We shall begin by defining the information of an operator \( A \) for a state \( \Psi \).

We shall be interested in discussing information and correlation for operators \( A, B, \ldots \) with respect to a state function \( \Psi \). These quantities are to be computed from the squares of the projections of \( \Psi \) upon the eigenstates of the operators, regarded as probability distributions over the eigenvalues, through the formulas of the preceding chapter. We have already seen that a state \( \Psi \) on an orthonormal basis \( \{ \phi_i \} \) leads to a square amplitude distribution of \( \Psi \) over the set \( \{ \phi_i \} \):

\[
P_i = |(\phi_i, \Psi)|^2 = \langle [\phi_i] \rangle \Psi
\]

so that we can define the information of the basis \( \{ \phi_i \} \) for the state \( \Psi \) as \( I_{\{ \phi_i \}} (\Psi) \) to be simply the information of this distribution relative to the uniform measure:

\[
I_{\{ \phi_i \}} (\Psi) = \sum \ln p_i = \sum |(\phi, \Psi)|^2 \ln |(\phi, \Psi)|^2
\]

We shall define the information of an operator \( A \) for state \( \Psi \) as \( I_A (\Psi) \), to be the information in the probability square amplitude distribution over its eigenvalues, i.e., the information of the probability distribution over the results of a determination of \( A \), which is prescribed in the probabilistic interpretation. For a non-degenerate operator \( A \), the distribution is the same as the distribution (2.1).
over the eigenstates, but because the information is
dependent only on the distribution involved on numerical
values, it is the same as the information of the
distribution over eigenvalues of $A$ is precisely the
information of the eigenvectors of $A$. 

\[ I_A (\Psi) = I_{\Psi, 3} (\Psi) = \sum_i \frac{\langle \Phi_i | \Psi \rangle^2}{\langle \Phi_i | \Phi_i \rangle} \ln \frac{\langle \Phi_i | \Psi \rangle^2}{\langle \Phi_i | \Phi_i \rangle} \] (known)

We see that for fixed $\Psi$, the information of all non-degenerate
operators having the same set of eigenstates is the same.

In the case of degenerate operators, it will
be convenient to take, as definition of information, the
information of the square amplitude distribution over
the eigenvalues relative to the information measure
which consists of the multiplicity of the eigenvalues,
rather than the uniform measure. This definition
preserves the choice of uniform measure over the
eigenstates, in distinction to the eigenvalues.

If $\Phi_i$ (i from 1 to $m_i$) are a complete orthonormal
set of eigenvectors for $A$, with distinct eigenvalues
$A_i$ (degenerate with respect to i), then the multiplicity
of the $i^{th}$ eigenvalue is $m_i$ and the information $I_A (\Psi)$
is defined to be:

\[ I_A (\Psi) = \sum_i \left( \sum_j \frac{\langle \Phi_{ij} | \Psi \rangle^2}{\langle \Phi_{ij} | \Phi_{ij} \rangle} \right) \ln \frac{\langle \Phi_{ij} | \Psi \rangle^2}{\langle \Phi_{ij} | \Phi_{ij} \rangle} \]

The usefulness of this definition lies in the fact
that any operator $A''$ which distinguishes further
between any of the degenerate states leads to a
refinement of the relative density, in the sense of Thm9,
and consequently has equal or greater information.
A non-degenerate operator thus represents the
maximal refinement and possesses maximal information.

Since we shall be primarily concerned with
expansions and consequent distributions over complete
orthonormal sets \( \{ \phi_i \} \), rather than distributions over
eigenvalues, we shall sometimes refer to "the corresponding
operator" for the set \( \{ \phi_i \} \), by which we mean
any non-degenerate Hermitian operator whose
eigenstates are the set \( \{ \phi_i \} \).

