Some Observations on Systems, Probability, Entropy and Management[1]

Myron Tribus

Defining a System

The first examination I ever flunked completely, thoroughly and without doubt, was given in 1937, in an honors course in Thermodynamics, taught by Professor W. F. Giau que, later to become a Nobel Laureate. Our assignment, beforehand, was to read Chapters I and II in the famous text by G. N. Lewis and Merle Randall, "Thermodynamics" (McGraw Hill, 1923) On page 8 a system is defined: "If we make an enclosure by means of physical walls, or if we imagine such an enclosure made by a mathematical surface, such an enclosing surface serves as the boundary of the system, which then comprises everything of thermodynamic interest contained within that boundary."

The examination consisted of just one question: "What is a system?" I responded by saying that a system was a method, or an approach, such as a "system for beating the odds at the race track." Failures, when abject enough, are powerful stimulants to the mind, and I never forgot the experience. Twenty one years later, in 1961, when I wrote my book on thermodynamics, "Thermostatics and Thermodynamics: An Introduction to Energy, Information and States of Matter" (D. Van Nostrand, 1961) I went to great lengths to define three classes of systems, each of which is a region of space bounded by an imaginary or real surface:

a) Closed systems, which do not exchange matter with their surrounds, but may still exchange energy and information.

b) Flow Systems, which exchange matter, energy and information with their surrounds, but the matter enters and leaves only in well defined streams, through well defined ports and under some degree of control, in a purposeful way.[2]

c) Open Systems, which exchange matter, energy and information in a diffuse manner, through various places on the surface, in the absence of well defined streams and not necessarily under precise control.[3]

---

[1] Thispaper was inspired by reading a letter from Dr. Horine to Dr. Deming. She raised a number of questions in the letter which prompted me to attempt to answer them.

[2] I likened such systems as similar to a prison in which the prisoners are transferred, under control, in packets. It is unlikely that those being thus transported are very different from those left behind.

[3] I likened these systems as similar to a prison from which prisoners could escape. The prisoners escaping are likely to be more energetic than those that remain behind.

These classifications are useful in chemical thermodynamics and in many other areas of study outside of chemistry and engineering. Note, however, that they represent a view taken from *outside* the system. Lewis and Randall (pg. 85) emphasized this outsider view by saying: "Thermodynamics exhibits no curiosity; certain things are poured into its hopper, certain others emerge according to the laws of the machine, no cognizance being taken of the mechanism of the process or of the nature and character of the various molecular species concerned."

This way of looking at systems is not appropriate for managing, but it is often used by those who try to manage by financial measures only. They treat the elements of a system as though it were a *thermodynamic system* and they turn the crank on their management machine, paying no attention to what happens inside the systems they manage.

It is evident to anyone who has studied economic theory and thermodynamics that the economists have been overly influenced by the thinking of thermodynamicists. Concepts such as "force", "pressure", and "equilibrium" have been taken over without change. In thermodynamics we understand that what we see is the result of averaging the behavior of many tiny particles, usually of order 10^{20}, which is equivalent to the population on a billion billion Earths! When I ask about the basis for economic predictions, those economists who are willing to talk seriously to someone outside their field, will refer to the "averaging effect" of many players in the marketplace. In thermodynamics we have a field called "statistical mechanics" which clarifies how the averaging process works. There is no such field of study in economics and thus there is no connection between what we know about the microeconomics of firm and the macroeconomics of a nation.

There is another very powerful difference. Atoms and molecules may be looked upon as having certain response characteristics. That is, given a stimulus, the response can be predicted. Human beings are more complex. How they respond to a stimulus is modified by what they *want* to do. This distinction is why the Skinnerian approach to psychology is so barren.

In short, there is no analog to statistical mechanics in economics, which explains why economics has such poor predictive ability. I pointed this out in an obscure publication for Tau Beta Pi, the Engineering Honor Society, "The Engineer in the Quality Revolution". This omission is important. What the economists omit are the various "ports" of business systems by which enterprises interact with the environment.

