Entropy and Its Inequalities

Last Update: 16th June 2008

1. von Neumann Versus Shannon Entropy

In classical statistical thermodynamics the entropy is defined as $S = k \log W$, where k is Boltzmann’s constant and W is the number of accessible microstates consistent with the known macrostate of the system. If all the microstates are equally probable, then each has probability p = 1/W. The Boltzmann entropy can then be written $S = -k \log p$. Alternatively, if the different microstates have differing probabilities, say p_i for the i^{th} state, then the entropy would be $-k \log p_i$ with a probability of p_i. So the ensemble average entropy would be $S_{Boltzmann} = -k \sum_i p_i \log p_i$. Dimensionless entropy is defined by dropping the Boltzmann constant factor. We shall assume dimensionless entropy from here on.

In information theory, the Shannon entropy (or information) is defined in the same way. Let’s say a message is transmitted using symbols x_i, and that these are known to occur with probability p_i. (This may be known, for example, because past messages have shown that x_1 occurs with a relative frequency p_1, etc.). How much information is there in a message N symbols long? Well, it is N times the average information per symbol transmitted, and the latter is defined as $S_{Shannon} = -\sum_i p_i \log_2 p_i$. Note that whereas entropy in physics is defined using the natural logarithm, in information theory $\log_2$ is used. This is natural because it means that one ‘evens’ binary choice corresponds to one unit of information (one bit). Some authors sometimes employ entropy defined using logs to the base of some other integer, for example the dimension of some relevant Hilbert space.

By analogy, von Neumann defined the entropy of a mixed quantum state as,

$$S_{vN} = -\mathrm{Tr}\left(\hat{\rho} \log_2 \hat{\rho}\right) \qquad (3.1.1)$$

Consider first of all that the density matrix has been put in diagonal form with respect to an orthonormal basis, $\{|\phi_i\rangle\}$. It can always be written $\hat{\rho} = \sum_i p_i |\phi_i\rangle\langle\phi_i|$ and the von Neumann entropy is then,

$$\begin{aligned}
S_{vN} &= -\mathrm{Tr}(\hat{\rho}\log_2\hat{\rho}) = -\mathrm{Tr}\left(\hat{\rho}\sum_i \log_2 p_i |\phi_i\rangle\langle\phi_i|\right) \\
&= -\mathrm{Tr}\left(\sum_j p_j |\phi_j\rangle\langle\phi_j| \sum_i \log_2 p_i |\phi_i\rangle\langle\phi_i|\right) \\
&= -\sum_k \langle\phi_k| \sum_j p_j |\phi_j\rangle\langle\phi_j| \sum_i \log_2 p_i |\phi_i\rangle\langle\phi_i||\phi_k\rangle \\
&= -\sum_{k,j,i} \delta_{kj} p_j \delta_{ji} \log_2 p_i \delta_{ik} \\
&= -\sum_i p_i \log_2 p_i
\end{aligned}$$

So the von Neumann entropy is the same as the Shannon entropy in this case. This is not surprising since we can imagine the classical message symbols, $\{x_i\}$, to be replaced by the quantum states $\{|\phi_i\rangle\}$. Since the latter are orthogonal they can be distinguished with certainty, as can the classical symbols, and hence there is no physical difference between the two situations.

It is important to realise that the von Neumann entropy is zero for any pure state. In the spectral representation of the density matrix, one p_i will be 1 and the rest zero. Note that this is true even if the pure state in question is expressed as a superposition of some basis states, e.g. $\alpha|\phi_1\rangle + \beta|\phi_2\rangle$. Of course this must be so: mathematically because we can always change basis so that $|\phi_1'\rangle = \alpha|\phi_1\rangle + \beta|\phi_2\rangle$, and physically because there is no more information to be had beyond the specification of the Hilbert state vector.

Note that the von Neumann entropy does not depend upon the basis chosen, because it depends only upon the eigenvalues of the density matrix. Thus changing basis so that $\hat{\rho} \to \hat{U}^{+}\hat{\rho}\hat{U}$ leaves the von Neumann entropy unchanged (and recalling that bases are always related by a unitary transformation). By the same token, a quantum state changes in time by unitary evolution. Consequently unitary transformations of the form $\hat{\rho} \to \hat{U}^{+}\hat{\rho}\hat{U}$ also represent temporal evolution, and hence the entropy is constant so long as unitary evolution applies. Quite how this is consistent with the classical concept of a relentless increase in entropy is discussed below.

