distribution of a “statistic” is eliminated; the evidence of the data is displayed fully in the likelihood function, which can be written down immediately. (B) One can eliminate nuisance parameters at the beginning of a calculation, thus reducing the dimensionality of a search algorithm. This can mean orders of magnitude reduction in computation over what would be needed with a least squares or maximum likelihood algorithm. The Bayesian computer programs of Bretthorst (1988) demonstrate these advantages impressively, leading in some cases to major improvements in the ability to extract information from data, over previously used methods.

A scientist who has learned how to use probability theory directly as extended logic, has a great advantage in power and versatility over one who has learned only a collection of unrelated ad-hoc devices. As the complexity of our problems increases, so does this relative advantage. Therefore we think that in the future, workers in all the quantitative sciences will be obliged, as a matter of practical necessity, to use probability theory in the manner expounded here. This trend is already well under way in several fields, ranging from econometrics to astronomy to magnetic resonance spectroscopy.

Finally, some readers should be warned not to look for hidden subtleties of meaning which are not present. We shall, of course, explain and use all the standard technical jargon of probability and statistics – because that is our topic. But although our concern with the nature of logical inference leads us to discuss many of the same issues, our language differs greatly from the stilted jargon of logicians and philosophers. There are no linguistic tricks and there is no “meta-language” gobbledygook; only plain English. We think that this will convey our message clearly enough to anyone who seriously wants to understand it. In any event, we feel sure that no further clarity would be achieved by taking the first few steps down that infinite regress that starts with: “What do you mean by ‘exists’?”

Acknowledgments: In addition to the inspiration received from the writings of Jeffreys, Cox, Pólya, and Shannon, I have profited by interaction with some 300 former students, who have diligently caught my errors and forced me to think more carefully about many issues. Also, over the years my thinking has been influenced by discussions with many colleagues; to list a few (in the reverse alphabetical order preferred by some): Arnold Zellner, George Uhlenbeck, John Tukey, William Sudderth, Stephen Stigler, John Skilling, Jimmie Savage, Carlos Rodriguez, Lincoln Moses, Elliott Montroll, Paul Meier, Dennis Lindley, David Lane, Mark Kac, Harold Jeffreys, Bruce Hill, Stephen Gull, Jack Good, Seymour Geisser, Willy Feller, Anthony Edwards, Morrie de Groot, Phil Dawid, Jerome Cornfield, John Parker Burg, David Blackwell, and George Barnard. While I have not agreed with all of the great variety of things they told me, it has all been taken into account in
one way or another in the following pages. Even when we ended in disagreement on some issue, I believe that our frank private discussions have enabled me to avoid misrepresenting their positions, while clarifying my own thinking; I thank them for their patience.

E. T. Jaynes
August 1993
Edwin Thompson Jaynes
July 5, 1922 - April 30, 1998

By: G. Larry Brethorst
Dept. of Chemistry
Washington University
St. Louis MO 63130

On July 5th, 1922 Edwin Thompson Jaynes Jr. was born in Waterloo, Iowa to Ethyl and Edwin Jaynes. At the time the Jaynes' lived near Cedar Falls, Iowa. Ed had two sisters, and an older brother. His father, a surgeon, practiced in Waterloo. By 1922 standards his family was fairly well off. However, his father died when Ed was very young, leaving his mother to care for four small children. Consequently, the Jaynes family moved to Parkersburg, Iowa where Ethyl's parents lived.

Ed grew up around Parkersburg. His first word was "kitty" which he uttered when he was seven months old. He took his first step at 13 months. He called his mother "Ma." His family did have some fine furniture, linen, silverware, a Knabe grand piano, a classical music collection, and a large library, including the Harvard Classics, housed in elegant glass-front mahogany bookcases - things left over from when his father was alive - but no money. Their only source of income was from renting half of the house. They had to grow most of their own food. Ed Jaynes has noted that his grandfather had to help them out financially many times [8].

The presence of the piano and library must have exerted a great influence on Ed, because he considered a career as a concert pianist. At his death he had a Bosendorfer grand piano in his home valued at approximately $100,000 dollars and hundreds of tapes of himself playing compositions from various classical composers. His library consisted of more than a thousand books, including statistics, physics, music, chemistry, biology, history, and philosophy.

Ed attended both grade, school and high, school in Parkersburg. However, Ed left Parkersburg in 1938 to enter Cornell College in Mount Vernon, Iowa. He returned to Parkersburg occasionally to visit his mother, and grandmother. Interestingly, his mother kept his grades from Cornell College. In his first semester he received a B in English, an A in mathematics, a B in German, and a B in chemistry. By the second semester he was doing better, receiving all A's except for a B in mathematics. By the end of his second year his grades were straight A's. That continued for the third and fourth year, with one exception. He received a C in philosophy. He finished his undergraduate schooling in 1942, receiving a B.A. in physics.