With these observations as prologue, we see that Dr. Deming's system of management (using the word "system" in the sense that earned me a failing grade 51 years ago!) regards systems from an insider's view. His emphasis on process, on internal interactions, on information exchange with the environment, viewed from inside the system boundaries, is in sharp contrast with the views emerging from the study of thermodynamics. (All of this is the more surprising since he and I started out to be chemists and got our first degrees in chemistry.)

Of course, not all the people in the generations before us had the outsider's view of systems. The first person of whom I am aware who studied systems with a larger view was A. M. Wellington whose book of 1887, "The Art of Railway Location" considered railway building as part of a larger system. He cautioned engineers to run railway lines close to the confluence of two rivers, for that is where he forecast the future cities would develop. He told them to take into consideration the fuel needs for the trains would some day run on the tracks and, to run the lines where the grades were less steep and, if possible, near a forest, to make it easier to conserve energy and to obtain wood for fuel. Shortly after World War II, Hall produced his famous book "Systems Engineering". (Someone has my copy and never returned it, so I cannot quote correctly from it.) Hall's work spawned a number of imitations. His book brought together much of the research done in the field of Operations Research (OR). OR people, during W.W.II started looking at operational parts of the military as *systems* and demonstrated how to optimize their effectiveness by avoiding sub-optimization. They worked according to "Given this system, what is the optimum way to use it?" Hall approached a different task: "Given this need (i.e., transportation), what is the best way to approach the design of the system?"

Hall, thus was a pioneer in bringing the insider view to the *design* of systems. He posed the question in this way: "If the environment is thus and so, and if the environment might change in these ways, how should the system be *designed* to meet its objectives, despite the changing environment?" The perspective changed from *analysis* to *design*.

Hall's work caused many people in the social sciences to try to extend these ideas to social systems, but in general they have not had the same rigor, because they often continued to use the same analogs to thermodynamics, speaking loosely of forces and equilibrium, although these do not have operational meanings in most of instances in which they were employed. Bertalanfy was one of those who participated in this extension of ideas, and while I do not wish to detract from his accomplishments, I did feel it useful to bring out the work of those who went before and to set Bertalanfy work in a wider context.

