æ

NOTES ON PRESENT STATUS AND FUTURE PROSPECTS

E. T. Jaynes
Wayman Crow Professor of Physics
Washington University
St. Louis, MO 63130, USA

1. Introduction

Listening to the talks presented at this meeting and seeing how the field is developing, I felt rather like the Sorcerer's Apprentice; having in a sense started all this, I am now unable to stop it or even steer it.

The qualification 'in a sense' only recognizes that Maxent is an idea whose time had come, and whether or not I had also come along, it would surely be recognized and used today. Several people have told me that they had the same idea at the same time, but were afraid to say so in public because it seemed such a radical idea then.

As soon as Claude Shannon's work appeared in 1948, there were bound to be readers who were already familiar with the work of Gibbs and Jeffreys. For any such reader it would be a small step to reverse the usual viewpoint and see it this way: the fact that a certain probability distribution maximizes entropy subject to certain constraints representing our incomplete information, is the fundamental property which justifies use of that distribution for inference; it agrees with everything that is known, but carefully avoids assuming anything that is not known. It is a transcription into mathematics of an ancient principle of wisdom; and it accomplishes automatically the needed synthesis of the viewpoints of Gibbs and Jeffreys.

That is, by using the viewpoint of Jeffreys, it gives a justification for the methods of Gibbs that is clearer conceptually, and simpler mathematically, than the long and inconclusive arguments offered in the textbooks of that time. Later we realized that it is also more generally applicable because of its freedom from such assumptions as ergodicity.

In fact, this also reversed the usual viewpoint toward ergodicity. When the goal of Statistical Mechanics was seen as predicting the laws of thermodynamics by deductive reasoning from the microscopic equations of motion, it was held that a proof of ergodicity (stated roughly, that the system "actually uses" the full phase space allowed by our Hamiltonian) was necessary in order to justify the use of the canonical or microcanonical ensembles. Now this was seen very differently; as soon as we see the goal as *inference from incomplete information* rather than *deduction*, then whether a system is or is not ergodic, the Maxent distribution still represents the best predictions we are able to make from the information we have.

Proof that a system is ergodic would be of interest from the standpoint of general dynamical theory; but it would not have the same relevance to the program of Statistical

Mechanics, because such a proof would not in any way change the Maxent predictions we are now making. But systematic failure of those predictions would give us cogent evidence for non-ergodicity and a clue as to which subspace of the full phase space Nature is using. This implies still another reversal of viewpoint; before, failure of the predictions was seen as a calamity to be avoided; now we look eagerly for such failures, because they would tell us new things about the underlying dynamics.

Quite generally in science, when predictions based on our present knowledge succeed, we are pleased but have not learned much. It is only when our best predictions fail that we acquire new fundamental knowledge. But all such subtleties are lost on those who do not comprehend the distinction between deduction and inference, and try to suppresss all mention of human information on the grounds that it is "subjective". Well, human information is all we have; and we had better recognize that fact.

Today, it is good to see the wide and expanding variety of subject matter to which Maxent and Bayesian methods of inference are being applied. This confirms our feeling that the logic of science is universal; the same principles of reasoning that work in statistical mechanics will work as well in astronomy, geophysics, biology, medical diagnosis and economics.

2. Beware of New Ideas

That this movement got off to a slow start is due to two factors. In the first, place, I did not realize that Maxent had important applications outside the statistical mechanics of irreversible processes until I had encountered the works of Burg (1967) and of Gull & Daniell (1978). The field did not get really moving until they had pointed the way. Secondly, every new conceptual idea (unlike a mathematical one) must go through a phase of facing opposition from two sides – the entrenched Establishment who thinks that its toes are being stepped on, and a lunatic fringe that springs up, seemingly by spontaneous generation, out of the idea itself.

Those whose fame and fortune are based on their very real accomplishments using previous methods have a strong vested interest in them and will raise strenuous opposition to any attempt to replace them. This phenomenon has been very well documented in many cases; the details would fill a separate volume.

In contrast to the Establishment which is protecting something that has some demonstrated value, the lunatic fringe has no vested interest in anything because it is composed of those who have never made any useful contribution to any field. Instead, they are parasites feeding on the new idea; while giving the appearance of opposing it, in fact they are deriving their sole sustenance from it, since they have no other agenda.