The difference between von Neumann and Shannon entropy arises when we consider a mixture of quantum states which are not orthogonal. Suppose now that $\hat{\rho} = \sum_i p_i |\tilde{\phi}_i\rangle\langle\tilde{\phi}_i|$ where the states $\{|\tilde{\phi}_i\rangle\}$ are *not* orthogonal. These states are therefore not distinguishable with certainty. If we ignore this fact we would once again get the Shannon entropy, $S_{\mathrm{Shannon}} = -\sum_i p_i \log_2 p_i$. But in truth the amount of information conveyed by a sequence of $\{|\tilde{\phi}_i\rangle\}$ states must be rather less than this, because of the ‘noise’ caused by the lack of perfect distinguishability of the ‘symbols’. The von Neumann entropy properly accounts for this. An example makes this clear.

Consider once again the mixture of two spin ½ particles which we introduced in Part 1 of these Notes, QM1. One particle is in the z-up spin state, and the other in the x-up state. From the classical (Shannon) point of view, we have a mixture with $p_1 = p_2 = 0.5$, giving an entropy of $-(0.5\log_2 0.5)\times 2 = 1$. The density matrix in the z-representation is,

$$\begin{aligned}\hat{\rho} &= 0.5|\uparrow\rangle\langle\uparrow| + 0.5\times\frac{1}{\sqrt{2}}\left[|\uparrow\rangle + |\downarrow\rangle\right]\frac{1}{\sqrt{2}}\left[\langle\uparrow| + \langle\downarrow|\right] \\ &= \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}\end{aligned}$$

To find the von Neumann entropy we need to diagonalise the density matrix, i.e. to find its eigenvalues. This is done by solving the secular equation $\left\|\begin{pmatrix} 0.75-\lambda & 0.25 \\ 0.25 & 0.25-\lambda \end{pmatrix}\right\| = 0$. This yields $\lambda = 0.1464$ or 0.8536. The von Neumann entropy is thus,

$$S_{vN} = -(0.1464\log_2 0.1464 + 0.8536\log_2 0.8536) = 0.6007$$

So, only ~0.6 of a bit of information would be conveyed per quantum ‘symbol’ transmitted in this example, compared with 1 bit per symbol in the classical Shannon case. Quite generally we find,

$$S_{vN} \le S_{Shannon} \qquad (3.1.2)$$

The equality holds only when the mixture is considered to consist of orthogonal quantum states. This, of course, is possible for any mixture. The von Neumann entropy does not depend upon the basis chosen. The Shannon entropy does. For a quantum mixture the Shannon entropy is just wrong.

Is the von Neumann entropy ‘right’? Certainly it is preferable to the Shannon entropy, which is basis dependent and incorrect on physical grounds. But the formula $S_{vN} = -\mathrm{Tr}(\hat{\rho}\log_2\hat{\rho})$ does have a degree of arbitrariness and some authors have proposed alternative definitions. However, I have seen claims (but not a proof) that the “sub-additivity inequality” (discussed below) is sufficient to imply the von Neumann formula. The sub-additivity inequality holds that, for a bipartite system, the entropy does not exceed the sum of the entropies of its parts: $S_{AB} \le S_A + S_B$. This inequality, together with the condition that the equality hold only for uncorrelated sub-systems ($AB = A\otimes B$), is claimed to yield the von Neumann formula. The inequality $S_{AB} \le S_A + S_B$ seems to be physically motivated, so this is a strong argument in favour of the von Neumann entropy.

2. The von Neumann Entropy Inequalities

We have seen one already, Equ.(3.1.2). There are several more…

2.1 Maximum Entropy Possible

2.1 Maximum Entropy Possible is $S_{vN} \le S_{vN}^{MAX} = \log_2 N$ where N is the number of non-zero eigenvalues of the density matrix. The maximum entropy is realised when there is an equal probability (1/N) for all the orthogonal eigenstates, i.e. maximum randomness. Within a given Hilbert space, the maximum is greatest when all states contribute, i.e. when N is the dimension of the Hilbert space.

2.2 Entropy Change on Mixing - Concavity

Suppose we have several separate mixtures over the same Hilbert space, $\hat{\rho}_1, \hat{\rho}_{2.}$, etc., where, $\hat{\rho}_1 = \sum_i p_{1i} |\phi_i\rangle\langle\phi_i|$, $\hat{\rho}_2 = \sum_i p_{2i} |\phi_i\rangle\langle\phi_i|$, etc. Then we can make a new mixture by combining these mixtures in the ratios (probabilities) q_1, q_2, etc., where $\sum_j q_j = 1$. The new mixture of mixtures is clearly,

$$\hat{\rho}_T = \sum_j q_j \hat{\rho}_j = \sum_{i,j} q_j p_{ji} |\phi_i\rangle\langle\phi_i| \qquad (3.2.2.1)$$

The von Neumann entropy of the mixture of mixtures will generally be greater than the average of the entropies of the constituent mixtures, i.e.,

$$S_{vN}(\hat{\rho}_T) = S_{vN}\left(\sum_j q_j \hat{\rho}_j\right) \ge \sum_j q_j S_{vN}(\hat{\rho}_j) \qquad (3.2.2.2)$$

This is referred to as “concavity”. (To be honest I’m confused as to why it isn’t called “convexity”).

