# "A few exciting words":  information and entropy revisited

**David Bawden and Lyn Robinson**
**Centre for Information Science**
**City University London**

## Abstract

A review is presented of the relation between information and entropy, focusing on two main issues: the similarity of the formal definitions of physical entropy, according to statistical mechanics, and of information, according to information theory; and the possible subjectivity of entropy considered as missing information. The paper updates the 1983 analysis of Shaw and Davis. The difference in the interpretations of information given respectively by Shannon and by Wiener, significant for the information sciences, receives particular consideration.

Analysis of a range of material, from literary theory to thermodynamics, is used to draw out the issues. Emphasis is placed on recourse to the original sources, and on direct quotation, to attempt to overcome some of the misunderstandings and over-simplifications which have occurred with these topics.

While it is strongly related to entropy, information is neither identical with it, nor its opposite. Information is related to order and pattern, but also to disorder and randomness. The relations between information and the 'interesting complexity', which embodies both pattern and randomness, are worthy of attention.

# "A few exciting words":  information and entropy revisited

> THOMASINA: When you stir your rice pudding, Septimus, the spoonful of jam spreads itself round making red trails like the picture of the meteor in my astronomical atlas. But if need stir backward, the jam will not come together again. Indeed, the pudding does not notice and continues to turn pink just as before. Do you think this odd?
> SEPTIMUS: No
> THOMASINA: Well, I do. You cannot stir things apart.
> SEPTIMUS: No more you can, time must needs run backward, and since it will not, we must stir our way onward mixing as we go, disorder out of disorder into disorder until pink is complete, unchanging and unchangeable, and we are done with it forever. This is known as free will or self-determination.
> Tom Stoppard, *Arcadia*, (Act 1, Scene 1), London: Faber and Faber, 1993

> If information is pattern, the non-information should be the absence of pattern, that is, randomness. This commonsense expectation ran into unexpected complications when certain developments within information theory implied that information could be equated with randomness as well as with pattern. Identifying information with *both* pattern and randomness proved to be a powerful paradox (Hayles 1999, p. 25)

## Introduction

This paper re-examines the relations between information and entropy, a topic of debate for nearly a century. It can be seen as updating, thirty years on, of the analysis of Shaw and Davis (1983), presenting new findings and new perspectives. In order to make proper sense of these, it is necessary to delve into the historical development of the issues. Where possible, we have used quotations from original sources, as this is an area where there has been much misunderstanding and oversimplification.  Our intention here is to try to understand what the original authors meant, by considering their own words in context, thus minimising the potential for misrepresenting them in our commentary. Such an exegetically cautious approach to source material may be a technique generally appropriate for studying the history of information science [a point graciously made by an anonymous referee]. We give particular attention to the differences between Claude Shannon and Norbert Wiener in interpretation of the information entropy formalism, and the question of the subjectivity of entropy according to personal knowledge, as both of these points are of particular interest for information science.

Necessarily, we have focused on certain aspects of a very broad topic. Good reviews of other aspects are given by Furner (2014), Logan (2012), Burgin (2010), and Schneider and Sagan (2005). This focus on one specific aspect of the information concept is both a strength and a weakness. It does not allow us to deal with other

central concepts of the information sciences, such knowledge, documents, and signs (see, for example, Capurro and Hjørland, 2003; Floridi, 2010; Robinson and Bawden, 2014; Mingers, 2014); nor does it address the idea that the information sciences do not need information as their fundamental concept (Furner, 2004). Nonetheless, we think it is of value to look critically and in detail at the relation between information and entropy, from the perspective of information science. We have three reasons for holding this view.

The first reason is that numerous writers have pointed out that consideration of the idea of entropy provides the strongest and most direct link between the conceptions of information in the physical and social realms: for example

> ".. there is one relationship that might give us cause to at least pause and think before concluding that only physical properties can shape physical reality. It lends some credibility to the notion that information itself might be considered as a *physical* thing. This is the relationship between information and entropy" (Baggott 2013, p. 241)

and

> "by identifying entropy with missing information, Boltzmann hurled the concept of information into the realm of physics" (von Baeyer 2003, p. 98).

Lossee (1997) insightfully referred to information and entropy as "cousins"; there is indisputably a family resemblance, but it is not necessarily a close one. Wicken (1987A, p. 18) wrote of the "semantic haze" that had developed over the decades concerning the meanings of information and entropy.

It may be that a clear and unequivocal link may not be found; perhaps the various meanings of information are simply incommensurable, and the link is only the use of the same English word (Bawden, 2001). However, we believe that a clarification of the information/entropy relationship should help to show, in more general terms, the most fruitful ways of exploring the relations between ideas of information in different realms, to the benefit of all the disciplines which have information as an important concept (Robinson and Bawden 2014, Bawden and Robinson 2013). Shaw and Davis (1983) addressed this issue thirty years ago, and many of the points they noted are still the subject of debate, although it may be said that Wicken's 'semantic haze' is dissipating.

The second reason is that recent development and ideas have cast light on a decades old problem; the debate as to whether information is one thing, or its exact opposite. We focus in particular on the way in which the relation between information and entropy has been interpreted in different, and sometime in diametrically opposite, ways. These divergences may be seen as an example of the way in which a theoretical framework may affect the meaning and significance of any given concept (see, for example, Hjørland, 2000, 2009, 2011). As a conceptual analysis, this study may add to discussions of the epistemological status of some of

the concepts of information science (see, for example, Bates, 2006; Hjørland, 2007; Furner, 2014).

Entropy has always has the reputation of being a rather mysterious and wide-reaching quality. Thus, Ford, writing from a physical science perspective, and arguing that "… entropy need not be enveloped in mystery. It is a thermodynamic property of a physical system, obtainable from measurements of the heat capacity", also acknowledges that "entropy and its increase are indeed culturally important because they present us with a deep perspective on events in the world" (Ford 2013 p. 256 and p. 258). The idea of entropy, in some contexts, became far-removed from its origins, becoming a social and literary trope, invoking ideas of complexity, patterns, disorder, and chaos; see, as examples, Moorcock (1972), Pynchon (1973), Rifkin (1980) and Greer (2009).  Shaw and Davis (1983) give numerous early examples of the application, or possibly sometimes misapplication, of the entropy concept, Simberloff (1978) analyses its uses in Thomas Pynchon's literary works, and its literary and cultural applications are analysed in detail by Hayles (1990, 1999), and the contributors to Hayles (1991), especially White (1991). This extension has not been without its critics, Kostic (2014, p. 954), for example, giving the, perhaps rather exaggerated, warning that "entropy is the most used and often abused concept in science, but also in philosophy and society". The American economist Paul Samuelson remarked that it was ".. the sign of a crank or a half-baked speculator in the social sciences is his search for something in the social system that corresponds to the physicist's notion of 'entropy' " (Samuelson 1965, p. 450); a somewhat ironic perspective, as Samuelson himself was a pioneer of the application to economics of concepts from physical chemistry generally, and thermodynamics particularly (Beinhocker 2006).

Even within a strictly scientific context, ideas of how entropy should be understood have changed with time, as Holt (2012, p. 61) summarises:

> "The concept of entropy is among the most fundamental in science. It explains why some changes are irreversible and why time has a direction, an "arrow" pointing from past to future. The notion of entropy arose in the nineteenth century from the study of steam engines, and originally concerned the flow of heat. Soon, however, entropy was rethought along more abstract lines, as a measure of the disorder or randomness of a system. In the twentieth century, entropy became still more abstract, merging with the idea of pure information."

And indeed, as we shall see, there is still controversy and debate about the interpretation of entropy, in its purely scientific sense. Greven, Keller and Warnacke 2003, p. xii), who consider entropy as quantifying various aspects of complexity of systems, comment on "the large number of different meanings attributed to this word" just in mathematics and closely related areas. Styer (2000) identifies 'disorder', 'randomness', 'smoothness', 'dispersion', 'homogeneity' and 'mixed-up-ness' as ways in which it has been understood, to which we will later add

'uncertainty', 'freedom', 'probability', 'information' and arguably 'negative information'.

The third reason is that the concept of information is inextricably interwoven with the concept of entropy. Given that the concept of information is, by definition, at the heart of information science, its relation to other fundamental concepts is of importance for the foundations of the discipline. The relation shows itself in two main respects: a formal similarity between the mathematical definitions of entropy and of information; and a vexing question as to whether the objective physical quantity of entropy is due to a subjective lack of information or knowledge. Conceptual analysis of these issues, using the approach followed here of a close reading of the original sources, may cast light on how foundational concepts of this kind can best be understood.

We will now consider these inter-related conceptions, beginning with entropy as it was first understood in the context of classical thermodynamics.

***Thermodynamic entropy***

The concept of entropy emerged in the mid-nineteenth century, from the study of the design of steam engines, with the very practical engineering aim of improving their efficiency (Müller 2007, Atkins 2007, 2010, Carroll 2011, Hemmo and Shenker 2012). The French engineer Nicolas Léonard Sadi Carnot initially showed that the most efficient engine possible cannot be perfect; some energy must be lost in its operation. The German physicist Rudolf Clausius coined the term 'entropy, from the Greek *trope*, meaning transformation, from an analysis of these phenomena (Clausius 1865), since he regarded it as a measure of the 'transformation content' of a body, its capacity for change (Denbigh 1981).

Consider a hot object that is cooling down, and losing heat to its surroundings; if, at each moment, the heat being lost is divided by the temperature, then the sum of this quantity over the whole process is the entropy. The evident fact that heat always flows spontaneously from hot bodies to cooler bodies, and never the other way round, is equivalent to the idea that the entropy of a closed system always increases, never decreases; the well-known Second Law of Thermodynamics. Entropy was generally regarded as a measure of the energy in a system that is in some way poor quality, 'lost', useless or unavailable. A more recent interpretation is that a concentration of energy which is dispersed, typically as heat, around a physical system cannot readily be retrieved to do useful work (see, for example, Sethna 2006).

It is represented most simply, in current notation, as

$$\Delta S = \Delta Q / T$$

$\Delta S$ being the entropy change associated with a change in heat $\Delta Q$, with T the absolute temperature. This can be thought of as the ratio of the quantity of energy to the intensity of energy (Schneider and Sagan 2005). In effect, energy changes its form over time, so that it becomes less useful; entropy quantifies the amount of energy within an isolated system that cannot be used without investing work. The energy of a system is divided into two parts: a part that is exploitable as work and a part that is not:

$$\Delta E = \int p dV - \int T dS$$

where $\Delta E$ is some change in the internal energy of a system between two equilibrium states, $\int p dV$ is the work produced during that transformation, and $\int T dS$, where T is the temperature of the system, is taken to be the change of energy to a form that is not exploitable as work, and dS is the entropy difference during the transformation.

Modern thermodynamics has more complex variants; for example, in the case of a gas differing in temperature, pressure and chemical potential (the tendency to exchange particles) with its surroundings, the entropy change is given as (Ford 2013):

$$dS_i = dS - dE/T - (P/T)dV + (\mu/T)dN$$

where $dS_i$ is the entropy change due to these factors, $dS$ is the total entropy change, $dE$ is the energy (heat) change, $T$ is the temperature, $P$ is the pressure, $dV$ is the change in volume, $\mu$ is the chemical potential, and $dN$ is the change in number of particles.

The details are unimportant. The point to be emphasised is that classical thermodynamic entropy was derived as a straightforward objective property of bulk matter, which can be measured with mundane instruments, typically a thermometer, without any of its later associations of disorder, uncertainty and subjectivity. And indeed, there are still some scientists who argue that this is where the entropy concept should remain: "[Entropy] is nothing by itself. It has to be seen and discussed in conjunction with temperature and heat, and energy and work. And, if there is to be an extrapolation of entropy to a foreign field, it must be accompanied by the appropriate extrapolations of temperature and heat and work" (Müller 2007, p. 126).

What entropy was, in itself, was unknown: an understanding of its nature first emerged in the development of statistical thermodynamics.

**Statistical entropy**

Statistical thermodynamics, pioneered by the Austrian physicist Ludwig Boltzmann in the late nineteenth century, recognizes that the properties of matter on a large scale are determined by the behaviour of the individual particles – generally atoms and molecules – of which it is composed (Ford 2013, Müller 2007, Atkins 2007, 2010, Carroll 2011). We cannot know the details of this behaviour – typical thermodynamic systems on a human scale contain around $10^{23}$ particles - but we can understand it statistically. Boltzmann developed what is now termed statistical mechanics in a number of publications over more then 20 years, but it is usually taken to have originated in his analysis of the thermal properties of gases (Boltzmann 1872). He showed that entropy was related to the probability of collisions between gas particles with different velocities, and could be equated to the probability distribution of the states of a system, expressed by the formula

$$S = k \log W$$

S being the entropy, k is Boltzmann's constant, and W is a measure of the number of states of a system; in general, the number of ways that molecules can be arranged, given a known total energy. (Although this equation, using an $\Omega$ symbol instead of the more modern W, is carved on Boltzmann's tombstone in the cemetery in Vienna, he never explicitly wrote the formula, in this form, which is due to Max Planck).

In modern statistical mechanics, W is usually taken as the Lebesgue measure, a mathematical generalization of the intuitive idea of size, but that does not affect the discussion here. The logarithm function is introduced simply to bring the very large numbers involved down to more understandable quantities, and to allow entropies to be added together. Strictly, the entropies of multiple systems are additive, but the total number of arrangements is the product of the numbers of arrangements in each system, not the sum; so it is the logarithms of the numbers of arrangements, not the arrangements themselves, which are additive (Albert 2000).

The values in Boltzmann's equation, counts of numbers of states, can be equated to measured thermodynamic entropy through the constant k; changes in one can be directly related to changes in the other, though they may seem conceptually very different

Boltzmann's idea is generally expressed through the concept of the *macrostate*, a physical situation which we can recognize and distinguish, and the *microstate*, the detailed situation at the level of individual particles, which we cannot. For example, if we have two beakers, each containing the same volume of pure water at room temperature, we will not be able to distinguish between them, and we will say that they are the same macrostate. However, the detailed pattern of positions and velocities of the water molecules in each glass, the microstates, will differ between the glasses, and within each glass from moment to moment. Boltzmann relates this directly to entropy, by showing that entropy was equivalent to the number of microstates in a macrostate.