The Establishment and the lunatic fringe have the common feature that they do not understand the new idea, and attack it on philosophical grounds without making any attempt to learn its technical features so they might try it and see for themselves how it works. Many will not even deign to examine the results which others have found using it; they *know* that it is wrong, whatever results it gives.

There is no really effective way to deal with this kind of opposition; one can only continue quietly accumulating the evidence of new useful results, and eventually the truth will be recognized. As we keep emphasizing, one can argue with a philosophy or even a theorem; it is not so easy to argue with a computer printout, which displays the facts of actual performance, independently of all philosophy and all theorems.

New ideas face other difficulties. We note the comments of Hermann Helmholtz, Max Planck, and Sigmund Freud, respectively, on the reception which their ideas got:

> "The discoverer of a new scientific truth finds it much harder to make out why others fail to understand him than it was to find the truth in the first place."

> "A new scientific truth does not become accepted by convincing its opponents and making them see the light, but rather because its opponents eventually die."

> "Every new idea in science must pass through three phases. In Phase 1 everybody says the man is crazy and the idea is all wrong. In Phase 2 they say that the idea is correct but of no importance. In Phase 3 they say that it is correct and important, but we knew it all along."

Many other such quotations could be offered; Louis Pasteur made one almost identical to that of Helmholtz. When our first Maxent workshop was held here in 1981 we were just moving into Freud's Phase 2; now there is good evidence that Phase 3 is beginning, as we see the Maxent *method* adopted, quietly and without comment, by authors who would never use its name or acknowledge its source.

In fact, even before our first meeting a whole book had been written, deriving the useful properties of Maxent distributions, without mentioning the Maxent principle (Barndorf–Nielsen, 1978). Thus Maxent is becoming the *de facto* method of calculation in many areas, even by those who continue to rail against its philosophy, simply because it produces better results than do the known alternatives. When this process has run its course these meetings will have accomplished their purpose.

Now let us turn to a few miscellaneous comments on things presented at this meeting and previous oues.

3. Maxent Applications

A useful account of the practical considerations that arise in running a Maxent computer program on a real problem, is given by G. J. Daniell & J. A. Potton (1989) in a tutorial explanation of the use of Maxent in finding liquid structure factors from neutron scattering data. Here many practical decisions (choice of measure, digitizing increments, range of output, stopping criterion) had to be made before fully satisfactory results were obtained.

Some have been disappointed at their first Maxent printout, and concluded that the Maximum Entropy principle "gives wrong answers". But that is not possible (assuming that one's computer program is working properly); it quite misses the point about what Maxent is supposed to do. The Maxent distribution is always the correct answer to a definite, well–posed question, namely: "What probability distribution has maximum entropy subject to the basic measure chosen and the constraints imposed?", or "What frequency distribution has maximum multiplicity subject to the conditions imposed?" Nothing in Maxent – or in any other mathematical principle – can do more than answer the question that the user asked.

Now it may be that, after seeing that answer and meditating about it, you decide that you really wanted the answer to a different question. It is always your privilege to do this. But then, instead of saying that the first Maxent run gave a 'wrong answer', the proper statement is that the first Maxent run served as a diagnostic tool that made you aware of how to formulate the right question.

Typically, the first printout reminds one that another constraint, that one had not thought to include in the first calculation, is essential to the problem. The most obvious constraint often seems too obvious to mention; but however obvious something may be, if we fail to state it explicitly, it may be missed. Or, the printout may make one aware that the measure chosen was inappropriate, perhaps from failure to include a volume element factor $4\pi r^2$ in a radial distribution $p(r)\,dr$.

It seems that no worker in probability theory is omniscient enough to think of every contingency and find the exactly right formulation of a problem on the first try; just as no computer programmer is able to write a nontrivial program that does everything right on the first run without debugging. The greatest experimental physicists often find that they must try a dozen times before actually seeing the effect that they are trying to study; the greatest musicians must still rehearse a new work many times before mastering it. Persons who deprecate any procedure because it did not do everything they wanted on the first try, only reveal their immaturity of outlook and lack of worldly experience; anything really worth doing requires practice.

4. Machine Translation

We were fascinated by the report of R. L. Mercer, about the problems of machine translation of French into English (perhaps he should have started by trying to translate British into American). Similar difficulties were noted long ago by Mark Twain, who came across a French translation of his short story, "The Jumping Frog of Calaveras County" and gleefully translated it back, word for word, into English. The result was hilariously funny, and told a quite different story.