Do not confuse a mixture, in this sense, with physically mixing two or more systems. The latter really means ‘adding’ systems together. In contrast, in a statistical mixture we still have just one system, but we don’t know for sure which one it is. Thus, physically mixing systems is an “AND” operation, whereas a quantum mixture is an “OR” operation.

(3.2.2.2) corresponds to the notion that information is lost on mixing. Physically this is because we have more information when we know all the quantities $\{q_j, \hat{\rho}_j\}$ than when we know only $\hat{\rho}_T$. Knowing only $\hat{\rho}_T$ we cannot re-create the original $\{q_j, \hat{\rho}_j\}$ because there are obviously many ways the total $\hat{\rho}_T$ can be decomposed into sub-mixtures. Thus, information has been lost and the entropy increases. In other words, the less we know about how the final mixture was prepared, the greater the entropy. The equality is achieved iff all the sub-mixture density matrices are the same.

Example: Consider just two sub-mixtures with density matrices,

$$\begin{pmatrix} 0.75 & 0 \\ 0 & 0.25 \end{pmatrix} \text{ and } \begin{pmatrix} 0.125 & 0 \\ 0 & 0.875 \end{pmatrix}$$

The vN entropies of these are respectively $-(0.75\log_2 0.75 + 0.25\log_2 0.25) = 0.8113$ and $-(0.125\log_2 0.125 + 0.875\log_2 0.875) = 0.5436$. Suppose these are combined 50%/50%, then the average entropy before mixing is 0.5 (0.8113 + 0.5436) = 0.6774.

The mixture of mixtures has density matrix $\begin{pmatrix} 0.5(0.75+0.125) & 0 \\ 0 & 0.5(0.25+0.875) \end{pmatrix}$ and this has vN entropy $-(0.4375\log_2 0.4375 + 0.5625\log_2 0.5625) = 0.9887$. Hence, the mixture of mixtures has a larger entropy than the original average entropy (0.9887 > 0.6774).

The general proof of (3.2.2.2) is given in Appendix 1.

2.3 Combined States - Subadditivity

The notion of "combined states" should not be confused with mixing sub-mixtures. Pure quantum states might be composed of two parts. For example, we may be considering deuterons, and each of the pure quantum states of a given deuteron might be considered as a combined state of a proton and a neutron. The simplest Hilbert space of combined states is the direct product of the constituent spaces: $H = H_A \otimes H_B$. In this case the A and B states are uncorrelated, i.e., given a state of the A component there is no preference for any particular B state. In this case the von Neumann entropy displays its "extensive" nature, i.e., the total entropy is just the sum of the entropies of the component sub-states,

For $H = H_A \otimes H_B$:

$$S_{vN}^{AB} = S_{vN}^{A} + S_{vN}^{B} \qquad (3.2.3.1)$$

This is obvious because the probability of the combined state i of A and j of B is clearly just $p_i^A p_j^B$, and hence,

$$\begin{aligned} S_{vN}^{AB} &= -\sum_{i,j} p_i^A p_j^B \log_2 p_i^A p_j^B = -\sum_{i,j} p_i^A p_j^B \left(\log_2 p_i^A + \log_2 p_j^B\right) \\ &= -\sum_{i,j} p_i^A p_j^B \left(\log_2 p_i^A\right) - \sum_{i,j} p_i^A p_j^B \left(\log_2 p_j^B\right) \qquad (3.2.3.2) \\ &= -\sum_{i} p_i^A \left(\log_2 p_i^A\right) - \sum_{j} p_j^B \left(\log_2 p_j^B\right) = S_{vN}^{A} + S_{vN}^{B} \end{aligned}$$

But what happens if we confine the combined state Hilbert space to some sub-set of the product space, i.e., $H \subset H_A \otimes H_B$. For example, a deuteron must be a spin 1 combination of a proton and a neutron, since the singlet spin state is not bound by the strong nuclear force. This means there are now correlations between the components states. In our deuteron example the spin states of the proton and neutron are correlated. There are therefore fewer states available to the combined system than were available in the product space $H_A \otimes H_B$. In view of (3.2.3.1) we therefore expect,

For $H \subset H_A \otimes H_B$: $$S_{vN}^{AB} \le S_{vN}^{A} + S_{vN}^{B} \qquad (3.2.3.3)$$

with equality holding only when $H = H_A \otimes H_B$. Physically, the correlations present within H create a degree of order, and hence reduce the entropy. Inequality (3.2.3.3) is known as “subadditivity”. Equality holds iff $H = H_A \otimes H_B$ with no correlations between the two parts.

