# Chapter 14

# The Whole is Less than the Sum of its Parts

***The entropy of quantum states and the inequalities it obeys. The quantum weirdness is here manifest in the fact that the information available in a multipartite system can be less than in its parts.***

Last Update: 4/2/12

## 1. von Neumann Entropy Versus Shannon Entropy

In classical statistical thermodynamics the entropy is defined as $S = k_B \log W$ , where $k_B$ 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 $S = -k_B \log p$ . Alternatively, if the different microstates have differing probabilities, say $p_i$ for the i$^{th}$ state, then the entropy would be $-k_B \log p_i$ with a probability of $p_i$. So the ensemble average entropy would be $S_{Boltzmann} = -k_B \sum_i p_i \log p_i$ . 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 classical 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 employ entropy defined using logarithms to some other integer base, for example the dimension of some relevant Hilbert space.

By analogy, von Neumann defined the entropy of a mixed quantum state in terms of its density matrix $\hat{\rho}$ as,

$$S_{vN} = -Tr(\hat{\rho} \log_2 \hat{\rho}) \tag{1}$$

Recall that a function of an operator is defined via the corresponding power series. This means that if an operator is represented by a diagonal matrix, with diagonal elements $p_i$ , a function $f$ of the operator is diagonal with elements $f(p_i)$. Consider then that the density matrix of some system has been put in diagonal form with respect to an orthonormal basis, $\{|\phi_i\rangle\}$. This is always possible because the density matrix is Hermetian. It can always be written $\hat{\rho} = \sum_i p_i |\phi_i\rangle\langle\phi_i|$ and the von Neumann entropy is then,

$$\begin{aligned}
S_{vN} &= -Tr(\hat{\rho}\log_2 \hat{\rho}) = -Tr\left(\hat{\rho}\sum_i \log_2 p_i |\phi_i\rangle\langle\phi_i|\right) \\
&= -Tr\left(\sum_j p_j |\phi_j\rangle\langle\phi_j| \sum_i \log_2 p_i |\phi_i\rangle\langle\phi_i|\right) \\
&= -\sum_k \langle\phi_k| \sum_j p_j |\phi_j\rangle\langle\phi_j| \sum_i \log_2 p_i |\phi_i\rangle\langle\phi_i||\phi_k\rangle \\
&= -\sum_{\mathrm{k,j,i}} \delta_{\mathrm{kj}} p_j \delta_{ji} \log_2 p_i \delta_{ik} \\
&= -\sum_i p_i \log_2 p_i
\end{aligned} \tag{2}$$

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.

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 because the eigenvalues are unchanged (and recall that bases are always related by a unitary transformation, $U$ ).

The Schrodinger equation shows that the quantum state evolves in time by a unitary transformation, $|\psi_t\rangle = U|\psi_0\rangle$, and hence $\hat{\rho} \to \hat{U}\hat{\rho}\hat{U}^+$, where $U = \exp\{-i\hat{H}t/\hbar\}$. Hence the entropy is constant over time so long as unitary evolution applies. This may seem inconsistent with the second law of thermodynamics, but the situation is no different from classical dynamics in this respect. For a perfectly known initial microstate, deterministic classical mechanics implies a perfectly known future microstate, and hence no increase of entropy. The second law will not arise from time-reversible dynamics in either case, whether classical or quantum.

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. Hence $-\sum_i p_i \log_2 p_i$ is 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. Furthermore, the von Neumann entropy is *only* zero for pure states. In a mixed state at least two of the $p_i$ are non-zero, and hence $-\sum_i p_i \log_2 p_i$ must be non-zero (noting that there can be no cancellation between terms in the sum since all terms contribute positive entropy).

Rick's Formulation of Quantum Mechanics QM6: Entropy and Its Inequalities

The difference between von Neumann and Shannon entropy arises when we consider a mixture of quantum states which are not orthogonal. Suppose now that $\hat{\rho} = \sum_i p_i \left|\tilde{\phi}_i\right\rangle\left\langle\tilde{\phi}_i\right|$ where the states $\left\{\left|\tilde{\phi}_i\right\rangle\right\}$ are *not* orthogonal. These states are therefore not distinguishable with certainty. If we ignore this fact we would once again get the Shannon entropy, $S_{Shannon} = -\sum_i p_i \log_2 p_i$ . But in truth the amount of information conveyed by a sequence of $\left\{\left|\tilde{\phi}_i\right\rangle\right\}$ 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 a mixture of two spin ½ particles, one particle 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,

$$\begin{aligned}\hat{\rho} &= 0.5\left|\uparrow\right\rangle\left\langle\uparrow\right| + 0.5\times\frac{1}{\sqrt{2}}\left[\left|\uparrow\right\rangle + \left|\downarrow\right\rangle\right]\frac{1}{\sqrt{2}}\left[\left\langle\uparrow\right| + \left\langle\downarrow\right|\right] \\ &= \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}\end{aligned} \tag{3}$$

