Notes on Quantum Entanglement

Jonathan Bain
Dept of Technology, Culture and Society; NYU-Tandon

1. Introduction

Quantum entanglement is characterized by a non-classical correlation between subsystems of a larger composite system. More precisely, when a composite system is in a quantum entangled state, there are properties that its subsystems possess that are correlated with each other in a non-classical way. To understand this, we first need to understand how quantum mechanics represents states and properties. A state of a physical system is a description of the system in terms of its properties at an instant in time. There are two ways to represent states in quantum mechanics--using vectors and using density operators--and each comes with a distinct definition of a quantum entangled state.

2. States as Vectors

In quantum mechanics, one way to represent a state with respect to a property that has n possible values is by an n-dimensional unit vector $|\psi\rangle$, which is a vector with unit length that is an element of an n-dim vector space $\mathcal{H}$; *i.e.*, a vector space that is spanned by n linearly independent basis vectors $\{|v_1\rangle, ..., |v_n\rangle\}$.[1] This means $|\psi\rangle$ can be expanded as a sum of basis vectors

$$|\psi\rangle = \sum_{i=1}^{n} c_i|v_i\rangle = c_1|v_1\rangle + \cdots + c_n|v_n\rangle \quad (1)$$

where the coefficients c_i are complex numbers. This sum is called a "superposition". In Dirac's notation, $|\psi\rangle$ is called a "ket" vector. It's dual is the "bra" vector $\langle\psi| = \sum_i c_i^*\langle v_i|$.

A vector space $\mathcal{H}$ comes with an inner-product on vectors: For $|\psi\rangle, |\phi\rangle \in \mathcal{H}$, the inner-product $\langle\psi|\phi\rangle$ is a number. The length of $|\psi\rangle$ is defined by $\sqrt{\langle\psi|\psi\rangle}$. Two vectors are

[1] The vector spaces used in quantum mechanics are "Hilbert spaces". A Hilbert space is a vector space $\mathcal{H}$ that has the property that every Cauchy sequence of vectors in $\mathcal{H}$ converges to a vector in $\mathcal{H}$. Ultimately, this allows you to take limits and perform differentiation and integration on vectors in $\mathcal{H}$.

orthogonal *if and only if* their inner-product vanishes. Basis vectors are required to be *orthonormal*, which means they have unit length, $\sqrt{\langle v_i|v_i\rangle} = 1$, for all i, and are mutually orthogonal, $\langle v_j|v_i\rangle = 0$, for $i \neq j$.

> <u>Note</u>: A necessary and sufficient condition for (1) to be a unit vector is $|c_1|^2+\cdots+|c_n|^2=1$.
> <u>Ex. 1</u>. A "qubit" is a state of a 2-state quantum system. A 2-state quantum system is a system with states that can be represented by 2-dim unit vectors. The corresponding vector space has two basis vectors typically called $|0\rangle$ and $|1\rangle$. Thus an arbitrary qubit $|Q\rangle$ can be expressed as $|Q\rangle = \alpha|0\rangle + \beta|1\rangle$, where α and β are complex numbers that satisfy $|\alpha|^2 + |\beta|^2 = 1$.

A *property* (or "observable") of a physical system is a quantifiable feature of the system (like momentum, position, spin, *etc*.). A property is quantifiable in the sense that it possesses *values*. One way to represent a property in quantum mechanics is by a linear operator O; *i.e.*, a map O that acts on vectors in a vector space $\mathcal{H}$ such that for any $|\psi\rangle \in \mathcal{H}$, $O|\psi\rangle$ is also a vector in $\mathcal{H}$. Linearity means that the action of O can be brought through sums of vectors, and multiplication of vectors by scalars α, β, so that $O(\alpha|\psi\rangle+ \beta|\phi\rangle) = \alpha O|\psi\rangle + \beta O|\phi\rangle$.

> <u>Note</u>: An operator can be represented by the "outer-product" $|\psi\rangle\langle\phi|$ of a ket vector $|\psi\rangle$ and a bra vector $\langle\phi|$, since when you act with this on any vector $|\varphi\rangle$ you get another vector $|\psi\rangle\langle\phi|\varphi\rangle$ (this latter expression is a vector since "$|\psi\rangle$" is a vector and "$\langle\phi|\varphi\rangle$" is a number).

An *eigenvector* of an operator O is a vector $|\lambda\rangle$ that does not change its orientation when O acts on it, which means $O|\lambda\rangle = \lambda|\lambda\rangle$, for some scalar λ. This scalar is called an *eigenvalue* of O. If O represents a property of a physical system, then an eigenvalue of O represents a possible value of that property, and the corresponding eigenvector of O represents the state of the system in which it possesses that value of that property. This is codified in:

Eigenvector/Eigenvalue Rule: A system possesses the value λ of the property represented by an operator O *if and only if* the system is in a state represented by the eigenvector $\lvert\lambda\rangle$ of O.

> <u>Note</u>: A *Hermitian* (or *self-adjoint*) operator on $\mathcal{H}$ is a linear operator O such that $\langle b|(O|a\rangle) = (\langle b|O)|a\rangle$, for any vectors $|a\rangle, |b\rangle \in \mathcal{H}$. If O is represented by a matrix, then this means that $O = O^\dagger$, where "$O^\dagger$" is the complex-transpose (the "Hermitian conjugate", or "adjoint") of O. One can show that the eigenvalues of a Hermitian operator are real numbers, and its eigenvectors form an orthonormal basis of $\mathcal{H}$ (which means that any vector $|\psi\rangle$ in $\mathcal{H}$ can be expanded in the eigenvector basis of a Hermitian operator). These features suggest that if we're going to use operators to represent properties, we should restrict our use to Hermitian operators.

There are experiments that suggest some properties of a quantum system cannot simultaneously possess values (examples include position and momentum, and various spin properties). This means that when a quantum system is in a state represented by an eigenvector of one of these properties (like position), it then possesses a value of that property, but cannot be said to possess a value of the other property (like momentum). Experiments suggest that, at most, all that we can predict about the other property is the probability of getting a particular value of it if we perform a measurement of it. These probabilities are assigned to states by the *Born Rule*:

> **Born Rule**: The probability $\Pr_\psi(b_i|B)$ that a quantum system in a vector state $|\psi\rangle$ possesses the value b_i of a property B is given by: $\Pr_\psi(b_i|B) \equiv |\langle\psi|b_i\rangle|^2 = |\alpha_i|^2$, where $|b_i\rangle$ is the eigenvector of B with eigenvalue b_i, and α_i is the expansion coefficient corresponding to $|b_i\rangle$ in the expansion of $|\psi\rangle$ in the eigenvector basis of B.[2]

In quantum mechanics, there are two ways a vector state can change. In the absence of a measurement, it evolves in time via the Schrödinger equation (the equation of motion for non-relativistic quantum mechanics). Schematically, $|\psi(t)\rangle = e^{-iHt}|\psi(t_0)\rangle$, for initial time t_0, where H is the Hamiltonian operator. When a measurement occurs, a vector state "collapses" according to:

> **Projection Postulate**: When a measurement of a property B is made on a system in a vector state $|\psi\rangle = \sum_i \alpha_i|b_i\rangle$ expanded in the eigenvector basis of B, and the result is the value b_i, then $|\psi\rangle$ collapses to the vector state $|b_i\rangle$.