2.4 Combined States - The Triangle Inequality & Entanglement

(3.2.3.3) gives the upper bound vN entropy for a combined (bipartite) system. Is there a lower bound? For a classical system, the maximum correlation between components A and B would be if, given any state of A then the state of B was fully determined, or vice-versa. Although we have not argued this rigorously, it is reasonable that this leads to a minimum entropy equal to the greater of that of system A and B. Hence, the classical (Shannon) entropy obeys $S_{Shannon}^{AB} \ge MAX\left(S_{Shannon}^{A}, S_{Shannon}^{B}\right)$. This corresponds to the very reasonable notion that, for a classical system, the combined system contains at least as much information as any of its components.

In quantum theory, the von Neumann equivalent is the Araki-Lieb inequality,

$$S_{vN}^{AB} \ge \left|S_{vN}^{A} - S_{vN}^{B}\right| \qquad (3.2.4.1)$$

This is difficult to prove, and was first proved only in 1970 – see Araki & Lieb (1970). (3.2.4.1) together with the sub-additivity inequality yield the triangle inequality,

$$\left|S_{vN}^{A} - S_{vN}^{B}\right| \le S_{vN}^{AB} \le S_{vN}^{A} + S_{vN}^{B} \qquad (3.2.4.2)$$

The name derives from the fact that if the entropies of the individual sub-systems are regarded as the lengths of two sides of a triangle, the entropy of the combined system is restricted to the possible lengths of the third side.

The Araki-Lieb lower bound entropy is remarkable and displays essentially quantum features. In contrast to the classical case, achieving the Araki-Lieb lower bound entropy implies that the combined system will have less entropy (i.e. less information) than either of its components. Classically this is incomprehensible. In quantum mechanics it comes about due to entanglement. Entanglement will be discussed in more detail in Part 7 of theses Notes, QM7. For now, consider the entangled state $\left(\left|\uparrow\right\rangle_A \left|\uparrow\right\rangle_B + \left|\downarrow\right\rangle_A \left|\downarrow\right\rangle_B\right)/\sqrt{2}$. Considered as a combined system it is a pure quantum state and hence has zero entropy.

But considered as separate sub-systems, each particle can be in one of two states. So the entropy of each sub-system (particle) is 1. The Araki-Lieb inequality is respected because $0 \geq |1-1|$. But notice that each sub-system has greater entropy (1) than the combined system (of zero entropy). There is more information in ***each*** of the parts than in the whole. This non-classical behaviour of quantum information is responsible for the quantum weirdness of entangled states, such as the EPR paradox.

2.5 Combined States – Strong Subadditivity

In 1973 Lieb and Ruskai proved a stronger inequality, which contains the ordinary subadditivity property as a special case,

$$S_{vN}^{ABC} + S_{vN}^{B} \leq S_{vN}^{AB} + S_{vN}^{BC} \qquad (3.2.4.1a)$$

This may also be written,

$$S_{vN}^{X\cup Y} + S_{vN}^{X\cap Y} \leq S_{vN}^{X} + S_{vN}^{Y} \qquad (3.2.4.1b)$$

Again it is a major mathematical endeavour to prove this, though simplified derivations have now been presented, e.g. Nielsen & Petz (2005).

2.6 The Entropy of Measurement & Maxwell’s Demon

With the exception of those curious processes called “measurements”, which involve the mysterious $\Re$-process, all other physical evolution is unitary according to quantum mechanics. But unitary evolution leaves the von Neumann entropy unchanged (as shown in Section 3.2.5.3). So all entropy increases would seem to be due to “measurements”, i.e. the $\Re$-process.

Suppose we have a mixed system with density matrix $\hat{\rho} = \sum_i p_i |\phi_i\rangle\langle\phi_i|$ wrt some orthonormal basis $\{|\phi_i\rangle\}$. Let us carry out a measurement of an observable Q on this mixed system. Any given $|\phi_i\rangle$ will give rise to the measurement outcome q_j with a probability, according to the Born Rule, of $|C_{ij}|^2$, where, $C_{ij} = \langle\phi_i|q_j\rangle$. But each state $|\phi_i\rangle$ occurs in the original mixture with relative frequency p_i, so that the overall probability of the measurement outcome q_j is,