As Carroll (2011 p. 36-37) puts it:

> ".. the great triumph of kinetic theory was its use by Boltzmann in formulating a microscopic understanding of entropy. Boltzmann realized that when we look at some macroscopic system, we certainly don't keep track of the exact properties of every single atom. If we have a glass of water in front of us, and someone sneaks in and (say) switches some of the water molecules around without changing the overall temperature and density and so on, we would never notice. There are many different arrangements of particular atoms that are *indistinguishable* from our macroscopic perspective. And then he noticed that low-entropy objects are more delicate with respect to such rearrangements. If you have an egg, and start exchanging bits of the yolk with bits of the egg-white, pretty soon you will notice. The situations that we characterize as "low-entropy" seem to be easily disturbed by rearranging the atoms within them, while "high-entropy" ones are more robust."

Entropy increases in physical changes because systems are more likely to move to macrostates with a larger number of possible microstates; there are simply more ways to be high entropy than to be low entropy. As an example, imagine a large glass tank filled with water, into which a small bottle of blue ink has been poured; the ink will initially form a small irregular blue cloud in the water. If we leave the situation for some considerable time, we will expect to find that the ink has spread throughout the tank, giving a uniform light blue colour; we would be very surprised to find that it has formed a very small cube of deep blue in one corner of the tank. This is simply explained in Boltzmann's terms. There are very many more microstates (specific arrangements of molecules) which will give the uniform distribution of ink that those which will give a small concentration of ink, and therefore the random movements of the ink and water molecules are overwhelmingly likely to bring about one of the uniform microstates. This is a statistical probability, not a deterministic law; if we watch the tank long enough, we will certainly find a small ink spot formed, but we will have to wait many times the age of the universe before this happens.

This statistical understanding of entropy, like the thermodynamic formulation, relates to objective physical properties of matter, and the two are related. As Hemmo and Shenker (2012, p. 57-58) put it:

> "Intuitively, the idea here is that the larger the size of a macrostate, the larger the set of microstates in that macrostate, so the less we know about which is the actual microstate within that set, and therefore the less control we have over the actual microstate within that macrostate, and therefore, in turn, the less exploitable is the energy in order to produce work – which is what entropy is about in thermodynamics. Notice that this expression is about the entropy of a state, and not about entropy differences as in thermodynamics. Moreover, this notion of entropy is defined for all macrostates, and not only for equilibrium macrostates as in thermodynamics. This is a conceptual generalization, involving a profound change that is brought about by replacing thermodynamics by mechanics."

This generalizing of the concept of entropy from its very specific thermodynamic origins to deal with probability and uncertainty is a recurring theme in the development of the understanding of the relationship between entropy and information.

Boltzmann introduced the idea that entropy may be equated with disorder; the high-entropy 'robust' systems described above by Carroll being thought of as disordered. This became a very familiar viewpoint, entropy being regarded as the influence behind everything tending to disorder and chaos. This is not altogether an accurate reading of statistical entropy, as will be discussed later. An alternative viewpoint is to regard entropy in terms of uncertainty at the microscopic level, since we can specify a system only at the macroscopic level.

There emerges, worryingly for the objective physical sciences, the possibility that entropy may be, to some extent, subjective. Entropy is a measure of the number of particular microscopic arrangements of atoms, or other constituents of a system, that appear indistinguishable from a macroscopic perspective. It relies on the identification of macrostates, since entropy is due to the number of microstates associated with them; but macrostates may be determined differently by different observers. To take the example above of the beakers of water in the same macrostate, if a second observer has a more accurate thermometer they may determine that the water in one beaker is slightly above room temperature; the macrostates will not be the same, and any calculation of entropy will be different. Or, as Carroll (2011, p. 159) asks, what if we encountered a race of super-observant extraterrestrials who could peer into a glass of liquid, and observe the position and momentum of every molecule? Would they think that there was no such thing as entropy? This subjective aspect to entropy has troubled physicists ever since the entropy/disorder relation was proposed, as will be discussed further below.

Further, and crucially for our purposes, Boltzmann's formalism allows entropy to be understood in terms of information, and more specifically of 'missing information'. It rests on our not being able, to a greater or lesser extent, to distinguish between the microstates in a given macrostate. The macrostate is what we know – the specification of the macrostate is the limit of our information. Entropy is the missing information – the uncertainty. If there were no missing information, no uncertainty, the macrostate and microstate would be the same, and there would be no entropy. If, that is to say, we had perfect information, then we could identify simply a state of the system, with microstate and macrostate being equivalent. As we do not, the size of the macrostate amounts to the extent of our ignorance, of the amount of missing information; states of high entropy have many possible microstates, and therefore may be said to correspond to a large information deficit. Statistical entropy can therefore be seen to arise from, and potentially be equated to our lack of information. As Albert (2000, p. 50-51) puts it:

> "The fact that the entropy of a macrocondition is in some sense a measure of *how many* microconditions are compatible with it means that entropy has

something to do with *information*. Entropy (that is) is a measure of *how much one can infer* about a system's microcondition from knowledge of its macrocondition. The higher the entropy of a macrocondition, the larger the volume of phase space which is compatible with it, the larger the number of microconditions which are compatible with it, the less information that macroconditions carries, the less a knowledge of that macrocondition can tell you"

"And entropy clearly has something to do with (at least) intuitive ideas of *randomness* and *disorder*. Conditions with higher entropies are in some sense less *structured*, less *arranged*, less *bunched-up*, more *dispersed*, more of *a mess*, than those with lower entropies."

It is for this reason that von Baeyer makes the dramatic statement, quoted above, about Boltzmann hurling information into physics, although it should be noted that Boltzmann never used the language of information in any of his writings. This issue will be discussed later.

Boltzmann's concept of entropy was based on the probabilities of individual particles in an isolated system having certain positions and momentums. This was generalised by the American chemist J. Willard Gibbs, who derived entropy in terms of probability functions of holistic states of systems rather than of collections of individual particles, while retaining Boltzmann's main insights (Gibbs 1902). (Gibbs, in fact, carefully termed his functions 'entropy analogues', to avoid the claim that they were identical to thermodynamic entropy (Denbigh 1981)). This approach allows us to know important aspects of thermodynamic regularities without needing to worry about defining macrostates. It is also more widely applicable in that, unlike Boltzmann's formula, it is applicable to systems interacting with their environment. Gibbs' formula for entropy is:

$$S_G = -k \sum p_x \log p_x$$

where $p_x$ is the probability that the system is in microstate x, k is Boltzmann's constant, and the summation is over all possible microstates.

This formulation is often used in statistical mechanics because it is easier to use in calculation than Boltzmann's formula, but it carries an important conceptual change. As Carroll (2011, p. 169-170) puts it:

"Instead of thinking of entropy as something that characterizes individual states – namely, the number of other states that look macroscopically similar – we could choose to think of entropy as characterizing *what we know* about the state. In the Boltzmann way of thinking about entropy, the knowledge of which macrostate we are in tells us less and less about the microstate as entropy increases: the Gibbs approach inverts this perspective and defines entropy in terms of how much we know…. it associates the idea of 'entropy' with our knowledge of the system, rather than with the system itself. This

has caused all kinds of mischief among the community of people who try to think carefully about what entropy really means."

The measures discussed so far have been derived for the domain of classical physics. Equivalents were introduced during the 1930s, most notably by von Neumann, for systems of quantum particles. Although these are of great practical importance for quantum computation (see, for example, Nielsen and Chang 2000 and Fayngold and Fayngold 2013), they do not bring any new conceptions of entropy relevant to this discussion.

A contribution to the debate on entropy and information, though it was not presented or recognized as such, came in a paper by the Hungarian-American physicist Leo Szilard  (1929), as a contribution to a long-running problem known as Maxwell's Demon; a thought experiment on the nature of entropy (Leff and Rex 1990). The demon is an imaginary tiny creature, who operates a frictionless trapdoor to separate high-energy and low-energy gas particles, and thus decrease entropy, breaking the second law or thermodynamics by using its intelligence and information. Szilard showed that the second law would not be violated if the demon's decision to open or close the shutter - a measurement of a binary state - was equivalent to an increase in the entropy of the overall system of $\Delta S = k \log 2$.  This exactly matches the decrease in entropy of the gas, due to the demon's action. The entropy of the system 'informed demon plus segregated gas' is equal to that of the system 'uninformed demon plus unsegregated gas'. Although Szilard did not state the matter in exactly these terms, this implies that the information gained by the demon can be expressed as $\Delta I = - k \log 2$; the negative sign implying that the increase in information corresponds to a decrease in the entropy of the gas. As Hargatti (2006, p. 46) puts it, "the solution included information transfer, although the term itself was not mentioned, as it was not yet in usage as a mathematical concept". This was later established, by Brillouin and others, as the minimum entropy production in any act of measurement, i.e. any gain of information (Denbigh and Denbigh 1985).

**Information entropy**
The final step in the generalization of the entropy concept, and simultaneously in its increasing involvement with information, came with the publication of Claude Shannon's well-known "mathematical theory of communication" (MTC) (Shannon 1948, Shannon and Weaver 1949). Shannon was dealing with a rather limited context: the efficiency of the transmission of 'meaning-free' messages by communication systems, building on the work of Nyquist and Hartley (Gleick 2011, Lossee 1997). However, his formalism, and his idea of a quantitative measure of information, soon took on a wider significance, Mackay arguing in 1953 that 'information theory' had a wider meaning than just Shannon's communication theory (and commenting that future historians of science might find it strange that despite the activity in the United States following the publication of Shannon and Weaver's book, the first international symposium on information theory was held in London in 1950 (MacKay 1969).

Shannon presented his formula for H, the measurement of information content, as

$$H = -K \sum p_i \log p_i$$

where $p_i$ is the probability of each symbol, and K is a constant defining the units. The minus sign is included to make the quantity of information, H, positive; necessarily, a probability will always be less than 1, and the log of such a number is always negative. He pointed out that formulae of the general form $H = -\sum p_i \log p_i$ appeared very often in information theory, as measures of information, choice and uncertainty; the three concepts seeming virtually synonymous for his purposes. Applied in a thermodynamic context, it follows the Gibbs approach in dealing with the probabilities of states which cannot be known with certainty. It is, however, much more general, in that it can be shown to be compatible with both the Boltzmann and Gibbs entropies, and can be applied more widely than either, to physical systems which are not in equilibrium (Ford 2013, p. 131).

The form of the equation is very similar to those for physical entropy discussed above, and, notoriously, Shannon then gave the name 'entropy' to his quantity H. It is usually stated that he was urged to do so by von Neumann, the original source for this being Myron Tribus, who quoting a private discussion between himself and Shannon in Cambridge, Massachusetts, on March 30th 1961, gives the following account:

> "When Shannon discovered this function he was faced with the need to name it, for it occurred quite often in the theory of communication he was developing. He considered naming it 'information' but felt that this word had unfortunate popular interpretations that would interfere with his intended uses of it in his new theory. He was inclined towards naming it 'uncertainty', and discussed the matter with the late John Von Neumann. Von Neumann suggested that the function ought to be called 'entropy' since it was already in use in some treatises on statistical thermodynamics. Von Neumann, Shannon reports, suggested that there were two good reasons for calling the function 'entropy'. 'It is already in use under that name', he is reported to have said, 'and besides, it will give you a great edge in debates because nobody really knows what entropy is anyway'. Shannon called his function 'entropy' and used it as a measure of 'uncertainty', interchanging between the two words in his writings without discrimination". (Tribus 1964, p. 354)

He slightly embellished the story it in a later version (Tribus and McIrvine 1971, p. 180), quoting Shannon as saying:

> "My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used. So I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one

knows what entropy really is, so in a debate you will always have the advantage."

Some commentators have doubted this story, including Gleick (2011) who calls it untrue, though plausible, and Vedral (2010), for whom it is a plausible urban myth. Smith (2001 p. 2) suggests that the "amusing anecdote is spoilt" by some pre-history: Tolman (1938) attributes this formula, with its application in statistical mechanics and relation to thermodynamic entropy, to Pauli (1933).

Others accepted it, and thought the less of von Neumann for it. Denbigh, for example, argued that "in my view von Neumann did science a disservice!" (Denbigh 1981, p. 172). Although information theory and statistical mechanics have a common basis in probability theory, and will require functions with the same mathematical structure, "this formal similarity does not imply that the functions necessarily signify or represent the same concepts…. It remains to be seen under what conditions, if any, thermodynamic entropy and information are mutually interconvertible" (Denbigh 1981, p. 172). Denbigh reported other suggestions, which he clearly found more acceptable, as more neutral names for Shannon's H function, such as 'spread' or 'dispersal'. Ben-Naim (2008, p. xviii) concurs with Denbigh's criticism arguing that information, choice and uncertainty are all simple, familiar, and meaningful terms, while "entropy merely corrupts the term information".

Donald Mackay was one of the first to express concern about this equivalence. In a talk broadcast on BBC radio in 1950, he noted the identical nature of the equations: "the units of course are different; but otherwise the only distinction is a difference in sign: where information is given out, entropy increases. It is possible to make too much of this resemblance" (MacKay 1969, p. 16). He later wrote that "the [Shannon] entropy of selective information-content of a selection should not be facilely identified with the physical entropy of thermodynamics. The two are equivalent only in the particular case where the ensemble from which the selection is made is a physical one defined for a state of thermodynamic equilibrium" (MacKay 1969 p174). Nonetheless MacKay was inspired to suggest in his 1950 talk that there might be an information equivalent to the second law of thermodynamics: that an isolated system can only lose or give out information, never gain it.

Schneider and Sagan (2005, p. 22) also consider that Shannon's use of the entropy term was confusing, since there was no simple correspondence between the thermodynamic and information entropies, and "no general equivalence of thermodynamics and information theory". Wicken (1987A, 1987B) argues that, although the basis of both Shannon entropy and thermodynamic entropy is uncertainty, the uncertainties are of a different nature, and it is wrong to suggest that Shannon's formula is a generalization, applicable also in the thermodynamic context. The entropy concept is also strictly applicable to ensembles of states (or messages), whereas Shannon's formula was applied to single messages. Complexity would have been a better term than entropy, since "what the Shannon formula measures, simply, is complexity of structural relationships" (Wicken 1987A, p. 24).