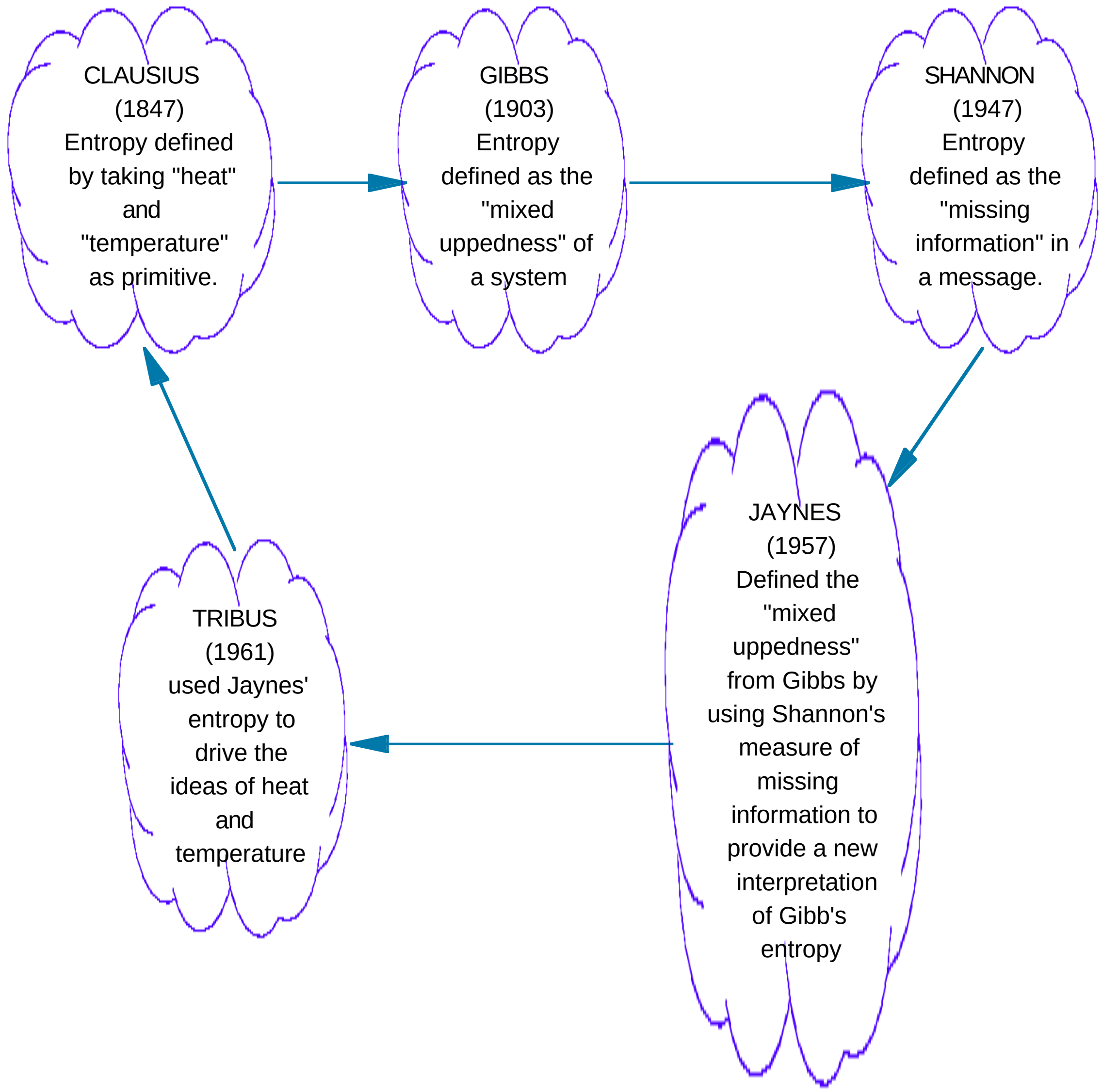
Rigorous thinking about systems in thermodynamics, in the early days, concentrated on the exchanges of matter and energy across the boundaries. But the systems in which we are interested also engage in the exchange of information. In that connection, we ought to pay heed to the work of Claude Shannon, who worked at the Bell Telephone Laboratories a trifle later than

Shewhart. Shannon's famous work "A Theory of Communication" (Bell Telephone Laboratories Journal, 1947) presented a new way to think about information. He introduced a new concept, "the entropy of information", which measures how much is left to learn when you are uncertain about the proper answer to a well defined question. He borrowed "entropy" from thermodynamics because, as he told me once, Von Neumann told him, "Your mathematical formula is similar to one used in statistical mechanics. People do not really understand entropy so if you use it in an argument, you will win every time, hands down." Shannon regarded the use of entropy in information theory as representing an entirely different concept and felt it would not introduce confusion if he used it in information theory.

A decade later, in 1957, in the hands of a physicist, Edwin T. Jaynes, a formal link was made between the concepts of entropy developed by Clausius in 1847 (the "outside view" of classical thermodynamics), the almost "inside view" of Josiah Willard Gibbs, in his work in statistical mechanics, and the "completely inside view" of the information theory concept of entropy. In 1958, because my doctoral committee eight years earlier had stumped me by asking how to connect all three of these concepts (but they passed me anyway because none of them could answer the question either), when I read the work of Ed Jaynes, I realized he had made a breakthrough in thinking. I therefore set about writing my textbook (referred to above) which used the idea of entropy from information theory as the basis to construct the ideas of Gibbs (microsystem averaging methods) and Clausius (the "outsider view"). These works are linked as in the figure on the following page.

None of the contributions since Clausius makes any of the previous works less true. What the new ideas do is to reconcile the outside and inside views and help us understand the power and limitations of each.

It was necessary for biologists, for example, to forego "outsider", thermodynamic thinking before they could begin to unravel the inner workings of DNA and even today well intentioned people confuse the issues. For example the Nobel acceptance lecture by Illya Progogine, in which he tries to discuss the use of "negentropy" in connection with living systems, basing his arguments on entirely classical reasoning (a'la Clausius) leads nowhere.