During the time he was attending Cornell College he supported himself by working and by loans from the Louise Foote Foundation. He noted with pride that these loans were repaid on time and that "my entire college education cost my mother a total of $25; one winter she bought me an overcoat" [8].

Between his junior and senior, year, he worked at the Warner Institute with Drs. Gustav Martin and Marvin R. Thompson. After receiving his B. A. he intended to return to the Warner Institute for
another summer, but his plans changed because of the war. From 1942 to 1944 he worked for the Sperry Gyroscope Company on Long Island helping to develop Doppler radar.

At the end of 1944, he became Ensign Jaynes, and worked at the Anacostia Naval Research Lab in Washington D. C. developing microwave systems. During his stay in the Navy he spent some time on Guam. There is one picture of him on Guam holding a machine gun. When he was discharged in 1946, he was a lieutenant (j.g.). There are two documents written by Ensign Jaynes: the first is a series of 9 lectures on solving circuit problems using Laplace and Fourier transforms [1]; the second, is titled “Theory of Microwave Coupling Systems” [2]. These two documents constitute the earliest known professional writings of Ed Jaynes.

Jaynes left the Navy in 1946 and headed for California. In the summer of 1946 he worked in the W. W. Hansen Laboratories of Physics at Stanford on the design of the first linear electron accelerator. At the end of the summer, he enrolled at the University of California at Berkeley. In “Disturbing The Memory” [7] Jaynes notes

I first met Julian Schwinger, Robert Dicke, and Donald Hamilton during the War when we were all engaged in developing microwave theory, measurement techniques, and applications to pulsed and Doppler radar; Schwinger and Dicke at the MIT Radiation Laboratory, Hamilton and I at the Sperry Gyroscope Laboratories on Long Island. Bill Hansen (for whom the W. W. Hansen Laboratories at Stanford are now named) was running back and forth weekly, giving lectures at MIT and bringing us back the notes on the Schwinger lectures as they appeared, and I accompanied him on a few of those trips.

I first met Edward Teller when he visited Stanford in the Summer of 1946 and Hansen, Teller, and I discussed the design of the first Stanford LINAC, then underway. After some months of correspondence I first met J. R. Oppenheimer in September 1946, when I arrived at Berkeley as a beginning graduate student, to learn quantum theory from him -- the result of Bill Hansen having recommended us strongly to each other. When in the Summer of 1947 Oppy moved to Princeton to take over the Institute for Advanced Study, I was one of four students that he took along. The plan was that we would enroll as graduate students at Princeton University, finish our theses under Oppy although he was not officially a Princeton University faculty member; and turn them in to Princeton (which had agreed to this somewhat unusual arrangement in view of the somewhat unusual circumstances). My thesis was to be on Quantum Electrodynamics.

But, as this writer learned from attending a year of Oppy’s lectures (1946-47) at Berkeley, and eagerly studying his printed and spoken words for several years thereafter, Oppy would never countenance any retreat from the Copenhagen position, of the kind advocated by Schrödinger and Einstein. He derived some great emotional satisfaction from just those elements of mysticism that Schrödinger and Einstein had deplored, and always wanted to make the world still more mystical, and less rational.

This desire was expressed strongly in his 1955 BBC Reith lectures (of which I still have some cherished tape recordings which recall his style of delivery at its best). Some have seen this as a fine humanist trait. I saw it increasingly as an anomaly - a basically anti-scientific attitude in a person posing as a scientist - that explains so much of the contradictions in his character.

As a more practical matter, it presented me with a problem in carrying out my plan to write a thesis under Oppy’s supervision, quite aside from the fact that his travel and other activities made it so hard to see him. Mathematically, the Feynman electromagnetic propagator made no use of those superfluous degrees of freedom; it was
equally well a Green’s function for an unquantized EM field. So I wanted to reformulate electrodynamics from the ground up without using field quantization. The physical picture would be very different; but since the successful Feynman rules used so little of that physical picture anyway, I did not think that the physical predictions would be appreciably different; at least, if the idea was wrong, I wanted to understand in detail why it was wrong.

If this meant standing in contradiction with the Copenhagen interpretation, so be it; I would be delighted to see it gone anyway, for the same reason that Einstein and Schrödinger would. But I sensed that Oppy would never tolerate a grain of this; he would crush me like an eggshell if I dared to express a word of such subversive ideas. I could do a thesis with Oppy only if it was his thesis, not mine.