It is convenient to introduce a new notation
for projection operators which are relevant for a
specified operator. Let \( \lambda \) be an operator \( \lambda \)
eigenfunctions \( \phi_i \) and eigenvalues \( \lambda_i \) (degenerate
with respect to \( i \)). Then define projections \( A_i \), the
projections on the eigenspaces of different eigenvalues, to be:

\[
A_i = \sum_{j=1}^{m_i} [\phi_i \phi_j^*]
\]

(2.5)

To each such projection there is associated a number \( M_i \),
the multiplicity of the degeneracy, which is the
dimension of the \( i \)th eigenspace. In this notation
the distribution over eigenvalues \( \lambda_i \) becomes simply:

\[
P(\lambda_i) = \langle A_i \rangle \psi
\]

(2.6)

and the information, given by (2.4), becomes:

\[
I_A = \sum_{i} \frac{\langle A_i \rangle \psi}{\psi} \ln \frac{\langle A_i \rangle \psi}{\psi}
\]

(2.7)
Similarly, for a pair of operators $A$ in $S_1, B$ in $S_2$ in a composite system with state $\Psi$, the joint distribution over eigenvalues is:

\[
\rho_{ij} = \rho(A_i, B_j) = \langle A_i B_j \rangle \Psi^5
\]

and the marginals:

\[
\rho_i = \sum_j \rho_{ij} = \langle A_i I \rangle \Psi^5 = \langle A_i I \rangle \Psi^5
\]

\[
\rho_j = \sum_i \rho_{ij} = \langle I I_j \rangle \Psi^5
\]

The joint information is given by:

\[
I_{AB} = \sum_{ij} \rho_{ij} \ln \frac{\rho_{ij}}{\rho_i \rho_j} = \sum_i \langle A_i I \rangle \Psi^5 \ln \frac{\langle A_i I \rangle \Psi^5}{\Psi^5}
\]

where $\rho_{ij}$ is the multiplicity of the eigenvalues $A_i$ and $B_j$.

The marginals are given by:

\[
I_A = \sum_i \langle A_i I \rangle \Psi^5 \ln \frac{\langle A_i I \rangle \Psi^5}{\Psi^5}
\]

\[
I_B = \sum_j \langle I I_j \rangle \Psi^5 \ln \frac{\langle I I_j \rangle \Psi^5}{\Psi^5}
\]

We then define the correlation $\{A, B\}$:

\[
\{A, B\} = \sum_{ij} \rho_{ij} \ln \frac{\rho_{ij}}{\rho_i \rho_j} = \sum_i \langle A_i B_i \rangle \Psi^5 \ln \frac{\langle A_i B_i \rangle \Psi^5}{\langle A_i I \rangle \Psi^5 \langle I B_i \rangle \Psi^5}
\]

and note that the expression does not involve the multiplicities, as do the information terms -- which simply reflects the idea of correlation on information.
We should like at this time to show the existence of a fundamental quantum correlation, \( \{ s_x, s_y \} \), between the two subsystems of \( S^5 \), and describe some of its properties. As we remarked earlier, a density matrix is Hermitian, so that there is a representation in which it is diagonal. In particular, for the decomposition of \( S^5 \) into \( S_x \) and \( S_y \), we can choose a representation in which both \( S_x^2 \) and \( S_y^2 \) are diagonal. This choice is always possible because \( S_x^2 \) is independent of the basis in \( S_y \) and \( S_y^2 \) is independent of the basis in \( S_x \), and hence, if both are diagonal, the basis will be both \( S_x^2 \) and \( S_y^2 \) diagonal. If we choose the basis \( \{ N_x, N_y \} \) for \( S_x \) so that \( S_y^2 \) is diagonal, \( (s_{yx}^2 = A; s_{fy} = 0) \), and let the \( S_y^2 \) be the relative states \( \beta \) and \( \gamma \) for \( N_x \) and \( N_y \), respectively. Then \( S^5 \) is represented in the form (2.13) where \( N_x, N_y \) are orthonormal by choice, and the \( N_x \) and \( N_y \) normal states they are all states. We therefore need only show that the states \( N_x, N_y \) are orthonormal.
\[
\left( \frac{\delta E}{\delta \phi_i} \right) = \left( \frac{\eta e}{2} \right) \phi_i \sum_k \phi^*_k \phi_k \rho_k \sum_m \phi_m \psi^*_m \psi_m
\]

\[
= \sum_{i,j,k} (\phi_i \eta_j \psi^*) (\phi_i \eta_j \psi^*) \delta_{m,n} = \sum_{i,j} (\phi_i \eta_i \psi^*) (\phi_i \eta_i \psi^*)
\]

\[
= \sum_{i,j} \delta_{i,j} = \lambda_k \delta_{i,j} N_i N_k = 0 \quad \text{for} \quad i \neq k
\]
Since we supposed \( P^{S_2} \) to be diagonal in this representation, we have thus constructed a canonical representation (2.13).

