Chapter 6

Quantum entropy

There is a notion of entropy which quantifies the amount of uncertainty contained in an ensemble of Qbits. This is the von Neumann entropy that we introduce in this chapter. In some respects it behaves just like Shannon's entropy but in some others it is very different and strange. As an illustration let us immediately say that as in the classical theory, conditioning reduces entropy; but in sharp contrast with classical theory the entropy of a quantum system can be lower than the entropy of its parts.

The von Neumann entropy was first introduced in the realm of quantum statistical mechanics, but we will see in later chapters that it enters naturally in various theorems of quantum information theory.

6.1 Main properties of Shannon entropy

Let X be a random variable taking values x in some alphabet with probabilities $p_x = \mathrm{Prob}(X = x)$. The Shannon entropy of X is

$$H(X) = \sum_x p_x \ln \frac{1}{p_x}$$

and quantifies the *average uncertainty* about X.

The joint entropy of two random variables X, Y is similarly defined as

$$H(X, Y) = \sum_{x,y} p_{x,y} \ln \frac{1}{p_{x,y}}$$

and the conditional entropy

$$H(X|Y) = \sum_y p_y \sum_{x,y} p_{x|y} \ln \frac{1}{p_{x|y}}$$

where

$$p_{x|y} = \frac{p_{x,y}}{p_y}$$

The conditional entropy is the average uncertainty of X given that we observe $Y = y$. It is easily seen that

$$H(X|Y) = H(X,Y) - H(Y)$$

The formula is consistent with the interpretation of $H(X|Y)$: when we observe Y the uncertainty $H(X,Y)$ is reduced by the amount $H(Y)$. The mutual information between X and Y is the complement of the remaining uncertainty $H(X|Y)$

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) = I(Y:X) \end{aligned} \tag{6.1}$$

It is easily seen that $I(X;Y) = 0$ iff $p_{x,y} = p_x p_y$.

The Kullback-Leibler divergence, or relative entropy, between two probability distributions p and q is a useful tool

$$D(p\|q) = \sum_x p_x \ln \frac{1}{q_x} - \sum_x p_x \ln \frac{1}{p_x} = \sum_x p_x \ln \frac{p_x}{q_x}$$

Note that this quantity is not symmetric, $D(p\|q) \neq D(q\|p)$. One can also check that

$$I(X;Y) = I(Y;X) = D(P_{X,Y}\|P_X P_Y)$$

Let us list the main inequalities of classical information theory and indicate which become true or false in the quantum domain.

- The maximum entropy state corresponds to the uniform distribution. For an alphabet with cardinality D we have

  $$0 \leq H(X) \leq \ln D$$

  with the upper bound attained iff $p_x = \frac{1}{D}$. Quantum mechanically this is still true.

- $H(X)$ is a concave functional of p_x. This means that if $p_0(x) = \sum_k a_k p_k(x)$, $a_k \geq 0$, $\sum_k a_k = 1$ then

  $$H_0(X) \geq \sum_k a_k H_k(X)$$

  QMly this is still true.

- Entropy is sub-additive,

  $$H(X,Y) \leq H(X) + H(Y)$$

  Equivalently conditioning reduces entropy $H(X|Y) \leq H(X)$, $H(Y|X) \leq H(Y)$, and mutual information is positive $I(X;Y) \geq 0$. QMly all this is true.

- The conditional entropy is positive, the entropy of (X,Y) is higher than that of X (or Y)

  $$H(X|Y) \geq 0, \qquad H(X,Y) \geq H(X), \qquad H(X,Y) \geq H(Y)$$

  with equality if $Y = f(X)$. QMly this is not true ! We will see that (again !) entanglement is responsible for this !