As Mercer noted, immediately after Shannon's 1948 work, some thought that this would provide the means for machine translation, but the first efforts were quite discouraging. Surely, one of the reasons for this was the lack of adequate computation facilities 40 years ago; but this was not the only reason. Some early efforts at MIT were recounted many years ago in a lecture by Professor Bar–Hillel, who also revealed, inadvertently, another of the reasons for their failure.

Bar–Hillel described his encounter with John von Neumann and his shock at what von Neumann tried to tell him. He could not believe his ears, because it seemed to him that von Neumann was saying that entropy is not a real physical property of anything, but only a measure of human information. Of course, that is exactly what von Neumann *was* trying to tell him; but his mind was unprepared to receive the message. Bar–Hillel and R. Carnap were unable to comprehend or accept von Neumann's statements; but they were so taken aback by them that they abandoned plans for publication of some of their own work.

With present hindsight, we can see that they were trapped in the Mind–Projection Fallacy; confusing a state of knowledge about reality with reality itself. von Neumann could not overcome that mental block. Of course, entropy was a relevant tool for dealing with their problem; but it was not the entropy of an imaginary 'statistical distribution' of French thought of as an objectively real property of the French language. It was the average entropy of the Maxent distribution of possible English equivalents, under constraints of the specific French messages at hand; this is a measure of how much additional information about the languages would be needed to achieve semantically/colloquially correct translations, and the job of the programmer is to find, from knowledge of the languages, further constraints that reduce this entropy as nearly as possible to zero.

We conjecture that efficient, reliable machine translation requires development of a Universal Intermediate Language (UIL) that has no ambiguity, no synonyms, no rules of grammar (therefore no long list of exceptions) and no special colloquial/idiomatic forms. That is, each word must contain internal evidence telling whether it is a verb, adjective, subject, object, preposition, etc. and each sentence has a unique meaning independent of the word order. Then we translate any language first into UIL, after which UIL can be translated into any other language. In a world with n languages, full translation capablities would require $2n^2$ computer programs without UIL; with UIL we would need only $2n$ programs, each considerably simpler.

5. A Psychological Hangup

The general principle of estimating parameters by maximizing expected utility has been familiar and accepted for decades. To estimate a frequency distribution by that criterion is only a special case of the general principle, so it should be equally acceptable. Now some seem to have a strange psychological hangup over the Maxent algorithm; to them, "information" sounds subjective and vague, while "utility" sounds objective and good. But the only thing you or I can ever know with certainty is not: "What is really true?" but only: "What is my state of knowledge?"

Any science that refuses to recognize that it is based, fundamentally and inevitably, on human knowledge, is in no way made more 'objective' by this; quite the contrary, it operates under a serious self-inflicted handicap. And it requires little thought to see that in real problems our utility is usually far more vague and ill-defined than is our information. Nevertheless, it seems that if we were only to call entropy a "utility measure" instead of an "information measure" this psychological hangup of some would be overcome.

Therefore, if anyone finds it comforting to replace the word "information" with "utility", by all means do so: to paraphrase John Parker Burg, it will not change a single number in the calculations. But it seems to us rather silly – as if an automobile mechanic refused to use a "wrench" because he disliked the sound of the word; but was happy to do so if it was called a "torque optimizer". A wrench by any other name will work as well, and so will the Maxent algorithm.

However, the most experienced scientists have been surprised by unexpected new developments, and so we want to keep an open mind here. It is conceivable that in the future some different procedure will be found which is more general or more powerful than Maxent. If any such procedure exists, we are most eager to hear about it; but we do not think it can lead to different results in the problems which we now solve by Maxent. For there are now several different theorems in which the Maxent algorithm is determined uniquely by requirements of a certain kind of logical consistency. It would seem, therefore, that any alternative procedure whose results conflict with those of Maxent must necessarily lack all those kinds of consistency. Thus it is not surprising that, to date, those who reject Maxent have not offered any usable alternative for the problems which Maxent solves; apparently, they propose to leave us with no solution at all.

Almost exactly the same arguments apply to the Bayesian methods (or more generally, the use of probability theory directly as logic). It is conceivable that some other procedure could be in some way more powerful or more general than the Bayesian one; but we do not think that it can lead to different results in the problems which we now solve by Bayesian methods. For we have Cox's theorems which derive our procedures uniquely from elementary qualitative desiderata of rationality. Therefore, any alternative procedure whose results conflict with those of our present Bayesian methods, must necessarily violate one of those desiderata. Again, it is not surprising that those who reject Bayesian methods offer no usable alternative except in problems so trivial that they scarcely need any formal theory at all; for problems at the level of those solved by Bretthorst (1988), all known alternative methods would be helpless.