The density matrix \( P^{S_2} \) is also automatically diagonal by the choice of representation making \( P^{S_2} \) diagonal, and relative states, \( S_i \) being orthonormal we have:

\[
(2.19) \quad P^{S_2}_{ij} = \sum_k (S_k \cdot S_j) (S_k \cdot S_i) = \sum_k \left( S_k \cdot \sum_m \sum_n a_{km}^* a_{mn} \right) \left( S_i \cdot \sum_m \sum_n a_{mn}^* a_{km} \right)
\]

\[
= \sum_{km} a_m^* a_n S_i S_k S_j S_l = \sum_{kj} a_i^* a_j \delta_{kj} \delta_{il} S_i S_j
\]

where \( P_i = a_i^* a_i \) is the marginal distribution over \( S_i \). Similar computation shows that the elements of \( P^{S_2} \) are the same:

\[
(2.18) \quad P^{S_2}_{kl} = a_k^* a_l \delta_{kl} = P_k \delta_{kl}
\]

Thus in the canonical representation, both density matrices are diagonal and have the same elements \( P_k \), which give the marginal square amplitude distributions over the sets \( S_i \) and \( S_j \), forming the basis of the representation.

Now, any pair of operators \( \hat{A} \) in \( S_2 \) and \( \hat{B} \) in \( S_2 \), which have as non-degenerate eigenfunctions the sets \( S_i \) and \( S_j \) (i.e., operators which define the canonical representation), are "perfectly" correlated in the sense that there is a one-one correspondence between their eigenvalues. The joint amplitude distribution for eigenvalues \( \lambda \) of \( \hat{A} \) and \( \mu \) of \( \hat{B} \) is:

\[
(2.19) \quad P(\lambda_i, \mu_j) = P(\hat{A}, \hat{B}) = \delta_{ij} = a_i^* a_j \delta_{ij} = P_i \delta_{ij}
\]
Therefore, the correlation between these operators, \( \{ \vec{A}, \vec{B} \} \), is:

\[
\{ \vec{A}, \vec{B} \} \psi^5 = \sum_{i,j} P(A_i, B_j) \ln \left( \frac{P(A_i, B_j)}{P(A_i) P(B_j)} \right) = \sum_{i,j} P_i P_j \delta_{ij} \ln \left( \frac{P_i P_j}{P_i P_j} \right) = -\sum_i P_i \ln P_i = \text{constant}
\]

We shall denote this quantity by \( \{ S_1, S_2 \} \psi^5 \) and call it the canonical correlation of the subsystems \( S_1 \) and \( S_2 \) for the system state \( \psi^5 \). It is the correlation between any pair of non-degenerate subsystem operators which define the canonical representation.

We note that in the representation, where the density matrices are diagonal \((6.13)\) and \((6.18)\), the canonical correlation is given by:

\[
\{ S_1, S_2 \} \psi^5 = -\sum_i P_i \ln P_i = -\text{Trace} \left( \rho \ln \rho \right) / \text{Trace} \left( \rho^2 \right)
\]

But the trace is invariant, so that \((2.21)\) holds generally independently of the representation, and we have therefore established the uniqueness of \( \{ S_1, S_2 \} \psi^5 \).