$$\tilde{p}_j = \sum_i p_i |C_{ij}|^2 \qquad (3.2.5.1)$$

So the density matrix after measurement will be $\tilde{\hat{\rho}} = \sum_j \tilde{p}_j |q_j\rangle\langle q_j|$. This may also be written $\tilde{\hat{\rho}} = \sum_j \hat{P}_j \hat{\rho} \hat{P}_j$ where $\hat{P}_j = |q_j\rangle\langle q_j|$, i.e., the $\hat{P}_j$ are the projection operators for the measurement, since,

$$\begin{aligned}\tilde{\hat{\rho}} &= \sum_j \hat{P}_j \hat{\rho} \hat{P}_j = \sum_j |q_j\rangle\langle q_j| \left|\sum_i p_i |\phi_i\rangle\langle\phi_i|\right| q_j\rangle\langle q_j| \\ &= \sum_{i,j} |q_j\rangle \left(Cij^* Cij\right) p_i \langle q_j| = \sum_j \tilde{p}_j |q_j\rangle\langle q_j|\end{aligned} \qquad (3.2.5.2)$$

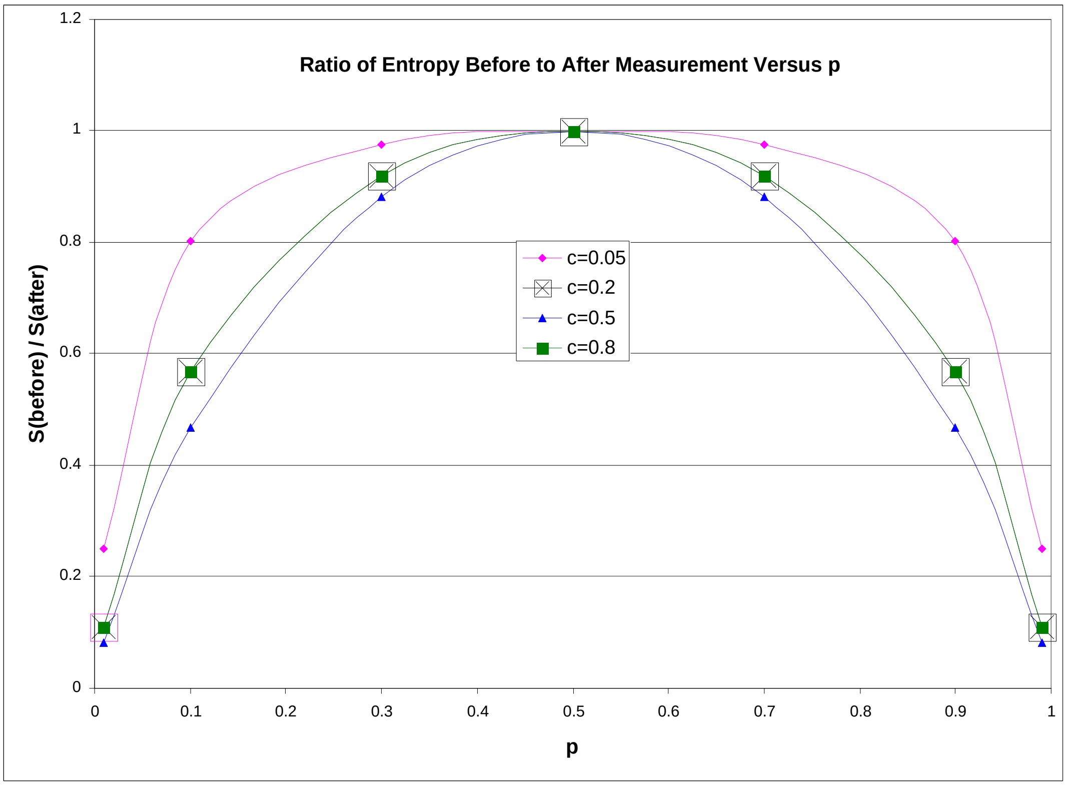
Of course, this density matrix applies only so long as we have not looked to see what measurement outcome has actually been realised. If we did that we would necessarily find just one of the states, $|q_i\rangle$, to which we would then assign an entropy of zero.

The density matrix (3.2.5.2) does, however, account for the “reduction of the wavepacket”; the irreversible, indeterminate, $\Re$-process. So whatever physical interactions are involved when the apparatus ‘measures’ the system, they have all happened prior to the density matrix becoming (3.2.5.2). This includes both any unitary evolution and the non-unitary $\Re$-process, and hence the decoherence implicit in the $\Re$-process. Indeed, it is only because of the non-unitary, irreversible, $\Re$-process that the measurement can result in an increase in entropy. And the measurement *does* increase the entropy. We have, as a mathematical consequence of (3.2.5.1,2),

$$S_{vN}\left(\tilde{\hat{\rho}}\right) = -\mathrm{Tr}\left(\tilde{\hat{\rho}} \log_2 \tilde{\hat{\rho}}\right) \geq S_{vN}(\hat{\rho}) = -\mathrm{Tr}(\hat{\rho} \log_2 \hat{\rho}) \qquad (3.2.5.3)$$

i.e., the von Neumann entropy after measurement is greater than, or equal to, the entropy before measurement. The general proof of this is given in Appendix 2.