<u>*Ex. 2*</u>. Let $|Q\rangle = \sqrt{½}(|0\rangle + |1\rangle)$ be the vector state of an electron in a basis associated with a z-axis spin-½ property.[3] Let $|0\rangle$ be the vector state in which the electron has the value "spin-up" of this property, and let $|1\rangle$ be the vector state in which the electron has the value "spin-down" of this property. According to the Eigenvector/Eigenvalue Rule, in the vector state $|Q\rangle$, the electron does not have a value of this property. According to the Born Rule, the probability that the outcome of a measurement of this property is "spin-up" is $|\langle Q|0\rangle|^2 = |\sqrt{½}|^2 = ½$, and the probability that it is "spin-down" is the same, $|\langle Q|1\rangle|^2 = |\sqrt{½}|^2 = ½$. According to the Projection Postulate, if a measurement of this property is made, and the outcome is "spin-up", then $|Q\rangle$ collapses to $|0\rangle$.

Finally, the *expectation value* of an observable O with respect to a vector state is its average value in that state. For a vector state $|\psi\rangle$, it is defined by:

> **Def. 1** (*Expectation value for vector state*). The **expectation value** $\langle O\rangle_\psi$ of an observable O with respect to a vector state $|\psi\rangle$ is given by $\langle O\rangle_\psi \equiv \langle\psi|O|\psi\rangle$.

---

[2] Recall that if $\alpha = x + iy$ is a complex number, then its absolute value is $|\alpha| = \sqrt{x^2 + y^2}$. This entails, for instance, that $\alpha^*\alpha = x^2 + y^2 = |\alpha|^2$.

[3] A spin-½ property with respect to an axis (*i.e.*, direction in space) has two possible values: "spin-up along the axis", and "spin-down along the axis". For every axis in 3-dim space, there's an associated spin-½ property. One way to think of this is in terms of 2 degrees of freedom (spin-up and spin-down) for any direction in 3-dim space.

$\langle O\rangle_\psi$ is the value of the property O that you get on average when you measure the system for this property when it's in the vector state $|\psi\rangle$. To see this, let the property be represented by B with eigenvectors $|b_i\rangle$ and eigenvalues b_i, and let $|\psi\rangle = \sum_i \alpha_i|b_i\rangle$. Then

$$\begin{aligned}\langle B\rangle_\psi \equiv \langle\psi|B|\psi\rangle &= (\textstyle\sum_{i,j} \alpha_j^*\langle b_j|)B(\sum_i \alpha_i|b_i\rangle)\\ &= \textstyle\sum_{i,j} b_i\alpha_j^*\alpha_i\langle b_j|b_i\rangle\\ &= \textstyle\sum_i b_i \Pr_\psi(b_i|B) \qquad \textit{since } \langle b_j|b_i\rangle = 1 \textit{ for } i=j\textit{, and } 0 \textit{ otherwise}\end{aligned}$$

where $\Pr_\psi(b_i|B) = |\alpha_i|^2 = \alpha_i^*\alpha_i$ is the probability (according to the Born Rule) for obtaining the value b_i of B when the system is in the vector state $|\psi\rangle$. Note that $\sum_i b_i\Pr_\psi(b_i|B)$ is the average value of the set $\{b_1, ..., b_n\}$ with the probabilities $\{\Pr_\psi(b_1|B), ..., \Pr_\psi(b_n|B)\}$ assigned to its members.

3. States as Density Operators

Another way to represent a state in quantum mechanics is by a density operator. This is important for cases in which a physical system is associated with an ensemble $\{|\psi_i\rangle, p_i\}$ of vector states $|\psi_i\rangle$, each with a given probability p_i.

Def. 2 (*Density operator*). The **density operator** ρ for a system in one of a number m of vector states $|\psi_i\rangle \in \mathcal{H}$ each with probablity p_i, is defined by

$$\rho \equiv \sum_{i=1}^m p_i|\psi_i\rangle\langle\psi_i| \quad \text{where } \sum_{i=1}^m p_i = 1$$

<u>Note 1</u>: The vector states $\{|\psi_i\rangle\}$ are not necessarily orthogonal to each other, and m is not necessarily the dimension n of $\mathcal{H}$. But ρ can always be re-expressed in terms of a basis $\{|\phi_i\rangle\}$ of $\mathcal{H}$ as $\rho = \sum_{i=1}^n \lambda_i|\phi_i\rangle\langle\phi_i|$, where $|\phi_i\rangle$ are eigenvectors of ρ with eigenvalues λ_i, such that each $\lambda_i \geq 0$ (this is a consequence of the fact that ρ is a Hermitian operator). And if ρ represents an ensemble $\{|\phi_i\rangle, \lambda_i\}$ of vector states $|\phi_i\rangle$ with probabilities λ_i, we require $\sum_i \lambda_i = 1$.

<u>Note 2</u>: A given density operator ρ can describe more than one ensemble of vector states.
<u>Ex. 3</u>. $\rho = \frac{1}{2}|a\rangle\langle a| + \frac{1}{2}|b\rangle\langle b| = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|$, where $|a\rangle = \sqrt{3/4}|0\rangle + \sqrt{1/4}|1\rangle$, $|b\rangle = \sqrt{3/4}|0\rangle - \sqrt{1/4}|1\rangle$. ρ is the density operator for a system in vector state $|a\rangle$ with probability ½ and vector state $|b\rangle$ with probability ½, and also for a system in vector state $|0\rangle$ with probability ¾ and vector state $|1\rangle$ with probability ¼.

One reason to use density operators to represent states is that they allow you to distinguish between "pure" states and "mixed" states:

Def. 3 (*Pure/mixed state*). If a density operator can be expressed as $\rho = |\psi\rangle\langle\psi|$, then it is called a **pure density operator state**. Otherwise, it is called a **mixed density operator state**.

Ignorance interpretation of mixed states: A pure density operator state $\rho = |\psi\rangle\langle\psi|$ corresponds to a vector state $|\psi\rangle$ with probability 1. A mixed density operator state has the form $\rho = p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2| + ...$, which, on the surface, means the system is either in the vector state $|\psi_1\rangle$ with probablity p_1, or vector state $|\psi_2\rangle$ with probability p_2, *etc*. So, on the surface, a mixed density operator state is associated with a degree of uncertainty as to which vector state the system is in. But care should be taken, since, as noted above, a given density operator can describe more than one ensemble of vector states. Moreover, an ignorance interpretation cannot be applied to a mixed density operator state of a subsystem of a larger composite system when the composite system is in an entangled pure density operator state. In this situation, the subsystem cannot be said to be in a definite pure density operator state (see Example 10 below).

Def. 4 (*Trace*). The **trace** of an operator O acting on a vector space $\mathcal{H}$ is the sum of the diagonal elements of a matrix representation of O:

$\mathrm{Tr}\, O \equiv \sum_i \langle w_i|O|w_i\rangle$, where $\{|w_i\rangle\}$ is a basis for $\mathcal{H}$.

Note: If O is a Hermitian operator on $\mathcal{H}$, then its eigenvectors $\{|\lambda_i\rangle\}$ form a basis for $\mathcal{H}$, and its trace is the sum of its eigenvalues: $\mathrm{Tr}\, O = \sum_i \langle\lambda_i|O|\lambda_i\rangle = \sum_i \lambda_i\langle\lambda_i|\lambda_i\rangle = \sum_i \lambda_i$.

Claim 1: $\mathrm{Tr}\,\rho = 1$ for both pure and mixed density operator states.
Proof: Since ρ is Hermitian, $\mathrm{Tr}\,\rho = \sum_i \lambda_i = 1$, where the λ_i are the eigenvalues of ρ.

Claim 2:
(a) ρ is a pure density operator state *if and only if* $\mathrm{Tr}\,\rho^2 = 1$.
(b) ρ is a mixed density operator state *if and only if* $\mathrm{Tr}\,\rho^2 < 1$.