Müller (2007, p. 124) is even more dismissive:

> "No doubt Shannon and von Neumann thought that this was a funny joke, but it is not – it merely exposes Shannon and von Neumann as intellectual snobs…. If von Neumann had a problem with entropy, he had no right to compound that problem for others … by suggesting that entropy has anything to do with information."

In fact, the American physicist Richard Tolman had, a decade earlier, used the same formula as presented by Shannon, in the context of thermodynamic entropy, carefully qualifying its use by noting:

> "the statement sometimes made that the entropy of a system is a measure of the degree of our ignorance as to its condition. From a precise point of view, however, it seems more clarifying to emphasise that entropy can be regarded as a quantity which is thermodynamically defined with the help of its relations to heat and temperature … and statistically interpreted with the help of the analogous relation between mechanical quantities" (Tolman 1938, p. 561).

Essentially the same point was made at about the same time by Slater (1939), who also related a Shannon-like formula to physical entropy.

Even earlier, the American chemist G.N. Lewis (1930, p. 573) had expressed essentially the same idea:

> "Gain in entropy always means loss of information, and nothing more. It is a subjective concept, but we can express it in a least subjective form, as follows. If, on a page, we read the description of a physico-chemical system, together with certain data which help to specify the system, the entropy of the system is determined by these specifications. If any of the external data are erased, the entropy becomes greater; if any essential data are added, the entropy becomes less."

Ben Naim (2008 p. 20) suggests that Lewis's insistence that entropy was subjective was due to his use of a general idea of information, not the specific and quantitative measure introduced by Shannon. Qvortrup (1993) has argued that Shannon and Weaver were unclear as to whether they regarded information as a substance or as a sign; but this apparent lack of clarity has not altered the general perception of Shannon' theory as dealing with objective, rather than individual and subjective, information. We will consider the debate about the subjectivity character of entropy below.

Despite these qualifications, Shannon's formula has been, and remains, widely used for the calculation of physical entropy, under the title of 'Shannon entropy' or 'information entropy' or 'statistical entropy', although it is strictly only a measure of quantity of information. Jaynes, as noted below, developed the link between

physical and statistical entropy mathematically a decade later. Katz gives a good example of the nature of the link: statistical mechanics involves calculation and prediction when we do not have full information, and probabilistic information theory allows us to make best estimates in such circumstances: "knowledge, or rather the lack of it, brings statistical mechanics into being" (Katz 1967, p. 12).

The analogy has been cogently expressed by Gleick (2011, p. 280):

> To the physicist, entropy is a measure of uncertainty about the state of a physical system: one state among all the possible states it can be. These microstates may not be equally likely, so the physicist writes $S = \sum p_i \log p_i$
>
> To the information theorist, entropy is a measure of uncertainty about a message: one message among all the possible messages that a communications source can produce. The possible messages may not be equally likely, so Shannon wrote $H = -\sum p_i \log p_i$.

The interpretation has been accepted in different ways. At the extremes, as we have seen, there are those who emphasise solely the physical nature of entropy: Lieb and Yngvason (2003), for example, deny that statistical mechanics, still less information theory, is needed for a full understanding of entropy, while Kostic (2014 p. 966) argues that "expanding [the] classical entropy concept to other types of disorder or information is 'overreaching' and could be the source of many misconceptions". Others, conversely, have sought to recast thermodynamics in largely or solely information terms (see, for example, Tribus 1961, Landsberg 1961, Katz 1969, Duncan and Semura 2004, 2007, Ben-Naim 2008 and Vieland, Das, Hodge and Seok 2013).

Warren Weaver, in his commentary on the first presentation of Shannon's MTC attempted an explanation of the information-entropy link, regarding the emergence of an entropy-like quantity in a mathematical theory of communication as "most significant". He nurtured the hope, as yet unfilled, that Shannon's theory could be extended to encompass meaning, encouraged by the view expressed by the leading British physicist Sir Arthur Eddington (1928) that physical entropy was more associated with concepts like beauty and melody, than with ideas like distance, mass and force. This idea was taken up enthusiastically by Landsberg who, in a book largely devoted to thermodynamics, nonetheless managed to suggest that Shannon entropy might be applied in biology, music, linguistics, economics, and psychology, and even suggested a thermodynamic analogy for Jeremy Bentham's utilitarianism, with entropy standing in for happiness (Landsberg 1961, pp. 234-240).

Noting that physical entropy was associated with randomness and 'shuffledness', Weaver argued that we could reasonably say:

> "This situation is highly organized, it is not characterized by large degree of randomness or of choice – that is to say, the information (or the entropy) is low" (Shannon and Weaver 1949, p. 103).

Or conversely, as Schneider and Sagan (2005, p. 23) put it:

> "Information … in information theory is equivalent not to order but to disorder in the sense that it takes more binary decisions – more ones and zeros, more bits of information – to describe disorderly situations of objects than orderly ones."

This view has come to be the most generally accepted, and its nature, power, and limitations are set out particularly concisely and by Ruelle (1991). As Smith (2001, p. 5) puts it "the three concepts of 'entropy', 'probability' and 'randomness' turn out to be equivalent, in the strict mathematical sense that establishment of any one leads to establishment of the other two." Indeed, the formula used by Boltzmann and Shannon had its origins in the eighteenth century study of probability in games of chance by the French mathematician deMoivre, and adapted to other domains where uncertainty is of major importance (Wicken 1987A, 1987B, Schneider and Sagan 2005).

Physical entropy is a measure of the amount of randomness in a system. Analogously, a message with a high Shannon information content is one which is very random; it is extracted from a large class of possible messages. This randomness may amount, in whole or in part, to useful information, or to nonsense. It is applicable to 'messages' such as musical melodies or works of art, whose information content may be measured, though it will in no way amount to a measure of quality. Nor can Shannon's theory address the "deep and complex" (Ruelle 1991, p. 134) issue of meaning.

A large information content is therefore, in a sense, 'the same as' a high entropy value. This viewpoint was very soon to be challenged.

**Shannon and Wiener, information and entropy**

At almost the same time as Shannon was publishing his MCT, Norbert Wiener (1948) published his introduction to the field of cybernetics. (For a historical account of this period, and the interactions between the main characters, including Shannon, Wiener and von Neumann, see Gleick 2011). Albeit starting from the very different perspective of an attempt to understand signaling and control in both animate and inanimate systems, Wiener's cybernetics also proposed a quantitative measure of information. This was essentially the same as Shannon's, with one significant difference, to be noted later. Weiner notes the simultaneity of the developments:

> ".. we had to develop a statistical theory of the *amount of information*, in which the unit amount of information was that transmitted as a single decision between equally probable alternatives. This idea occurred at about the same time to several writers, among them the statistician R.A. Fisher, Dr. Shannon of the Bell Telephone Laboratories, and the author. Fisher's motive in studying this subject is to be found in classical statistical theory; that of Shannon in the problem of coding information; and that of the author in the problem of noise and message in electrical filters" (Weiner 1948, p. 104).

It seems clear that Wiener regarded the quantitative definition of information in his work to be equivalent to that of Shannon. A distinction is that Wiener immediately makes a link with entropy, in a seemingly more direct manner than did Shannon:

> "The notion of the amount of information attaches itself very naturally to a classical notion in statistical mechanics: that of *entropy*. Just as the amount of information in a system is a measure of its degree of organization, so the entropy of a system is a measure of its degree of disorganization; and the one is simply the negative of the other… We have said that amount of information, being the negative logarithm of a quantity which we may consider as a probability, is essentially a negative entropy…. It will be seen that the processes which lose information are, as we should expect, closely analogous to the processes which gain entropy. … No operation on a message can gain information on the average. Here we have a precise application of the second law of thermodynamics in communication engineering. Conversely, the greater specification of an ambiguous situation, will … generally gain information, and never lose it." (Wiener 1948, pp. 18 and 78-79)

And similarly in a slight later work:

> "A measure of information is a measure of order. Its negative will be a measure of disorder … This measure of disorder is known to the statistical mechanist as entropy .. [which concept] is associated with that of pattern, and represents the disorder in a class of patterns. Amount of information is a measure of the degree of order … [The] amount of information is a quantity which differs from entropy merely by its algebraic sign and a possible numerical factor. Just as entropy tends to increase spontaneously in a closed

system, so information tends to decrease; just as entropy is a measure of disorder, so information is a measure of order. (Wiener 1950, pp. 18-19, 20-21 and 129)

The difference in the Shannon and Wiener formulae was simply a minus sign. Their calculations gave the same numerical value but apparently opposite meanings. For Shannon, information was equivalent to entropy; for Wiener it was equivalent to the opposite, to negative entropy or 'negentropy'.

Hayles (1999, pp. 102 and 305) explains the distinction in detail :
"although [Wiener and Shannon] conceived of information in similar ways, Wiener was more inclined to see information and entropy as opposites … Like Brillouin and many others of his generation, Wiener accepted the idea that entropy was the opposite of information. The inverse relation made sense to him because he thought of information as allied with structure and viewed entropy as associated with randomness, dissipation and death… Claude Shannon took the opposite view and *identified* information and entropy rather than opposing them. Heuristically, Shannon's choice was explained by saying that the more unexpected (or random) a message is, the more information it conveys"

Gleick (2011, p. 281) sums it up in this way:

"Where Shannon identifies information with entropy, Wiener said it was *negative entropy*. Wiener was saying that information meant order, but an orderly thing does not necessarily embody much information. Shannon himself pointed out their difference and minimized it, calling it a sort of "mathematical pun". They get the same numerical answers, he noted: "I consider how much information is *produced* when a choice is made from a set – the larger the set the *more* information. You consider the larger uncertainty in the case of a large set to mean less knowledge of the situation and hence *less* information."

Put another way, H [in Shannon's sense] is a measure of surprise. Put yet another way, H is the average number of yes/no questions needed to guess the unknown message. Shannon had it right – at least, his approach proved fertile for mathematicians and physicists a generation later – but the confusion lingered for some years. Order and disorder still needed some sorting."

It may be said that in the information sciences, where Shannon's approach has proved distinctly unfertile (Capurro and Hjørland  2003, Bawden and Robinson 2012, chapter 4), this question is not so readily answered. And, as we note below, issues of order and disorder relating to entropy itself still cause controversy. While Gleick is correct in saying that most scientists have accepted Shannon's understanding, others still do not: the physicist Lee Smolin, for example, recently declaring simply that information is the inverse of entropy (Smolin 2013). Gregory Bateson carefully

states that, in certain instances, information and negative entropy "overlap" (Bateson 1971, p. 231). Hayles (1990, pp. 57-58) analysed a set of textbooks on information theory published up to 1990, and concluded that they showed a clear divide on the information/entropy issue. Those written by electrical engineers follow Shannon, in concluding that the more uncertain a message is, the more information it conveys. But they ignore the obvious conclusion that maximum information is gibberish, nor do they give much space to the relation between information and thermodynamic entropies. Those written by physical scientists followed the Wiener/Brillouin approach, and gave much more discussion on the information / thermodynamic entropy relation.

Hayles (1990, pp. 59 and 270) endeavours to explain the difference in terms of personal experience, though using Brillouin, rather than Wiener, as the opponent to Shannon. Shannon worked for a telecommunications company that made its living by satisfying people's curiosity. In more uncertain times, people want more information - news, financial information, weather reports, etc. and they send more telegrams, make more phone calls, etc.  Shannon would naturally have thought information and uncertainty were allied. Physical scientists, especially those involved in thermodynamics, would see entropy as an inhuman chaos, and disorder as an evil, the opposite of understanding and knowledge. She also suggests that it was the experience of the second world war that made the association of information with uncertainty so compelling to Shannon. Intriguing as these ideas are, they are not fully convincing, given that Wiener had a similar background to Shannon. They do remind us, however, that historically contingent sensibilities play a part in understanding these issues.

Hayles (1990) also notes that one of the books she quotes as following Shannon's viewpoint, that of Raisbeck (1964), was "steeped in the engineering tradition" and was written by Wiener's son-in-law (p57). This is indeed a purely engineering text, follows Shannon, and does not discuss the Shannon/Wiener issue, nor physical entropy; and mentions Wiener only to list his *Cybernetics* as a reference. This is an indication that Hayles recognized a rather different distinction: to Shannon's view (information as entropy)she recognized the opposite (information as negative entropy) but associated this viewpoint with Brillouin rather than Wiener.

Donald MacKay argued that Shannon's formula measured, not information *per se*, but information-content, defined in terms of unexpectedness of messages selected from an ensemble, but noted, in a chapter written in 1953, that it took "the varying names of 'selective information-content', 'selective entropy', 'negentropy', or even simply 'entropy'" (MacKay 1969, p. 133). He took the view that Shannon information was the opposite of entropy, and "that which determines form", but noted "some ambiguity of sign in the literature" (MacKay 1969, p160 and 172).

Gell-Mann and Lloyd (1996, p45) argue that entropy is associated with uncertainty, and hence with ignorance. Information is the opposite of ignorance. Yet Shannon information, I, "measures ignorance or uncertainty. However, I also measures information. But ignorance and information are opposites: how can they be

measured by the same quantity?" Their answer is that the amount of information conveyed by a message (a coin toss or a telegram) is measured by the uncertainty of the tosser of the coin, or the reader of the telegram, before the event.

Kaufmann (2000, p. 88) explains it in this way: "The sum of these "*plogp*" terms for the total set of messages at the source is the entropy of the source. Reception of the signal reduces the receiver's uncertainty about what is being sent from the source, hence is a negative entropy. Shannon's information measure is, thus, just the negative of the normal entropy measure".

