Recall: H(X) measures the ignorance on the rv X.

Let X,Y be two rv’s, with distribution p(xy).
H(XY) = H(p) as before (treat XY as a composite rv).
Fix a particular outcome for Y, say y, with X unknown.
Define q_y = p(X|Y=y) as the distribution of X given Y=y.
q_y(x) = p(xy) / \sum_x p(xy)
H(q_y) is the uncertainty of X when Y=y.
Def: Conditional entropy H(X|Y) = \sum_y p(y) H(q_y).
Meaning: average (over unknown Y) uncertainty of X:
Fix a particular outcome for Y, say y, with X unknown.
H(q_y) is the uncertainty of X when Y=y.
Fact: H(X|Y) = H(XY)-H(Y)
easy to remember
aka "Chain rule."  Proof: def+algebra

Properties of H(X), H(XY), I(X:Y):
1. 0 \leq H(X) \leq \log |\Omega|  [obvious, but useful]
2. H(XY) \leq H(X) + H(Y)  [called Subadditivity]
equivalent to I(X:Y) \geq 0
equivalent to H(X|Y) \leq H(X)
[meaning: conditioning reduces uncertainty
knowing Y cannot hurt]
"*" iff X, Y independent (MI=0, conditioning useless)
Ideas: (i) define relative entropy H(p||q) = \sum_x p(x) \log(p(x)/q(x)),
(ii) show that it is nonnegative [since –(ln 2) (log z) \geq 1-z,
H(p||q) = \sum_x p(x) \log(q(x)/p(x)) \geq \sum_x p(x) (1-q(x)/p(x))/(ln 2)=0,
with equality hold only iff q(x)=p(x) \forall x].  (iii) rewriting I(X:Y) as
H(p(x,y)||p(x)q(y)).

Properties of H(X), H(XY), I(X:Y):
4. H(X|Y) \geq 0  [follows from Def: average over nonnegative entropies]
5. H(XY) = H(Y)+H(X|Y)  [Chain Rule, extends to multiple rv’s]
6. H(X) \geq H(Y)  [follows from 4&5]
7. H(Z) + H(XYZ) \leq H(XZ) + H(YZ)
[called Strong Subadditivity SSA]
Note that Z special, XY symmetric.
As if Z added to each term in SA.  (Thus the name)
equiv to H(Y|XZ) \leq H(Y|Z) or H(X|YZ) \leq H(X|Z)
Conditioning (on a new rv) decreases conditional entropy.
Here: H(Y|Z) on X or H(Z|Y) on Y.

Quantum analogues:
Recall S(\rho) = H(spec(\rho))
Let A,B be two quantum systems
\rho density matrix representing state on AB
S(AB) = S(\rho), S(A) = S(tr_B(\rho)), S(B) = S(tr_A(\rho)).
Classical: H(X|Y) = H(XY)-H(Y)
In quantum setting, no obvious meaning to condition
on one of the two systems.
Def: S(A|B) = S(AB)-S(B)  Imitate classical expression
but not the meaning.
Quantum analogues:

Classical: \( I(X;Y) = H(X) - H(X|Y) \)

Def [quantum mutual information]:
\[
S(A:B) = S(A) - S(A|B) = S(A) + S(B) - S(AB).
\]

Imitate classical expression, due to \( S(A|B) \), meaning of \( S(A:B) \) not immediately clear. (Investigate later.)

Example 1:
Suppose we have a pure state \( |\psi\rangle \) on \( AB \).
There is always a Schmidt decomposition:
\[
|\psi\rangle = \sum_x \sqrt{p(x)} |e_x\rangle_A |f_x\rangle_B
\]
where \( \{e_x\} \) is orthonormal in \( C^A \), \( \{f_x\} \) o.n. in \( C^B \).

Note that
\[
\rho_A = \sum_x p(x) |e_x\rangle_A \langle e_x|,
\rho_B = \sum_x p(x) |f_x\rangle_B \langle f_x|.
\]

\( S(AB) = 0 \).
\( S(A) = S(B) = H(p) \).
\( S(A:B) = 2 H(p), \ S(A|B) = -H(p) \)

Properties of \( S(A), S(A|B), S(A:B) \):

1. \( S(\rho) = S(U\rho U^\dagger) \), \( S(A:B) = S(\rho_{AB}) \)
2. \( S(\rho_{AB}) \geq S(A) + S(B) \) [subadditivity]
3. Let \( \tau_1, \tau_2, ... \) be states on the same system and \( \{p_k\} \) a distribution. Then,
\[
S(\sum_k p_k \tau_k) \geq \sum_k p_k S(\tau_k)
\]
[entropy of the average \( \geq \) average entropy]

Why: consider \( \rho = \sum_k p_k \tau_k \otimes |k\rangle \langle k| \).
\( S(\rho) = H(p) + \sum_k p_k S(\tau_k) \).
Follows from 2.