6. Why Sample Spaces?

For many years we have been bemused by a question which is still unanswered. Perhaps by pointing it out, someone may have the right inspiration to answer it. In our equations, $P(A|I)$ is not defined numerically unless alternatives A', A'', etc. are specified (by I). This enumeration of A and all the alternatives to be considered if A should be false, is called the *sample space* or as we prefer to call it because it implies less, the *hypothesis space*. Then it is essentially the principle of indifference or its generalization to Maxent, on that space that assigns our initial probabilities.

It is not entirely clear why we need this; the basic product and sum rules derived by Cox make no reference to any sample space, and we know of no theorem proving that a sample space is necessary in order to get numerical values of probabilities. We use it because no other way seems to be known; but perhaps this signifies only our lack of imagination. Are there are other ways of assigning numerical values of probabilities directly our of our information without setting up any sample space?

The status of sample spaces becomes more puzzling when we acquire new information B. This will in general change (update) the probability of other propositions according to the basic product and sum rules of probability theory; but again those rules make no direct reference to any hypothesis space. In principle, we expect such updating to be accomplished by Bayes' theorem in problems where we have an hypothesis space and a model, while Maxent is the appropriate tool when we have an hypothesis space but no model.

Our bemusement is at the fact that in problems where we do not have even an hypothesis space, we have at present no officially approved way of applying probability theory; yet intuition may still give us a strong preference for some conclusions over others. Is this intuition wrong; or does the human brain have hidden principles of reasoning as yet undiscovered by our conscious minds? We could use some new creative thinking here.

7. Artificial Intelligence (AI)

At the 1989 Dartmouth Maxent workshop we noted briefly the situation that has developed in AI, through its failure to use Bayesian methods in situations where rational inference requires them. This resulted in some correspondence offering additional horror stories. Although they only make the same point, one is worth noting here because it emphasizes the importance, not only within the AI community, but for all of us, of correcting this situation.

The AI theory of certainty factors was explained to physicians at some length by Shortliffe & Buchanan (1975). Then the AI program MYCIN (Shortliffe, 1976) was developed at Stanford to help in the diagnosis and treatment of bacterial infection diseases. This uses 'certainty factors' instead of probabilities, which are manipulated according to different rules which can generate not only quantitative inconsistencies, but qualitatively wrong relative ordering of hypotheses. Their unsatisfactory nature was pointed out by Spiegelhalter & Knill-Jones (1983) and Spiegelhalter (1985).

Then Michaelsen, Mitchie, and Boulanger (1985) explained the AI rules to microcomputer users, pointing to available programs like MYCIN and TAXADVISOR. In response, P. Goetz (1985) pointed out that these used the irrational rule of assigning a certainty factor $c(A, B)$ to the conjunction of propositions as simply the minimum of their separate certainty factors; $c(A)$, $c(B)$; and equally bad, for the disjunction they used simply the maximum.

It is evident that this can mislead, in a potentially dangerous way. For example, let $A \equiv$ "Joe's right eye is blue", and $B \equiv$ "his left eye is brown", with probabilities $p(A) = p(B) = 1/2$. The AI rule would then assign $p(AB) = 1/2$; an even chance that he has one blue and one brown eye. The correct rule is $p(AB) = p(A)\,p(B|A) = p(B)\,p(A|B)$, which can be anywhere in $0 \le p(AB) \le 1/2$, depending on the conditional probabilities $p(B|A)$, $p(A|B)$. In fact, we can have $p(AB) = 1/2$ only in the degenerate case where the proposition AB is redundant (in the sense that one of the propositions A, B implies the other). Thus the AI rule is virtually certain to overestimate $p(AB)$, and in a way that can have serious consequences; it can put high credence in a disease even though it is logically ruled out by the evidence.

Likewise, for the disjunction a Bayesian will hardly take $p(A + B) = 1/2$; the correct rule is $p(A + B) = p(A) + p(B) - p(AB)$, which can be anywhere in $1/2 \le p(A + B) \le 1$, depending again on those conditional probabilities. But again we can have $p(A+B) = 1/2$ only in the degenerate case where the proposition $A + B$ is redundant in the sense that one of the propositions implies the other. The AI rule is virtually certain to underestimate $p(A+B)$ with equally serious consequences; it may give only partial credence to a conclusion that is logically required by the evidence.

