

FURTHER QUESTIONS RE THE CONCEPT OF ENTROPY.

Jeremy Dunning-Davies,
Departments of Mathematics and Physics (retd),
University of Hull, England
and
Institute for Basic Research,
Palm Harbor, Florida, U.S.A.

email: masjd@masjd.karoo.co.uk

Key Words: entropy, thermodynamics, statistical mechanics, information theory, statistics.

Abstract.

As has been mentioned previously [1], the idea of entropy has caused, and continues to cause, problems of real understanding for all but especially for students. Several issues were raised in an earlier article [1] but here attention is focused on issues occurring through work in information theory. Again, it is hoped these points will provoke thought among interested parties towards producing a clear and accurate understanding of precisely what entropy is physically.

Introduction.

In a recent article [1] attention was drawn to the unanswered question of ‘What is Entropy?’ Much of the article was devoted to a discussion of the notion of entropy in thermodynamics and in statistical mechanics/thermodynamics. It was pointed out that the property known as the entropy was introduced in classical thermodynamics via the well-known Second Law - using quite naturally one of the two basic forms of that law due to either Kelvin or Clausius. This introduction naturally leads to a physical link with the entropy via the realization that its change occurs as a result of a change in heat divided by the absolute temperature at which that change occurs. The notion is then extended to cover situations covered by statistical mechanics/thermodynamics via a variety of routes using well established results from classical thermodynamics. This latter material is well covered in many traditional books such as that by Hill [2]. In fact, it is both interesting and informative to note Hill’s approach. He looks at several different ensembles in detail and, in each case, derives expressions for the usual thermodynamic functions, including the entropy but in no case does he make use of the Boltzmann entropy formula as a starting point, although in each case considered he is able to show that the Boltzmann formula may be deduced. Hence, the introduction of the idea of entropy into statistical mechanics/thermodynamics comes via introducing statistical mechanical ideas into well-established thermodynamic relations and deriving the entropy in each individual case. In a specific example this overall approach is well illustrated in the case of the ideal relativistic quantum gases [3]. However, it seems real issues arise when the situation in information theory is considered. Whatever the situation was earlier, it seems that much, if not most, of modern information theory originated with the seminal work of Shannon which resulted in the publication of both an article [4] and a book [5]

Shannon and Information theory.

Possibly the best way to approach the introduction of the notion of entropy in information theory is to consider the way Shannon himself introduces the topic [5]. To do this, consider Shannon's own wording beginning on page 48 of the mentioned book:- he begins by pointing out that, up to that point in his presentation, a discrete information source has been represented by a Markov process. He then poses the question of whether or not it is possible to '*define a quantity which will measure, in some sense, how much information is "produced" by such a process, or better, at what rate information is produced*'.

He continues by considering a set of possible events whose probabilities of occurrence are $p_1, p_2, \dots, p_n$. These probabilities are all that is known concerning which event will occur. The question then is whether or not a measure of how much "choice" is involved in the selection of the event or of how uncertain the outcome is may be found. If such a measure, say $H(p_1, p_2, \dots, p_n)$, exists it is claimed that it is reasonable to require the following properties for it:

- (i) H should be continuous in the p_i
- (ii) If all the p_i are equal, $p_i = 1/n$, then H should be a monotonically increasing function of n
- (iii) If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H .

On the basis of these required properties, it is shown in appendix 2 (pages 116-8) of the quoted book that the only H satisfying these three assumptions is of the form:

$$S = -K \sum_{i=1}^n p_i \log p_i$$

where K is a positive constant which merely amounts to a choice of a unit of measure. Shannon points out that expressions of this form '*play a central role in information theory as measures of information, choice and uncertainty*'. He also goes on to point out at this very stage in his discussion that 'the form of H will be recognised as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of a system being in cell i of its phase space. Of course, in statistical mechanics/ thermodynamics, the constant K is the Boltzmann constant represented by k . Again at this point in his discussion, Shannon proceeds to say that '*we shall call*' the expression $H = -\sum p_i \log p_i$ the entropy of the set of probabilities $p_1, \dots, p_n$. He also says that, if x is a chance variable, we will write $H(x)$ for its entropy; thus x is not an argument of a function but a label for a number, to differentiate it from $H(y)$ say, the entropy of the chance variable y .