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,

$$S_{vN} = -(0.1464\log_2 0.1464 + 0.8536\log_2 0.8536) = 0.6008$$

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,

$$\mathrm{S_{vN}} \le \mathrm{S_{Shannon}} \tag{4}$$

where by the Shannon entropy we mean the function $S_{Shannon} = -\sum_i p_i \log_2 p_i$ when the density matrix is expressed in the particular basis $\left\{\left|\tilde{\phi}_i\right\rangle\right\}$, i.e., $\hat{\rho} = \sum_i p_i \left|\tilde{\phi}_i\right\rangle\left\langle\tilde{\phi}_i\right|$ . The equality in (4) holds only when the mixture is considered to consist of orthogonal quantum states. This, of course, is possible for any mixture by suitably choosing the basis. The von Neumann entropy does not depend upon the basis chosen. The Shannon entropy does. For a quantum state the Shannon entropy is just wrong, see for example Brukner and Zeilinger (2001).

There is a more general version of (4), sometimes referred to as the "ensemble inequality". Denote the von Neumann entropy of a density matrix $\rho$ as $S_{vN}(\rho)$ and the denote the Shannon entropy of a set of probabilities as $S_{Shannon}(p_i) = -\sum_i p_i \log_2 p_i$ .

Suppose we have an ensemble of states, each described by a density matrix $\rho_i$, such that the i[th] state occurs with relative frequency $p_i$ in the ensemble. The density matrix of the ensemble is therefore $\rho = \sum_i p_i \rho_i$. It can be shown that,

$$S_{vN}\left(\sum_i p_i \rho_i\right) \le S_{Shannon}(p_i) + \sum_i p_i S_{vN}(\rho_i) \qquad (5)$$

In the case that each of the individual states $\rho_i$ in the ensemble is a pure state, (5) reduces to (4) because $S_{vN}(\rho_i) = 0$.

If $p_i$ are the eigenvalues of a density matrix in a Hilbert space of dimension *N*, so that the von Neumann entropy is $-\sum_i p_i \log_2 p_i$, its maximum possible value is $\log_2 N$ and this is achieved only if all the probabilities (eigenvalues) are equal, $p_i = 1/N$. This is simple to prove. The maximum of $-\sum_i p_i \log_2 p_i$ subject to the constraint $\sum_i p_i = 1$ is found from the turning point of $-\sum_i p_i \log_2 p_i + \lambda \sum_i p_i$ where $\lambda$ is a Lagrange multiplier. Taking the derivative wrt $p_i$ gives $1 + \lambda + \log_2 p_i = 0$ which means that all the $p_i$ must be equal, and hence equal to $1/N$. So we have,

$$S_{vN} \le \log_2 N \qquad (6)$$

The same inequality applies for Shannon entropy. However, in the case of Shannon entropy, *N* is the number of distinct classical symbols/letters/objects/microstates, whereas for von Neumann entropy *N* is the dimension of the Hilbert space, i.e., the number of reliably distinguishable (orthogonal) quantum states.

## 2. The von Neumann Entropy Inequalities

There are a tranche of inequalities concerning von Neumann entropy of which (5) is one example. Some of these are discussed without proof below. Proofs can be found in many places, see for example Nielsen (1998) or Araki and Lieb (1970).

### 2.1 Entropy of an Ensemble - Concavity

Suppose we have several separate mixed states over the same Hilbert space, $\hat{\rho}_1, \hat{\rho}_2$, etc. Then we can make a new mixture (an ensemble) by combining these mixtures in the ratios (probabilities) $p_1, p_2, \ldots$ The density matrix of the ensemble is $\hat{\rho} = \sum_j p_j \hat{\rho}_j$.