<u>Example</u>: Consider a 2D density matrix in diagonal form wrt an orthonormal basis and with diagonal components p and 1-p before measurement. The matrix in (3.2.5.1) which transforms the probabilities before measurement to those after measurement consists of real, positive numbers such that every row and every column add to unity. Hence the most general form of this matrix in 2D is $\begin{pmatrix} c & 1-c \\ 1-c & c \end{pmatrix}$, where $0 \leq c \leq 1$. Hence, the probabilities after measurement are $\tilde{p}_1 = 1 + 2cp - c - p$ and $\tilde{p}_2 = 1 - \tilde{p}_1 = c + p - 2cp$. Hence, the vN entropy before measurement is $-\left(p \log_2 p + (1-p)\log_2(1-p)\right)$ and the vN entropy after is $-\{(1+2cp-c-p)\log_2(1+2cp-c-p) + (c+p-2cp)\log_2(c+p-2cp)\}$. The following graph plots the ratio of the entropy before to that after measurement against p for a number of different values of c. It is readily seen that the entropy after measurement is always greater except when equality holds at p = 0.5.



What about entropy of measurement on a pure state?

We have noted that measurement produces a pure state outcome – so long as we look – and that this has zero entropy. But the entropy of measurement for a mixed state, as given by (3.2.5.1,2,3), requires that we don’t look! If we do look, then even for an initially mixed state we will find a pure state and hence zero entropy. So we haven’t compared like with like.

What if we perform our measurement of Q on an initial pure state which is not an eigenstate of Q? Consider the state $|\phi_1\rangle = \sum_i \langle q_i|\phi_1\rangle|q_i\rangle = \sum_i C^*_{1i}|q_i\rangle$. Before measurement the density matrix is $|\phi_1\rangle\langle\phi_1| = \sum_{i,j} C^*_{1i}C_{1j}|q_i\rangle\langle q_j|$. Note that in the orthonormal basis provided by $\{|q_i\rangle\}$ this density matrix has non-zero off-diagonal components. These establish the correlations that mean that diagonalisation [which, of course, is accomplished in basis $\{|\phi_i\rangle\}$] results in a pure state, i.e. all but one of the diagonal elements are zero.

If we now perform the measurement but do not look at the outcome, we will have a mixed state – because we will know that each state $|q_i\rangle$ has a probability of $|C_{1i}|^2$ of being revealed when we *do* look. So the density matrix after measurement on a pure state, but before looking at the outcome, is $-\sum_i |C_{1i}|^2 \log_2 |C_{1i}|^2$. So measurement creates entropy even when the initial state is pure, so long as we haven’t looked at the outcome.

I have not seen this stated in any texts, but I suspect that the entropy of measurement on a mixed state will always exceed that on a pure state. That is,

$$S_{after}^{pure} = -\sum_i |C_{1i}|^2 \log_2 |C_{1i}|^2 \le S_{after}^{mixture} = -\sum_j \sum_i p_i |C_{ij}|^2 \log_2 \left(\sum_k p_k |C_{kj}|^2 \right) \qquad (3.2.5.4)$$

This may be demonstrated quite easily for the 2D case. The most general form of the matrix of elements $|C_{ij}|^2$ is $\begin{pmatrix} c & 1-c \\ 1-c & c \end{pmatrix}$ where $0 \le c \le 1$. The entropy of the pure state after measurement is thus $S_{after}^{pure} = -(c\log_2 c + (1-c)\log_2(1-c))$. The entropy of the mixture after measurement has already been derived in the previous example and is, $S_{after}^{mixture} = -\{(1+2cp-c-p)\log_2(1+2cp-c-p) + (c+p-2cp)\log_2(c+p-2cp)\}$. The above graph plots the ratio of $-(p\log_2 p + (1-p)\log_2(1-p))$ to $S_{after}^{mixture}$ against p. But $S_{after}^{mixture}$ is symmetrical under exchange of p and c. So the graph of the ratio of $-(c\log_2 c + (1-c)\log_2(1-c))$ to $S_{after}^{mixture}$ against c will look identical. Hence $S_{after}^{pure} / S_{after}^{mixture}$ is always less than unity except for c = 0.5, when the two entropies are equal.

The general proof of (3.2.5.4) is given in Appendix 3.