A System of Red Beads....And the Meaning of Probability

Before I examined a few data points, I thought I ought to see red beads on the paddle in proportion to their presence in the mixture. This is not what I saw. Even after averaging some results, I did not see what I had expected, a-priori, to see. This tells me that something else is at work I had not considered. Maybe the thickness of paint is different, red vs. white beads. Maybe they develop different amounts of electric charge, which will influence whether they settle on the paddle in proportion to their numbers. There are a thousand reasons this might be so.

There is more to the issue than just a reasonable person's anticipation that the number of beads might be describable by a Normal distribution (or a Poisson). What fascinates me is that concurrent with the 114 year journey from Clausius' conception of entropy, and starting about 85 years earlier, there has been an important paradigm shift in the field of statistics itself. This shift involves the meanings to be attached to the word "probability" and the interpretations which should be placed on our mathematical manipulations of the symbol "p", which we use to represent the concept of probability.

LaPlace's dictum of "equal a-priori" versus the long run frequency interpretation of people like Von Mises, have presented logicians with unsolvable difficulties as they tried to say what "probability really is." The question was resolved in a little known work by Richard T. Cox in a paper published in 1947 and expanded into a book "The Algebra of Probable Inference" (Johns Hopkins University Press, Baltimore 1961). Jaynes, whom I mentioned earlier, based his work in entropy and statistical mechanics on this little known work of Cox, so I began to study it in earnest. Building on their work I produced a second book in 1969, with the purpose of showing that what had been done in the earlier thermodynamics text was an example of a very powerful method of reasoning which not only could produce thermodynamics, but could be used to resolve many other problems as well. In the second book, "Rational Descriptions, Decisions and Designs" (RD^3) (Pergamon, 1969), using the work of Cox and Jaynes, it is demonstrated that there is one, and only one, interpretation of the meaning of the word "probability" which allows us to understand, in a unified way, what the word means in all of its uses. When I say, "one and only one", I am relying on a rigorous mathematical derivation which uniquely connects the result to the equations of constraint applied in the developemnt.

The definition which emerges is this:

> **PROBABILITY is a numerical encoding of incomplete information.**
>
> **The task of probability theory is to develop consistent and practical means to convert incomplete information of various sorts and combinations into an assignment of numbers to probabilities associated with well defined answers to well defined questions.**

What RD^3 does is to demonstrate how these rules apply. One of the rules is:

> **When assigning values to a set of probabilities, make the assignment consistent with what you know and maximally non-committal to what you do not know. In other words, do not assume as true anything you do not know is true and do not neglect what you do know.**
>
> **PROBABILITY provides a means to tell someone else neither more nor less than you truly know.**

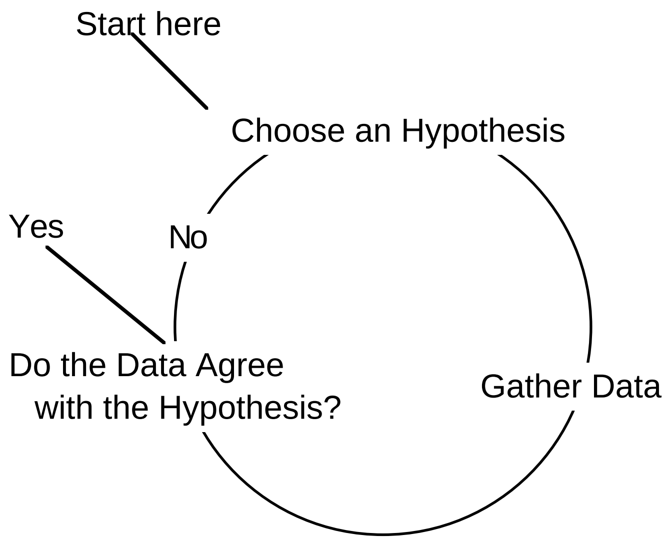
Looking at probability theory as a way to communicate incomplete information, it is no surprise that the entropy function, as developed by Shannon, should play a significant role in a wide variety of applications. I am not alone in giving it special status. Today there exists an informal society of researchers around the World who gather once a year to discuss new applications of this concept. You will find a dozen books have been published from the 12 years of meetings of this group, which calls its conferences "The Maximum Entropy and Bayesian Methods Workshops". About a hundred people show up each year at the annual meeting, held in Seattle, or Laramie, or Paris, or Cambridge (UK) or some other place. The publisher of the series is D. Reidel Publishing Company of the Kluwer Academic Publishers Group. In the papers of the conference you will find examples in biology, physics, medicine, image reconstruction, butterfly populations, as well as epistemology, all treated by a common methodology, extending statistical methods in all directions.