Proof: First recall that any density operator state can be expressed by $\rho = \sum_{i=1}^{n} \lambda_i|\phi_i\rangle\langle\phi_i|$, where $|\phi_i\rangle$ are eigenvectors of ρ that form a basis for $\mathcal{H}$, and λ_i are eigenvalues of ρ such that $\sum_i \lambda_i = 1$. Then

$$\begin{aligned}\rho^2 &= \textstyle\sum_{i=1}^{n} \lambda_i|\phi_i\rangle\langle\phi_i| \sum_{j=1}^{n} \lambda_j|\phi_j\rangle\langle\phi_j| \\ &= \textstyle\sum_{i,j} \lambda_i \lambda_j|\phi_i\rangle\langle\phi_i|\phi_j\rangle\langle\phi_j| \\ &= \textstyle\sum_i \lambda_i{}^2\, |\phi_i\rangle\langle\phi_i| \qquad \textit{since } \langle\phi_i|\phi_j\rangle = 1 \textit{ for } i = j, \textit{ and } 0 \textit{ otherwise}\end{aligned}$$

So the eigenvalues of ρ^2 are $\lambda_i{}^2$, and thus $\mathrm{Tr}\,\rho^2 = \sum_i \lambda_i{}^2$. Note that $\rho^2 = \rho$ if and only if ρ is pure, and $\rho^2 \neq \rho$ if and only if ρ is mixed.

(a) Suppose ρ is a pure density operator state. Then $\rho^2 = \rho$. So $\mathrm{Tr}\,\rho^2 = \mathrm{Tr}\,\rho = 1$, from Claim 1. Now suppose ρ is a density operator state and $\mathrm{Tr}\,\rho^2 = 1$. Then $\sum_i \lambda_i{}^2 = 1$. But we also have $\sum_i \lambda_i = 1$. Now $\sum_i \lambda_i{}^2 = \sum_i \lambda_i = 1$ if and only if one of the λ_i is 1 and the rest are 0. And this means $\rho = |\phi\rangle\langle\phi|$; i.e., ρ is pure.

(b) Suppose ρ is a mixed density operator state. From the above, $\mathrm{Tr}\rho^2 = \sum_i \lambda_i{}^2$, and in general $\sum_i \lambda_i{}^2 \leq \sum_i \lambda_i = 1$. So $\mathrm{Tr}\rho^2 \leq 1$, with equality if and only if ρ is pure. So if ρ is mixed, then $\mathrm{Tr}\rho^2 < 1$. Finally, suppose ρ is a density operator state and $\mathrm{Tr}\rho^2 < 1$. From the above, $\mathrm{Tr}\rho^2 = \sum_i \lambda_i{}^2$, so $\sum_i \lambda_i{}^2 < 1$. Together with $\sum_i \lambda_i = 1$, this excludes the case of one of the λ_i being 1 and the rest 0 (i.e., the pure case). So ρ must be mixed.

Note: A mixed density operator state is not the same as a vector state in a superposition. The density operator $\rho_{\mathrm{I}} = ½|0\rangle\langle 0| + ½|1\rangle\langle 1|$ is a mixed density operator state of a system I that has prob ½ of being in the vector state $|0\rangle$ and prob ½ of being in the vector state $|1\rangle$. This is different from a system II in a superposed vector state $\sqrt{½}\{|0\rangle + |1\rangle\}$. The density operator for system II is $\rho_{\mathrm{II}} = ½\{|0\rangle + |1\rangle\}\{\langle 0| + \langle 1|\} = ½\{|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|\}$. Note that this is a pure density operator state.

Finally, the expectation value of an observable O with respect to a density operator state is defined as follows:

Def. 5 (*Expectation value for density operator state*). The **expectation value** $\langle O\rangle_\rho$ of an observable O with respect to a density operator state ρ is given by $\langle O\rangle_\rho \equiv \mathrm{Tr}\rho O$.

As with Def. 1, $\langle O\rangle_\rho$ is the value of the property O that you get on average when you measure the system for this property when it's in the density operator state ρ. Since ρ represents an ensemble $\{|\psi_j\rangle, p_j\}$ of vector states, the average value of O should be a weighted sum of the average value of O in each vector state $|\psi_j\rangle$, weighted by the probability distribution p_j. In other words $\langle O\rangle_\rho = \sum_j p_j \langle O\rangle_{\psi_j}$. To see this, let the property be represented by B with eigenvectors and eigenvalues $|b_i\rangle$ and b_i, and let $|\psi_j\rangle = \sum_k \alpha_{kj}|b_k\rangle$, so $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j| = \sum_{j,k,l} p_j \alpha_{jk}\alpha^*_{jl}|b_k\rangle\langle b_l|$□Then

$$\begin{aligned}\langle B\rangle_\rho \equiv \mathrm{Tr}\rho B &= \sum_i \langle b_i|\rho B|b_i\rangle \\ &= \sum_{i,j,k,l} p_j b_i \alpha_{jk}\alpha^*_{jl}\langle b_i|b_k\rangle\langle b_l|b_i\rangle \\ &= \sum_j p_j \sum_i b_i \alpha_{ji}\alpha^*_{ji} \\ &= \sum_j p_j \sum_i b_i \mathrm{Pr}_{\psi_j}(b_i|B) = \sum_j p_j \langle O\rangle_{\psi_j}\end{aligned}$$

4. Multipartite States and Quantum Entanglement

Quantum entanglement involves subsystems of a larger "multipartite" (i.e., "multiple parts") system which is characterized by a *product vector space*. Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be n-dim and m-dim vector spaces. The *product vector space* $\mathcal{H}_A \otimes \mathcal{H}_B$ is an $(n \times m)$-dim vector space with an inner-product given by

$$\langle \phi_A \otimes \phi_B | \psi_A \otimes \psi_B \rangle \equiv \langle \phi_A | \psi_A \rangle \langle \phi_B | \psi_B \rangle$$

where $|\psi_A\rangle, |\phi_A\rangle \in \mathcal{H}_A$, $|\psi_B\rangle, |\phi_B\rangle \in \mathcal{H}_B$, and $|\psi_A \otimes \psi_B\rangle, |\phi_A \otimes \phi_B\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$.[4]

If $\{|v_1\rangle, ..., |v_n\rangle\}$ and $\{|w_1\rangle, ..., |w_m\rangle\}$ are bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, then a basis for $\mathcal{H}_A \otimes \mathcal{H}_B$ is given by $\{|v_1 w_1\rangle, |v_1 w_2\rangle, ..., |v_1 w_m\rangle, |v_2 w_1\rangle, ..., |v_n w_m\rangle\}$. Any vector $|\psi\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$ can be expanded in this basis: $|\psi\rangle = a_{11}|v_1 w_1\rangle + a_{12}|v_1 w_2\rangle + \cdots + a_{nm}|v_n w_m\rangle$.

A product space may be the product of more than two lower-dim spaces, and may admit more than one decomposition into lower-dim spaces. For instance, a 16-dim product space can be decomposed into a product of four 2-dim spaces, or a product of two 4-dim spaces.[5]

4.1. Multipartite Vectors and Quantum Entanglement

There are 2 types of vectors in a product vector space: *product vectors*, and *non-product vectors*.

Def. 6 (*Product/non-product vector*). A **product vector** in a product space $\mathcal{H}$ with respect to a decomposition $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ is a vector $|\psi\rangle$ that can be written as a product of n vectors $|\psi\rangle = |v_1\rangle \otimes \cdots \otimes |v_n\rangle$, where $|v_i\rangle \in \mathcal{H}_i$. A **non-product vector** in $\mathcal{H}$ is a vector that is not a product vector.