Shannon and Wiener agree about how to calculate uncertainty – about the exact microstate of a system, or the exact message which will be set – but they disagree about how this uncertainty relates to information. For Shannon a high uncertainty is a large amount of information, since the receipt of a message, or the identification of a microstate, will remove this great uncertainty. Indeed, as Weaver acknowledged, the logical, and seemingly strange, conclusion must be that when extraneous noise is introduced into a message, the increased uncertainty makes the message more informative (Wicken 1987B). For Wiener, this same high uncertainty is a small amount of information, since it is our lack of information which causes the uncertainty. In a sense, these are two sides of the same coin: Shannon focuses on the amount information which will be gained by receipt of a message, or the identification of a microstate, while Weiner considers the amount of information implicit in our current knowledge. Or, as Hayles (1990, pp. 558-559) puts it, the relation "is not especially complex or difficult … The difference between the two viewpoints can now be succinctly stated. *Shannon considers the uncertainty in the message at its source, whereas [Wiener] considers it at the destination*. To ask which is correct is like asking whether a glass is half empty or half full. The answer is important not because it is correct but because it reveals an orientation towards the glass and, by implication, an attitude towards life. Similarly, the [two] heuristics reveal different attitudes towards chaos by their orientations toward the message."

Shannon says that the more certain a message is, the less information it contains, Wiener that the more surprising a message is, the less information it contains. Both agree that entropy correlates with uncertainty. Shannon concentrates on the inherent uncertainty before the message is received, Wiener on the uncertainty that remains after the message is received. For this reason, Shannon entropy is often, and arguably better, regarded as 'prior uncertainty'. It is generally used in this sense when Shannon's formula is used as to assess uncertain outcomes, as in a study of interface design, where the users were uncertain, to various degrees as to the system's response to their input (Spiekermann and Korunovska 2014).

Wiener's concept of negentropy as an active force, associated with order, had been anticipated by Erwin Schrödinger, one of the pioneers of quantum mechanics, in a book based on a series of lectures given in Dublin in 1943 (Schrödinger 1944). In this, he attempted, as no-one had done before , to give an account of living systems in physical terms, and thereby arguably inaugurated the disciplines of biophysics and molecular biology.  Addressing the issue of how living organisms maintained their

orderly nature in a universe of increasing disorder, he invoked the thermodynamic entropy concept, expressed, following Boltzmann and Gibbs, as

entropy = k log D

where k is the Boltzmann constant, and D is 'a quantitative measure of the atomistic disorder of the body in question'. (Schrödinger 1944, p. 73). [Presumably D is here equivalent to W in the entropy formalism noted above, and Schrödinger used the D symbol for disorder in addressing a lay audience.]

He then argued that living organisms feed upon 'negative entropy'; plants obtaining this from sunlight, and animals from consuming 'well-ordered' materials as food. Finally, he concluded that "the awkward expression 'negative entropy' can be replaced by a better one: entropy, taken with the negative sign, is itself a measure of order" (Schrödinger 1944, p. 75).

Another early introduction of the idea that living creatures could create complex organization, and hence bring about a local reduction in entropy, was put forward by Blum (1951), who took the view that entropy, determined by the degree of randomness, was the opposite of complexity and organisation.

An early critical analysis of the relation between Shannon information content and thermodynamic entropy, and the first to specifically consider whether they were the same or opposite, was given by D.A. Bell, a British professor of electrical engineering at the universities of Birmingham and Hull. His contribution, which has largely gone unnoticed, was presented in his book on the engineering applications of information theory, first published in 1953, and running into four editions over fifteen years.

His view was stated unequivocally from the first. In the first two editions (1953 and 1956) he stated "Information is the negative of entropy. This means that the information content (or better the potential information content) of any waveform, collection of code symbols, or any other *pattern* can be assessed mathematically by the same process used to define the entropy of such a system" [author's italics] (Bell 1953, p. 1).

In the third edition (1962) and fourth edition (1968), this trenchant view was stated in a slightly more nuanced way: "Information in a certain sense is a measurable quantity which is independent of the physical medium by which it is conveyed: in this respect it may be compared with "pattern". The most appropriate measure of information is mathematically similar to the measure of entropy, but there are good reasons for reversing the sign and stating that information is the negative of entropy in nature as well as in mathematical formulation …. The measure of information is mathematically related to the measure of entropy, and there is also a physical argument for a relationship between information and the negative of entropy" (Bell 1962, pp.1 and 4-5).

His basic view never changed over the fifteen years covering the four editions of the book, the preface all editions stating: "Some space is devoted to an important question of principle, whether information should be equated to *negative* or *positive* entropy, and the author is firmly in favour of the former [author's italics] (bell 1953, p. v).  And in all four editions, the same explanation is offered: "It is then rather disturbing to find that in a paper which is one of the major works on information theory for communication engineers, Shannon appears to equate information to entropy *without the negative sign*. But there is in fact no direct comparison between our application of entropy and Shannon's, since he uses it as a measure of the statistical characteristics of a source of information, not as a step towards finding the information value of any given waveform or function. *The entropy of a message source is then proportional to the number of different combinations (messages) which that source can produce*" [author's italics] (Bell 1953, p. 27).

In all editions, Bell gives a homely example – the breaking of a set of type used to print the text of *Hamlet*, which increases entropy as it reduces information – and then develops the idea mathematically in an appendix, beginning "That there is a similarity of mathematical form between information and negative entropy is undoubted, but there is room for argument whether there is any 'real' or 'physical' connection between the two. Much of the published work on communication theory glosses over this question" (Bell 1953, p. 120). His argument is that physical entropy is intrinsically linked with energy and temperature, and that to relate it to information, a more abstract dimensionless quantity must be found. In the second edition of 1956, he introduced the idea that such a pattern was also to be found in the physical representation of information, predating by some years the concept that information is physical, usually attributed to Landauer.

Negative entropy, or negentropy was used in a wider scientific context very soon after Shannon's introduction of the idea, most notably by French physicist Léon Brillouin, whose influential book *Science and Information Theory* appeared in 1956, with a second edition in 1962, following the earlier introduction of his ideas (Brillouin 1949, 1951A, 1951B, 1953). Other who took up the idea were Raymond (1951), Rothstein (1951; 1952A; 1952B) and Branson (1953).

Rothstein (1952A, p. 1281) sums up their viewpoint: "statistical entropy has long been recognized as a measure of missing information or lack of organization (disorder). Precise formulation of the information concept leads, via the measurement-communication analogy, to recognizing that physical information is essentially negative entropy."

Information theory was applied by these scientists to a variety of physical problems, though – as Denbigh and Denbigh (1985) point out – with a surprisingly limited treatment of thermodynamic entropy. While citing Shannon as the originator of information theory, Brillouin's perspective is clearly that of Wiener:

"The origin of our modern ideas about entropy and information can be found in an old paper by Szilard, who did the pioneer work but was not well understood at the time. The connection between entropy and information was rediscovered by Shannon, but he defined entropy with a sign just opposite to that of the standard thermodynamical definition. Hence what Shannon calls entropy of information actually represents negentropy … Information and physical entropy are of the same nature. Entropy is a measure of the lack of detailed information about a physical system. The greater is the information, the smaller will be the entropy. Information represents a negative term in the entropy of a system, and we have stated a negentropy principle of information." (Brillouin 1962, pp. 161 and 293).

In Brillouin's understanding of Maxwell's Demon, the Demon's information allows him to sort molecules, thus decreasing the system's entropy; but this information had to be paid for by an increase in entropy elsewhere in the system. So information and entropy are opposites and must have opposite signs; hence the idea of negentropy for information.

The use of 'negentropy' has been criticized by some as confusing the issue, Ben-Naim (2008), for example, suggesting that it is a 'corruption', and that it would be better to call entropy neg-information. Wicken (1987B) argues that Brillouin failed to distinguish between information and order, the latter being better related to negentropy. Blum (1968) similarly took negentropy to be a probabilistic concept, equivalent to an increase in order, and associated with arrangement and pattern, but emphasized that it was not simply the opposite of thermodynamic entropy, not related to meaning in any sense. And indeed, it has been shown that Brillouin's proofs of scientific concepts in information theoretic terms work equally well if the signs are reversed (Wilson 1968).

The concept of negentropy, like that of entropy itself as noted above, has since been applied, some would say unwisely, in a variety of disciplines (see the commentaries and references by Shaw and Davis 1983, Hayles 1990, White 1991, Logan 2012, and Robinson and Bawden 2014), and as Müller rather sourly puts it "has fired the imagination of physicists of the more esoteric type and theologians" (Müller 2007, p. 322). Müller also notes the enthusiastic take-up of entropy and negentropy concepts among such groups as biologists, economists, ecologists and sociologists, and warns against "a lack of intellectual thoroughness in such extrapolations. Each one ought to be examined properly for mere shallow analogies" (Müller 2007, p. 73). The same is surely true for applications in the information sciences.

Following Brillouin, other scientists followed this trend of using information theory to explain scientific issues, in biology particularly. Following early consideration of these ideas across wide areas of biology (Quastler 1953), Gatlin (1972) and Brooks and Wiley (1988) both attempted an analysis of biology in information terms, focusing on the relation between information entropy and evolution. All such attempts required new concepts of information and/or entropy, Gatlin taking information to be the difference between the maximum and observed entropies,

and Brooks and Wiley introducing the concept of 'instructional information'. This is an early indication of the difficulties of applying simple ideas of information and entropy to dealing with issues of complexity and organization in the physical and biological worlds: "some "entropies" and "informations" are physically interpretable", lament Brooks and Wiley (1988, p. 66), "and some are not". Wicken (1987A) criticizes what he sees as a proliferation of new and unneeded entropies, and unwarranted use of the idea of 'information content', where complexity would be a better term. Generally, however, these approaches all follow Wiener's perspective: increasing information stems from processes which decrease entropy, and information is associated with organisation. A recent contribution analyses the issue in terms of different contributions of thermodynamic entropy and information entropy (Mitrokhin, 2014).

Conversely, in the physical sciences Shannon's approach was used, in both formalism and in interpretation, to cement the links between information content and physical entropy, most notably by E.T. Jaynes (1957A; 1957B).

Jaynes' basic concept is that the least biased assignment of probabilities to available data is that which maximizes the Shannon entropy. His approach allows the derivation of all of Gibbs' results without the conceptual difficulties associated with Gibbs' approach (Tribus 1961). What has become known as Jaynes' principle, the maximizing of Shannon's H, forms a general method of statistical inference in statistical problems from many fields; a way of making the best estimate in the face of insufficient information. It is entirely to do with the probabilities, and not with information in any other sense, and certainly not with physical entropy; "Entropy is not just a physical quantity for measuring the degree of randomness or information. It is a reasoning tool too" (Tseng and Tuszynski, 2014, p. 3756).

Hobson (1971) states clearly that "E.T. Jaynes was the first to establish a clear and useful connection between information theory and statistical mechanics" (p viii), and that "If one takes the view that statistical mechanics is the study of incompletely specified mechanical systems, then it becomes natural to try to applying the mathematical theory of information to statistical mechanics. E.T Jaynes seems to have been the first to make a clear, quantitative connection between information theory and statistical mechanics, although several authors established a qualitative connection prior to the work of Jaynes." (pp. 10-11). These "several authors" he names as Brillouin, Raymond, Rothstein, Landsberg and Tribus.

Emphasing the entirely probabilistic nature of Jaynes' approach, without any reference either to the thermodynamic context or to any ideas of order or structure, Hobson (1971, p. 35) notes that "… information theory concepts should be relevant to any field in which inductive probabilities are useful. The reason is that inductive probabilities arise whenever the given information is not sufficient to permit deductive inferences: any theory which purports to study information quantitatively is likely to be useful in such a situation." Shannon's measure is a measure of the *missing* information in a *single* probability statement, and "Jaynes's principle extends the information concept from the idea of the information content in a

probability distribution to the more direct idea of the information content in a body of data" (Hobson 1971, p. 49). This usage of information entropies as a tool dealing with probabilities in any context has proven useful in many fields; see, for example, a recent example in ecology (Harte and Newman 2014).

Myron Tribus provided the first textbook to base the laws of thermodynamics on information theory, rather then on physical arguments, using Jaynes' ideas to show a link between thermodynamic and information entropies (Tribus 1961). He set out his ideas unambiguously at the start.

> "In 1957, E.T. Jaynes of Stanford University published a remarkable paper. Ever since 1948, the work of Claude Shannon of the Bell Telephone Laboratories on the mathematical theory of communication had shown a formal relation between information theory and thermodynamics. Jaynes … showed that if one took the idea of information theory as primitive and more basic than thermodynamics, all of the formulae of statistical mechanics could easily be derived … It seemed useful therefore to attempt the writing of a textbook for undergraduate students in engineering, presenting thermostatics and thermodynamics from Jaynes' point of view."
> (Tribus 1961, p. v111)

Tribus (1999) recalls that he was asked at his doctoral viva about the link between Shannon and Clausius entropies but could not explain it satisfactorily. He read up everything he could and then was directed by a student to Jaynes' two-part 1957 paper. He saw this as the Rosetta stone, and studied with Jaynes. Tribus' 1961 book was based on work by Jaynes which was never published, and so was cited as 'in print'.  In 1961, Tribus gave a seminar on these topics at MIT and was heavily criticized. Among these critics was Benoit Mandelbrot, who consulted Shannon; the latter disspelled Mandelbrot's concerns, but criticized the use of Shannon's definition of entropy outside communication channels.

In similar vein, Katz (1967), argued that using the information theoretic approach to statistical mechanics could not give any new results, but that it had pedagogical value as an aid to understanding the concepts; entropy among others. Ben Naim (2008) argues similarly that such approaches did not solve any thermodynamic problem, but allowed the subject to be presented consistently and logically. A number of studies have derived mathematical equivalences between the three entropies – thermodynamic, statistical and informational – on the basis of probability, more recently in a quantum mechanical framework; as an example see Baumgartner (2014).

The British statistician R.A. Fisher had, as noted above, derived a measure for information very similar to that of Shannon (Fisher 1922, 1935, 1950). It has been widely used in the statistics of measurement theory, dealing with the maximum amount of information which may be obtained by any measurement. Fisher's 'amount of information' is a measure of precision or weight of evidence, equivalent for a normal distribution to the reciprocal of the variance (MacKay 1969). However,

there is a clear distinction between Fisher information and Shannon (or Wiener) information, put clearly at an early stage by Baer (1953 p. 22):

> "we can distinguish sharply between the Shannon-Wiener functional "information entropy" and another functional defined as "information" by R.A. Fisher. Fisher also based his definition on the logarithm of the probability distribution but the "information" he defines has to do with the statistics of sampling and the determination of the parametric dependence of the corresponding probability distribution"

Shannon's is a simple scale function, not depending on any particular probability distribution, and hence more generally applicable.