- Conditioning reduces conditional entropy

  $$H(X|Y,Z) \leq H(X|Y)$$

  This inequality is also called “strong sub-additivity and is equivalent to

  $$H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$$

  Equality is attained iff $X - Y - Z$ form a Markov chain. This means that $p_{x,z|y} = p_{x|y}p_{z|y}$ or equivalently $p_{x,y,z} = p_{z|y}p_{y|x}p_x$ (a Markov chain is reversible: $Z - Y - X$ is also a Markov chain). We will see that QMly strong sub-additivity still holds. In view of the great gap in difficulty between the classical and quantum proofs it is fair to say that this fact is subtle and remarkable. However the notion of Markov chain is not obvious in the quantum case (there is no natural notion of conditional probability) so it is not easily asserted when equality holds.

- A consequence of strong sub-additivity is the data processing inequality obeyed by Markov chains $X - Y - Z$

  $$H(X|Z) \geq H(X|Y)$$

  Indeed $H(X|Z) \geq H(X|Z,Y) = H(X|Y)$ where the first inequality is strong sub-additivity and the equality follows from the fact Z and X are independent given Y. Since the notion of Markov chain is not clear QMly the quantum version of the data processing inequality is a subtle matter.

- The relative entropy is positive

  $$D(p\|q) \geq 0$$

  This is basically a convexity statement which is also true QMly.

- A very useful algebraic identity which follows immediately from definitions, is the chain rule

  $$H(X_1, ..., X_n|Y) = \sum_{i=1}^{n} H(X_i|X_{i+1}, ..., X_n, Y)$$

  and

  $$I(X_1, ..., X_n|Y) = \sum_{i=1}^{n} I(X_i|X_{i+1}, ..., X_n, Y)$$

6.2 Von Neumann entropy and main properties

We assume that the system of interest is described by its density matrix ρ and furthermore we restrict ourselves to the case of a finite dimensional Hilbert space $\dim \mathcal{H} = D$. the von Neumann entropy is by definition

$$S(\rho) = -Tr\rho \ln \rho$$

In physics this quantity gives the right connection between quantum statistical mechanics and thermodynamics when $\rho = e^{-\beta H}/Z$ is the Gibbs state describing a mixture at thermal equilibrium. In quantum information theory this entropy enters in many theorems (data compression, measures of entanglement etc...) and thus acquires a fundamental status.

For the moment we just note that the definition is reasonable in the following sense. Suppose the quantum system is prepared in a mixture of states $\{|\phi_x\rangle; p_x\}$ so that its density matrix is

$$\rho = \sum_x p_x |\phi_x\rangle\langle\phi_x|$$

For the special case where $|\phi_x\rangle$ form an orthonormal basis of $\mathcal{H}$, this is a diagonal operator, so the eigenvalues of $\rho \ln \rho$ are $p_x \ln p_x$, and $S(\rho) = -\sum_x p_x \ln p_x = H(X)$, where X is the random variable with distribution p_x. In an orthogonal mixture all states can be perfectly distinguished so the mixture behaves classically: the quantum and classical entropies coincide.

We emphasize that for a general mixture the states $|\phi_x\rangle$ are not orthonormal so that $S(\rho) \neq H(X)$. In fact we will see that the following holds in full generality

$$S(\rho) \leq H(X)$$

where X is the random variable associated to the "preparation" of the mixture. This bound can be understood intuitively: since the states $|\phi_x\rangle$ cannot be perfectly distinguished (unless they are orthogonal, see chap 2) the quantum uncertainty associated to ρ is less than the classical uncertainty associated to X.

In the case of a pure state $\rho = |\Psi\rangle\langle\Psi|$ we see that the eigenvalues of ρ are 1 (multiplicity one) and 0 (multiplicity $D-1$). Thus

$$S(|\Psi\rangle\langle\Psi|) = 0$$

The entropy of a pure state is zero because there is no uncertainty in this state (in line with the Copenhagen interpretation of QM).