Ex. 4. Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be 2-dim vector spaces with bases $\{|0\rangle_A, |1\rangle_A\}$ and $\{|0\rangle_B, |1\rangle_B\}$. Then the 4-dim product space $\mathcal{H}_A \otimes \mathcal{H}_B$ is spanned by the basis $\{|0\rangle_A|0\rangle_B, |0\rangle_A|1\rangle_B, |1\rangle_A|0\rangle_B, |1\rangle_A|1\rangle_B\}$. Any vector $|Q\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$ can be expanded as $|Q\rangle = a|0\rangle_A|0\rangle_B + b|0\rangle_A|1\rangle_B + c|1\rangle_A|0\rangle_B + d|1\rangle_A|1\rangle_B$. An example of a *non-product vector* in $\mathcal{H}_A \otimes \mathcal{H}_B$ is

$$\sqrt{½}\{|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\}$$

This vector cannot be factored into a product of two vectors with one in $\mathcal{H}_A$ and the other in $\mathcal{H}_B$. Examples of *product vectors* in $\mathcal{H}_A \otimes \mathcal{H}_B$ are:

(i) $\sqrt{¼}\{|0\rangle_A|0\rangle_B + |0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\}$
(ii) $\sqrt{½}\{|0\rangle_A|0\rangle_B + |1\rangle_A|0\rangle_B\}$
(iii) $|0\rangle_A|0\rangle_B$

---

[4] Instead of "$|\psi_A \otimes \psi_B\rangle$" we can alternatively write "$|\psi_A \psi_B\rangle$" or "$|\psi_A\rangle|\psi_B\rangle$" or "$|\psi_A\rangle \otimes |\psi_B\rangle$".

[5] As long as the dimension n of a vector space isn't a prime number, it will admit at least one decomposition into the product of lower-dim spaces. How many decompositions it admits will depend on the prime factorization of n.

(i) can be factored into $\sqrt{1/4}\{|0\rangle_A + |1\rangle_A\}\{|0\rangle_B + |1\rangle_B\}$, and (ii) can be factored into $\sqrt{1/2}\{|0\rangle_A + |1\rangle_A\}|0\rangle_B$. In both cases, we have a product of two vectors, one in $\mathcal{H}_A$ and the other in $\mathcal{H}_B$. (iii) is already in this form.

There are two important facts about product vectors and non-product vectors.

Fact 1. A vector in a product vector space $\mathcal{H}$ can be a non-product vector with respect to one decomposition of $\mathcal{H}$, and a product vector with respect to another decomposition.

Fact 1 says that whether or not a vector is a non-product vector is relative to a decomposition of the product vector space it's an element of.

Ex. 5. (Rieffel & Pollak 2011, 39–40.) Let $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \otimes \mathcal{H}_4$ be a decomposition of a 16-dim product space $\mathcal{H}$ into four 2-dim spaces, and let $\{|0\rangle_i, |1\rangle_i\}$, $i = 1...4$, be a basis for each of the 2-dim spaces $\mathcal{H}_i$. Then a basis for $\mathcal{H}$ in this decomposition is given by the 16-member set,

$$\{|0\rangle_1|0\rangle_2|0\rangle_3|0\rangle_4,\ |0\rangle_1|0\rangle_2|0\rangle_3|1\rangle_4,\ ...,\ |1\rangle_1|1\rangle_2|1\rangle_3|1\rangle_4\}$$

Another decomposition is given by $\mathcal{H} = \mathcal{H}_{13} \otimes \mathcal{H}_{24}$, where $\mathcal{H}_{13}$ and $\mathcal{H}_{24}$ are 4-dim spaces. Let $\{|00\rangle_{13}, |01\rangle_{13}, |10\rangle_{13}, |11\rangle_{13}\}$ and $\{|00\rangle_{24}, |01\rangle_{24}, |10\rangle_{24}, |11\rangle_{24}\}$ be bases for $\mathcal{H}_{13}$ and $\mathcal{H}_{24}$, respectively. Then a basis for $\mathcal{H}$ in this decomposition is given by the 16-member set

$$\{|00\rangle_{13}|00\rangle_{24},\ |00\rangle_{13}|01\rangle_{24},\ ...,\ |11\rangle_{13}|11\rangle_{24}\}$$

Now consider the vector in $\mathcal{H}$ given by

$$\begin{aligned} |\psi\rangle &= \tfrac{1}{2}\{|0\rangle_1|0\rangle_2|0\rangle_3|0\rangle_4 + |0\rangle_1|1\rangle_2|0\rangle_3|1\rangle_4 + |1\rangle_1|0\rangle_2|1\rangle_3|0\rangle_4 + |1\rangle_1|1\rangle_2|1\rangle_3|1\rangle_4\} \\ &= \sqrt{1/2}\{|00\rangle_{13} + |11\rangle_{13}\} \otimes \sqrt{1/2}\{|00\rangle_{24} + |11\rangle_{24}\} \end{aligned}$$

$|\psi\rangle$ is a non-product vector with respect to the decomposition $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \otimes \mathcal{H}_4$, and a product vector with respect to the decomposition $\mathcal{H} = \mathcal{H}_{13} \otimes \mathcal{H}_{24}$.

The second important fact about product vectors and non-product vectors has to do with correlations between observables represented by *product operators*.

Def. 7 (*Product operator*). Let O_A and O_B be operators on vector spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, and let $|\psi_A\rangle \in \mathcal{H}_A$, $|\psi_B\rangle \in \mathcal{H}_B$, and $|\psi_A \otimes \psi_B\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. Then the **product operator** $O_A \otimes O_B$ is defined by $(O_A \otimes O_B)\,|\psi_A \otimes \psi_B\rangle \equiv O_A|\psi_A\rangle \otimes O_B|\psi_B\rangle$.

Note: This definition generalizes to n-partite product vector spaces $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$, and product operators $O_1 \otimes \cdots \otimes O_n$, where O_i is an operator on $\mathcal{H}_i$.

Two observables O_A, O_B are *correlated* in a vector state just when their joint expectation value cannot be factored into a product of their expectation values taken separately:

> **Def. 8** (*Correlated observables for vector state*). Let O_A and O_B be operators on vector spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ with identity operators I_A and I_B, and let $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. Then the observables represented by O_A and O_B are **correlated** in vector state $|\psi\rangle$ just when $\langle O_A \otimes O_B\rangle_\psi \neq \langle O_A \otimes I_B\rangle_\psi \langle I_A \otimes O_B\rangle_\psi$.

<u>*Claim*</u>: Under the Born Rule, the observables represented by O_A and O_B are correlated *if and only if* they are *statistically dependent*, which means:

$$\Pr_\psi(a_i, b_j | O_A, O_B) \neq \Pr_\psi(a_i | O_A)\Pr_\psi(b_j | O_B), \quad \text{for all } i, j$$

where $\Pr_\psi(a_i, b_j | O_A, O_B)$ is the joint probability of obtaining the values a_i, b_j of O_A and O_B in the state $|\psi\rangle$, and $\Pr_\psi(a_i|O_A)$, $\Pr_\psi(b_j|O_B)$ are the probabilties of obtaining these results separately.

The second important fact then is:

> **Fact 2**. Observables represented by the operators O_i appearing in a product operator $O_1 \otimes \cdots \otimes O_n$ are uncorrelated in a product vector state and correlated in a non-product vector state.