It is also interesting to note that the quantity \\
- \( -\text{Trace} (\rho \ln \rho) \) is just the entropy of a mixture of states characterized by the density matrix. \( \rho \). Therefore the entropy of the mixture characteristic of a subsystem \( S_1 \) for state \( \psi^5 \) \( \psi^5_{S_1} \) is exactly matched by a correlation information \( \{ S_1, S_2 \} \), which represents the correlation between any pair of operators \( \vec{A}, \vec{B} \) which define the canonical representation. The situation is thus quite similar to that of classical mechanics.
Another special property of the canonical rep. is that any operators \( A, B \) defining a canonical \( \varepsilon \) representation with discrete spectrum have minimum marginal information (i.e., \( I_A \geq I_B \) all \( \varepsilon \)). The canonical rep. in (2.13) with diagonal elements \( P = \sum_k \phi_k \phi_k \) is not appropriate for this rep., and \( A, B \) any \( \varepsilon \)-representation operators with eigenfunctions \( \{ \phi_k \}, \{ \phi_l \} \)

where \( \psi = \sum_k \phi_k \phi_k \), \( \eta = \sum_l \phi_l \phi_l \), then \( P \) in \( \psi \), \( \phi \) rep.

\[
(2.22) \quad \psi = \sum_{k \in \mathbb{R}} \alpha_k \phi_k \quad \phi = \sum_{l \in \mathbb{R}} \alpha_l \phi_l \quad \phi = \sum_{k \in \mathbb{R}} (\sum_{a \in \mathbb{R}} c_{a,k} \phi_a \phi_k) \quad \phi = \sum_{k \in \mathbb{R}} (\sum_{a \in \mathbb{R}} c_{a,k} \phi_a \phi_k)
\]

and the joint \( \varepsilon \)-amplitude distribution for \( \phi_k \), \( \phi_n \) is:

\[
(2.27) \quad P_{ke} = \left( \sum_{i,j} \alpha_i \alpha_j \phi_i \phi_j \right)^2 \quad P_{ke} = \sum_{i,j} \alpha_i \alpha_j \phi_i \phi_j \phi_i \phi_j
\]

while the marginals are:

\[
(2.27) \quad P_k = \sum_k P_{ke} = \sum_{i,j} \alpha_i \alpha_j \phi_i \phi_j \phi_i \phi_j
\]

\[
(2.75) \quad P_i = \sum_k P_{ke} = \sum_{i,j} \alpha_i \alpha_j \phi_i \phi_j \phi_i \phi_j
\]

and similarly.

2.25 Then the marginal information \( I_A \) is:

\[
(2.76) \quad I_A = \sum_k P_k \ln P_k = \sum_k \left( \sum_{i,j} \alpha_i \alpha_j \phi_i \phi_j \phi_i \phi_j \right) \ln \left( \sum_{i,j} \alpha_i \alpha_j \phi_i \phi_j \phi_i \phi_j \right)
\]

where \( T_{ik} = C_{ik}^* C_{ik} \), is doubly-stochastic (\( \sum_i T_{ik} = \sum_k T_{ik} = 1 \)).

Therefore, by Theorem 2.2 (general):
\[ I_A = \sum_k \left( \sum_{x_k} \alpha_k^* T_{ik} \right) \ln \left( \sum_{x_k} \alpha_k^* T_{ik} \right) \]
\[ \leq \sum_{x_k} \alpha_k^* \ln \alpha_k^* \alpha_k = I_A \]

and we have proved that \( A \) has maximal marginal information. Identical proof holds also for \( B \).

While this result was proved only for non-degenerate operators \( A \), it is immediately extended to the degenerate case, since as a consequence of our definition for the degenerate operators, \( (2.7) \) its information is still less than that of an operator which removes the degeneracy. We have thus proved:

\[ I_A^{(3)} \leq I_A^{(4-s)} \]

where \( A \) is any non-degenerate operator defining canonical representation and \( A \) is any operator (finite spectrum).

We should like to conclude the discussion of the canonical representation by conjecturing that in addition to the maximum marginal information properties of \( A, B \), which define the representation, they are also maximally correlated, by which we mean that for any pair of operators \( C \) in \( S_1 \), \( D \) in \( S_2 \):

\[ (2.78) \]
\[ \text{Conjecture: } \{ C, B \}^{(s)} \leq \{ A, B \} = \{ S_1, S_2 \} \]

For all operators \( C \in S_1, D \in S_2 \).