3. But Isn’t Measurement Supposed to *Provide* Information?

There is a paradox involved in the notion that measurement increases entropy. Measurement is intended to provide us with information about the system. But entropy is a measure of our ignorance about the system. Greater system entropy means that the system could be in a greater number of states as far as we know, roughly speaking. But inequality (3.2.5.3) says that measurement has increased the system entropy, and hence *decreased* our knowledge about the system! That’s a pretty useless measurement if it decreases our knowledge of the system. And yet this applies to all measurements.

The resolution of the paradox is that (3.2.5.3) only applies before we have actually looked at the result of the measurement. At this stage we have not yet taken benefit from the knowledge gained from the measurement. The increase in entropy in (3.2.5.3) is solely that due to the physical interaction between the system (mixture) and the measurement apparatus. If we were considering a classical process, in which measurements could always be accomplished in principle without disturbing the system, then, so long as we were skilled experimentalists, we could always arrange for the equality in (3.2.5.3) to hold. For classical systems the physical process of measurement need not affect the entropy.

After the physical process of measurement has happened, i.e. after the interaction between apparatus and system/mixture is complete, the system/mixture can be regarded as having essentially the same status whether classical or quantum physics is considered. In either case we have a classical (incoherent) mixture of perfectly distinguishable alternatives. Any given physical realisation of the system is now in a definite state, even though we do not yet know what that state is because we have not looked. This applies also to the quantum system, because the measurement has been made and the wavepacket has already collapsed[1]. Consequently, what happens to the entropy when we now "look at what result we got" is exactly the same for the classical and quantum cases. It is at this point that the entropy suddenly nose-dives to zero. Of course it must, because now a single outcome is apparent – there is no longer any uncertainty.

So, ultimately, measurement does reduce entropy (to zero). But this is a completely separate part of the overall measurement process to that considered in Section 2.6. The situation is summarised as,

1) Physical interaction with apparatus, no peeking! - Quantum entropy increases (except for a pure state of the measurement operator).

2) Physical interaction with apparatus, no peeking! - Classical entropy can always be contrived to remain unchanged with sufficient experimental care.

3) Look at measurement outcome: Either classical or quantum: Entropy decreases to zero.

The irreversible, indeterminate $\Re$-process, which applies only in quantum mechanics, is responsible for the entropy increase in (1) which does not apply in (2). If quantum measurement were accomplished via a purely unitary interaction with the apparatus, then quantum measurement would lead to no entropy increase either. The actual entropy decrease which applies eventually, at step (3), is no different for classical or quantum systems/mixtures. It is solely at step (3) that the entropy deceases and information about the system is gained.

So this is one paradox resolved. Unfortunately, it merely raises another. According to the much revered Second Law of Thermodynamics, the entropy of a closed system cannot decrease. How, then, is step (3) possible?

4. How is Measurement Possible?

It is important to realise that step (3) is the same for classical and quantum measurements. Hence this paradox of measurement was always present in classical physics too. In fact, it turns out that it is essentially the same paradox as Maxwell's Sorting Demon.

---

[1] This perspective on things holds that the collapse of the wavepacket is the result of a real physical interaction with the apparatus and does not require the involvement of our consciousness.

Maxwell’s demon is envisaged as sitting by a trapdoor in a panel which divides a gas into two compartments. The gas is initially in thermal equilibrium, the two compartments containing gas at the same pressure, temperature and density. The demon is supposed to spot the speeds of oncoming molecules and open the trapdoor to let fast molecules pass from left to right, but not slow ones. Similarly, the demon lets only slow molecules pass from right to left. In this way the demon engineers a temperature and pressure difference to arise between the gas in the two compartments.

It might be supposed that this apparent violation of the Second Law is rationalised by the demon’s requirement to expend work in making these measurements of speed and in operating the trapdoor. However, it turns out that these operations can, in principle, be carried out with arbitrarily small expenditure of work and hence arbitrarily small entropy increase. It turns out that the resolution of the paradox is identical to that of our classical measurement paradox. The current received wisdom is that both problems come down to the entropy of record erasure [see Leff and Rex (1990), Feynman (1996), Lubkin (1987) or Szilard (1929), for example].

The Accepted Answer relates to how we, as observers, or the Maxwell demon, who is also an observer, records – and subsequently re-records – information. In both cases it is necessary that we have some sort of register which records the result of the observation. This may be simply a computer style register which contains a number. The resolution of the paradox comes by considering a continuing cycle of observations. On each new observation it is necessary to over-write the result of the previous observation which is currently stored in the register. This is an irreversible process. Once the previous record is over-written, that information is lost (i.e. entropy is produced). Since the register must have at least as many possible states as the number of states of the system being measured, it follows that the amount of information lost by over-writing must at least balance the entropy reduction in step (3).