<u>*Ex. 6*</u>. Let's look at the simple case of a bipartite product operator $O_A \otimes O_B$ on $\mathcal{H}_A \otimes \mathcal{H}_B$. Let $|\psi_{\text{prod}}\rangle = |\psi_A\psi_B\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a product vector, and let $|\psi_{\text{non}}\rangle = \sqrt{½}\{|\psi_A\phi_B\rangle + |\phi_A\psi_B\rangle\} \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a non-product vector. Then

$$\begin{aligned}
\langle O_A \otimes O_B\rangle_{\psi_{\text{prod}}} &= \langle \psi_A\psi_B|(O_A \otimes O_B)|\psi_A\psi_B\rangle \\
&= \langle \psi_A\psi_B|(O_A|\psi_A\rangle \otimes O_B|\psi_B\rangle) && \textit{definition of product operator} \\
&= \langle \psi_A|O_A|\psi_A\rangle\langle \psi_B|O_B|\psi_B\rangle && \textit{definition of inner-product for product space} \\
&= \langle \psi_A|O_A|\psi_A\rangle(\langle \psi_B|I_B|\psi_B\rangle)(\langle \psi_A|I_A|\psi_A\rangle)\langle \psi_B|O_B|\psi_B\rangle \\
&= \langle \psi_A\psi_B|(O_A \otimes I_B)|\psi_A\psi_B\rangle\langle \psi_A\psi_B|(I_A \otimes O_B)|\psi_A\psi_B\rangle \\
&= \langle O_A \otimes I_B\rangle_{\psi_{\text{prod}}}\langle I_A \otimes O_B\rangle_{\psi_{\text{prod}}}
\end{aligned}$$

$$\begin{aligned}
\langle O_A \otimes O_B\rangle_{\psi_{\text{non}}} &= ½\{\langle \psi_A\phi_B| + \langle \phi_A\psi_B|\}(O_A \otimes O_B)\{|\psi_A\phi_B\rangle + |\phi_A\psi_B\rangle\} \\
&= ½\{\langle \psi_A|\langle \phi_B| + \langle \phi_A|\langle \psi_B|\}\{O_A|\psi_A\rangle O_B|\phi_B\rangle + O_A|\phi_A\rangle O_B|\psi_B\rangle\} \\
&= ½\{\langle \psi_A|O_A|\psi_A\rangle\langle \phi_B|O_B|\phi_B\rangle + \langle \phi_A|O_A|\psi_A\rangle\langle \psi_B|O_B|\phi_B\rangle \\
&\qquad + \langle \psi_A|O_A|\phi_A\rangle\langle \phi_B|O_B|\psi_B\rangle + \langle \phi_A|O_A|\phi_A\rangle\langle \psi_B|O_B|\psi_B\rangle\} \\
&= ½\{\langle O_A\rangle_{\psi_A}\langle O_B\rangle_{\phi_B} + \langle \phi_A|O_A|\psi_A\rangle\langle \psi_B|O_B|\phi_B\rangle + \langle \psi_A|O_A|\phi_A\rangle\langle \phi_B|O_B|\psi_B\rangle \\
&\qquad + \langle O_A\rangle_{\phi_A}\langle O_B\rangle_{\psi_B}\}
\end{aligned}$$

$$\begin{aligned}\langle O_A \otimes I_B\rangle_{\psi_{\text{non}}} &= ½\{\langle\psi_A\phi_B| + \langle\phi_A\psi_B|\}(O_A \otimes I_B)\{|\psi_A\phi_B\rangle + |\phi_A\psi_B\rangle\}\\ &= ½\{\langle\psi_A|\langle\phi_B| + \langle\phi_A|\langle\psi_B|\}\{O_A|\psi_A\rangle|\phi_B\rangle + O_A|\phi_A\rangle|\psi_B\rangle\}\\ &= ½\{\langle\psi_A|O_A|\psi_A\rangle\langle\phi_B|\phi_B\rangle + \langle\phi_A|O_A|\psi_A\rangle\langle\psi_B|\phi_B\rangle\\ &\qquad + \langle\psi_A|O_A|\phi_A\rangle\langle\phi_B|\psi_B\rangle + \langle\phi_A|O_A|\phi_A\rangle\langle\psi_B|\psi_B\rangle\}\\ &= ½\{\langle O_A\rangle_{\psi_A} + \langle O_A\rangle_{\phi_A}\}\end{aligned}$$

Similarly, $\langle I_A \otimes O_B\rangle_{\psi_{\text{non}}} = ½\{\langle O_B\rangle_{\phi_B} + \langle O_B\rangle_{\psi_B}\}$. So

$$\langle O_A \otimes I_B\rangle_{\psi_{\text{non}}}\langle I_A \otimes O_B\rangle_{\psi_{\text{non}}} = ¼\{\langle O_A\rangle_{\psi_A}\langle O_B\rangle_{\phi_B} + \langle O_A\rangle_{\psi_A}\langle O_B\rangle_{\psi_B} + \langle O_A\rangle_{\phi_A}\langle O_B\rangle_{\phi_B} + \langle O_A\rangle_{\phi_A}\langle O_B\rangle_{\psi_B}\}$$

Hence $\langle O_A \otimes O_B\rangle_{\psi_{\text{non}}} \neq \langle O_A \otimes I_B\rangle_{\psi_{\text{non}}}\langle I_A \otimes O_B\rangle_{\psi_{\text{non}}}$

This generalizes to multipartite product operators and multipartite product vector states. This movitates the following definition of a quantum entangled vector state:

Def. 9 (*Entangled vector state*). A state represented by a multipartite vector $|\psi\rangle$ is **quantum entangled** just when $|\psi\rangle$ is a non-product state.

Recall Ex. 4's non-product vector $\sqrt{½}\{|0_A\rangle|0_B\rangle + |1_A\rangle|1_B\rangle\}$ in $\mathcal{H}_A \otimes \mathcal{H}_B$. According to Def. 9, this is a quantum entangled two qubit state in which qubit A is in a superposition of $|0\rangle$ and $|1\rangle$, and qubit B is in a superposition of $|0\rangle$ and $|1\rangle$, and the two superpositions are "entangled"; *i.e.*, they cannot be separated into a qubit A part and a qubit B part. According to the Eigenvector/Eigenvalue Rule, neither qubit has a definite value. If we measure qubit A and get the value "0", then according to the Projection Postulate, the two qubit state collapses to $|0_A\rangle|0_B\rangle$, which is a two qubit state in which qubit A has the value "0" and qubit B has the value "0". Note that we did not measure qubit B. So prior to the measurement on qubit A, qubit B had no definite value, and after the measurement on qubit A, both qubits have definite values. Qubits A and B thus exhibit a correlation.[6] Einstein, Podolsky and Rosen (1935) found this disturbing: A measurement on qubit A instantaneously affects the state of qubit B, no matter how far apart they might be. Einstein called this "spooky action at a distance" and thought it violated special relativity's prohibition on superluminal signalling. But it's not that spooky for two reasons:

(a) While the correlation can't be explained by a causal signal that A sends to B, an explanation in terms of a common cause might still be possible: perhaps qubits A and B interacted in the past and this interaction established the correlation. (It turns out that a common cause explanation can also be ruled out, but this was only established by Bell in 1964. See Appendix 2 for details.)

(b) The correlation cannot be used to send signals faster than the speed of light. This is the essence of the "No Signalling Theorem" (Theorem A1.2 in Appendix 1).

[6] More precisely, they exhibit a correlation between two single-qubit observables each of which has the possible values "0" and "1".

Our example of an entangled vector state, call it $|\Phi^+\rangle$, has 3 variations:

$$|\Phi^+\rangle = \sqrt{½}\{|0_A\rangle|0_B\rangle + |1_A\rangle|1_B\rangle\}$$
$$|\Phi^-\rangle = \sqrt{½}\{|0_A\rangle|0_B\rangle - |1_A\rangle|1_B\rangle\}$$
$$|\Psi^+\rangle = \sqrt{½}\{|0_A\rangle|1_B\rangle + |1_A\rangle|0_B\rangle\}$$
$$|\Psi^-\rangle = \sqrt{½}\{|0_A\rangle|1_B\rangle - |1_A\rangle|0_B\rangle\}$$

These are called "Bell states", or sometimes "EPR states". Recall that a basis for $\mathcal{H}_A \otimes \mathcal{H}_B$ is given by $\{|0_A\rangle|0_B\rangle, |0_A\rangle|1_B\rangle, |1_A\rangle|0_B\rangle, |1_A\rangle|1_B\rangle\}$. Another basis is given by $\{|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle\}$.