**Shannon and Wiener, information and (dis)organization**

Hayles (1999, p. 103) points out an interesting consequence of Wiener's perspective:

> … In retrospect, identifying entropy with information can be seen as a crucial crossing point, for this allowed entropy to be reconceptualised as the thermodynamic motor driving systems to self-organisation rather than as the heat engine driving the world to universal heat death … chaos went from being associated with dissipation in the Victorian sense of dissolute living and reckless waste to being associated with dissipation in a newly positive sense of increasing complexity and new life"

Shannon's use of the entropy term "led to the (metaphoric) knotting together of concepts that are partly similar and partly dissimilar" (Hayles 1990, p. 50). If randomness is understood as maximum information, then it is easy to see chaos as the source of novelty; then complexity can be seen as rich in information, rather than as deficient in order.

Although many texts and review articles within information science and related areas mention Shannon's theory, few mention Wiener, and the opposing sign. And where they do, this is usually just a mention, without discussion. It appears that Qvortrup (1993) was the first author from LIS and related areas to draw explicit attention to the discrepancy between the positions of Shannon (information as chaos and entropy) and Wiener, and those scientists who took his position, notably Bell, Brillouin and Stonier (information as order and negentropy).

This is particularly strange, as we imagine that many in the information disciplines would instinctively regard information as associated with meaning, order and organization, more than with the opposite. A few authors in LIS and cognate disciplines who seek to link physical and social conceptions of information, take Wiener's viewpoint: for example Lossee (2012, p.45): "… the average information, in an information theoretic sense, may be viewed as the negative of the entropy as computed by a physicist". MacKay (1969), as noted above, associated Shannon's information with negentropy and with the determination of form. Brier (2013, 2010), noting that, from the perspective of entropy, Shannon's information is the opposite of Wiener's information, sees the distinction as being that for Shannon entropy, i.e. randomness, contains more information than does structure, while for Wiener

information amounts to a structured piece of the world. However, he finds both lacking as a satisfactory basis for an account of consciousness and meaning. Stonier, who regarded information as synonymous with organization, held that "the idea that information and entropy are the same was subsequently replaced with the idea that information was equivalent to negentropy" (Stonier 1990, p. 56). Although he held that his view of information was diametrically opposite to that of Shannon, and hence essentially that of Wiener, Stonier does not cite Wiener, crediting Schrödinger with the negentropy concept; see Furner (2014) for a clear analysis of Stonier's ideas. Beer (1972) argued similarly that negentropy was the active information content of a system, while Simon (1962), in his analysis of complexity and hierarchy, took the Wiener position of information as the opposite of entropy, as does Kauffman (2000) in his conceptions of self-organising systems. Bates regards information as synonymous with pattern, and entropy as "pattern-less", so that "the only thing in the universe that does not contain information is total entropy: that alone is pattern-free" (Bates 2006, p. 1033).

In a recent review, Logan writes that it is "extraordinary that [Shannon's] definition of information, limited in scope by his own admission, became the standard by which almost all forms of information were gauged" (Logan 2012, p.75). Shannon himself would also have found it odd; as he himself commented "It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field" (Shannon 1953, p. 105). Logan takes the side of Wiener "arguing, as have many physicists before us, that information and entropy are opposites and not parallel as suggested by Shannon" (Logan 2012, p. 69), and regarding information as equivalent to organization. Furner (2014) suggests that Logan's idea of information is closely related to those of Stonier and of Bates, but that the precise nature of the relation is unclear. (Furner also points out that various misprints and misstatements in Logan's article make the arguments difficult to follow.)

Logan also draws attention to a 'counter revolution' within the information science community, which attempted to build on the ideas of Shannon and Wiener, by formulating a definition of information including semantic meaning. Donald Mackay, notably, wished to identify information as that which brings about a change in a recipient's internal representation, and hence linked with meaning (MacKay 1969, Hayles 1999). This appears very similar to Brookes' (1980) 'fundamental equation', which also presents incoming information as changing a person's knowledge state, but Brookes and MacKay do not seem to have acknowledged each other's insights.

Shannon information, for MacKay (1969) was 'selective', concerned with how message elements are chosen from a set of possible elements; semantic information is 'structural' (introducing new features) or 'metrical' (increasing our confidence in features already known), and concerned with how selective information is to be understood. Typical of the definitions advanced were MacKay's (1969, p. 158) "makes a difference … to what we believe to be the case" - so that for MacKay, as for many later authors, including Hjørland (2007) and Wallace (2012, p. 31), information must be about something - and Bateson's later and very similar "a difference that

makes a difference". Strictly speaking, Bateson's definition, though usually quoted in this way, was more specific: information is "any difference which makes a difference in some later event" (Bateson 1971, p. 231), which he considered equivalent to an elementary 'idea' (Bateson 1970).  Logan suggests that "distinction" is more closely tied to meaning than "difference", but the disparity is subtle. Logan, together with other writers such as Burgin (2010), notes that the Shannon and Wiener formulations, though seemingly opposite, are similar in that they are defined by what they are (a thing), whereas Mackay's and similar formulations define information by what it does (a process).

However, MacKay's information and similar formulations could not be quantified, and hence, despite their significance for LIS and related areas, are not germane to the information/entropy discussion. However, the idea of structural information has echoes in some information-related measures of complexity (Bawden and Robinson 2015).

We might also agree with Logan that MacKay, the originator of the term 'information theory', and one of the first to challenge Shannon's formula on the grounds that it did not recognize that the measure of the kind of information of interest in the social world - to LIS, although MacKay did not specifically mention our discipline - is its ability to make a difference through meaning, was "certainly a scholar who made a difference and deserves more credit and attribution than he usually receives" (Logan 2012, p. 77).


**Shannon and Wiener: an architectural aside**
Smolin (2013 pp. 196-197), in a popular treatment of the physics of time, gives an interesting example of the Wiener interpretation, in the context of architecture. For a conventional brick building, the macrostate is the architect's drawings, while the microstate is the description of the position of each brick. "The architect needs to specify only that brick walls of such-and-such dimensions are built, with openings for windows and doors. He doesn't need to say which bricks go where. Most bricks are identical, so it doesn't affect the structure if two identical bricks are swapped. Thus there are a huge number of different microstates that give the same macrostate." He contrasts this with a building whose outer surface is composed of individual elements, each unique, so that each element must go into a specific place, exemplified by Frank Gehry's Guggenheim Museum in Bilbao, with its metal sheet construction. "In this case, the architectural drawing again specifies the macrostate, and where each sheet goes is the microstate. But unlike the brick building, there is no freedom to tamper with the microstate. There is only one microstate that gives the intended macrostate."

We may take Battersea Power Station in London – the largest brick building in Europe, with over 60 million bricks - as a specific example of a brick building.  (Gehry is one of the architects working on its transformation)

The entropy of a building, Smolin says, is a measure of the number of different ways to put the parts together to realize the drawing of the architect. The conventional brick building has a very high entropy; a building like Gehry's has a low entropy, zero if only one microstate instantiates the macrostate. Smolin then argues that "We can then see from this example that *entropy is inverse to information*" [Smolin's italics] It takes a lot more information to specify the design for a Gehry building because you need to tell exactly how to fabricate each piece and exactly where each piece goes. It takes much less information to specify the design of a normal brick building because all you need to know are the dimensions of its walls". Extending this argument to physical systems generally, he denotes all improbable arrangements – those with few microstates corresponding to the macrostate – as having low entropy and high information. He gives the example of a cat – there are many more ways for the molecules comprising the cat to be scattered around the room that to be in the organized form of the living cat.

Shannon would say "I consider how much information is *produced* when a choice is made from a set – the larger the set the *more* information". The brick building has a large set from which the particular microstate of individual bricks is chosen – so Shannon would say it has a lot of information". Smolin, and Wiener, are interested in how much information has to be put in, to specify the particular microstate. Wiener would "consider the larger uncertainty in the case of a large set to mean less knowledge of the situation and hence *less* information." In the brick building, we have less knowledge of exactly what microstate is involved, and so less information.

The figure is calculated in the same way – it is the uncertainty as to which microstate it is and is determined by the size of the set – but Shannon looks at the information gained when the microstate is known, when the uncertainty is resolved, while Wiener looks at the uncertainty per se as a lack of information.

And if the microstate is *interesting* – such as an artistic building or a cat – then we will think of the information as making it interesting or meaningful. The choice of microstate for the brick building, though informative from Shannon's perspective, is uninteresting, although the brick building is more ordered and structured.

**Algorithmic entropy**

The decades after the introduction of Shannon's theory saw the advent of information engineering, with the design and widespread use of digital computers, processing physical instantiations of information. And, as studies of steam engines processing energy had led to the first recognition of the entropy concept, so the information revolution led to new perspectives on the relations between energy, entropy and information, and to Landauer's (1991) famous claim that 'information is physical'. (Leff and Rex 1990, 2002, Gleick 2011).

We will restrict ourselves here to the central point of importance for this discussion; a coda to the concept introduced, though not explicitly stated, by Szilard, that there is a direct connection between information and entropy. The work of Landauer and Bennett in particular established that, while logical operations on information elements are entropy-free, at the point when information is erased during the process of a computation, heat is generated, and the entropy of the surroundings is increased; for an accessible presentation, see Bennett and Landauer (1985). Logically irreversible operations such as erasure necessarily involve entropy increase by $k\log2$ per bit of erased information. In one sense, the link between information and entropy is definitively established.

Having completed this sketchy and partial review of the historical development of the information/entropy relationship, we now go on to analyse some particular aspects, and recent developments.

**Entropy and order; oddities, paradoxes and new views**

The developments described in outline above have led to general agreement that information is indeed related to physical entropy. But the relation is more subtle and complicated that the ways in which it is commonly presented.

The simple, and until recently generally accepted, viewpoint is that physical entropy is equivalent to disorder; Slater, for example, stating the commonly held 'pre-Shannon' position that entropy measured "randomness or disorder" (Slater 1939, p. 9). If we take Shannon's position that information is equivalent to entropy, and reaches its maximum in randomness and disorder, then this is a strange situation; the idea that 'information equals disorder' is not a viewpoint that usually makes sense from the perspective of the information disciplines. (The term 'random' is perhaps unfortunate here, as it has associations of 'meaningless' and 'useful'; Shannon's theory is silent on the meaning and utility of messages whose information content is measured by probability.) But equally, the idea that 'information equals negentropy equals order' does not seem appropriate; the quotation from Gleick (2011, p. 281) above reminds us that order is not necessarily informative, and orderly things do not necessarily embody much information.

At this point, one might be excused for concluding that the information/entropy relationship is not really relevant to the kind of information, and information issues, of interest to information science, taking the kind of position espoused by Hjørland (2007) and Cornelius (2014). But we should remember that the entropy/disorder relation is by no means as simple as it has been presented in the past. (And even in science, there have been authors such as Denbigh and Denbigh (1985) who have criticized it, and have given examples where the relation between entropy and order is ambiguous, particularly when 'disordered' systems are simple and uniform.)

Let us first consider four examples which illustrate this.

1. An example of the entropy/disorder relationship often presented by library/information science tutors seeking to make the idea accessible is that of a disordered book collection; Shaw and Davis (1983) quote several such analogies. A diligent librarian sorts and orders the book, doing both mental and physical work in the process. The entropy of the collection reduced, while to compensate the entropy of the wider universe is increased, as a result of the heat generated by the librarian's work. Sadly, this appealing and homely analogy is rather lacking. The work done, and universal entropy generated, would be exactly the same if the librarian were further disordering the books, or just moving them from one disordered state to another. Any claim to entropy reduction as a result of ordering relies on having a definition of order, such that we can show that the final result is a smaller macrostate, in Boltzmann's terms, than the original; this is surprisingly difficult to do, lacking a formal objective definition of order suitable for this situation. The librarian may arrange the books by colour; however a colour-blind reader will discern no change in the order and hence no change in the entropy. The librarian may organize a collection according to a classification scheme; but the resultant increase in order, and in information content, will be evident only to someone who is familiar with the

intellectual rationale for the scheme, and considers that it is a helpful arrangement for their purposes.

Wicken (1987B) criticizes a similar example given by Gatlin (1972), of the ordering of disordered Scrabble pieces, on the grounds that entropy, disorder and disorganization have very different meanings which are often conflated. Lambert (1999) emphasizes that moving macroscopic objects around does not affect their thermodynamic entropy, though it may affect their information content, if a satisfactory way can be found to measure this.

2. This leads to the familiar tidy/messy desk example, a tidy desk being claimed to be of lower entropy than a messy desk, and hence its tidying an example of entropy reduction. There are several issues here. Most obviously, as Ford (2013, pp. 1-2) and Schneider and Sagan (2005, p. 21) point out, a messy desk does not have more entropy than a tidy one; they are simply two specific states of the system. To bring entropy into it, we would have to define 'tidy' and 'untidy' in an unambiguous way - Ford suggests by the fraction of desk top visible – then we could assess the statistical likelihood of how such a starting configuration would change over time, if essentially random changes were made. We would expect that a 'tidy' desk, having fewer microstates, would become more untidy over time; it would diverge in an uncertain way from its initially well-defined state. In that sense, there is an analogy with statistical entropy, but it should not be taken too literally. And again there is the need for a definition of 'ordered' or 'tidy'; particularly tricky in this context, when research has shown that 'messy desk' owners can retrieve items just as quickly as keepers of tidy desks (Lansdale 1991) and that apparent messiness in an office environment may enhance creativity (Vohs, Redden and Rahinel 2013). There may well be a 'hidden order' in such situations, which may confound simplistic untidy/entropic assumptions. And, while it is true that tidying a messy desk requires work, causing an increase of temperature, and hence entropy, in the room in which the desk is located, so also does an enthusiastic messing up of a tidy desk (Schneider and Sagan 2005, p. 21). Nonetheless, the idea that a messy desk or room is an example of high entropy, and that information may reduce entropy, in the form of a knowledge of where things should go in the tidying process, remains a popular teaching tool; see, for instance, Peterson (2014).