A quantity that plays an important role is also the relative entropy defined by analogy with the KL divergence

$$S(\rho||\sigma) = Tr\rho \ln \rho - Tr\rho \ln \sigma$$

Let us set up some notation concerning the entropy of composite systems and their parts. For a bipartite system $\mathcal{AB}$ with density matrix $\rho_{\mathcal{AB}}$ we write

$$S(\mathcal{AB}) = -Tr\rho_{\mathcal{AB}} \ln \rho_{\mathcal{AB}}$$

and for its parts described by the reduced density matrices $\rho_{\mathcal{A}} = Tr_{\mathcal{B}}\rho_{\mathcal{AB}}$ and $\rho_{\mathcal{B}} = Tr_{\mathcal{A}}\rho_{\mathcal{AB}}$,

$$S(\mathcal{A}) = -Tr\rho_{\mathcal{A}} \ln \rho_{\mathcal{A}}, \qquad S(\mathcal{B}) = -Tr\rho_{\mathcal{B}} \ln \rho_{\mathcal{B}}$$

One could try to pursue further the analogies with the classical case and define conditional entropies as $S(\mathcal{A}|\mathcal{B}) = S(\mathcal{AB}) - S(\mathcal{B})$, $S(\mathcal{B}|\mathcal{A}) = S(\mathcal{AB}) - S(\mathcal{A})$ and mutual information as $I(\mathcal{A};\mathcal{B}) = I(\mathcal{B};\mathcal{A}) = S(\mathcal{A}) + S(\mathcal{B}) - S(\mathcal{AB})$. However it is not clear that these are of any fundamental use since they do not enter (yet) in any theorem of quantum information theory. Perhaps two more serious argument for suspicion are that first, as we will see $S(\mathcal{AB}) - S(\mathcal{B})$ can be negative, and second it is not at all clear how to define the quantum analog of conditional probabilities.

Let us now proceed to the statements and proofs of the basic inequalities satisfied by von Neumann's entropy.

- **Uniform distribution maximizes entropy.** Any ρ can be diagonalized and has positive eigenvalues ρ_x which sum to one. Thus $S(\rho) = -\sum \rho_x \ln \rho_x$, a quantity which is maximized for the distribution $\rho_x = \frac{1}{D}$ (as in the classical case). Thus in the basis where it is diagonal $\rho = \frac{1}{D}I$, and this is also true in any basis. We conclude

  $$0 \leq S(\rho) \leq \ln D$$

  where the upper bound is attained for the "fully mixed" (or most disordered, or uniform) state $\rho = \frac{1}{D}I$. The lower bound is attained for pure states (check !).

- **Concavity.** Let ρ and σ be two density matrices. then

  $$S(t\rho + (1-t)\sigma) \geq tS(\rho) + (1-t)S(\sigma), \qquad 0 \leq t \leq 1$$

  The proof follows the same lines as the classical one which uses convexity of $x \to x \ln x$. We prove below that $\rho \to Tr \rho \ln \rho$ is a convex functional and this immediately implies concavity of von Neumann's entropy.

  Lemma 1 (Klein's inequality). *Let A and B self-adjoint and f convex from $\mathbf{R} \to \mathbf{R}$. We have*

  $$Tr(f(A) - f(B) - (A-B)f'(B)) \geq 0$$

  Proof. Let $A|\phi_i\rangle = a_i|\phi_i\rangle$ and $B|\psi_i\rangle = b_i|\psi_i\rangle$. Then

  $$Tr(f(A)-f(B)-(A-B)f'(B)) = \sum_i \langle\phi_i|f(A)-f(B)-(A-B)f'(B)|\phi_i\rangle \tag{6.2}$$

  Each term in the sum equals

  $$f(a_i) - \langle\phi_i|f(B)|\phi_i\rangle - a_i\langle\phi_i|f'(B)|\phi_i\rangle + \langle\phi_i|Bf'(B)|\phi_i\rangle \tag{6.3}$$

  Using the closure relation

  $$1 = \sum_j |\psi_j\rangle\langle\psi_j|$$

  equation (6.2) can be rewritten as

  $$\sum_j |\langle\phi_i|\psi_j\rangle|^2 \Big(f(a_i - f(b_j) - (a_i - b_j)f'(b_j)\Big)$$