4.2. Multipartite Density Operators and Quantum Entanglement

There are 3 types of density operators that act on a product vector space: *product density operators*, *separable density operators*, and *non-separable density operators*.

Def. 10 (*Product/separable/non-separable density operator*). Let $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ be an n-partite product space. A **product density operator** on $\mathcal{H}$ is a density operator ρ_{prod} that can be written as the product of n terms $\rho_{\text{prod}} = \rho^1 \otimes \cdots \otimes \rho^n$, where ρ^k is a pure density operator on $\mathcal{H}_k$. A **separable density operator** on $\mathcal{H}$ is a density operator ρ_{sep} that can be written as a sum of product density operators $\rho_{\text{sep}} = \sum_i p_i\,(\rho_i^1 \otimes \cdots \otimes \rho_i^n)$, where $\sum_i p_i = 1$, and ρ_i^k is a pure density operator on $\mathcal{H}_k$. A **non-separable density operator** on $\mathcal{H}$ is a density operator that is not separable.

Note: A product density operator $\rho_{\text{prod}} = \rho^1 \otimes \cdots \otimes \rho^n = |\phi^1\rangle\langle\phi^1| \otimes \cdots \otimes |\phi^n\rangle\langle\phi^n| = |\phi^1 \cdots \phi^n\rangle\langle\phi^1 \cdots \phi^n|$ corresponds to a product vector $|\phi^1 \cdots \phi^n\rangle$. A separable density operator $\rho_{\text{sep}} = \sum_i p_i\,(\rho_i^1 \otimes \cdots \otimes \rho_i^n) = \sum_i p_i\,|\phi_i^1 \cdots \phi_i^n\rangle\langle\phi_i^1 \cdots \phi_i^n|$ then represents an ensemble of product vectors $\{|\phi_i^1 \cdots \phi_i^n\rangle; p_i\}$. So a non-separable density operator is a density operator that cannot be written as an ensemble of product vectors.

Ex. 7. The mixed density operator $½|00\rangle\langle 00| + ½|11\rangle\langle 11|$ is separable: it corresponds to an ensemble $\{|00\rangle, |11\rangle; ½, ½\}$ of product vectors. The pure density operator $½\{|00\rangle + |11\rangle\}\{\langle 00| + \langle 11|\}$ is non-separable: it corresponds to the non-product vector $\sqrt{½}\{|00\rangle + |11\rangle\}$.

Does Fact 2 hold for density operator states? Note first that we can modify Def. 8 for density operator states by replacing expectation values with respect to vector states $\langle \ldots \rangle_\psi$ with expectation values with respect to density operator states $\langle \ldots \rangle_\rho$. Note second that there are now two types of "non-product" density operator: separable and non-separable density operators. Do both types exhibit correlations for product operators?

Let's look at the simple bipartite case. Let $\rho_{\text{prod}} = \rho^A \otimes \rho^B$ and $\rho_{\text{sep}} = \sum_i p_i\,(\rho_i^A \otimes \rho_i^B)$ be a product density operator and a separable density operator on $\mathcal{H}_A \otimes \mathcal{H}_B$. Then for any product operator $O_A \otimes O_B$, we have

$$\begin{aligned}
\langle O_A \otimes O_B\rangle_{\rho_{\text{prod}}} &= \mathrm{Tr}\,\rho_{\text{prod}}(O_A \otimes O_B) \\
&= \mathrm{Tr}\,(\rho^A \otimes \rho^B)(O_A \otimes O_B) \\
&= \textstyle\sum_i \langle \psi_i^A \psi_i^B |(\rho^A \otimes \rho^B)(O_A \otimes O_B)|\psi_i^A \psi_i^B\rangle \qquad \textit{biorthogonal decomposition theorem} \\
&= \textstyle\sum_i \langle \psi_i^A |\rho^A O_A|\psi_i^A\rangle\langle \psi_i^B |\rho^B O_B|\psi_i^B\rangle \\
&= (\mathrm{Tr}\rho^A O_A)\,(\mathrm{Tr}\rho^B O_B) \\
&= \langle O_A\rangle_{\rho^A}\langle O_B\rangle_{\rho^B} \\
&= \langle O_A \otimes I_B\rangle_{\rho_{\text{prod}}}\langle I_A \otimes O_B\rangle_{\rho_{\text{prod}}}
\end{aligned}$$

So product density operator states do not exhibit correlations for product operators. This should be obvious, since a product density operator state corresponds to a product vector state, and there are no correlations between observables with respect to a product vector state.

On the other hand, note that:

$$\begin{aligned}
\langle O_A \otimes O_B\rangle_{\rho_{\text{sep}}} &= \mathrm{Tr}\,\rho_{\text{sep}}(O_A \otimes O_B) \\
&= \textstyle\sum_i p_i\,\mathrm{Tr}\,(\rho_i^A \otimes \rho_i^B)(O_A \otimes O_B) \\
&= \textstyle\sum_i p_i\,(\mathrm{Tr}\,\rho_i^A O_A)\,(\mathrm{Tr}\,\rho_i^B O_B) \\
&\neq \langle O_A \otimes I_B\rangle_{\rho_{\text{sep}}}\langle I_A \otimes O_B\rangle_{\rho_{\text{sep}}}
\end{aligned}$$

So separable density operator states exhibit correlations for product operators. This is because the sum in the third line above prevents the expectation value from factorizing into an A-part and a B-part. The same thing happens for a non-separable density operator state.[7] So non-separable density operator states also exhibit correlations for product operators. However, the correlations present in a separable density operator state can be interpreted as due to classical procedures: they involve assigning a probability distribution p_i to a collection of product vector states to get an ensemble $\{|\phi_i^1 \cdots \phi_i^n\rangle; p_i\}$. The correlations present in a non-separable density operator state cannot be interpreted in this way. This motivates describing separable density operator states as exhibiting "classical correlations" and non-separable density operator states as exhibiting "non-classical correlations". Since quantum entanglement is supposed to involve correlations that are not

[7] This is harder to explicitly demonstrate; but think about it: a non-separable density operator cannot be written as a separable density operator, which means it can't be written as a product density operator, either. And product density operators are the only type of density operator for which expectation values of product operators factorize.

present in classical systems, we're led to identify non-separable density operators with quantum entangled states:

Def. 11 (*Entangled density operator state*). A state represented by a multipartite density operator ρ is **quantum entangled** just when ρ is non-separable.

Ex. 8. (Rieffel & Polak 2011. pg. 224). The mixed density operator $\frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$ is separable and exhibits classical correlations. The pure density operator $|\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}\{|00\rangle + |11\rangle\}\{\langle 00| + \langle 11|\}$ is entangled and exhibits non-classical correlations.

Warning! A density operator state that corresponds to an ensemble of non-product (i.e., entangled) vector states is not necessarily non-separable. Consider the ensemble of entangled vector states $\{|\Phi^+\rangle, |\Phi^-\rangle; \frac{1}{2}, \frac{1}{2}\}$, where $|\Phi^+\rangle = \sqrt{\frac{1}{2}}\{|00\rangle + |11\rangle\}$ and $|\Phi^-\rangle = \sqrt{\frac{1}{2}}\{|00\rangle - |11\rangle\}$. The density operator state that corresponds to this ensemble is $\rho = \frac{1}{2}|\Phi^+\rangle\langle\Phi^+| + \frac{1}{2}|\Phi^-\rangle\langle\Phi^-|$, which can also be written as $\frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$. Expressed in this latter way, ρ is separable.