Intuitively we would expect the entropy of the ensemble to be greater than the average of the entropies of its constituent mixed states – because our uncertainty is compounded by not even knowing which of the states, $\rho_i$, the system is in. We have more information when we know all the quantities $\{p_j, \hat{\rho}_j\}$ than when we know only the ensemble density matrix, $\hat{\rho}$. Knowing only $\hat{\rho}$ we cannot re-create the original $\{p_j, \hat{\rho}_j\}$ because there are obviously many ways the ensemble $\hat{\rho}$ can be decomposed into mixed sub-states. 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 following inequality expresses this intuition and is rigorously demonstrable,

$$S_{vN}\left(\sum_j p_j \hat{\rho}_j\right) \geq \sum_j p_j S_{vN}\left(\hat{\rho}_j\right) \tag{7}$$

The equality is achieved iff all the sub-mixture density matrices are the same. (7) is referred to as “concavity” because it is the definition of a concave function. (Personally I would have called it convex, but of course this merely depends upon whether you look at the graph of the function from above or below! (7) is concave from below, convex from above).

Note that (5) and (7) taken together provide both a lower and an upper bound on the entropy of the ensemble,

$$\sum_i p_i S_{vN}\left(\hat{\rho}_i\right) \leq S_{vN}\left(\sum_i p_i \rho_i\right) \leq S_{Shannon}\left(p_i\right) + \sum_i p_i S_{vN}\left(\rho_i\right) \tag{8}$$

Consider the following example. Take just two sub-states 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 $-\left(0.75\log_2 0.75 + 0.25\log_2 0.25\right) = 0.8113$ and $-\left(0.125\log_2 0.125 + 0.875\log_2 0.875\right) = 0.5436$. Suppose these are combined 50%/50%, then the average entropy is 0.5 (0.8113 + 0.5436) = 0.6774. The ensemble of these states 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 $-\left(0.4375\log_2 0.4375 + 0.5625\log_2 0.5625\right) = 0.9887$. Hence, the ensemble has a larger von Neumann entropy than the average entropy of the constituent states, i.e., 0.9887 is greater than 0.6774.

## 2.2 Multipartite Systems and Additivity

A multipartite system can be regarded as consisting of more than one part. A bipartite system can be regarded as comprising two parts, for example the electron and the proton in a hydrogen atom. This should not be confused with an ensemble of states. In a bipartite system, every state comprises both a part A *and* a part B. In an ensemble, the system may be in state $\rho_1$ *or* state $\rho_2$ *or* state $\rho_3$, etc.

The Hilbert space of pure bipartite states is the direct, or tensor, product of the constituent spaces: $H = H_A \otimes H_B$. Do not misunderstand this to mean that the general bipartite pure state is a product state, $|i\rangle_A |j\rangle_B$. Rather the general pure state is a linear superposition of such product states, $\sum_{i,j} a_{ij} |i\rangle_A |j\rangle_B$. The most general mixed state is,

$$\hat{\rho}_{AB} = \sum_{i,j,k,l} C_{ijkl} |i\rangle_A |j\rangle_B \langle k|_A \langle l|_B \tag{9}$$

However it may be that the state in question can be written as a direct product of density matrix states of parts A and B, i.e., $\hat{\rho}_{AB} = \hat{\rho}_A \otimes \hat{\rho}_B$. In this case the probability of any observable of the A subsystem taking a given value is unrelated to the probability of any observable of the B subsystem taking a given value. So we would expect the information (entropy) in the bipartite state to be just the sum of those in the constituent states. It is,

For $\hat{\rho}_{AB} = \hat{\rho}_A \otimes \hat{\rho}_B$:
$$S^{AB}_{vN} = S^{A}_{vN} + S^{B}_{vN} \tag{10}$$

This is the additive property of von Neumann entropy. It results from the uncorrelated nature of the A and B parts, and the logarithm in the definition of entropy. It is the quantum mechanical generalisation of $\log(N_A \times N_B) = \log N_A + \log N_B$, which would apply in the classical case of equal probabilities for $N_A$ possibilities for A and $N_B$ possibilities for B. The proof is trivial,

$$\begin{aligned} S^{AB}_{vN} &= -\sum_{i,j} p^A_i p^B_j \log_2 p^A_i p^B_j = -\sum_{i,j} p^A_i p^B_j \left(\log_2 p^A_i + \log_2 p^B_j\right) \\ &= -\sum_{i,j} p^A_i p^B_j \left(\log_2 p^A_i\right) - \sum_{i,j} p^A_i p^B_j \left(\log_2 p^B_j\right) \\ &= -\sum_{i} p^A_i \left(\log_2 p^A_i\right) - \sum_{j} p^B_j \left(\log_2 p^B_j\right) = S^{A}_{vN} + S^{B}_{vN} \end{aligned} \tag{11}$$