His dissertation was a calculation of the electrical and magnetic properties of ferroelectric materials. Ferroelectric materials are crystalline substances which have a permanent electric polarization (an electric dipole moment per unit volume) that can be reversed by an electric field. His dissertation “Ferroelectricity” was finished in 1950 and he received his Ph.D. in physics. He published his first paper in 1950 while still at Princeton. It was titled “The Displacement of Oxygen in $\text{BaTiO}_3$” [1]. This paper is essentially a one page summary of some of his thesis results. The paper is so short that it does not begin to hint at the amount of work and original thought that went into his thesis calculations. Joel Snow, one of Jaynes’ early graduate students, refers to “Ferroelectricity” as a “tour de force”. [4] Jaynes’ thesis was extensively modified and later published by the Princeton University Press in 1953 [3] in the series Investigations In Physics. It was vol. I of this prestigious series. Investigations In Physics has featured many famous authors, including John Von Neumann, Eugene Wigner, and Eugene Feenberg.

After finishing his degree, Jaynes returned to Stanford in 1950. He stayed through 1960. In his early work Jaynes was both theoretician and experimentalist. For example, his fourth paper was on the observation of a paramagnetic resonance in a single crystal of barium titanate [4], essentially an experimental paper. His second paper, on the concept and measurement of the impedance in periodically loaded wave guides [2], had both theoretical and experimental aspects. Jaynes continued to maintain an active research laboratory well into the 70’s (many years after moving to St. Louis). Indeed, among his papers there was a copy of the Sunday magazine supplement of the St. Louis Globe-Democrat dated May 30, 1967 containing a photograph, of Ed Jaynes in his laboratory in Crow Hall on the Washington University campus. The photograph shows him working with a high energy laser. During this period his students were testing some of the predictions of neoclassical and quantum theory.

While he was an Associate Professor at Stanford he also supported himself consulting with Varian Associates, the U. S. Army Corp of Engineers, and the University of California at Livermore. While consulting he wrote a number of reports for both Varian and the U. S. Army. Many of the U. S. Army reports still survive, but are not available for general release; a condition that will change shortly. Two of the reports done for Varian still survive, but are only available from the main Varian corporate library. Varian, at that time, was a young upstart company that could not afford to pay Jaynes in cash, so they paid him in stock. Additionally, Jaynes’ records indicate that he continued to buy Varian stock throughout most of this period. At the time of his death this stock constituted about one fourth of Jaynes’ total wealth.

Prior to 1957 Jaynes published a total of 6 articles [1, 2-8]. These articles essentially grew out of his thesis work. However, his interest were varied as illustrated by the fact that these 6 papers covered such diverse research areas as solid state, classical electrodynamics, electron spin resonance, and
nuclear magnetic resonance. Yet these papers are all related, they are all applications of classical
electrodynamics to real physical problems. Jaynes had essentially four different areas of research: his
first could be called applied classical electrodynamics; his second, information theory (entropy as a
measure of information); his third, probability theory; and finally, semiclassical and classical
radiation theory.

During the years preceding 1957, Jaynes was preparing a set of lecture notes on probability theory.
This material eventually presented to the Field Research Laboratory of the Socony-Mobil Oil
Company. This group in turn published, at least internally, a collection of five of these lecture notes
[7]. Jaynes did try to publish the first of these lectures, “How Does The Brain Do Plausible
Reasoning,” in 1960. However, this work was also rejected by the referee and Jaynes eventually gave
up on publishing it. It was later rediscovered in the Stanford Microwave Laboratory library and, with
Jaynes’s permission, it was published in 1988 [64]; some 28 years after Jaynes first tried to publish it.

In 1957 Jaynes published his first articles in information theory, “Information Theory and Statistical
Mechanics,” [9,10]. In these two articles Jaynes reformulated statistical mechanics in terms of
probability distributions derived by the use of the principle of maximum entropy. This reformulation
of the theory simplified the mathematics, allowed for fundamental extensions of the theory, and
reinterpreted statistical mechanics as inference based on incomplete information. These articles were
published over the objection of a reviewer. (Jaynes comments on this review in “Where do we Stand
on Maximum Entropy,” [37]). Jaynes kept that review, framed and hanging on the wall of his office
for more than 40 years.

The two 1957 articles, by themselves, would have been a career for most scientists; but Jaynes was
far from finished. In the three years he remained at Stanford he published articles on wave guides [11,
14], relativity [12], information theory [13], masers [15] and 50 others after moving to Washington
University in St. Louis.