Closer analysis reveals that it is actually the erasure of the previous record which accounts for the entropy production (i.e. information loss). The process of over-writing consists first of erasure to produce some standard “blank” condition for the register, followed by a writing operation starting from this standard “blank” condition. The latter step can be a unitary (reversible) process, whereas the erasure cannot be. So the erasure is responsible for the entropy production and keeping the entropy book-keeping of the universe in accord with the Second Law.

If the information-theoretic explanation leaves you wondering what is happening to the physical entropy, bear in mind that there must be a physical mechanism which causes the erasure. Szilard (1929) considered a model in which erasure was implemented by dunking the register in a heat bath. The production of entropy is then readily apparent.

A possible objection to this explanation is that, on the first observation, no erasure has yet been necessary. But we still loose information from the register when we write to it, unless we have been assured that it is pre-prepared in the “blank” state. For that matter, we can envisage being supplied with a continual stream of new registers which have been pre-prepared in the “blank” state (a blank tape), thus perpetuating the objection across all cycles. The objection to this is that the system is no longer closed. We are being provided with a crucial resource from “outside”: an arbitrarily long blank tape. This blank tape is in a highly ordered state and hence may be regarded as providing “negative entropy”, or functioning as a “garbage can” (Lubkin, 1987).

Thus, a Maxwell Sorting Demon could in principle be built which would cause a gas to be divided into two parts out of equilibrium. But, to function, the demon would require to be supplied with the equivalent of “blank tape” or a means of erasing its register. Either way the total entropy, including the source of blank tape and/or the erasing device, would not decrease. That is the current received wisdom.

5. Difficulties

Personally, I still find the situation unsatisfactory. The root cause of this may be the observation (I think first made by Dirac) that the $\Re$-process is actually the only source of entropy increase in the universe. According to quantum theory, all other physical change comes about by unitary evolution. But unitary evolution leaves entropy unchanged because the states evolve according to $|\psi\rangle \to |\psi'\rangle = U|\psi\rangle$ and hence the density matrix evolves according to,

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \to \hat{\rho}' = \sum_i p_i |\psi'_i\rangle\langle\psi'_i| = \sum_i p_i U|\psi_i\rangle\langle\psi_i|U^+ = U\hat{\rho}U^+$$

and hence,

$$S'_{vN} = -\mathrm{Tr}\left(\hat{\rho}'\log_2 \hat{\rho}'\right) = -\mathrm{Tr}\left(U\hat{\rho}U^+ \log_2 U\hat{\rho}U^+\right) = -\sum_i p_i \log_2 p_i = S_{vN} \quad (3.2.5.3.1)$$

(where we have used the fact that the eigenvalues of a matrix are unchanged by a unitary similarity transformation).

Another expression of my difficulty is that when entropy is expressed in terms of information, it is our lack of knowledge about the particular microstate of the system which causes the system to “possess information”. The role of “our knowledge” in this view renders entropy horribly subjective. This is revealed in step (3) in the following way: suppose a second observer has already peeked at the measuring device before we do. He will now regard the system as having zero entropy, whereas we will not. This subjective interpretation is not at all satisfactory, in my view, because entropy would appear to be a physical property of the system. The maximisation of entropy as a system moves towards thermal equilibrium, from an initial state which is out of equilibrium, is not subjective. The entropy is about how many states the system has available to it, and which it can occupy (and perhaps does, transiently). Our state of knowledge does not come into it.

The objections of both the preceding two paragraphs suggest that step (3) does not involve any entropy change. How can it when the $\Re$ -process, the only physical source of entropy increase, has already happened? Step (3) itself does not involve any $\Re$ -process – *unless* you assume, like Wigner, that the $\Re$ -process does not take place until we perceive the outcome of the measurement, i.e. that our consciousness is crucial to the $\Re$ -process. I am strongly disinclined to believe this. In any case that view seems contradictory. We need only place a human (Wigner’s friend) inside the measurement apparatus. That will implement the $\Re$ -process for us. But we still do not know the measurement outcome yet, because this is merely a statement about our knowledge, and has nothing to do with quantum mechanical niceties.

If the $\Re$ -process is a physical process, albeit non-unitary and currently rather mysterious, then the state of the system post-measurement is already determined, and the fact that we do not know what it is makes no difference to the system. Instead, the entropy changes have already occurred in step (1) or (2). This must be so, since the $\Re$ -process takes place in step (1), and this alone can change entropy.

But in that case, how does entropy arise in classical physics, which has no $\Re$ -process? Perhaps, in a careful analysis, entropy in classical physics cannot consistently be interpreted in terms of information.

< Appendices 1, 2 and 3 to be added >