## 2.3 Bipartite Systems and Subadditivity

In the general case the density matrix of a bipartite system is not a product state, $\hat{\rho}_{AB} \neq \hat{\rho}_A \otimes \hat{\rho}_B$, but of the general form (9). In this case, if we are given a bipartite state $\hat{\rho}_{AB}$, what do we mean by the individual states $\hat{\rho}_A$ and $\hat{\rho}_B$ of the sub-systems? We mean the reduced density matrix obtained by “tracing out” the other part (see Chapter 2),

$$\hat{\rho}_A = Tr_B\left(\hat{\rho}_{AB}\right) \equiv \sum_B \langle B|\hat{\rho}_{AB}|B\rangle \text{ and } \qquad \hat{\rho}_B = Tr_A\left(\hat{\rho}_{AB}\right) \equiv \sum_A \langle A|\hat{\rho}_{AB}|A\rangle \tag{12}$$

Since additivity, (10), results from the absence of correlation between A and B, which maximises the number of possibilities, it is reasonable to expect that any correlation between the parts will reduce the entropy of the bipartite state. This is indeed the case and the following subadditivity inequality can be proved,

$$S^{AB}_{vN} \le S^{A}_{vN} + S^{B}_{vN} \tag{13}$$

Equality holds iff $\hat{\rho}_{AB} = \hat{\rho}_A \otimes \hat{\rho}_B$. Hence the equality holds for unentangled pure states. However equality may or may not hold for unentangled mixed states (see Chapter 33). Equality definitely does not hold for entangled states, mixed or not.

Subadditivity says that, as far as information is concerned, the whole is generally less than the sum of its parts, and never more.

## 2.4 Bipartite States and the Triangle Inequality

(13) gives the upper bound von Neumann entropy for a 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.

This leads to a minimum entropy equal to the greater of that of system A and B. Hence, the classical entropy obeys $S^{AB}_{Shannon} \geq MAX\left(S^{A}_{Shannon}, S^{B}_{Shannon}\right)$. 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. It may be perfectly reasonable in the classical world, but in quantum mechanics it is wrong as a general statement. In quantum theory, the von Neumann equivalent is the Araki-Lieb inequality,

$$S^{AB}_{vN} \geq \left|S^{A}_{vN} - S^{B}_{vN}\right| \tag{14}$$

This is difficult to prove, and was first proved only in 1970 – see Araki & Lieb (1970). Together (13) and (14) constitute the triangle inequality,

$$\left|S^{A}_{vN} - S^{B}_{vN}\right| \leq S^{AB}_{vN} \leq S^{A}_{vN} + S^{B}_{vN} \tag{15}$$

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. The remarkable thing, of course, is that the lower bound can actually be achieved. A system with the Araki-Lieb lower bound entropy has less entropy (i.e. less information) than at least one of its components. So, not only is the whole less than the sum of its parts, but in some cases the whole can also be less than the individual parts. Classically this is incomprehensible. In quantum mechanics it comes about due to entanglement (see Chapter 33).

As an example consider the entangled state $\left(|\uparrow\rangle_A|\uparrow\rangle_B + |\downarrow\rangle_A|\downarrow\rangle_B\right)/\sqrt{2}$. 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. This can be confirmed algebraically: the reduced density matrix of each part is $\frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$. The Araki-Lieb inequality is respected because $0 \geq |1-1|$.

But notice that each sub-system has greater entropy (i.e., 1) than the combined system (which has 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,

$$S^{ABC}_{vN} + S^{B}_{vN} \leq S^{AB}_{vN} + S^{BC}_{vN} \tag{16}$$

This may also be written,

$$S^{X\cup Y}_{vN} + S^{X\cap Y}_{vN} \leq S^{X}_{vN} + S^{Y}_{vN} \tag{17}$$

Again it is a major mathematical endeavour to prove this, though a simplified derivation has been given by Nielsen and Petz (2005). Closely related inequalities are,

$$S_{vN}^{A} + S_{vN}^{C} \le S_{vN}^{AB} + S_{vN}^{BC} \tag{18}$$

$$S_{vN}^{AB} \le S_{vN}^{BC} + S_{vN}^{AC} \tag{19a}$$

$$S_{vN}^{BC} \le S_{vN}^{AB} + S_{vN}^{AC} \tag{19b}$$

$$S_{vN}^{AC} \le S_{vN}^{AB} + S_{vN}^{BC} \tag{19c}$$

Equs.(19) are also known as a triangle inequality, but not to be confused with (15).