The AI rules here commit barbaric violations of common sense. A conditional certainty factor of the form $c(B|A)$ or $c(A|B)$ is clearly necessary for any rational assignment of certainty factors involving a combination of two propositions. But merely pointing this out does not seem to get the point across; for Michaelson (1985) proceeded to reply to Goetz as follows:

> "The people who built MYCIN were aware that they were not using 'probabilities.' Instead, they created their own calculations for dealing with uncertainty, based on confirmation theory, and called them 'certainty factors.' They rejected the use of probabilities in their expert system because the system violated the assumptions about statistical independence and prior probabilities that are necessary with Bayes' rule."

Instead of recognizing the defects of the AI rule, he takes refuge in criticizing Bayes' rule (which Goetz had not mentioned). But we protest that Bayesian methods do not make assumptions about 'statistical independence' and Bayesian prior probabilities are not 'assumptions' at all; they are representations of our prior information, and as the above examples make clear, there is no hope of solving medical diagnosis problems until that prior information is taken into account. Thus we have here another example of what we have noted before: those who reject Bayesian methods only reveal, by the arguments they use, their ignorance of what Bayesian methods are.

But even if Bayesian methods did not exist at all, those uncertainty factor rules would still be unacceptable. Why is this not evident to their users on grounds of plain common sense? Do they never try out simple examples like the above? It is appalling to think of such AI algorithms being used to automate important decisions, medical or otherwise.

We have been told that in practice MYCIN performs about as well as do the human experts; presumably, this is intended to tell us something about MYCIN. Perhaps it really tells us something about those experts; if people trained in a subject matter like bacteriology but not in scientific inference do not reason any better than does MYCIN, then the world may be in desperate need of Bayesian expert systems to replace them.

Many other cases could be given; we have a long list of references in which one tried to apply AI methods in problems of inference where Bayesian methods are the natural tool; yet none of them mentions Bayesian methods.

C. A. Zraket (President and CEO of the MITRE Corp.) summed it up thus: "For the past 25 years the hype on AI has far exceeded its accomplishments." (Graubard, 1988). Likewise, the DATACON 90 Conference (St. Louis, October 1990) has a session entitled "AI – Expert Systems: The Magic Bullet That Wasn't", to discuss "vendor promise vs. customer reality". It is not only Bayesians who are skeptical of AI.

Of course, the blame for this poor performance does not lie exclusively in failure to recognize the validity of probability theory as logic; the problems are genuinely difficult. Yet the fact that the human brain solves problems of the same kind effortlessly when they do not involve enormous masses of detail, shows that the solutions cannot be intractable if approached with proper understanding of the reasoning needed.

We think that not only is the adoption of Bayesian methods essential for the technical solution of these problems, Bayesian–oriented thinking is an equally essential part of that general understanding. That is, probability theory as logic is just the appropriate mathematical model of human reasoning, with the inconsistencies removed. The only reason for calling on a computer is that it can carry out those operations faster than we can.

We recognize also that current AI activity involves much that has nothing to do with Bayesian analysis *per se*, such as elicitation of knowledge from those human experts by carefully planned questioning. But this is an activity that the Bayesian may find as necessary as anybody else; it depends on the kind of problem one works on.

8. Some Objections

Let us consider briefly a few of the objections to Maxent alluded to above, to show their general flavor. Penrose (1979) complains that we do not specifiy which piececs of information are to be taken into account as constraints. Well, we had thought it rather obvious that one should always take into account *all of the relevant information one has*; and find it incredible that anyone could have supposed differently. Again, however obvious something may be, if you fail to state it explicitly someone is sure to miss the point.

Seidenfeld (1979, pp. 432-433) thought that entropy was supposed to have the semantic meaning of the word "information" in colloquial language (although nearly every writer on the subject starting with Shannon had warned against this error), and accused us of inconsistency because new information about one quantity can lead to a probability distribution of higher entropy for a different quantity. Well, we had thought it rather obvious that this phenomenon can and should happen in inference: for example, learning the results of a new poll represents more 'information' in the colloquial sense of the word, but it can make us either more certain or less certain about the result of the election. In Jaynes (1957,

p. 186) I took pains to point this out and warn against this misinterpretation of entropy.