If all we're concerned with are pure states (*i.e.*, states that can be represented by vectors), then a quantum entangled state is just a non-product vector state. But if we want to take into account mixed states (*i.e.*, states that have to be represented by mixed density operators), then a quantum entangled state is a non-separable density operator state.[8]

5. Reduced Density Operators and Entanglement Entropy

Entanglement entropy is a way of measuring the extent to which a density operator state is quantum entangled. To understand it, we need to understand the notion of a reduced density operator, which first requires the notion of the *partial trace*:

Def. 12 (*Partial trace*). Let $O = O_A \otimes O_B$ be an operator acting on a product vector space $\mathcal{H}_A \otimes \mathcal{H}_B$, and let $\{|w_{B_i}\rangle\}$ be a basis for $\mathcal{H}_B$. The **partial trace** $\mathrm{Tr}_B(O_A \otimes O_B)$ of $O_A \otimes O_B$ over $\mathcal{H}_B$ is defined by $\mathrm{Tr}_B(O_A \otimes O_B) \equiv \sum_i \langle w_{B_i}|O_A \otimes O_B|w_{B_i}\rangle$.

Note: Tracing out the degrees of freedom of $\mathcal{H}_B$ from a product operator $O_A \otimes O_B$ results in an operator on $\mathcal{H}_A$: $\mathrm{Tr}_B(O_A \otimes O_B) = \sum_i \langle w_{B_i}|O_A \otimes O_B|w_{B_i}\rangle = \sum_i O_A \langle w_{B_i}|O_B|w_{B_i}\rangle = O_A \mathrm{Tr}(O_B)$, which is an operator on $\mathcal{H}_A$.

Def. 13 (*Reduced density operator*). Let ρ_{AB} be a density operator for a composite system with subsystems A and B. The **reduced density operator** for system A is defined by $\rho_A \equiv \mathrm{Tr}_B(\rho_{AB})$.

[8] It's not quite as simple as this. It turns out that for multipartite mixed density operator states for systems with more than just two subsystems, the notion of quantum entanglement is hard to formalize! See Earman (2015) for an extended discussion.

ρ_A is supposed to be the density operator for subsystem A obtained by tracing out the degrees of freedom of subsystem B from the composite system AB.

<u>Ex. 9</u>. Let $\rho_{AB} = \tau \otimes \sigma$. Then $\rho_A = \mathrm{Tr}_B(\tau \otimes \sigma) = \tau \mathrm{Tr}(\sigma) = \tau$, and $\rho_B = \sigma$.

<u>Ex. 10</u>. For the entangled vector state $|\psi_{AB}\rangle = \sqrt{½}\{|0_A0_B\rangle + |1_A1_B\rangle\}$, we have:

$$\begin{aligned}
\rho_{AB} &= |\psi_{AB}\rangle\langle\psi_{AB}| = ½\{|0_A0_B\rangle + |1_A1_B\rangle\}\{\langle 0_A0_B| + \langle 1_A1_B|\} \\
&= ½\{|0_A0_B\rangle\langle 0_A0_B| + |1_A1_B\rangle\langle 0_A0_B| + |0_A0_B\rangle\langle 1_A1_B| + |1_A1_B\rangle\langle 1_A1_B|\} \\
&= ½\{|0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B| + |1_A\rangle\langle 0_A| \otimes |1_B\rangle\langle 0_B| + |0_A\rangle\langle 1_A| \otimes |0_B\rangle\langle 1_B| \\
&\qquad + |1_A\rangle\langle 1_A| \otimes |1_B\rangle\langle 1_B|\}
\end{aligned}$$

$$\begin{aligned}
\rho_A &= \mathrm{Tr}_B(\rho_{AB}) \\
&= ½\{\mathrm{Tr}_B(|0_A0_B\rangle\langle 0_A0_B|) + \mathrm{Tr}_B(|1_A1_B\rangle\langle 0_A0_B|) + \mathrm{Tr}_B(|0_A0_B\rangle\langle 1_A1_B|) \\
&\qquad + \mathrm{Tr}_B(|1_A1_B\rangle\langle 1_A1_B|)\} \\
&= ½\{|0_A\rangle\langle 0_A|\langle 0_B|0_B\rangle + |1_A\rangle\langle 0_A|\langle 0_B|1_B\rangle + |0_A\rangle\langle 1_A|\langle 1_B|0_B\rangle + |1_A\rangle\langle 1_A|\langle 1_B|1_B\rangle\} \\
&= ½\{|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|\} = ½ I_A
\end{aligned}$$

$$\rho_B = \mathrm{Tr}_A(\rho_{AB}) = ½\{|0_B\rangle\langle 0_B| + |1_B\rangle\langle 1_B|\} = ½ I_B$$

<u>*Comments on Example* 10</u>:

(a) The 2-qubit entangled density operator state ρ_{AB} is pure, since $\mathrm{Tr}(\rho_{AB}{}^2) = 1$. The density operator states ρ_A, ρ_B of qubits A and B are mixed, since $\mathrm{Tr}(\rho_A{}^2) = \mathrm{Tr}(\rho_B{}^2) < 1$.

(b) Suppose we wrote the 2-qubit vector state as $|\psi_{AB}\rangle = \sqrt{½}\{a|0_A\rangle(1/a)|0_B\rangle + b|1_A\rangle(1/b)|1_B\rangle\}$, for arbitrary constants a, b. Then there's an ambiguity over what vector state qubit A is in: is it $\sqrt{½}\{|0_A\rangle + |1_A\rangle\}$, or $\sqrt{½}\{a|0_A\rangle + b|1_A\rangle\}$? But there's no ambiguity over what reduced density operator state it's in. It's still $½ I_A$. So using vectors to represent states of composite systems can be ambiguous in a way that using density operators is not (another reason to use density operators to represent states).

(c) The qubit A reduced density operator ρ_A looks like the density operator for a system in vector state $|0_A\rangle$ with probability ½ and in vector state $|1_A\rangle$ with probability ½. And similarly for ρ_B. Moreover, qubits A and B are correlated: whenever one of them is 0, so is the other; and whenever one of them is 1, so is the other. Thus, if they are in definite states, the 2-qubit system should be in the state corresponding to $\rho_{AB} = ½\{|0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B| + |1_A\rangle\langle 1_A| \otimes |1_B\rangle\langle 1_B|\}$. But by assumption, this in not the case. So an ignorance interpretation of ρ_A and ρ_B cannot be applied. (This is an example of subsystems in mixed density operator states that are part of a larger composite system in a pure entangled state.)

Def. 14 (*Von Neumann entropy*). The **von Neumann entropy** $S_{\mathrm{vN}}(\rho)$ of a state ρ is defined by $S_{\mathrm{vN}}(\rho) \equiv -\mathrm{Tr}(\rho \ln \rho)$.

<u>Note 1</u>: In this definition, "$\ln \rho$" is a function "ln" that acts on an operator ρ. We know what $\ln x$ does to a number x, but what does it do to an operator? Let B be an operator such that $B|\psi_i\rangle = b_i|\psi_i\rangle$, and let $f(x)$ be a function on the real numbers. Then we can

define the corresponding *operator function* $f(B)|\psi_i\rangle \equiv f(b_i)|\psi_i\rangle$. In other words, the operator function $f(B)$ is defined to be the operator with eigenvalue $f(b_i)$. Thus if $\rho = \sum_{i=1}^{n} \lambda_i|\phi_i\rangle\langle\phi_i|$ so that $\rho|\phi_i\rangle = p_i|\phi_i\rangle$ then $\ln \rho|\phi_i\rangle = \ln \lambda_i|\phi_i\rangle$. This means we can rewrite Def. 14 as: $-\mathrm{Tr}(\rho \ln \rho) = -\sum_i \langle\phi_i|(\rho \ln \rho)|\phi_i\rangle = -\sum_i \lambda_i \ln \lambda_i \langle\phi_i|\phi_i\rangle = -\sum_i \lambda_i \ln \lambda_i$.

Note 2: From Def. 5, $-\mathrm{Tr}(\rho \ln \rho) = \langle -\ln \rho\rangle_\rho$. In other words, $S(\rho)$ is the expectation value of the operator $-\ln \rho$ in the state ρ.