  Now since $f : \mathbf{R} \to \mathbf{R}$ is convex we have

  $$f(a_i) - f(b_j) \geq (a_i - b_j)f'(b_j)$$

  which proves the statement. □

Corollary 2. *Let A and B self-adjoint and positive (positive means that all eigenvalues are positive or equivalently that all diagonal averages $\langle\psi|A|\psi\rangle$ are positive for any $|\psi\rangle$). Then*

$$TrA\ln A - TrA\ln B \geq Tr(A-B)$$

Proof. Take $f(t) = t\ln t$ and apply Klein's inequality. □

Now choose $A = \rho$ and $B = t\rho + (1-t)\sigma$. From the corollary

$$Tr\rho\ln\rho - Tr\rho\ln(t\rho + (1-t)\sigma) \geq (1-t)Tr(\rho-\sigma) = 0$$

Choose $A = \sigma$ and $B = t\rho + (1-t)\sigma$. Then

$$Tr\sigma\ln\sigma - Tr\sigma\ln(t\rho + (1-t)\sigma) \geq tTr(\sigma-\rho) = 0$$

Multiplying the first inequality by t and the second by $(1-t)$ and adding them yields

$$Tr(t\rho + (1-t)\sigma)\ln(t\rho + (1-t)\sigma) \leq tTr\rho\ln\rho + (1-t)Tr\sigma\ln\sigma$$

which proves the concavity of entropy.

- **Positivity of relative entropy.** Choose $A = \rho$ and $B = \sigma$ and apply the corollary,

$$S(\rho||\sigma) = Tr\rho\ln\rho - Tr\rho\ln\sigma \geq Tr(\rho-\sigma) = 0$$

- **Sub-additivity.** In the classical case one has

$$H(X) + H(Y) - H(X,Y) = D(p_{x,y}||p_x p_y) \geq 0$$

In the quantum case the proof is similar, but we detail the steps

$$\begin{aligned}
S(\mathcal{A}) + S(\mathcal{B}) - S(\mathcal{AB}) &= -Tr_{\mathcal{A}}\rho_{\mathcal{A}}\ln\rho_{\mathcal{A}} - Tr_{\mathcal{B}}\rho_{\mathcal{B}}\ln\rho_{\mathcal{B}} + Tr\rho_{cab}\ln\rho_{\mathcal{AB}} \\
&= -Tr\rho_{\mathcal{AB}}\ln\rho_{\mathcal{A}}\otimes I_{\mathcal{B}} - Tr\rho_{\mathcal{AB}}\ln I_{\mathcal{A}}\otimes\rho_{\mathcal{B}} + Tr\rho_{\mathcal{AB}}\ln\rho_{\mathcal{AB}} \\
&= Tr\rho_{\mathcal{AB}}\ln\rho_{\mathcal{AB}} - Tr\rho_{\mathcal{AB}}(\ln\rho_{\mathcal{A}}\otimes I_{\mathcal{B}} + \ln I_{\mathcal{A}}\otimes\rho_{\mathcal{B}}) \\
&= Tr\rho_{\mathcal{AB}}\ln\rho_{\mathcal{AB}} - Tr\rho_{\mathcal{AB}}\ln\rho_{\mathcal{A}}\otimes\rho_{\mathcal{B}} \\
&= S(\rho_{\mathcal{AB}}||\rho_{\mathcal{A}}\otimes\rho_{\mathcal{B}}) \geq 0
\end{aligned}$$

Note that sub-additivity can also formally be written as $S(\mathcal{A}|\mathcal{B}) \leq S(\mathcal{A})$ in terms of the naive conditional entropy. We may say that conditioning reduces quantum entropy, as in the classical case.

Exercise: check the identity $\ln\rho_{\mathcal{A}}\otimes I_{\mathcal{B}} + \ln I_{\mathcal{A}}\otimes\rho_{\mathcal{B}} = \ln\rho_{\mathcal{A}}\otimes\rho_{\mathcal{B}}$ by using spectral decompositions.