Kalman (1981) accuses "the physicists" of committing that same misinterpretation, and in consequence getting, in Maxent Spectrum Analysis, a result that can in some cases differ from the rational fraction form that he wanted us to get. Well, we had thought it rather obvious that the purpose of inference is not to get any preconceived rational fraction form, but rather to represent honestly the import of the data, whatever analytical form it might take. If one thinks that he already knows the correct form of the solution, why use any statistical theory at all – Maxent or any other?

Curiously, the Maxent specrum analysis work was done entirely by John Parker Burg; I did not even know about it until several years later. Yet Kalman attacks me repeatedly and never mentions Burg at all! I tried to reply and explain more things about Maxent spectrum analysis in Jaynes (1982), but let us recall briefly what Maxent actually does here. The problem we are concerned with is that of reasoning from $m + 1$ given autocovariance values $\{R_0, R_1, \ldots R_m\}$, to a power spectrum $P(f)$ satisfying

$$R_k = \int_{-1/2}^{1/2} P(f)\, exp(2\pi i k f)\, df$$

With $m < \infty$ this is not enough to determine $P(f)$ uniquely, so Maxent is used to resolve the ambiguity, with the result that we make the estimate

$$\hat{P}(f) = \left[\sum_{k=-m}^{m} \lambda_k \exp(-2\pi i f k) \right]^{-1}$$

where the λ_k are the Maxent Lagrange multipliers. In Burg's thesis one can find a proof that this does, after all, correspond to the shortest minimum phase prediction error filter that agrees with the $\{R_k\}$. In general this is of length m; Kalman states that in some cases it should be less than m. But the Maxent length may be less than m, because as explained in Jaynes (1982), the datum R_m may be redundant (*i.e.*, just what Maxent would have predicted from the other data), in which case $\lambda_m = 0$.

The Maxent solution to this problem has so many other desirable properties, both mathematical and conceptual, that nobody who has taken the trouble to read Burg's thesis would ask for a different one unless we had different information, in which case we would be concerned with a different problem. What puzzles us is this: what could motivate a person to publish attacks on a work which he has not even bothered to read? Does he not realize what he is doing to his own professional reputation?

But the champion objector of all is a disciple of Kalman, C. A. Los (1989), who denounces not only our methods, but virtually every useful thing ever done in data analysis, going back to R. A. Fisher; we shall devote a separate article (Jaynes, 1991) to answering him.

9. How Does it all End?

When a new method finally becomes mature and generally understood, its sociology undergoes a change. This has happened in many different fields, but let us illustrate it by a single example, Fourier Analysis.

In the last Century Fourier analysis was a marvelous, mysterious new process, misused and/or under attack from those who did not understand it. On the one hand, many considered it absurd to suppose that an arbitrary function could be represented by sine waves. On the other hand, many who accepted the Fourier theorem as valid mathematically, still thought that its application in scientific problems was unjustified because those sine waves were only 'subjective artifacts', and not 'real'. As some put it, "We know that the phenomena do *not* consist of sine waves."

Early in this Century, the respected economist Wesley Mitchell opposed the use of Fourier analysis on the grounds that "periodicities in the data are not real, because they can be reproduced easily merely by correlations in the data". Today economists are sophisticated enough to understand that "periodicity" and "correlations in the data" are only two different ways of saying the same thing. At the same time, the brilliant and respected experimental physicist R. W. Wood (1911) did not understand the meaning of the Fourier theorem any better, and proposed optical experiments to determine whether the sidebands of frequencies $(\omega \pm \nu)$ in an amplitude modulated wave

$$(1 + m \cos \nu t) \cos \omega t = \cos \omega t + \frac{m}{2} \cos(\omega + \nu)t + \frac{m}{2} \cos(\omega - \nu)t$$

are 'physically real frequencies actually present'; or whether 'there is in reality only one frequency ω, but with varying amplitude.'

Today, all of our students are sophisticated enough to understand that the above equation is a *mathematical identity*; the issue is not one of physical fact at all, but only of the meaning of words (the two statements which Wood sought to distinguish by performing experiments are, again, only two different ways of saying the same thing), and no physical experiment could have any bearing on it.