3. As a further thought experiment, let is return to our example of the blue ink poured into a large glass tank of water, slightly amending an example given by Carroll (2011). We noted that we would expect after some time the ink cloud to disperse, giving a uniform light blue colour; a straightforward example of spontaneous entropy increase. We expect this situation to remain indefinitely, but are surprised to find that some time the ink has coalesced to form an accurate representation of the Statue of Liberty. We go to find a camera to record this strange event, but on our return the ink has further coalesced into a small featureless blob. This is, as said earlier, something which we would expect to happen, but only on extremely long time scales; the point is, however, that in moving from diffuse blue tank, to Statue of Liberty, to small blob, entropy has steadily increased, simply because the volume occupied by the ink, and hence the number of possible

microstates, has drastically reduced. This seems intuitively wrong, if entropy is associated with order; the Statue of Liberty conveys meaning, it is interesting, it must surely be, in some way, more organized and order-rich than a small featureless blob. This reminds us that meaning and interest are subjective, and not directly associated with the kind of order that is the opposite of entropy. That order reflects the simple notion that dispersed distributions correspond to larger numbers of possible arrangements - vastly larger if the number of elements is involved is large – than do concentrated distributions do, so they take up correspondingly larger chunks of phase space (Albert 2000).

4. Finally, we may refer to an example given by both Carroll (2011) and Ford (2013); the totality of the physical universe. According to our current best knowledge, this has progressed from a smooth hot gas, through the current era of structure and organization – galaxies, stars, planets, people, books – to end in a sea of cold featureless uniformity. And yet we are told that entropy increases steadily throughout this process, which again seems intuitively wrong. Further, if the smooth hot gas at the beginning of the universe is indeed of high entropy, then it should be high in Shannon information, which again seems intuitively strange.

One partial explanation is simply that the early universe was very hot, and hot materials have more entropy than cold materials. Another is the influence of gravity. Simply stated, although the details are not fully understood, irregularities and structures formed under the influence of gravity have a high entropy (for brief accounts, see Gribbin 2009, Carroll 2011 and Smolin (2013), and for more detail see Wallace 2010); in a very real sense, gravity puts information into the universe. As Davies (1987, p. 135) puts it "Gravity in the early universe can therefore be seen as "the fountainhead of all cosmic organization ... triggering a cascade of self-organizing processes".

This is a dramatic illustration of the extent to which physical entropy involves considerations which can have no equivalent in applications of entropy in the information sciences. There is a second point. While the initial and final states of the universe are very different in physical entropy, they are very similar in another respect; they are both very simple, in contrast to the complexity of the present era. Denbigh and Denbigh (1985), among others, give similar examples for chemical systems. This is an indication that it will be necessary to look to ways to include complexity in extending and illuminating the information/entropy relationship; see Bawden and Robinson (2015).

All four examples raise questions of subjectivity. The question as to whether physical entropy is subjective, in a way which is worrying for a scientific concept, has been of concern since the idea that it was related to missing or incomplete information was introduced by Boltzmann and Gibbs. Maxwell, the originator of the Demon, may have been the first to recognize this (Schneider and Sagan 2005), since he wrote in a *Encyclopedia Britannica* article on diffusion in 1878 that "dissipated energy is energy which we cannot lay hold of and direct at pleasure, such as the energy of the confused agitation of molecules which we call heat. Now, confusion, like the

correlative term order, is not a property of material things in themselves, but only in relation to the mind which perceives them" (Maxwell 1878, p. 216).

Those approaching the issue from the physical sciences have typically wanted to insist on its objectivity as a physical quantity. Hobson (1971) wrote a clear statement of this position:

> "The view that statistical mechanics is the study of incompletely specified mechanical systems is sometimes criticized on the grounds that it is subjective (i.e. involves the observer), whereas science is supposed to be objective. Without getting into a discussion of whether or not science is actually objective, the author would like to point out that this viewpoint is as objective as any physical theory can be expected to be. According to the operational philosophy of physics, physical theories should depend upon the measuring instruments of the observer. Mechanics deals with the ideal case of perfect and complete measurements. Statistical mechanics makes predictions based on information obtained from imperfect measurements. The predictions naturally depend upon the information, i.e. upon the measuring instruments. The theory does not, however, depend upon the subjective views of the observer: two observers, equipped with identical measuring instruments and obtaining identical readings with these instruments, will make precisely the same statistical mechanical predictions about the outcomes of future measurements … [if two observers with instruments of different sensitivity reach different conclusions, then] … the difference is subjective in the sense that it is different for the two observers, but objective in the sense that the difference depends only on the measuring instruments available."  (Hobson 1971, pp. 6-7)

For Hobson, entropy is a measure of the observer's uncertainty about the exact state, knowing only some macroscopic data. It is meaningless to speak of the entropy of a particular state .. it only has meaning in terms of a probability distribution. "It might be helpful to speak of the 'entropy of the data' or the 'entropy relative to the observer' rather than the entropy of the system … For any measured values of [macroscopic parameters] the entropy takes on a perfectly well-defined, measurable value S … the same for all observers possessing the same data. Thus the entropy is not really subjective (i.e. relative to the observer), but is instead relative to observer's *data*" (Hobson 1971, p. 90).

This, Hobson tells us, explains well-known Gibbs paradox. If two volumes of gases A and B, at first kept apart, mix, then here is an increase in entropy – the entropy of mixing – provided A and B are distinguishable. If they are not distinguishable, there is no entropy increase. The increase is the same, whether A and B are almost the same in nature , or totally different. This is not explicable physically, but is explicable on the basis of a discontinuous jump in the data – distinguishable to indistinguishable. (We might note that Slater (1939) had explained the paradox four decades earlier in similar terms, but without invoking the 'entropy of the data' concept, relying rather on quantum theory to make the point that entities are either identical or not, a sharp binary distinction.)

Objectivity here means that several observers, given the same data, would reach the same conclusions about entropy. Apparent subjectivity comes simply from the fact that different observers may have different data, as in the example quoted earlier that a rearrangement of books by colour or by classification scheme will not be an apparent reduction in entropy to readers who are respectively colour-blind or are ignorant of the intellectual basis of the classification.

Others made similar points. "This all-pervasive capacity for change within the universe is surely independent of man's presence and is thus fully objective" wrote Denbigh (1981, p. 185), while Albert (2000, p. 103) insisted that entropy "is an objective physical characteristic of the individual microconditions of individual thermodynamic systems. The entropy of a microcondition is the logarithm of the standardly calculated phase-space volume of the macrocondition to which the microcondition in question belongs."

Atkins (1984, p. viii) argued similarly, noting that he had deliberately omitted any reference to the relation between information theory and entropy, although he conceded that the principles and mathematics of information theory could substantially contribute to the formulation of thermodynamics. Atkins was concerned that:

> "there is a danger, it seems to me, of giving the impression that entropy requires the existence of some cognizant entity capable of possessing "information" or of being to some degree "ignorant". It is then only a small step to the presumption that entropy is all in the mind, and, hence is an aspect of the observer. I have no time for this muddleheadedness and intend to keep such accretions at bay".

Schneider and Sagan (2005, p. 22) consider that this objection is "understandable but overdone" since information theory is about much more than messages exchanged between conscious beings. As developed by scientists such as Jaynes (1957A, 1957B), it focuses on information obtained from experiments, and is as applicable to thermodynamics as to any other area.

The issue was addressed thoughtfully by Denbigh (1981), who concluded that this was not an issue provided that a number of caveats were taken into account. The entropy which we calculate is the entropy of our data on a physical system, and – like many other quantities - its value will change when our data changes. Also in some circumstances we will have a choice of how to calculate entropy, depending on the assumptions which we make about physical systems. As Landsberg had put it, two decades earlier, considering both physical entropy and information entropy of messages, "the entropy of a system or text depends not only on the system or text, but also on our knowledge of it, and on the questions we ask about it" (Landsberg 1961, p. 237).

A very full treatment was given by Denbigh and Denbigh (1985), to provide a counter to what they regarded as a widely accepted, though erroneous, view that "entropy

really signifies nothing more than a lack of human knowledge" (p. vii). Such ideas, they suggest, began with Gibbs, were first presented explicitly by Lewis, in the 1930 quotation above, and gained general credence with the use of information theory to address thermodynamic problems. Their rebuttal relies on the idea that incomplete information does not equate to subjectivity. A calculated entropy may change, like any other physical quantity, when available data changes, but this is not subjective in the usual meaning of the word, for entropy any more than for any other quantity. Given the change in data, and an agreed formula, any scientist would agree with the changed value, which may reasonably be called objective. They therefore prefer the approach of Hobson (1971), noted above, that we should speak of 'the entropy of the data', which will naturally change when the data change.

Nor is it the case that gaining more accurate information necessarily reduces the calculated entropy of a system; they give several physical examples where the calculated entropy increases. Finally, they are argue that using the formalism  and terminology of information for physical entropy allows the introduction of terms such as 'ignorance' and 'surprise' which have no meaning in the physical context. Their contention is that physical entropy is an objective quantity, which is unaffected by our knowledge, though any calculated numerical value is subject to our ignorance.

Ben-Naim (2008, p. 29) agrees with Denbigh and Denbigh that there is no "problem about whose information is measured by H", provided that we use Shannon's objective measure of information.

Ford (2013), Carroll (2011), Ben-Naim (2008) and Georgii (2003) also present arguments against being overly concerned about subjectivity. As long as we agree on what system we are measuring, and a framework for what is to be included in our description of it and with what degree of detail, then entropy should be a well-defined and objective measure of remaining uncertainty, calculated from measurements made within that framework. We should not be misled by the use of terms like uncertainty and ignorance to believe that it is subjective. This does however mean that there will be different, perhaps many, entropies, depending the nature of framework and what variables are included within it (Lebowitz and Maes 2003). This is in line with the general view of objective interpretations and judgments; that they make sense only when viewed within a constant set of criteria (Gaukroger 2012).

The need for such a framework is well illustrated by an example due to Brillouin (1949). If we take a textbook, or a volumes of printed newspapers, and an equal weight of waste paper, do they have the same entropy? Brillouin suggested that they do if we are making a fire with them, but not if we propose to read them. Or, to adapt an example from Lebowitz and Maes (2003), we may measure the entropy change when a collection of coloured balls – let us say red, green and blue – initially separated by colour are allowed to mix; but the assessment will be different for an observer who is red-green colour-blind, or who is watching the process on a black-and –white monitor. This does not make the value of entropy subjective; it is just

being measured within a framework of different variables, or, more generally, within a different conceptual perspective.

Of course, the problem largely disappears if we regard entropy as equivalent to, or the opposite of information. The idea that we may have a different amount of information about a system according to the context and circumstances does not trouble us, and does not make us want to insist that the information cannot therefore be considered objective.

We might also note *en passant* that the concept of entropy as missing information is rooted in classical physics, whereby gaining perfect information is always possible, at least in principle. In quantum physics, uncertainty and probability becomes a fixed and fully objective attribute of the word, removing in a different way concerns about subjectivity (Vedral 2012). Recent studies also suggest that entropy may result from quantum entanglement and decoherence causing a loss of objective information, rather than from human ignorance, again restoring the objectivity of entropy; for an accessible review, see Wolchover (2014), and also Tegmark (2014) and Gharibyan and Tegmark (2013).

The matter is no means entirely settled however, and different opinions are still expressed: "there are several different answers to such questions, none of which is found satisfactory by everyone working in the field of statistical mechanics" (Carroll 2011, p.160). For example, Hemmo and Shenker (2012), deal with it by adopting the philosophical position of physicalism, and regarding the, allegedly subjective, recognition of macrostates by an observer as simply an objective microstate of the observer's brain; the observer is a part of the mechanical system, while Baranger (2001) argues that entropy is entirely subjective; it is a measure of disorder, but the disorder is in our heads, in our knowledge of a system. The eminent British physicist Sir Roger Penrose (2004) worries about what he perceives as a subjective element to entropy, and suggests that this means that it may not be a truly fundamental physical entity.

There has been a general move away from equating entropy with disorder in scientific contexts in recent years. This interpretation, introduced by Boltzmann as essentially synonymous with a state of high probability, has been criticized by a number of science educators as unhelpful, if not actively misleading, following the suggestions of earlier commentators such as Denbigh and Denbigh (1985). Leff (1996) and Lambert (2002) have recommended its replacement by the idea of entropy increase as 'energy dispersal', among a large number of possible microstates; Kostic (2014) makes similar points. Leff (2012) argues that disorder is a poor metaphor for entropy. Spreading or dispersal of energy is better, as this explicitly includes the energy aspect of physical entropy, but uncertainty, or more specifically missing information, is also a useful metaphor. Both metaphors refer to choice, since as a system can spread over many microstates, there is more uncertainty as to what the state actually is; "spreading and missing information provide complementary, useful views of entropy" (Leff 2012, p. 276). Disorder is only

loosely related to this, and may be misleading. Leff suggests that the symbol S, traditionally used for entropy, should indicate 'Spreading'.

Styer (2000) also notes several arguments against the disorder interpretation, including its invitation to focus on a single state rather than a class of states, and its vagueness, with no precise definition of disorder. However, recognizing that it is at times appropriate, he recommends a joint use of the similes of 'entropy as disorder' and 'entropy as freedom'; the latter essentially the same as energy dispersal. Brissaud (2005) comes to similar conclusions.

One rather technical point is that physical entropy is an extensive property; that is to say crudely that if an amount of substance has a certain entropy, then twice as much substance will have twice as much entropy. While this seems sensible for the physical quantity, it does not seem intuitively sensible to suggest that the larger amount of material is 'twice as disordered'. To deal with this problem, and the fact that entropy is not always additive, a variety of 'extended entropies', most notably those of Lansberg, Tsallis, Renyi and Vedral, have been developed, which characterize disorder in somewhat different ways (Davison and Shiner 2005); Shterenberg similarly considers new formulations of the entropy concept to relate to ideas of order, orderliness and organization (2013), while Schneider and Sagan (2005) note such proliferations as metric, topological, algorithmic and Galois entropies.