Footnote: The relations \( \{ C, B \} \leq \{ A, B \} = \{ S_1, S_2 \} \) and \( \{ C, D \} \leq \{ S_1, S_2 \} \) for all \( C, D \) can be proved similarly in a manner similar to \( (2.77) \).
As a final topic for this section we point out that

\[ \nabla_x^2 \nabla_k^2 \geq \frac{1}{4} \quad \text{for all } \psi(x) \]

where

\[ \nabla_x^2 = \frac{\delta^2}{\delta x^2} \quad \text{and} \quad \nabla_k^2 = \frac{\delta^2}{\delta k^2} \]

\[ \left( \langle k^2 \rangle - \langle k \rangle^2 \right) \quad \text{and} \quad \left( \langle p^2 \rangle - \langle p \rangle^2 \right) \]

\[ \Pi_k \] is the variance of \( k \) and \( \Phi \) is the variance of \( \Phi \). The conjectured information form of the uncertainty principle is

\[ I_x + I_k \leq \ln \frac{1}{\hbar} \quad \text{for all } \psi(x) \]
Although this inequality has not yet been proved with complete rigor, it is made probable by the circumstances that equality holds for $V(x)$ of the form $V(x) = \left(\frac{1}{2} x^2 \right)^{1/2}$, the so-called "minimum uncertainty factors" which give normal distributions, and that furthermore the first variation of $(I_x^2 + I_y^2)$ vanishes for such $V(x)$. (See App.)

There, although $\ln \Omega$ has not been proved an absolute maximum for $I_x^2 + I_y^2$, it is at least a stationary value.

The principle (2.14) is stronger than (7.13), since it implies (2.13) but is not implied by it. To see that it implies (2.13) we use the well-known fact (easily established by a variation calculation)
These quantities are to be computed through the formulas of the preceding chapter, from the square amplitudes of the coefficients of the expansion of \( \mathcal{T} \) in terms of the eigenstates of the operators.

For a family of states ordered by increasing information, as the information tends to maximum (zero), the states go to eigenstates.

We can phrase this more precisely by considering a family of state functions ordered according to increasing information for an operator \( \mathcal{A} \).

\[
\mathcal{A} (\mathcal{A} \leq \mathcal{B}) = \mathcal{A}^2 + \mathcal{A} \mathcal{B} = \mathcal{B}^2 + \mathcal{A} \mathcal{B}
\]

\[
\sum_{\mathcal{B}} \mathcal{A} = \sum_{\mathcal{B}} (\mathcal{A} \leq \mathcal{B}) \mathcal{B}
\]

Measurement always increases certain quantities (significance).

Qualitatively unchanged. (Non-discover)
In addition to the correlation between subsystem operators, given by (2.12), there always exists a unique quantity \( \{ S^l_1, S^l_2 \} \), the canonical correlation, which has some special properties and may be regarded as the fundamental correlation between the two subsystems \( S^1 \) and \( S^2 \) of the composite system \( S \).

(2.14) \[
P_{S^2} = \Lambda^S \cdot S_{ij}
\]

and let the \( S^j \) be the relative states in \( S^j \) for the \( \Lambda^j \) in \( S^j \).

(2.15) \[
\{ i \} = N^S \sum_j \left( \frac{\partial \Lambda^j}{\partial \nu} \right) \phi_j \quad \text{(some basis } \{ \phi_j \} \text{)}
\]

Then...

Thus...

In the sense that for all other discrete spectrum operators \( \Lambda^S \) on \( S^j \) and \( \Lambda^S \) on \( S^j \), \( I_A \equiv I_{A^S} \) and \( I_B \equiv I_{B^S} \),

where
\[
\begin{align*}
\nabla_x^2 &= \langle x^2 \rangle - \left[ \langle x \rangle \right]^2 \\
\nabla_k^2 &= \left( -\frac{\partial^2}{\partial x^2} \right) \langle k \rangle^2 - \left[ \langle \hat{k} \rangle \right]^2 = \langle \hat{k}^2 \rangle - \left[ \langle \hat{k} \rangle \right]^2
\end{align*}
\]