These misunderstandings are not really different in kind from those surrounding Maxent today. We can imagine that, in the face of this kind of confusion, scientists learning how to use Fourier analysis properly in several fields might have gathered together three generations ago – just as we do now – for their mutual protection and defense as well as for exchange of technical information.

But today we do not have annual "Fourier Analysis Workshops" where people in all kinds of different fields get together to sing the praises of Fourier analysis. The reason is that Fourier analysis is now understood and it is all explained in many textbooks; we have no Establishment deploring it, no incompetents misusing it, and no lunatic fringe attacking it. Engineers and Economists both make constant use of Fourier analysis, but nobody sees that as a reason for them to get together to exchange ideas about Fourier analysis.

The same thing will surely happen to Maxent and Bayesian methods when they become generally understood and adopted as standard tools of scientific inference. We will go back to our separate fields to use them routinely, just as we now do for Fourier analysis.

10. Conclusions

Looking ahead to the future in a more general way, we note that ideas are much like living creatures; they may thrive and grow, they may wither and die for lack of support; or they may even be actively destroyed by intellectual barbarians. But this analogy, like any other, fails when pushed too far; dead ideas may be resurrected. The French mathematician Jean Dieudonné (1971) noted that the theory of invariants has been pronounced dead many times, but it keeps rising from its ashes like the Phoenix.

An even better example is the Bernoulli – Laplace vision of probability theory, which has been pronounced dead a thousand times in this Century, but which is today producing important new results far beyond the powers of any other method, and at a rate faster than ever before. Among others, Zellner (1987) and Bretthorst (1988) give many useful solutions to problems so complex that they can hardly be formulated at all in terms of conventional probability theory. In this respect our subject is surely alive and well.

But in another respect there is cause for concern. As Eugene Wigner (1949) put it, just because you have some healthy young cabbage plants, it does not follow that you will harvest any cabbages; a goat way wander along and eat them. Contemplating his cabbages (the then state of understanding of quantum theory) with the goat that he saw eyeing them (the attitude of society toward fundamental knowledge versus practical applications) he was pessimistic, and concluded, "It is almost impossible to believe that the goat will not eat the cabbages".

Forty years later, it appears to us that his cabbages were indeed nearly all eaten, but by a different goat. It was not external pressures, but a complacent, uncritical attitude within science, that put a stop to progress in fundamental knowledge about quantum theory. We may sum it up thus: "Show me a field where theoreticians have been fumbling about for forty years without producing any really significant advance; and I will show you a field where the thinking is aimed in the wrong direction."

What is the mechanism at work here? Darwin's principle of "Survival of the Fittest" was a good beginning, but it seems to us faulty in its expression because the species which survives is not necessarily the one which is fittest to do so. Obviously, the species which survives is simply the one that is most successful at reproducing itself, however unfit it may be in every other respect. An inferior species can easily kill off a superior one merely by outbreeding it.

Likewise, as history teaches us over and over again, a barbarian society can easily destroy a civilized one which does not pay sufficient attention to its own defense. The civilized society tends to be less vigorous externally, because it is occupied internally with more worthy pursuits than fighting its neighbors (education, art, literature, science, medicine, engineering). The very quality which makes it more fit to survive is also the one that makes it less likely to survive.

In the intellectual sphere, the ideas which survive and grow are not necessarily the ones most deserving of this; they are the ones which are taught most aggressively to the next generation of students. For forty years the Copenhagen interpretation of quantum theory has been taught to all physics students – aggressively, dogmatically, and to the exclusion of all other views – with the results just noted.

And in this respect, I fear that Bayesians also face a major crisis. While we are occupied among ourselves with the more worthy pursuits of extending our methods of inference to useful solutions of larger and larger problems, our opponents, with the usual energy and zeal of barbarians, are out–reproducing us by a large factor without producing any results beyond the level of Mickey Mouse trivia. Ideological slogans are so much easier to teach than actual thinking.

Just for that reason, they have captured the attention of the media hucksters who constantly bring them to the general public attention by writing an incredible number of sensational books and articles showing the usual technique of advertising: gush about how *wonderful* it is without ever getting around to explaining *what* it is, and what it is actually accomplishing. The hype on AI and Fuzzy Sets, mass–produced like sausages for twenty years, would now fill many library shelves; to the best of our knowledge, not a single such work has ever been written on Bayesian or Maximum Entropy principles.