Ben-Naim (2008, 2009, 2011) noting that the interpretation of entropy as disorder has prevailed for over 100 years, finds two difficulties: the fact that there are cases where disorder is clearly observed without a corresponding energy change; and, more fundamentally, that the concept of order is not well-defined, so that it may be difficult, if not impossible, to say which of two states is more or less ordered. The idea of disorder is "at best a vague, qualitative and highly subjective one" (Ben-Naim 2008, p. 10). He recommends, and defends in detail, an interpretation in terms of Shannon's measure, since this concept of information is very general, and encompasses both objective and subjective types of information. He argues that entropy, in all senses in which the word is used, including but not limited to thermodynamics, is best understood as uncertainty or, more fully, the amount of missing information.

Ford (2013) makes a similarly convincing case that entropy is better regarded as uncertainty, or missing information, than as disorder. We do not possess all the microscopic information about a complex system, and that missing information constitutes its entropy:

> "Entropy is uncertainty commodified. .. We do not or cannot measure all the details of the present state of the world and so when processes occur we are not quite sure what will happen … Our certainty about the future is less then our certainty about the present … The increased uncertainty is quantified as an increase in the total entropy of the world, and that is what entropy is. The

most remarkable thing is that we can measure it with a thermometer." (Ford 2013, p. xiii).

The growth of disorder, says Ford (2013, p. 135) is a useful shorthand description for the growth of entropy, and does capture some of Boltzmann's insight, since describing a system as disorderly implies uncertainty in its future behaviour; but it is not the whole story. There are too many counter-examples, and, as noted already, we have no good definition of order that will fit all circumstances. It also spares us worrying about whether the final state of the universe is ordered (in that there is uniformity) or disordered (in that it must be high-entropy); it simply has a maximally uncertain microscopic state.

This idea of the link between entropy as disorder (disorder increases over time, because there are more ways to be ordered than disordered) and entropy as missing information (we can have, in principle, complete knowledge of a system in the past, or in the present, but our knowledge decreases the further into the future we predict) has been made by Katz (1967) and by Tribus: "if we recognize that entropy only measures the extent of our ignorance about the detailed behavior of a system, we can see that the increase is in our confusion about the system" (Tribus 1961, p. 146).

Greene (2011, p. 251-253) gives intuitively clear examples of the idea of physical entropy as equating to missing, or, as he puts it, hidden, information. Entropy, the number of rearrangements of a system's microscopic constituents that leave its overall macroscopic features unchanged, is a measure of the gap in information between the data we have, about the macroscopic arrangements, and the data we do not have, about the particular microstate. The microscopic details include information which is hidden when we consider only macroscopic properties; we can quantify this in Shannon's terms as the number of yes/no questions that could be answered if we knew all the unknown microscopic details. Entropy is the measure of hidden information, measured in this way.

**Conclusions**
"Information is information", wrote Norbert Wiener (1961, p. 132), "not matter or energy". It is not entropy either.

In an editorial, warning of the uncritical and over-enthusiastic use of what even he was by then calling information theory, Claude Shannon (1956, p.3) wrote that "the use of a few exciting words like *information*, *entropy*, *redundancy*, do not solve all our problems" [his italics].

Consideration of the relation between information and physical entropy over the past three decades, since the Shaw and Davis paper, has been very productive for the physical sciences, introducing information as a fundamental physical quantity.

But it has not been so not so productive for the information sciences; here is it the wrong question, since information is not entropy or negentropy, though it is related

to both. The interesting question for the information sciences is the relation between information and 'interesting complexity'. This has also developed considerably since Shaw and Davis's review (Bawden and Robinson 2015).

Information is not entropy, other than in a purely formalistic or metaphorical sense. And physical entropy is not information, since it involves, heat, temperature, gravity and other physical quantities. Entropy can therefore only be equated with information if we accept the controversial argument that all reality is at root information (Bawden and Robinson, 2013). The success of the 'information entropy' formalism in physical studies does not affect this conceptual point. Similarly, there is now general agreement, with some remaining concerns, that entropy is not intrinsically subjective, dependent only on our state of knowledge.

Information is associated with both pattern and randomness. We can now see clearly that there are two aspects to information role in creating complexity. For the unpredictable, and arguably creative, aspects, Shannon information is appropriate. For the aspects of structure and organisation, Wiener's negative entropy concept is inappropriate, and too simplistic. We need a better way of understanding and formalising this aspect of information, not more wrangling over the relation between information and entropy.

Physical entropy is best understood as uncertainty, and, to a certain extent, disorder. Shannon's theory is the best general formula for entropy, in all its instantiations, since all depend on the probabilities laws. Perhaps Von Neumann was right all along.

Wiener's insight that information is related in some way to order needs to be dealt with, not by amending Shannon's ideas but by adding to them. Arguably, negentropy would be better defined as difference between maximum entropy and the actual Shannon value, rather than the negative of Shannon.

The philosophers of information, Dretske (1981) by implication, and Floridi (2010) explicitly, take Shannon's viewpoint. But Floridi (2013), reminding us that Shannon's theory is best understood as a constraint on what we can say about information and its role, speaks of an 'Archipelago of information', different definitions of information for different purposes. Similarly, there are different forms of entropy for different purposes (Frigg and Werndl 2011). This reinforces the value of studies of information and related concepts in different domains (Furner, 2014; Robinson and Bawden, 2014).

The relation between information and entropy, order and disorder then, is far from clear. Perhaps it is more helpful to consider information and complexity (or simplicity), rather than information and order (or disorder), though entropy is also involved in any such comsideration (Bawden and Robinson, 2015)

**Acknowledgements**

# References


Albert, D.Z. (2000), Time and chance, Cambridge MA: Harvard University Press

Atkins, P. (2010), *The laws of thermodynamics: a very short introduction*, Oxford: Oxford University Press

Atkins, P. (2007), *Four laws that drive the universe*, Oxford: Oxford University Press

Atkins, P. (1984), *The second law: energy, chaos and form*, San Francisco: W.H. Freeman

Baer, R.M. (1953), Some general remarks on information theory and entropy, in Quastler H. (ed.), *Essays on the use of information theory in biology*, Urbana IL: University of Illinois Press, pp 21-24

Baggott, J. (2013), *Farewell to reality: how fairytale physics betrays the search for scientific truth*, London: Constable

Baranger, M. (2001), Chaos, complexity and entropy, unpublished presentation, available at http://necsi.edu/projects/baranger/cce.pdf

Bates, M.J. (2006), Fundamental forms of information, *Journal of the American Society for Information Science and Technology*, 57(8), 1033-1045

Bateson, G. (1970), Form, substance and difference, *General Semantics Bulletin*, 37, 221-245, reprinted in Bateson, G. (2000), *Steps to an ecology of mind*, Chicago IL: University of Chicago Press

Bateson, G. (1971), A re-examination of "Bateson's rule", *Journal of Genetics*, 60(3), 230-240, reprinted in Bateson, G. (2000), *Steps to an ecology of mind*, Chicago IL: University of Chicago Press

Baumgartner, B. (2014). Characterizing entropy in statistical physics and in quantum information theory. *Foundations of Physics*, 44(10), 1107-1123

Bawden, D. (2001). The shifting terminologies of information. *Aslib Proceedings*, 2001, 53(3), 93-98

Bawden, D. and Robinson, L. (2015). "Waiting for Carnot": information and complexity. In preparation.

Bawden, D. and Robinson, L. (2013) "Deep down things": in what ways is information physical, and why does it matter for LIS?, *Information Research*, 18(3), paper C03 [online], available at http://InformationR.net/ir/18-3/colis/paperC03.html

Bawden, D. and Robinson, L. (2012), *Introduction to information science*, London: Facet

Beer, S. (1972), *The brain of the firm*, London: Allen Lane

Beinhocker, E.D. (2006), *The origin of wealth: evolution, complexity, and the radical remaking of economics*, Cambridge MA: Harvard Business School Press

Bell D.A. (1953), *Information theory and its engineering applications*, London: Pitman

Bell D.A. (1956), *Information theory and its engineering applications (2$^{nd}$ edn.)*, London: Pitman

Bell D.A. (1962), *Information theory and its engineering applications (3$^{rd}$ edn.)*, London: Pitman

Bell D.A. (1968), *Information theory and its engineering applications (4$^{th}$ edn.)*, London: Pitman

Ben-Naim, A. (2011), Entropy: order or information, *Journal of Chemical Education*, 88(5), 594-596

Ben-Naim, A. (2009), An informational-theoretical formulation of the second law of thermodynamics, *Journal of Chemical Education*, 86(1), 99-105

Ben-Naim A. (2008), *A farewell to entropy: statistical thermodynamics based on information*, Singapore: World Scientific

Bennett, C.H. and Landauer, R. (1985), The fundamental physical limits of computation, *Scientific American*, 253 (July 1985), 48-56

Boltzmann, L. (1872), Weitere Studien über das Wärmgleichgewicht unter Gasmolekülen [Further studies on the thermal equilibrium of gas molecules], *Sitzungsberichte der Akademie der Wissenschaften Wien*, 66(3), 275-370

Blum, H.F. (1951), *Time's arrow and evolution*, Princeton NJ: Princeton University Press

Blum, H.F. (1968), *Time's arrow and evolution (3$^{rd}$ edn)*, Princeton NJ: Princeton University Press

Branson, H.R. (1953), A definition of information from the thermodynamics of irreversible processes and its application to chemical communication in biological systems, in Quastler H. (ed.), *Essays on the use of information theory in biology*, Urbana IL: University of Illinois Press, pp 25-40

Brier, S. (2013), The transdisciplinary view of information theory from a cybersemiotic perspective, in *Theories of information, communication and knowledge: a multidisciplinary approach*, Ibekwe-SanJuan, F. and Dousa, T. (eds.), Dordrecht: Springer, pp 23-49

Brier, S. (2010), Cybersemiotics: an evolutionary world view going beyond entropy and information into the question of meaning, *Entropy*, 12(8), 1902-1920

Brillouin, L. (1962), *Science and information theory (2nd ed)*, New York NY: Academic Press

Brillouin, L. (1953), The negentropy principle of information, *Journal of Applied Physics*, 24(9), 1152-1163

Brillouin, L. (1951A), Maxwell's Demon cannot operate: information and entropy, 1. *Journal of Applied Physics*, 22(3), 334-337

Brillouin, L. (1951B), Physical entropy and information. II, *Journal of Applied Physics*, 22(3), 338-343

Brillouin, L. (1949), Life, thermodynamics and cybernetics, *American Scientist*, 37(4), 554-568

Brissaud, J-B. (2005), The meanings of entropy, *Entropy*, 7(1), 68-96

Brookes, B.C. (1980), The foundations of information science. Part 1: philosophical aspects, *Journal of Information Science*, 2(3/4), 125-133

Brooks, D.R. and Wiley, E.O. (1988), *Evolution as entropy: towards a unified theory of biology (2nd edn.)*, Chicago IL: University of Chicago Press

Burgin, M. (2010), Theory of information: fundamentality, diversity and unification, World Scientific: Singapore

Capurro, R. and Hjørland, B. (2003), The concept of information, *Annual Review of Information Science and Technology*, 37, 343-411

Carroll, S. (2011), *From eternity to here: the quest for the ultimate theory of time*, Oxford: Oneworld Publications

Clausius, R. (1865), Über verschiedene für die Anwendugen bequeme Formen der Haupgleichungen der mechanischen Wärmtheorie [On different forms of the fundamental equations of the mechaical theory of heat and their convenience for application], *Pogendorff's Annalen der Physik*, 125(2), 353-399

Cornelius, I. (2014). Epistemological challenges for information science: constructing information, in F. Ibekwe-SanJuan and T.M. Dousa (eds.), *Theories of information, communication and knowledge*, Dorderecht: Springer, pp. 181-203

Davies, P. (1987), *The Cosmic Blueprint: Order and Complexity at the Edge of Chaos*, Penguin, London.

Davison, M. and Shiner, J.S. (2005), Extended entropies and disorder, *Advances in Complex Systems*, 8(1), 125-158

Denbigh, K. (1981), How subjective is entropy?, *Chemistry in Britain*, 17(4), 168-185

Denbigh, K.G. and Denbigh, J.S. (1985), *Entropy in relation to incomplete knowledge*, Cambridge: Cambridge University Press

Dretske, F.I. (1981). *Knowledge and the flow of information.* Oxford: Blackwell.

Duncan, T.L. and Semura, J.S. (2007), Information loss as a foundational principle for the second law of thermodynamics, *Foundations of Physics*, 37(12), 1767-1773

Duncan, T.L. and Semura, J.S. (2004), The deep physics behind the second law: information and energy as independent forms of bookkeeping, *Entropy*, 6(1), 21-29

Eddington, A.S. (1928), *The nature of the physical world*, Cambridge: Cambridge University Press

Fayngold, M. and Fayngold, V. (2013), *Quantum mechanics and quantum information*, Weinheim: Wiley-VCH

Fisher, R.A. (1922), On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society of London*, Section A 222, 309-368

Fisher, R.A. (1935), *The design of experiments*, Edinburgh: Oliver and Boyd

Fisher, R.A. (1950), *Contributions to mathematical statistics*, New York, Wiley

Floridi, L. (2010), *Information – a very short introduction*, Oxford: Oxford University Press

Floridi, L. (2013). Unpublished keynote presentation at the 8th Conceptions of Library and Information Science Conference, Copenhagen, August 2013.