It is quite clear from this abbreviated introduction to a basic formula of information theory that its introduction is purely mathematical and does not rely at all on any genuine physical input – certainly nothing remotely connected with notions of classical thermodynamics. Interestingly, the outline of this derivation also serves to emphasise that the same is basically true of the introduction of the entropy concept into statistical mechanics/thermodynamics. Also it seems the introduction of a link between the entropy functions of information theory and classical thermodynamics followed the route outlined in Brillouin's book [6] and that is that it was simply noted that the entropy expressions of information theory and statistical mechanics/thermodynamics had the same mathematical form and so were assumed to be the same. The link with classical thermodynamics then followed via the statistical mechanics link with that subject. However, in both Shannon's and Brillouin's introductions to the subject, the constant appearing is just a constant; it is not Boltzmann's constant k which appears from the outset in classical thermodynamics and does, of course, possess quite definite units - ergs/degree. Hence, when examining the origins of the three entropy functions, it quickly becomes apparent that as progression occurs from that for classical thermodynamics to

statistical mechanics/thermodynamics to information theory, actual physical ideas play a gradually decreasing role and heat only appears in the first of these derivations. It is also important to remember that, in both statistical mechanics/thermodynamics and information theory, the entropy function often appears initially as a purely mathematical function, just as it does in mathematical statistics itself, with little immediate physical interpretation whereas, in classical thermodynamics it is irrevocably linked with heat exchange from the outset. However, as indicated above, as far as statistical mechanics/thermodynamics is concerned, the approach adopted by Hill makes the link with classical thermodynamics both clear and reasonable so that it seems clear that the entropy functions in these two cases are the same.

As far as information theory is concerned, after Shannon's early initiative, the subject appeared to take off in its own right. A vitally important contribution at this point was due to Rényi [7]. In this article he pointed out that Shannon's result for the entropy could be derived from various different sets of postulates. He suggested the simplest set of postulates be that proposed by Fadeev [8] but noted that, if Fadeev's fourth postulate is replaced by one giving the entropy the property of additivity then the result would be that many quantities could satisfy them. As an example he cites

$$H_\alpha = \frac{1}{(1 - \alpha)} \log_2 \left(\sum_{k=1}^n p_k^\alpha \right)$$

where $\alpha > 0$ and $\alpha \neq 1$. This point, mentioned early in the article by Rényi, immediately raises several issues so important when considering the thorny question of 'What is entropy?' The first being the very serious one concerning whether or not entropy in classical physics is the same function as the entropy of information theory.

Further Thoughts.

Since the postulates leading to the derivation of Shannon's information entropy can lead also to other expressions for that quantity, the question must arise as to whether or not entropy in information theory is the same quantity as that which is referred to as entropy in thermodynamics – in each case the entropy formula can look the same but that needn't be the case as Rényi so clearly showed. It also raises the question of why did Shannon refer to his quantity as entropy? Was it simply because its mathematical form was the same as the quantity named entropy in thermodynamics? This whole question is all the more disturbing when it is remembered that, in 1988, Tsallis [9] published an article in which he supposedly considered a possible generalisation of Boltzmann-Gibbs statistics. In this piece, after a very short introductory paragraph he *postulated*, as a generalisation of the standard expression for the entropy in information theory, that

$$S = k \frac{1 - \sum_{i=1}^W p_i^q}{q - 1}$$

where W is a real number and is the total number of possible (microscopic) configurations and $\sum_{i=1}^W p_i = 1$. However, there seemed to be no justification for even thinking this formula to be a possible generalisation of the Boltzmann-Gibbs expression and no link with physics was mentioned in the article. Furthermore, as was pointed out by Lavenda and Dunning-Davies [10], a close analysis of the literature shows that a normalised form of this expression appeared in an article by Havrda and Charvat in 1967 [11]. Further details of this entire field of study may be found in the book by Aczél and Daróczy [12]. However, the main point raised here refers to whether or not these entropy expressions in information theory are referring to the same function as the entropy of thermodynamics and statistical mechanics. As indicated above, Hill has shown how the Boltzmann entropy expression does fit in with that of thermodynamics

but, other than the visual similarity of Shannon's expression with that of Boltzmann, there seems little reason to suppose the entropy expressions of information theory the same as that of physics. This is yet another point which simply leads to confusion in the minds of students and possibly of researchers in the physical sciences as well. In the physical sciences, the starting point must be that the entropy appears via its change being given by the change in heat (either positive or negative) divided by the absolute temperature at which that change occurs. This has to be the starting point for all discussion relating to the idea of entropy in the physical sciences.