People are very resistant to the introduction of new paradigms. When I introduced the new view in thermodynamics I was met with unbelievable hostility. To justify what I had done I wrote the second book, which had no persuasive power on thermodynamicists, who read only their own works, but instead got me in trouble with classically trained statisticians, who don't wish to have anyone disturb *their* paradigms. In the intervening 36 years, many other people have come to the same perspective regarding the foundations of statistical inference (though still a small minority) and the editors of journals are now less hostile to the new views than they were.

Symbolically, this way of looking at probability relies upon a notation, such as p(A|BE}, to express the idea: "p is a number representing our knowledge of the truth of a proposition, represented by A, <u>given</u> that some other proposition, B, is true and that E represents all the other evidence we have which might be relevant." To assign a defensible number to p, requires that the <u>process</u> used to make the assignment be unambiguous, open, consistent, honest, candid and not self-contradictory. It also ought to be so general as to allow people to apply the method to any set of A, B and E, regardless of the meanings associated with A, B and E. Finally, the process should not invoke sleight of hand or leaps of faith. What is so wonderful about the work of Cox and Jaynes is that the <u>method</u> for making assignments they have developed meets all these tests! This statement is not true of the methods used in classical statistics, which is one of the reasons there are so many books on the subject.

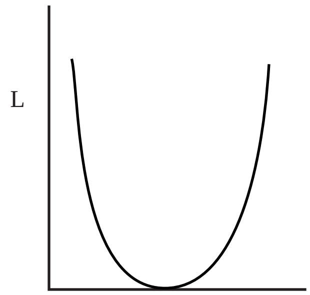
Jaynes' contributions are gathered in the book "E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics", edited by R. D. Rosenkrantz (D. Reidel Publishing Co., Dordrecht, Holland, 1983)

Now, how does this relate to your choice of the Normal distribution? Or to my guess that the red beads would be present in proportion to their numbers? Each of us, using what little we know, made an initial assignment, but it does not necessarily hold up to the observations. We need to think of ourselves as involved in a cycle:

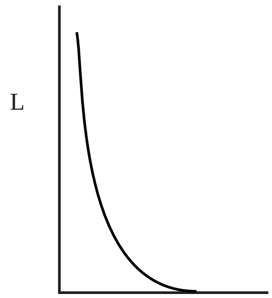


Unfortunately, if one follows this line there is no reason to ever stop. One may go on striving for better and better hypotheses as long as someone else will put up the money. The statistical literature overflowith with useless papers generated by this attitude.

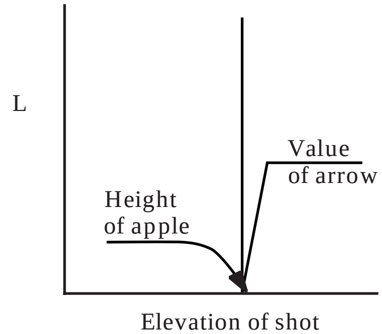
What is required is to develop descriptions which may be readily used with "value functions" (or if you are a pessimist, "loss functions"). What is needed is a way of describing incomplete knowledge so it may be easily combined with value or loss functions and, therefore, lead to better decision making. But classical statistics hasn't been developed in a way to make this easily possible. That is why there has been so much controversy about Dr. Taguchi's introduction of the loss function. Although Taguchi is not a trained statistician, he is a good engineer and when he approached the problem of designing more robust systems, he simply invented ad-hoc procedures, trusting to his intuition that he wasn't doing something wasteful. He introduced the loss function which had already been developed in the field of decision analysis. I do not know who first introduced that concept. It is in the early works of Howard Raiffa of Harvard University, who pioneered the use of decision analysis in business. You will find the loss function defined and used in RD[3]. The problem with Taguchi is that he did not know of the earlier work which shows how to use loss functions properly. To compensate for this he introduces what he calls the "signal to noise ratio" which is appropriate only for symmetrical, quadratic, loss functions. But there are many other kinds of loss functions, as suggested in the figures below. (These were first shown to me by Ed Jaynes, by the way).