Ford, I. (2013), *Statistical physics: an entropic approach*, Chichester: Wiley

Frigg, R. and Werndl, C. (2011), Entropy – a guide for the perplexed, in Beisbart, C. and Hartmann, S. (eds.), *Probabilities in physics*, Oxford: Oxford University Press, pp 115-142

Furner, J. (2004). Information studies without information. *Library Trends*, 52(3), 427-446

Furner, J. (2014), Information without information studies, in *Theories of information, communication and knowledge: a multidisciplinary approach*, Ibekwe-SanJuan, F. and Dousa, T. (eds.), Dordrecht: Springer, pp 143-179

Gatlin, L. (1972), *Information theory and the living system*, New York NY: Columbia University Press

Gaukroger, S. (2012), *Objectivity: a very short introduction*, Oxford: Oxford University Press

Gell-Mann, M. and Lloyd, S. (1996), Information measures, effective complexity and total information, *Complexity*, 2 (1), 44-52

Georgii, H-O. (2003), Probabilistic aspects of entropy, in Greven, A., Keller, G. and Warnacke, G. (eds.), *Entropy*, Princeton NJ: Princeton University Press, pp 37-54

Gibbs, J.W. (1902), *Elementary principles in statistical mechanics*, New Haven, CT: Yale University Press

Gharibyan, H. and Tegmark, M. (2013), Sharpening the second law of thermodynamics with the quantum Bayes' theorem, arXiv:1309.7349 [quant-ph], retrieved from http://arxiv.org/abs/1309.7349 30 April 2014

Gleick, J. (2011), *The information: a history, a theory, a flood*, London: Fourth Estate

Greene, B. (2011), *The Hidden Reality*, London: Allen Lane

Greer, B. (2009), *Entropy*, Hove: Picnic Publishing

Greven, A., Keller, G. and Warnacke, G. (2003), Preface to Greven, A., Keller, G. and Warnacke, G. (eds.), *Entropy*, Princeton NJ: Princeton University Press, pp xi-xii

Gribbin, J. (2009), *In search of the multiverse*, London: Penguin

Hargatti. I. (2006). *The Martians of science: Five physicists who changed the twentieth century*, Oxford: Oxford University Press

Harte, J. and Newman. E.A. (2014). Maximum information entropy: a foundation for ecological theory. *Trends in ecology and evolution*, 29(7), 384-389.

Hayles, N.K. (1999), *How we became post-human: virtual bodies in cybernetics, literature and informatics*, Chicago IL: University of Chicago Press

Hayles, N.K. (ed) (1991), *Chaos and order: complex dynamics in literature and science*, Chicago IL: University of Chicago Press

Hayles, N.K. (1990), *Chaos bound: orderly disorder in contemporary literature and science*, Ithaca NY: Cornell University Press

Hemmo, M. and Shenker, O.R. (2012), *The road to Maxwell's Demon: conceptual foundations of statistical mechanics*, Cambridge: Cambridge University Press

Hjørland, B. (2000). Library and information science: practice, theory and philosophical basis. *Information Processing and Management*, 36(3), 501-531

Hjørland, B. (2007). Information: objective or subjective/situational? *Journal of the American Society for Information Science and Technology*, 58(10), 1448-1456

Hjørland, B. (2009). Concept theory *Journal of the American Society for Information Science and Technology*, 60(8), 1519-1536

Hjørland, B. (2011). The importance of theories of knowledge: indexing and information retrieval as an example. *Journal of the American Society for Information Science and Technology*, 62(1), 72-77

Hobson, A. (1971), *Concepts in statistical mechanics*, New York NY: Gordon and Breach

Holt J. (2012), *Why does the world exist? An existential detective story*, London: Profile Books

Jaynes, E.T. (1957A), Information theory and statistical mechanics, *Physical Review*, 106(4), 620-630

Jaynes, E.T. (1957B), Information theory and statistical mechanics II, *Physical Review*, 108(2), 171-190

Katz, A. (1967), *Principles of statistical mechanics: the information theory approach*, San Francisco; W.H. Freeman

Kauffman, S.A. (2000), *Investigations*, New York NY, Oxford University Press

Kostic, M.M. (2014), The elusive nature of entropy and its physical meaning, *Entropy*, 16(2), 953-967

Lambert F.L. (1999), Shuffled cards, messy desks and disorderly dorm rooms – examples of entropy increase? Nonsense!, *Journal of Chemical Education*, 76(10), 1385-1387 [an updated and annotated version is available online at http://entropysite.oxy.edu/shuffled_cards.html]

Lambert F.L. (2002),  Disorder – a cracked crutch for supporting entropy discussions, *Journal of Chemical Education*, 79(2), 187-192 [an updated and annotated version is available online at http://entropysite.oxy.edu/cracked_crutch.html]

Landauer, R. (1991). Information is physical, *Physics Today*, 44(5), 23-29

Landsberg, P.T. (1961), *Thermodynamics with quantum statistical illustrations*, New York NY: Interscience

Lansdale, M.W. (1991), Remembering about documents: memory for appearance, format and location, *Ergonomics*, 34(8), 1161-1178

Lebowitz, J.L and Maes, C. (2003), Entropy: a dialogue, in Greven, A., Keller, G. and Warnacke, G. (eds.), *Entropy*, Princeton NJ: Princeton University Press, pp 269-276

Leff, H.S. (2012), Removing the mystery of entropy and thermodynamics – part V, *The Physics Teacher*, 50(2), 274-276

Leff, H.S. (1996), Thermodynamic entropy: the spreading and sharing of energy, *American Journal of Physics*, 64(10), 1261-1271

Leff, H.S. and Rex, A.F. (1990), *Maxwell's Demon: entropy, information, computing*, Bristol: IOP Publishing

Leff, H. S., and Rex, A.F. (2002), *Maxwell's Demon 2: Entropy, classical and quantum information, computing*, Bristol: IOP Publishing

Lewis, G.N. (1930), The symmetry of time in physics, *Science*, 71(1849), 569-577

Lieb, E.H. and Yngvason, J. (2003), The entropy of classical thermodynamics, in Greven, A., Keller, G. and Warnacke, G. (eds.), *Entropy*, Princeton NJ: Princeton University Press, pp 147-195

Logan, R.K. (2012), What is information?: why is it relativistic and what is its relationship to materiality, meaning and organization?, *Information*, 3(1), 68-91

Lossee, R.M. (2012), *Information from processes: the science of information*, University of North Carolina, Chapel Hill [electronic book], http://InformationFromProcesses.org

Lossee, R.M. (1997), A discipline independent definition of information, *Journal of the American Society for Information Science*, 48(3), 254-269

MacKay, D.M. (1969), *Information, mechanism and meaning*, Cambridge MA: MIT Press

Maxwell, J.C. (1878), Diffusion, *Encyclopaedia Britannica*, 9th edn. vol. 7, 214-221, Edinburgh: A&C Black

Mingers, J. (2014). What is Information such that there can be Information Systems? University of Kent: Working Paper No. 302. http://www.kent.ac.uk/kbs/documents/res/working-papers/2014/302_2014.pdf

Mitrokhin, Y. (2014). Two faces of entropy and information in biological systems, *Journal of Theoretical Biology*, 359, 192-198

Moorcock, M. (1972), *The English assassin: a romance of entropy*, London: Allison and Busby

Müller, I. (2007), *A history of thermodynamics: the doctrine of energy and entropy*, Berlin: Springer

Nielsen, M.A. and Chang, I.L. (2000), *Quantum computation and quantum information*, Cambridge: Cambridge University Press

Pauli, W. (1933), *Handbuch der Physik, Band 24*, Berlin: Verlag von Julius Springer

Penrose, R. (2004), *The road to reality: a complete guide to the laws of the universe*, London: Vintage

Peterson, J. (2014), Evolution, entropy and biological information, *The American Biology Teacher*, 76(2), 88-92

Pynchon, T. (1973), *Gravity's rainbow*, New York NY: Viking

Quastler H. (ed.) (1953), *Essays on the use of information theory in biology*, Urbana IL: University of Illinois Press

Qvortrup, L. (1993). The controversy over the concept of information. An overview and a selected and annotated bibliography. *Cybernetics and human knowing*, *1*(4), 3-24.

Raisbeck, G. (1964), *Information theory: an introduction for scientists and engineers*. Cambridge MA: MIT Press.

Raymond, R.C. (1951), The well-informed heat engine, American Journal of Physics, 19(2), 109-112

Rifkin, J. (1980), *Entropy: a new world view*, New York NY: Viking

Robinson, L. and Bawden, D. (2014), Mind the gap: transitions between concepts of information in varied domains, in *Theories of information, communication and*

*knowledge: a multidisciplinary approach*, Ibekwe-SanJuan, F. and Dousa, T. (eds.), Dordrecht: Springer, pp 121-141

Rothstein, J. (1951). Information, measurement and quantum mechanics, *Science*, 114, 171-175

Rothstein, J. (1952A). Organization and entropy, *Journal of Applied Physics*, 23(11), 1281-1282

Rothstein, J. (1952B), Information and thermodynamics, *Physical Review,* 85(1), 135

Ruelle, D. (1991), *Chance and chaos*, Princeton NJ: Princeton University Press

Samuelson, P. (1965), Some notions on causality and teleology in economics, in Lerner, D. (ed.), *Cause and effect*, New York NY: The Free Press, 1965, reprinted in Merton, R.C. (ed.), *The collected scientific papers of Paul A. Samuelson*, Cambridge Mass: The M.I.T. Press, 1972, pp 428-472.

Schneider, E.D. and Sagan, D. (2005), *Into the cool: energy, flow, thermodynamics and life*, Chicago IL: University of Chicago Press

Schrödinger, E. (1944), *What is life? The physical aspect of the living cell*, Cambridge: Cambridge University Press

Sethna, J.P. (2006), *Entropy, order parameters and complexity*, Oxford: Oxford University Press

Shannon, C.E. (1953), The lattice theory of information, *IRE Transactions on Information Theory*, 191), 105-107

Shannon, C.E. (1956), The bandwagon, *IRE Transactions on Information Theory*, 2(1), 3

Shannon, C.E. (1948), A mathematical theory of communication, *Bell System Technical Journal*, 27(3), 379-423

Shannon, C.E. and Weaver, W. (1949), *The mathematical theory of communication*, Urbana IL: University of Illinois Press

Shaw, D., and Davis, C. H. (1983), Entropy and information: A multidisciplinary overview. *Journal of the American Society for Information Science*, 34(1), 67-74

Shterenberg, M.I. (2013), Basic concepts of entropy, order, organization, information, knowledge and meaning, *Scientific and Technical Information Processing*, 40(3), 13-18

Simberloff, D. (1978), Entropy, information and life: biophysics in the novels of Thomas Pynchon, *Perspectives in Biology and Medicine*, 21(4), 617-625

Simon, H.A. (1962), The architecture of complexity, *Proceedings of the American Philosophical Society*, 106(6), 467-482

Slater, J.C. (1939), *Introduction to chemical physics*, New York NY: McGraw Hill

Smith, J.D.H. (2001), Some observations on the concepts of information-theoretic entropy and randomness, *Entropy*, 3(1), 1-11

Smolin, L. (2013), *Time reborn*, London: Allen Lane

Spiekermann, S. and Korunovska, J. (2014), The importance of interface complexity and entropy for online information sharing, *Behaviour and Information Technology*, 33(6), 636-645

Stonier, T. (1990), *Information and the internal structure of the universe*, London: Springer Verlag

Styer, D.F. (2000), Insight into entropy, *American Journal of Physics*, 68(12), 1090-1096

Szilard, L. (1929), Über die Entropiever-minderung in einem hermodynamischen System bei Eingriffen intelligenter Wesen [On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings], *Zeitschrift für Physik*, 53(6), 840-856, translated into English by A Rapoport and M. Knoller, and reproduced in Leff and Rex (1990), pp 124-133

Tegmark, M. (2014), *Our mathematical universe*, London: Allen Lane

Tolman, R.C. (1938), *The principles of statistical mechanics*, Oxford: Oxford University Press

Tribus, M. (1961). Thermostatics and thermodynamics: an introduction to energy, information and states of matter, Princeton NJ: van Nostrand

Tribus, M. (1964), Information theory and thermodynamics, in Johnson, H.A. (ed.), *Heat transfer, thermodynamics and education: Boelter anniversary volume*, New York NY: McGraw Hill,  pp 348-368

Tribus, M. (1999). A tribute to ET Jaynes, in W. von der Linde, V. Dose, R. Fischer and R. Preuss (eds.), *Maximum entropy and Bayesian Methods. Proceedings of the 18th International Workshop on Maximum entropy and Bayesian Methods of statistical analysis, Garschang, Germay 1998,* Berlin: Springer, pp 11-20

Tribus, M. and McIrvine, E.C. (1971), Energy and information, *Scientific American*, 224(3), 179-188

Tseng, C-Y. and Tuszynski, J. (2014). Maximum entropy in drug discovery. *Entropy*, 16(7), 3754-3768.

Vedral, V. (2010), *Decoding reality: the universe as quantum information*, Oxford: Oxford University Press

Vedral, V. (2012), In from the cold, *New Scientist*, 216(no. 2886), 33-37

Vieland, V.J., Das, J., Hodge, S.E. and Seok, S. (2013), Measurement of statistical evidence on an absolute scale following thermodynamic principles, *Theory in Biosciences*, 132(3), 181-194

Vohs, K.D., Redden, J.P. and Rahinel, R. (2013), Physical order produces healthy choices, generosity and conventionality, whereas disorder produces creativity, *Psychological Sciences*, 24(9), 1860-1867

Von Baeyer, H.C. (2003), *Information: the new language of science*, London: Weidenfeld and Nicholson

Wallace, D. (2012), *The emergent multiverse: quantum theory according to the Everett interpretation*, Oxford: Oxford University Press

Wallace, D. (2010), Gravity, entropy and cosmology: in search of clarity, *British Journal for the Philosophy of Science*, 61(3), 513-540

White, E.C. (1991), Negentropy, noise and emancipatory thought, Chapter 12 of Hayles, N.K. (ed) (1991), *Chaos and order: complex dynamics in literature and science*, Chicago IL: University of Chicago Press, pp 263-277

Wicken, J.S. (1987A), *Evolution, thermodynamics and information: extending the Darwinian program*, New York NY: Oxford University Press

Wicken, J.S. (1987B), Entropy and information: suggestions for common language, *Philosophy of Science*, 54(2), 176-193

Wiener, N. (1948), *Cybernetics, or control and communication in the animal and the machine*, New York, NY: John Wiley and Sons

Wiener, N. (1961), *Cybernetics, or control and communication in the animal and the machine (2nd edn.)*, Cambridge MA: MIT Press

Wiener, N. (1950), *The human use of human beings: cybernetics and society*, London: Eyre and Spottiswood

Wilson, J.A. (1968), Entropy, not negentropy, *Nature*, 219(5153), 535-536

Wolchover, N. (2014), Time's arrow traced to quantum source, *Quanta magazine*, April 16 2014, [online] available at
https://www.simonsfoundation.org/quanta/20140416-times-arrow-traced-to-quantum-source