- **Araki-Lieb bound.** Classically $H(X,Y) \geq H(X)$ (the whole is more disordered than the parts). But quantum mechanically this can be completely wrong as the following counterexample shows. In quantum mechanics it is not true that the naive conditional entropy is always non-negative. Let
$$\rho_{\mathcal{AB}} = |B_{00}\rangle\langle B_{00}|$$
This is a pure state so $S(\mathcal{AB}) = 0$. However we have for the two parts
$$\rho_{\mathcal{A}} = \frac{1}{2}I_{\mathcal{A}}, \qquad \rho_{\mathcal{B}} = \frac{1}{2}I_{\mathcal{B}}$$
which have maximal entropies $S(\mathcal{A}) = S(\mathcal{B}) = \ln 2$. The two parts of the EPR pair when looked upon locally are as disordered as they can be, however the global state is highly correlated.

  Is there a general good lower bound for $S(\mathcal{AB})$ in terms of the entropies of the parts ? The answer is provided by

  Theorem 3 (Araki-Lieb)**.**
$$S(\mathcal{AB}) \geq |S(\mathcal{A}) - S(\mathcal{B})|$$

  Proof. The proof is a nice application of the purification idea and the Schmidt decomposition theorem. We introduce a third system R such that $\mathcal{ABR}$ is a purification of $\mathcal{AB}$. That is
$$\rho_{\mathcal{ABR}} = |\mathcal{ABR}\rangle\langle\mathcal{ABR}|, \qquad Tr_{\mathcal{R}}\rho_{\mathcal{ABR}} = \rho_{\mathcal{AB}}$$
By sub-additivity
$$S(\mathcal{AR}) \leq S(\mathcal{A}) + S(\mathcal{R}) \tag{6.4}$$
Now since $\rho_{\mathcal{ABR}}$ is a pure state the non-zero eigenvalues of $\rho_{\mathcal{AB}}$ and $\rho_{\mathcal{R}}$ are equal; and also the non zero eigenvalues of $\rho_{\mathcal{AR}}$ and $\rho_{\mathcal{B}}$ are equal (Schmidt theorem). Thus
$$S(\mathcal{AB}) = S(\mathcal{R}), \qquad S(\mathcal{AR}) = S(\mathcal{B})$$
Replacing in (6.4) we get
$$S(\mathcal{B}) - S(\mathcal{A}) \leq S(\mathcal{AB})$$
Since $\mathcal{A}$ and $\mathcal{B}$ play a symmetric role we can exchange them which ends the proof. □

- **Strong sub-additivity.** let $\mathcal{ABC}$ be a quantum system formed of three parts $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B} \otimes \mathcal{H}_\mathcal{C}$. We have similarly to the classical case

  $$S(\mathcal{ABC}) + S(\mathcal{B}) \leq S(\mathcal{AB}) + S(\mathcal{BC})$$

  This can be written also as $S(\mathcal{C}|\mathcal{AB}) \leq S(\mathcal{C}|\mathcal{B})$ in terms of "naive" conditional entropies. So one may say that further conditioning reduces conditional entropy (although the "conditional" entropy is not necessarily positive). As in classical information theory, this inequality plays an important role.

  Classically the proof of this inequality is based on the positivity of the KL divergence. It turns out that quantum mechanically the proof is much more difficult. We will omit it here except for saying that one can base it on the *joint concavity* of the functional

  $$f(A, B) = TrM^\dagger A^s MB^{(1-s)}$$

  for any matrix M (not necessarily self-adjoint) and any $0 \leq s \leq 1$. This fact was a conjecture of Wigner-Yanase-Dyson for many years until Lieb found a proof (1973). Later, Lieb and Ruskai realized that it implies strong sub-additivity.

6.3 Useful bounds on the entropy of a mixtures

This section is devoted to the proof of the following important theorem

Theorem 4. *Let X be a random variable with distribution p_x and $\rho = \sum_x p_x \rho_x$ where ρ_x are mixed states. We have*

$$S(\rho) \leq \sum_x p_x S(\rho_x) + H(X)$$

This inequality has a clear interpretation: the uncertainty about ρ cannot be greater than the average uncertainty about each ρ_x plus the uncertainty about the classical preparation described by X. If in particular $\rho_x = |\phi_x\rangle\langle\phi_x|$ are pure states we have $S(\rho_x) = 0$ so, as announced at the beginning of the chapter,

$$S(\rho) \leq H(X)$$

Proof. First we deal with a mixture of pure states. For convenience we call this mixture $\mathcal{A}$ and set

$$\rho_{\mathcal{A}} = \sum_x p_x |\phi_x\rangle_{\mathcal{A}} \langle \phi_x|_{\mathcal{A}}$$