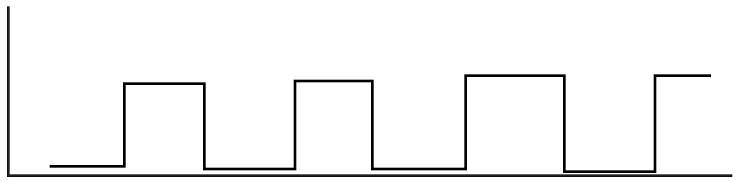
Taguchi Quadratic



One sided



"William Tell"



Shooting at ducks in a shooting gallery.

Taguchi gets around all this by inventing different kinds of loss functions to be used in different situations. The proper way to deal with any loss function is described in the paper "An Alternative View of the Taguchi Approach" (Quality Progress, May 1989, Vol. XXII, N5, Pg. 46)

What comes out of all this is that, according to these views, one does not attempt to find the best possible hypothesis but rather attempts to find the hypothesis which is most appropriate to use with the loss function. In other words, there is a point at which it does not pay to get more information, which might be used to develop a "better hypothesis". Instead, one looks for the economically justifiable information to gather and use to develop the hypothesis most useful to the loss function.

In the case of a quadratic loss function, all that matters is the expectations of the mean and variance. For such cases, the Normal distribution is the most useful fit. This is a powerful idea. Since a quadratic loss function is so often the proper one to use (for reasons which Taguchi has demonstrated in several of his publications) we are interested in being able to predict the mean and the variation about this mean. The distribution most appropriate for that purpose, provided the loss function is symmetrical, is the Normal Distribution. What the theory of maximum entropy then tells us is that *even if the Normal distribution is not the best fit to the data,* the Normal distribution, fitted to the data, is adequate for decision making. All we really care about is the moments of the distribution, not how well it fits the data.

With respect to the red bead experiment the following strategy is justifiable:

1. If you have **no** information other than the fraction of red beads in the bowl, the only defensible prediction is that the number of beads will follow the binomial distribution, with a mean value equal to the fraction of red beads in the bowl.

2. If you have access to data, then the data will suggest a distribution which, whatever it is, should be incorporated in your general information. The methods in RD^3 show how to do this. You should use the data at hand (however limited) to make a better prediction about what will occur in the future, subject to the condition that what is done in the future is a replication of what was done when the data were taken, i.e., the process is under statistical control.

3. If you are to make a decision on the basis of the information gathered, then the data should be used to develop a statistical distribution appropriate to the loss function. Dr. Deming uses this basic idea in "Out of the Crisis" when discussing whether to test an entire lot or not at all, but the formulation he gives does not explicitly invoke the loss function. (Chapter 15, "Out of the Crisis")

When following the approach in RD^3 the concepts of enumerative and analytical statistics merge into one comprehensive approach in which the method depends upon the purpose.

Epistemologists ask such questions as "What is Knowledge?" and "What is Truth?"

Physicists, in general, are not as good as chemists in dealing with questions such as "What is knowledge?" or "What is Truth?" Of course there are exceptions, but their training leads them away from such issues. It seems that now and then someone like Hawkins appears, who provides an exceptional view. The main mode of physicists, which derives from their training, is deductive. They just love to start with a grand equation, such as Newton's equation, or Maxwell's equations, or Einstein's equations and then deduce everything.

Chemists, on the other hand, work with induction. They begin, as I did as a freshman, trying to figure out "What is this unknown substance?" As a senior I was asked to deduce a molecular structure. Biochemists try to unravel the structure of ever more complex molecules. Chemists must develop improved methods to describe situations in which their knowledge will always be incomplete.

Conclusion

Dr. Deming's "System of Profound Knowledge" has been presented as being "all of one piece". You cannot accept one part of it and neglect the others. I hope that this rambling essay has provided examples of why it is so.