Example 2: Consider a density matrix \( \rho = \sum_x \lambda_x |e_x\rangle \langle e_x| \)
Suppose we measure in some basis (WLOG the computation basis) and the outcome is \( y \).
Note that
\[
p(y) = \sum_x \lambda_x |V_{xy}|^2 \quad \text{and} \quad p(y|x) = |V_{xy}|^2.
\]

Let \( |e_x\rangle = \sum_y V_{xy} |y\rangle \).
Make a matrix \( V \) where \( V_{xy} \) is the entry for the \( x \)-col & \( y \)-row so \( V \) transforms the comp basis to the eigenbasis of \( \rho \).
The distributed given by \( p(y) \) (as a vector labeled by \( y \)) is obtained from the distribution \( \lambda_x \) (as a vector labeled by \( x \)) by applying the matrix \( D \) (where \( D_{xy} = |V_{xy}|^2 \)).

In general, we say that \( D \) is a stochastic map taking \( X \) to \( Y \) if \( D \) has nonnegative entries with columns sum to 1.
Here, the rows of \( D \) also sum to 1, and we call it doubly stochastic. It is known to be entropy nondecreasing.

Therefore \( S(\rho) = H(X) \leq H(Y) \) (meas outcomes are more random than the prep).

Properties in the quantum setting:

4. \( S(A|B) \geq 0 \) or \( S(A|B) = 0 \)
5. \( S(AB) = S(B) + S(A|B) \) [by def]
6. \( S(AB) \geq S(B) \) or \( S(AB) \leq S(B) \)
7. Strong subadditivity (for any tripartite state on \( ABC \))
\[
S(C) + S(ABC) \leq S(AC) + S(BC)
\]
equiv to \( S(A|BC) \geq S(A|B) \)

(i) equiv to \( S(A|BC) \leq S(A|B) \)
\[
\text{(so, } S(A|C:B) = S(A|B) - S(A|BC)) \geq 0 \)
\]
(ii) equiv to \( S(A|BC) \leq S(A|BC) \)

Proof of equivalences:

(i) \( S(ABC) = S(AB) - S(BC) \)
\[
\text{(ii) } S(A|BC) = S(A|BC) - S(ABC) \geq 0 \]

Recall any TCP map \( E \) can be realized by an isometry \( B \to B'E \) where \( E \) is a suitable environment initially in a pure state, followed by discarding the environment.
The von Neumann entropy is invariant under a unitary change of basis. Thus \( S(A|B) = S(A|B') \).

Conversely, discarding is a TCP map.
How much info can we learn about a quantum state by measuring it?

Given an ensemble \( \mathcal{E} = \{ \rho_1, \rho_2 \} \), consider a game:

\[
\begin{array}{ccc}
X & \rightarrow & A \\
\rho_x & \rightarrow & B \\
& \rightarrow & Y
\end{array}
\]

Alice draws \( x \) with probability \( p(x) \), prepares \( \rho_x \), and sends to Bob. Bob performs a measurement \( \{ M_i \} \) with operators \( \{ M_i \} = \{ \sum_i \rho_i \} \).

Probability to obtain outcome \( y \) if state is \( \rho_x \):

\[
p(y|x) = \text{tr}(M_y \rho_x)
\]

Joint distribution \( p(y|x) = p(y|x) p(x) \)

Classical mutual info \( I(X:Y) \) quantifies the information on which state \( X \) given by the outcome \( Y \)

Def: \( I_{\text{acc}}(\mathcal{E}) = \max_{\rho_y} I(X:Y) \) [accessible info of \( \mathcal{E} \)]

I\(_{\text{acc}}\) is a natural quantity to define but difficult to compute.

Examples (proof of optimality of meas left as Ex/HW):

e.g.1 \( \rho_1 = |\psi_1\rangle \langle \psi_1| \) for \( |\psi_1\rangle = \cos \theta |0\rangle + \sin \theta |1\rangle \)

\( \rho_2 = |\psi_2\rangle \langle \psi_2| \) for \( |\psi_2\rangle = \cos \theta |0\rangle - \sin \theta |1\rangle \)

drawn with \( \rho_1 = \rho_2 = \frac{1}{2} \)

\( x \) \( y \) \( \text{opt meas} \) \( |\psi_1\rangle \langle \psi_1| \)

\( |\psi_2\rangle \) \( |\psi_3\rangle \)

\( M_1 = |+\rangle\langle +| \)

\( M_2 = |-\rangle\langle -| \)

\( M_3 = |\psi_3\rangle \langle \psi_3| \)

\( M_{\text{opt}} = 2(1-|\psi_1\rangle \langle \psi_1|)/3 \)

\( M_{\text{opt}} \) takes the outcome that maximizes the accessible info

\( M_{\text{opt}} = \max_{M_i} I(X:Y) = (\log 3) - 1 \)

\( I_{\text{acc}} = I(X:Y) = (\log 3) - 1 \approx 0.5850 \)

With \( n=2 \), we take the basis as the BB84 states

Each state drawn with uniform probability \( 1/2n \).