Let $\mathcal{H}_{\mathcal{R}}$ a space whose dimension is equal to the number of terms in the mixture and with orthonormal basis labeled as $|x\rangle_{\mathcal{R}}$. The pure state

$$|\mathcal{AR}\rangle = \sum_x \sqrt{p_x} |\phi_x\rangle_{\mathcal{A}} \otimes |x\rangle_{\mathcal{R}}$$

is a purification of $\rho_{\mathcal{A}}$ because

$$Tr_{\mathcal{R}} |\mathcal{AR}\rangle\langle\mathcal{AR}| = \sum_x p_x |\phi_x\rangle\langle\phi_x|_{\mathcal{A}} = \rho_{\mathcal{A}}$$

We also have that

$$\rho_{\mathcal{R}} = Tr_{\mathcal{A}} |\mathcal{AR}\rangle\langle\mathcal{AR}| = \sum_{x,x'} \sqrt{p_x}\sqrt{p_{x'}} \langle\phi_x|\phi_{x'}\rangle_{\mathcal{A}} |x\rangle_{\mathcal{R}}\langle x'|_{\mathcal{R}} \tag{6.5}$$

By the Schmidt theorem we know that $\rho_{\mathcal{A}}$ and $\rho_{\mathcal{R}}$ have the same non zero eigenvalues, thus

$$S(\rho_{\mathcal{A}}) = S(\rho_{\mathcal{R}})$$

Consider now

$$\rho'_{\mathcal{R}} = \sum_x p_x |x\rangle_{\mathcal{R}}\langle x|_{\mathcal{R}}$$

and look at the relative entropy

$$S(\rho_{\mathcal{R}}||\rho'_{\mathcal{R}}) = Tr\rho_{\mathcal{R}} \ln \rho_{\mathcal{R}} - Tr\rho_{\mathcal{R}} \ln \rho'_{\mathcal{R}} \geq 0$$

Thus

$$S(\rho_{\mathcal{A}}) = S(\rho_{\mathcal{R}}) \leq -Tr\rho_{\mathcal{R}} \ln \rho'_{\mathcal{R}} \tag{6.6}$$

It remains to compute the right hand side. Since $|x\rangle_{\mathcal{R}}$ is an orthonormal basis

$$\ln \rho'_{\mathcal{R}} = \sum_x (\ln p_x) |x\rangle_{\mathcal{R}}\langle x|_{\mathcal{R}}$$

which implies

$$Tr\rho_{\mathcal{R}} \ln \rho'_{\mathcal{R}} = \sum_x (\ln p_x) Tr\rho_{\mathcal{R}} |x\rangle_{\mathcal{R}}\langle x|_{\mathcal{R}} = \sum_x (\ln p_x) \langle x|\rho_{\mathcal{R}}|x\rangle$$

From the expression of $\rho_{\mathcal{R}}$ (6.5) we remark that

$$\langle x|\rho_{\mathcal{R}}|x\rangle = p_x$$

Thus (6.6) becomes

$$S(\rho_{\mathcal{A}}) \leq -\sum p_x \ln p_x = H(X)$$

Consider now the general case of a mixture of mixed states $\rho = \sum_x p_x \rho_x$. each mixed state has a spectral decomposition

$$\rho_x = \sum_j \lambda_j^{(x)} |e_j^{(x)}\rangle\langle e_j^{(x)}|$$

so

$$\rho = \sum_{x,j} p_x \lambda_j^{(x)} |e_j^{(x)}\rangle\langle e_j^{(x)}|$$

Note that this is a convex combination of one dimensional projectors so that we can apply the previous result

$$\begin{aligned} S(\rho) &\leq -\sum_{x,j} p_x \lambda_j^{(x)} \ln p_x \lambda_j^{(x)} \\ &= -\sum_{x,j} p_x \lambda_j^{(x)} \ln p_x - \sum_{x,j} p_x \lambda_j^{(x)} \ln \lambda_j^{(x)} \\ &= -\sum_x p_x \ln p_x - \sum_x p_x \sum_j \lambda_j^{(x)} \ln \lambda_j^{(x)} \\ &= H(X) + \sum_x p_x S(\rho_x) \end{aligned}$$