In every case $S_{vN}^{AB}$ means the von Neumann entropy of density matrix $\hat{\rho}_{AB}$, and the density matrices are related by trace-out, e.g., $\hat{\rho}_{AB} = Tr_C\left(\hat{\rho}_{ABC}\right)$, etc.

**2.6 The Entropy of Measurement**

With the exception of those curious processes called "measurements", which involve the collapse of the wavefunction, all other physical evolution is unitary according to quantum mechanics. Unitary evolution leaves the von Neumann entropy unchanged. So how does a "measurement" result in an entropy increase, and by how much?

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,

$$\tilde{p}_j = \sum_i p_i |C_{ij}|^2 \tag{20}$$

When the measurement outcome is $q_j$ the system will be left in the state $|q_j\rangle$ so the density matrix after measurement will be $\hat{\hat{\rho}} = \sum_j \tilde{p}_j |q_j\rangle\langle q_j|$. Of course, this density matrix applies only so long as we have not looked to see what measurement outcome has actually been realised. There is nothing strange about this as long as you recall that the density matrix accounts for our ordinary lack of knowledge about purely deterministic systems, as well as quantum mechanical effects. If we did look we would necessarily find just one of the states, $|q_j\rangle$, with probability $\tilde{p}_j$.

After the measurement the entropy is $S_{vN}\left(\hat{\hat{\rho}}\right)$ and it can be shown that as a consequence of (20),

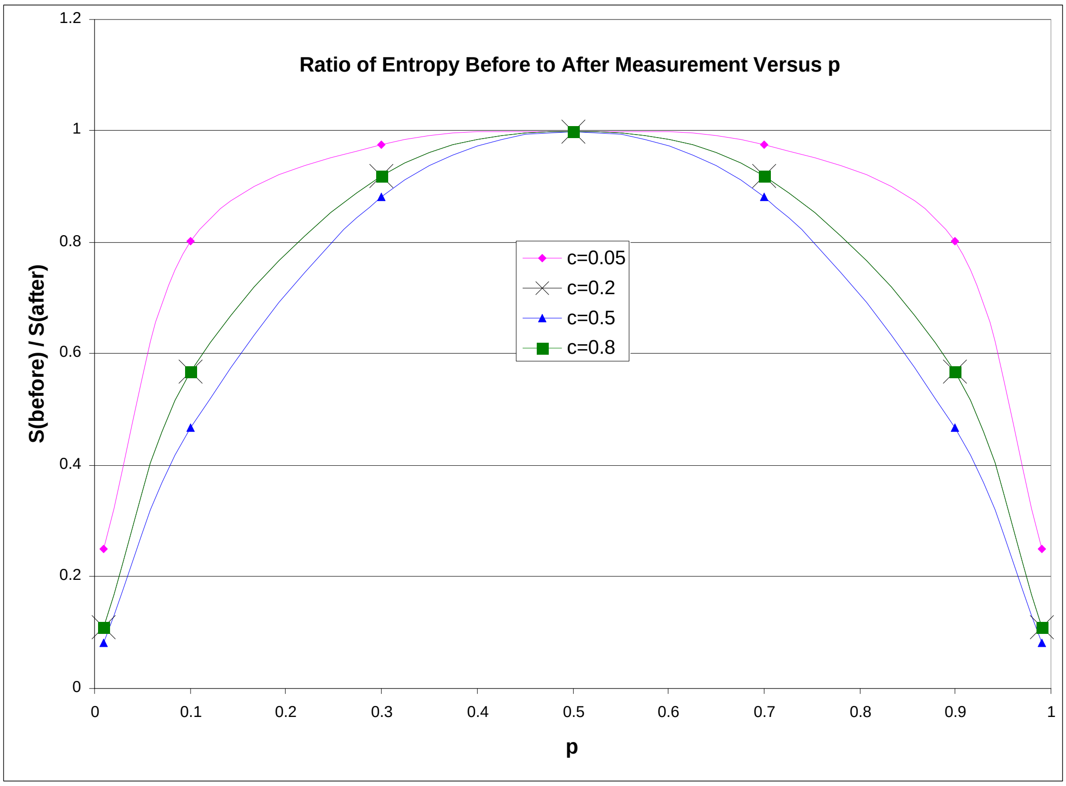
$$S_{vN}\left(\hat{\tilde{\rho}}\right)=-\sum_j \tilde{p}_j \log_2 \tilde{p}_j \geq S_{vN}\left(\hat{\rho}\right)=-\sum_j p_j \log_2 p_j \qquad (21)$$

The entropy after the measurement is greater than that before the measurement.