It might also be noted at this point that, as mentioned above, from its introduction via the Second Law of Thermodynamics, it is clear that entropy changes can be positive or negative. If that were not so then, for example, it would be impossible to have a Carnot cycle since, in any cyclic process, the end point must exactly the same as the starting point. It follows that the claim that the Second Law states that, in every physical process, the entropy stays constant or increases [13, for example], is fundamentally incorrect. This notion that the Second Law tells us that entropy either increases or remains constant is true only in clearly specified circumstances. The Second Law must always be used and remembered in one of two original forms – either that due to Kelvin or that due to Clausius, - although the more mathematical form due to Carathéodory is also acceptable even though its use has been limited somewhat due to the emergence of its close links with the other two mentioned forms of that law. [14]

Deductions and Conclusions.

The above reflections undoubtedly raise a number of serious questions which need answers if a real understanding of the meaning of entropy is to be achieved. The apparent link with the physical notion of entropy was forged by the appearance of Shannon's work [4,5], together with the book by Brillouin [6]. It seems possible that, because Shannon's expression for some sort of measure for information had the same mathematical form as that of Boltzmann's expression for the entropy in statistical mechanics, the name entropy was chosen for it. On the face of it, this does not seem unreasonable and physical reasons may be assumed to support this – for example, the idea of associating order and disorder with entropy. However, the appearance of Rényi's work [7] and the realisation that one set of initial assumptions in information theory can lead to more than one expression for the so-called entropy raises more questions than it answers. One huge problem for the physical sciences follows especially from the idea advanced by Tsallis [9] of an entropy form supposedly generalising Boltzmann-Gibbs statistics. It is difficult to find clear justification for this claim and certainly no-one has yet shown a direct link between his 'new' expression and the traditional one in classical thermodynamics, although queries have been raised concerning some of this work and uses of it [15]. All this, however, leaves students and probably many others in academia in a quandary and does little to help produce a clear cut answer to the question 'What is entropy?'

It is always hoped to end any article with a definite statement or answer to a posed question but such cannot be the case here. In a previous article [1] this same question was raised in the hope that someone could produce a good clear cut answer but the realisation of the content of Rényi's seminal article [7] sheds further doubt on the way the discussion seemed to be leading. Now the first question to be clearly answered must be 'Is the entropy of information theory the same function as the entropy of classical thermodynamics or are they different entities which happen to have the same mathematical form if one particular set of initial conditions is chosen in the information theory case?' It seems highly likely that this could be a far from trivial question to answer but it seems one that requires an answer for all in both areas of academic endeavour.

References.

- [1] Dunning-Davies, J., 2026, "Questions re the concept of entropy", *Parana J. Sci. Educ.*, v.12, n.3, (6-14), June 1
- [2] Hill, T.L., 1956, *Statistical Mechanics*, McGraw-Hill, New York.
- [3] Landsberg, P.T. & Dunning-Davies, J., "Statistical thermodynamics of the ideal relativistic quantum gas"
in *Statistical Mechanics of Equilibrium and Non-equilibrium*, ed. J.Meixner; (North-Holland, Amsterdam, 1965)
- [4] Shannon, C.E., 1948, Bell System Tech. J., **27**, 379
- [5] Shannon, C.E. & Weaver, W., 1949, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana
- [6] Brillouin, L., 1962, *Science and Information Theory*, Academic Press, New York
- [7] Rényi, A., 1961, Proc. 4th Berkeley Symp. Math. Statist. Probability 1960, **1**, 547
- [8] Fadeev, 1957, Arbeiten zur Informationtheorie I, Berlin, deutscher Verlag der Wissenschaften, 85
- [9] Tsallis, C., 1988, J. Stat. Phys. **52**, 470
- [10] Lavenda, B.H. & Dunning-Davies, J., 2005, J. of Applied Sciences, **5**, 315
- [11] Havrda, J. & Charvat, F., 1967, Kybernetika (Prague) **3**, 30
- [12] Aczél, J. & Daróczy, Z., *On Measures of Information and their Characterizations*, Academic Press, New York.
- [13] Ruelle, D., 1991, *Chance and Chaos*, Princeton University Press, New Jersey
- [14] Landsberg, P.T., 1964, Nature, **201**, 485
Dunning-Davies, J., 1965, Nature, **208**, 576; 1969, Il Nuovo Cimento **64B**, 82
- [15] Lavenda, B.H. & Dunning-Davies, J., 2005, J. of Applied Sciences, **3**, 30; **5**, 920; **5**, 928