In the years immediately preceding his departure from Stanford (1960) he was becoming increasingly
dissatisfied with the publish or perish mentality plaguing Stanford, a condition he talked about in
“Backward Look to the Future” [74]. So in 1960 he packed his belongings, sold his house, and
moved to St. Louis, Missouri where he joined the physics faculty of Washington University.

Upon arriving in St. Louis, Jaynes set out on his remaining research interest, reformulating quantum
electrodynamics to avoid quantization of the electromagnetic field. Jaynes published his first paper on
this subject in 1963 with Fred Cummings. It was titled “Comparison of Quantum and Semiclassical
Radiation Theory with Application to the Beam Maser” [18]. Jaynes continued to publish articles on
both semiclassical and classical radiation theory more or less continuously until he retired [18, 26,
27, 31, 34, 35, 39]. However, much of this research is in the theses of his graduate students and has
never been published in the open literature.

Sometime in early 1980 Ray Smith and Tom Grandy, both at the University of Wyoming, organized a
conference on the use of Maximum Entropy and Bayesian methods. They solicited Jaynes’ advice on
whom to invite. The meeting was held in June of 1981 at the University of Wyoming. The list of
attendees included such people as John Parker Berg, John Skilling, John Shore, Tom Grandy, Ray
Smith, B. R. Frieden, Ted Ulyrhc, and David McGovery. While at this meeting Jaynes presented two
papers “Where Do We Go From Here,” [50] and “Entropy and Search Theory” [52]. This meeting
was to be the first in a series of meetings that have continued to this day. Jaynes regularly attended
these meetings until roughly 1990.

While his graduate students were working on semiclassical and classical radiation theory, he was
continuing with his other research interests. For example, in statistical mechanics he published 14
articles from 1960 onward. In probability theory, especially during the 80’s, he published 21 articles.
These articles addressed fundamental questions within these theories and often extended them to new application areas. Many of these articles are published in proceedings volumes. It was commonplace for mainstream journals to reject his manuscripts. Consequently, he often had to wait many years to respond in print to a critic.

In 1982 Jaynes took a two year sabbatical. He spent the first year at the University of Wyoming as an Adjunct Professor. While there he taught statistics, gave a few colloquia, generally renewed old friendships, enjoyed himself, thought, and wrote. Ray Smith told me that he, his son, and Gary Erickson took Jaynes camping in the Snowy mountains. Apparently Jaynes enjoyed this experience so much that when he returned to St. Louis he went out and bought a tent and sleeping bag in the hopes of going camping again some day.

The second year of his sabbatical was spent as a Fellow at St. John’s College in Cambridge, England. The time Jaynes spent at St. John’s college was the highlight of his life. Had it been possible, he would have stayed. During the time at Cambridge he attended the weekly meetings John Skilling had with his graduate students. He wrote a number of papers related to the discussions that were going on at these meetings; “Monkeys Kangaroos and N” being the most notable of these [58]. He wandered the campus thinking about the history and magnitude of human accomplishments that are associated with Cambridge. He mentioned spending parts of several weeks trying to find the tomb of Rev. Thomas Bayes. He succeeded at this, because in his possessions were a number of pictures of Bayes’ tomb.

Jaynes retired in 1992 after a long and productive career. Jaynes’ contributions to science were of the highest caliber. His work in reformulating statistical mechanics has illuminated the foundations of that theory and enabled extensions to non-equilibrium systems. His dedication to rooting out contradictions in quantum mechanics is legendary. He must have single-handedly sparked more debate in quantum mechanics than any other person in the last 50 years. The verdict on his neoclassical radiation theory is still not in, and may not be for many more years. It may yet prove to be a better description of nature than quantum electrodynamics. He also helped take an interpretation of probability theory from being virtually unknown to a healthy research area that is being applied daily in economics, biology, physics, nuclear magnetic resonance and many other disciplines. His writing helped to clarify the foundations of probability theory in a way never achieved before. He wrote profusely, in a warm and friendly way that enabled one to see complex points as if they were intuitively obvious. He spoke as he wrote. When he criticized someone’s work, he always stuck to the facts; he never reverted to name calling. His friendship was hard to earn, and hard to keep, for he had little tolerance for incompetence. He would undoubtedly be uncomfortable with all of the attention being lavished on him now that he is dead.

Edwin Thompson Jaynes
July 5, 1922 - April 30, 1998

- Bibliography
- Unpublished Works
- Book Reviews
Next: Bibliography

This site is being maintained by:
Larry Bretthorst
Dept. Of Chemistry and Radiology
Washington University
St. Louis MO 63130
larry@bayes.wustl.edu
Phone: 314 362-9994