This does not mean that we should join in the hype; in the long run we will look much better because we did not stoop to that. But it does mean that, unless we ourselves take the education of the next generation very seriously, and see to it that the needed textbooks and courses of instruction come into existence in our Libraries and Universities, we may be overwhelmed by the sheer number of barbarians being bred today.

REFERENCES

Adams, J. B. (1976), "A Probability Model of Medical Reasoning and the MYCIN Model", Math. Biosciences **32**, 177-186.

Barndorf-Nielsen, O. (1978), *Information and Exponential Families in Statistical Theory*, J. Wiley & Sons, New York.

Burg, John Parker (1975), *Maximum Entropy Spectrum Analysis*, Ph.D. Thesis, Stanford University.

Daniell, G. J. & Potton, J. A. (1989), "Liquid Structure Factor Determination by Neutron Scattering – Some Dangers of Maximum Entropy", in Skilling (1989), pp. 151 - 162.

Dieudonné, J. & Carrel, J. B. (1971), *Invariant Theory, Old and New*, Academic Press, New York.

Dreyfus, H. L. (1979), *What Computers Can't do, the Limits of Artificial Intelligence*, Harper & Row Publishers, New York.

Goetz, P. (1985),"Calculating Probabilities", BYTE magazine, November, p. 14

Graubard, S. R. (1988), *The Artificial Intelligence Debate: False Starts, Real Foundations*, MIT Press, Cambridge MA.

Hahn, G. J. (1985), "More Intelligent Statistical Software and Statistical Expert Systems: Future Directions" The American Statistician, **38**, 1-16.

Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983), *Building Expert Systems*, Addison–Wesley, Raeading, MA.

Jaynes, E. T. (1957), "Information Theory and Statistical Mechanics II", Phys. Rev. **108**, 171–190; p. 186.

Jaynes, E. T. (1982), "On the Rationale of Maximum Entropy Methods", Proc. IEEE, **70**, 939–952.

Jaynes, E. T. (1991), "Commentary on Two Articles by C. A. Los", to be published in Vol. 3 of special issues, *On System-theoretic Methods in Economic Modelling*, S. Mittnik, Editor, in Computers and Mathematics with Applications; and subsequently as a monograph by Pergamon Press.

Kalman, R. E. (1981), "Realization of Covariance Sequences", in Proceedings of the Toeplitz Memorial Conference, Tel Aviv University; pp. 331–342.

Los, C. A. (1989), Computers Math. Applic. **17**, 1269 – 1304.

Michaelsen, R., Michie, D. & Boulanger, A. (1985), BYTE magazine, April, p. 111.

Michaelsen, R. (1985), Reply to Goetz, BYTE, November, p. 14.

Penrose, O. (1979), "Foundations of Statistical Mechanics", Rep. Prog. Phys. **42**, 1937–2006.

Pregibon, D. & Gale, W. A. (1984), "REX: an Expert System for Regression Analysis", in Havranek, T., Sidak, Z., & Novak, M. (editors), *Proceedings of COMPSTAT 84*, Physica–Verlag, pp. 242–248.

Seidenfeld, T. (1979), "Why I am not an Objective Bayesian", Theory and Decision, **11**, 413-440.

Shortliffe, E. & Buchanan, B. (1975), "A Model of Inexact Reasoning in Medicine", Mathematical Biosciences, **23**, 351.

Shortliffe, E. H. (1976), *Computer Based Medical Consultations: MYCIN*, American Elsevier Press, New York.

Skilling, J. (1989), Editor, *Maximum Entropy and Bayesian Methods* Proceedings of the Eighth MAXENT Workshop, Cambridge, England, August 1988., Kluwer Academic Publishers, Dordrecht, Holland.

Spiegelhalter, D. J. & Knill-Jones, R. P. (1984), "Statistical and knowledge–based approaches to clinical decision-support systems", J. Roy Stat. Soc. (B), **147**, 35-77.

Spiegelhalter, D. J. (1985), "A statistical view of uncertainty in expert systems", in *Proceedings of the Workshop on Artificial Intelligence and Statistics*, AT&T Bell Laboratories, Princeton NJ.

Tversky, A. & Kahneman, D. (1981), "The Framing of Decisions and the Psychology of Choice", Science **211**, 453–458.

Wigner, E. (1949) Remarks made in an evening informal Commons Room discussion at the Graduate College, Princeton, attended by the writer.

Wood, R. W. (1911), *Physical Optics*, Mac Millan, N. Y.; 3rd edition, 1934.