As an example 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 (20) 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 entropy before measurement is $-\left(p\log_2 p + (1-p)\log_2(1-p)\right)$ and the entropy after measurement is $-\{(1+2cp-c-p)\log_2(1+2cp-c-p)+(c+p-2cp)\log_2(c+p-2cp)\}$. Figure 1 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.

**Figure 1 Ratio of entropies before and after measurement**



What about the entropy after a measurement on a pure state? Suppose a pure state is written in terms of the eigenstates of the observable to be measured as $|\psi\rangle = \sum_i a_i |q_i\rangle$.

The entropy before measurement is zero, of course. But the density matrix after measurement is diagonal with elements $p_i = |a_i|^2$. So the entropy after measurement is non-zero (assuming the initial state was not an eigenstate) namely $-\sum_i |a_i|^2 \log_2 |a_i|^2$, so inequality (21) is respected.

Why does measurement increase entropy? It is because we loose information. For example, in the case of a pure state, after measurement we can no longer obtain any information about the coefficients $a_i$ of the initial state, $|\psi\rangle = \sum_i a_i |q_i\rangle$, because the wavefunction has collapsed to be just one of the states $|q_i\rangle$. This is an irreversible loss of information and hence causes an increase in entropy.

This seems paradoxical because a measurement is supposed to increase our information, not reduce it. But the entropy associated with the post-measurement mixed state applies only before we look at the result of the measurement. When we do we find the specific eigenstate selected by the measurement and the entropy of the system becomes once again zero.

This change, from a multiplicity of possibilities (entropy) to our knowledge of the unique outcome (zero entropy) appears to present another paradox, namely a reduction of entropy. This might be thought to violate the second law of thermodynamics. The resolution of this conundrum lies in the *physical* nature of “our knowledge”. Taking this into account reveals a hidden increase of entropy elsewhere. There is no need to bring consciousness or our brain into the matter. The knowledge (information) can be considered as recorded in a computer register. But such recording is inevitably also associated with erasure of the register’s previous contents, and erasure, being irreversible, involves an increase of entropy. This is discussed in more detail in Plenio and Vitelli (2001) and is essentially identical to the resolution of the famous Maxwell Sorting Demon paradox. These insights are due primarily to Rolf Landauer and Charles Bennett. For further details see Leff and Rex (1990) and Feynman (1999).

**References**

C.Brukner and A.Zeilinger (2001), “Conceptual Inadequacy of the Shannon Information in Quantum Measurements”, Phys.Rev.A63:022113,2001, arXiv:quant-ph/0006087.

M.A.Nielsen (1998), “Quantum Information Theory”, Doctoral dissertation, The University of New Mexico, Albuquerque. Available as arXiv: quant-ph/0011036.

M.A.Nielsen and D.Petz (2005), “A simple proof of the strong subadditivity inequality”, Quantum Information and Computation **6**, 507.

Huzihiro Araki and Elliott H. Lieb, *Entropy Inequalities,* Communications in Mathematical Physics, vol 18, 160-170 (1970).

Elliott H. Lieb and Mary Beth Ruskai, *Proof of the Strong Subadditivity of Quantum-Mechanical Entropy,* Journal of Mathematical Physics, vol 14, 1938-1941 (1973).

Rick's Formulation of Quantum Mechanics QM6: Entropy and Its Inequalities

M.B.Plenio and V.Vitelli (2001) "The physics of forgetting: Landauer's erasure principle and information theory", Contemporary Physics **42**, 25 – 60. arXiv:quant-ph/0103108v1

R.P.Feynman, edited by J.G.Hey and R.W.Allen (1999) "Feynman Lectures on Computation", Penguin, GB.

H.S.Leff and A.F.Rex, editors, (1990) "Maxwell's Demon: Entropy, Information, Computing", Adam Hilger.