(\( x \) is encoded in the computational or conjugate basis wp \( 1/2 \) each)

Optimal measurement turns out to be \( M_y = \frac{1}{2} \rho_y \)
i.e. randomly measure in one of the two possible bases

Let \( TY \) denote Bob's entire data set, where \( T \) is the coin toss specifying his measurement basis, and \( Y \) is the outcome of that measurement.

With prob \( \frac{1}{2} \), Bob's random basis equals the actual one, giving \( Y=X \), so \( I(X:Y) \) (correct) \( = \log n \). With prob \( \frac{1}{2} \), he measures in the "conjugate basis" so \( Y \) is random and independent of his quantum state (elaborate). So, \( I(X:Y) \) (wrong) \( = 0 \). So, \( I(X:Y) = \frac{1}{2} \log n \).

A lower bound to accessible information

For a density matrix \( \rho \) in \( d \) dimensions with eigenvalues \( \{ \lambda_k \} \), define the "subentropy":

\[
Q(\rho) = -\sum \lambda_k \log \lambda_k /
\]

\( \log \lambda_k \)

For the ensemble \( \mathcal{E} = \{ \rho_1, \rho_2 \} \),

\( I_{\text{acc}}(\mathcal{E}) \geq \sum \lambda_k \log \lambda_k \)

achieved by measuring in the random basis

If \( \rho_y \) are pure and \( 1/d = \sum \rho_y \rho_y \) (an ensemble of pure states that averages to the maximally mixed state),

\( I_{\text{acc}} \geq \log(d) - (\log e)/2 + 1/3 + ... + 1/d \) (in bits)

For \( d = 2 \), \( I_{\text{acc}} \geq 0.2787 \), for \( d \rightarrow \infty \), \( I_{\text{acc}} \geq 0.60995 \). 

When \( n=2 \), these are the BB84 states
An upper bound to accessible information

For the ensemble \( E = \{p_x, \rho_x\} \), define
Holevo information
\[ \chi(E) = S(\rho) - \sum p_x S(\rho_x) \]

Theorem: \( I_{\text{acc}}(E) \cdot \chi(E) \)

Proof:
The ensemble can be represented by the "CQ" state
\[ \tau_{XQ} = \sum p_x |x\rangle \otimes \rho_x \]
We interpret Alice as using the info \( x \) in system \( X \) to prepare the state \( \rho_x \) in system \( Q \) which is then transmitted to Bob.
Bob makes a measurement with POVM \( \{M_y\} \) and outcome \( y \) stored in \( Y \), and discards the system \( Q \). The joint system is
\[ \tau'_{XY} = \sum p_x |x\rangle \langle x| \otimes \sum_y \text{tr}(M_y \rho_x) |y\rangle \langle y| \]

Proof (ctd):
\[ \rho_1 = |\psi_1\rangle \langle \psi_1| \text{ for } |\psi_1\rangle = \cos \theta |0\rangle + \sin \theta |1\rangle \]
drawn with \( p_1 = p_2 = 1/2 \)
\[ \rho = \cos^2 \theta |0\rangle \langle 0| + \sin^2 \theta |1\rangle \langle 1| \]
\[ \chi(E) = S(\rho) \]

\[ e.g.1 \quad p_1 = |\psi_1\rangle \langle \psi_1| \text{ for } |\psi_1\rangle = |0\rangle \]
\[ p_2 = |\psi_2\rangle \langle \psi_2| \text{ for } |\psi_2\rangle = |0\rangle - i|1\rangle \]
drawn with \( p_1 = p_2 = 1/2 \)
\[ \rho = \cos \pi |0\rangle \langle 0| + \sin \pi |1\rangle \langle 1| \]

Note that in general, there are many many 1-qubit states, and to specify one such state takes many bits.

Preparing the quantum state (and not knowing the classical label) less than \( S(I/2) = 1 \) bit of info can be extracted.

It is highly irreversible.

Holevo’s bound also says that we cannot use 1 qbit cannot transmit more one 1 bit of data.