In the last equality we used $S(\rho_x) = \sum_j \lambda_j^{(x)} \ln \lambda_j^{(x)}$.

□

6.4 Measuring without learning the measurement outcome cannot decrease entropy

Suppose we are given a mixed state ρ and a measurement apparatus with measurement basis $\{|x\rangle\langle x|\}$. According to the measurement postulate the possible outcomes are pure states

$$|x\rangle, \qquad \text{with probability } p_x = \langle x|\rho|x\rangle$$

[Note that $\sum_x \langle x|\rho|x\rangle = 1$]. If we observe the measurement result we know that we have some $|x\rangle$ with zero entropy.

Now imagine that we do the measurement but do not record the measurement result (subsequently we will call this a "blind" measurement). Then our description of the state of the system is a mixture $\{|x\rangle, p_x\}$ with diagonal density matrix

$$\rho_{\text{blind}} = \sum_x \langle x|\rho|x\rangle |x\rangle\langle x|$$

Note that this diagonal density matrix is equivalent to a classical state. If we look at the relative entropy

$$S(\rho||\rho_{\text{blind}}) \geq 0$$

we find, by a small calculation[1],

$$S(\rho) \leq H(\langle x|\rho|x\rangle) = S(\rho_{\text{blind}})$$

Thus blind measurements can only increase the entropy or leave it constant.

To conclude the chapter consider again a composite system $\mathcal{AB}$ where Alice and Bob are very far apart and do not communicate. A local measurement (with an apparatus $\{|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}}\}$) is done by Alice on part $\mathcal{A}$ which is blind to Bob. Thus according to the previous inequality $S(\rho_{\mathcal{B}}^{\text{blind}}) - S(\rho_{\mathcal{B}}) \geq 0$. However a true (immediate) increase would violate locality and it is very reassuring to check that $S(\rho_{\mathcal{B}}^{\text{blind}}) = S(\rho_{\mathcal{B}})$

After Alice's measurement the possible outcomes for the total system are

$$\frac{(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})\rho_{\mathcal{AB}}(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})}{Tr(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})\rho_{\mathcal{AB}}(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})}$$

or equivalently

$$\rho_{\mathcal{AB}}^{(i)} = \frac{(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})\rho_{\mathcal{AB}}(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})}{\langle i|\rho_{\mathcal{A}}|i\rangle}$$

with probability (we set $|i\rangle_{\mathcal{A}} = |i\rangle$ to alleviate the notation)

$$\langle i|\rho_{\mathcal{A}}|i\rangle$$

Since this is a blind measurement for Bob the reduced density matrix is (a mixture of mixed states)

$$\rho_B^{\text{blind}} = \sum_i \langle i|\rho_{\mathcal{A}}|i\rangle Tr_A \rho_{\mathcal{AB}}^{(i)}$$

[1] identical to the one in the proof of the upper bound in the previous section

A short calculation shows that this equals

$$\rho_B^{\text{blind}} = \sum_i \langle i|\rho_{\mathcal{A}}|i\rangle \frac{\langle i|\rho_{\mathcal{AB}}|i\rangle}{\langle i|\rho_A|i\rangle} = \sum_i \langle i|\rho_{\mathcal{AB}}|i\rangle = Tr_A \rho_{\mathcal{AB}} = \rho_B$$

So after Alice's measurement not only Bob's entropy is unchanged but even his density matrix is left the same as it was before the measurement. This provides a completely general proof that Bob does not notice Alice's measurements. On Alice's side if she does not record her measurement outcome her entropy is greater.