Note 3: If ρ is a density operator state on an n-dim vector space $\mathcal{H}$, then the maximum value of $S_{\mathrm{vN}}(\rho)$ is $\ln n$.
Proof: The maximum value of $S_{\mathrm{vN}}(\rho) = S_{\mathrm{vN}}(\lambda_i)$ occurs for that value λ_i^* of λ_i which makes the derivative of $S_{\mathrm{vN}}(\lambda_i)$ vanish. The derivative of $S_{\mathrm{vN}}(\lambda_i)$ is given by $(d/d\lambda_i)S_{\mathrm{vN}}(\lambda_i) = -\sum_i \ln \lambda_i$, or $dS_{\mathrm{vN}}(\lambda_i) = -\sum_i \ln \lambda_i d\lambda_i$. So we need to solve for λ_i^* in:

$$dS_{\mathrm{vN}}(\lambda_i^*) = -\sum_i \ln \lambda_i^* d\lambda_i = 0$$

Note that $\sum_i \lambda_i = 1$, so $\sum_i d\lambda_i = 0$, and this holds even if we multiply $\sum_i d\lambda_i$ by an arbitrary constant α. So, in general:

$$-\sum_i (\ln \lambda_i^* - \alpha) d\lambda_i = 0$$

Thus $\ln \lambda_i^* = \alpha$, or $\lambda_i^* = e^\alpha$. If we substitute this back into $\sum_i \lambda_i = 1$, we get $\sum_i e^\alpha = ne^\alpha = 1$, so $\alpha = \ln(1/n)$, and hence $\lambda_i^* = (1/n)$. So the maximum value of $S_{\mathrm{vN}}(\lambda_i)$ is $S_{\mathrm{vN}}(1/n) = -\sum_i (1/n)\ln(1/n) = -n(1/n)\ln(1/n) = \ln n$.

Note 4: A "maximally mixed" density operator state $\rho = \sum_{i=1}^{n} \lambda_i|\phi_i\rangle\langle\phi_i|$ is a density operator state in which all the λ_i are equal, which means $\lambda_i = 1/n$. (Think of this as the "farthest" away ρ can be from the pure state case in which just one λ_i is 1 and all the rest are 0.) And this entails $\rho_{\max} = \sum_{i=1}^{n} (1/n)|\phi_i\rangle\langle\phi_i| = (1/n)I_n$, where I_n is the identity operator. One can now show that $S_{\mathrm{vN}}(\rho)$ varies from zero, for a pure density operator, to $\ln n$ (its maximum value), for a maximally mixed density operator state.
Proof: Suppose ρ is a pure density operator state. Then $S_{\mathrm{vN}}(\rho) = -\sum_{i=1}^{n} \lambda_i \ln \lambda_i = -\ln 1 = 0$. Now suppose $\rho_{\max}$ is a maximally mixed density operator state. Then $S_{\mathrm{vN}}(\rho_{\max}) = -\sum_{i=1}^{n} (1/n)\ln(1/n) = -\ln(1/n) = \ln n$.

Note 4 indicates that $S_{\mathrm{vN}}(\rho)$ is a measure of the degree to which the density operator ρ is mixed.

Def. 15 (*Entanglement entropy*). For a bipartite system AB in density operator state ρ_{AB}, the **entanglement entropy** S_A of subsystem A is defined to be the von Neumann entropy of ρ_A, so $S_A \equiv S_{\mathrm{vN}}(\rho_A) = -\mathrm{Tr}(\rho_A \ln \rho_A)$.

S_A is a measure of the degree to which ρ_A is mixed. What does S_A have to do with entanglement? One can show the following (Neilson & Chuang 2010, pg. 514):

Claim: Let $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ be a pure density operator state on a product vector space $\mathcal{H}_A \otimes \mathcal{H}_B$. Then $|\psi_{AB}\rangle$ is an entangled vector state *if and only if* $S_A > 0$ (i.e., ρ_A is mixed).

Proof:

(a) "⇐". We'll first show that if $|\psi_{AB}\rangle$ is not entangled (i.e., if it is a product vector state), then ρ_A is not mixed (i.e., it is pure). This means that if ρ_A is mixed, then $|\psi_{AB}\rangle$ is entangled. So suppose $|\psi_{AB}\rangle = |\varphi_A \varphi_B\rangle$ is a product vector state, where $|\varphi_A\rangle \in \mathcal{H}_A$, and $|\varphi_B\rangle \in \mathcal{H}_B$, and let $\{|w_{B_i}\rangle\}$ be a basis of $\mathcal{H}_B$. Then

$$\begin{aligned}\rho_A = \mathrm{Tr}_B(\rho_{AB}) &= \textstyle\sum_i \langle w_{B_i}|\rho_{AB}|w_{B_i}\rangle \\ &= \textstyle\sum_i \langle w_{B_i}|\varphi_A\varphi_B\rangle\langle\varphi_A\varphi_B|w_{B_i}\rangle \\ &= |\varphi_A\rangle\langle\varphi_A| \textstyle\sum_i \langle\varphi_B|w_{B_i}\rangle\langle w_{B_i}|\varphi_B\rangle \\ &= |\varphi_A\rangle\langle\varphi_A|\langle\varphi_B|\varphi_B\rangle = |\varphi_A\rangle\langle\varphi_A|\end{aligned}$$

which means ρ_A is pure (i.e., not mixed).

(b) "⇒". Now we'll show that if ρ_A is pure, then $|\psi_{AB}\rangle$ is a product vector state (i.e., not entangled). This means that if $|\psi_{AB}\rangle$ is entangled, then ρ_A is not pure (i.e., mixed). Strategy: If $\{|w_{A_i}\rangle\}$ and $\{|w_{B_i}\rangle\}$ are bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, then $|\psi_{AB}\rangle$ can be expanded as $|\psi_{AB}\rangle = \sum_i \alpha_i |w_{A_i} w_{B_i}\rangle$ (this is called a "biorthogonal expansion"). We want to show that if ρ_A is pure, then there's only one term in this biorthogonal expansion of $|\psi_{AB}\rangle$, which makes it a product vector.

$$\begin{aligned}\rho_A = \mathrm{Tr}_B(\rho_{AB}) &= \textstyle\sum_i \langle w_{B_i}|\rho_{AB}|w_{B_i}\rangle \\ &= \textstyle\sum_i \langle w_{B_i}|\sum_j \alpha_j |w_{A_j} w_{B_j}\rangle \sum_k \alpha_k^* \langle w_{A_k} w_{B_k}|w_{B_i}\rangle \\ &= \textstyle\sum_{i,j,k} \alpha_j \alpha_k^* |w_{A_j}\rangle\langle w_{A_k}|\langle w_{B_i}|w_{B_j}\rangle\langle w_{B_k}|w_{B_i}\rangle \\ &= \textstyle\sum_{i,j} \alpha_j \alpha_i^* |w_{A_j}\rangle\langle w_{A_i}|\langle w_{B_i}|w_{B_j}\rangle \\ &= \textstyle\sum_i \alpha_i \alpha_i^* |w_{A_i}\rangle\langle w_{A_i}|\end{aligned}$$

If ρ_A is pure, then all the α_i are zero except for one; and this entails $|\psi_{AB}\rangle$ is a product vector state.

So: When a bipartite system AB is in a pure density operator state, S_A is a measure of the degree to which its subsystems are entangled. However, recall that a mixed density operator state of a subsystem of a composite system in an entangled state cannot be given an ignorance interpretation; *i.e.*, we can't read off the statistics for the subsystem from its mixed density operator state. This means that S_A cannot be a measure of uncertainty for subsystem A. (It can be a measure of the degree to which subsystem A is correlated with subsystem B, however.)

Appendix A1: Entanglement Theorems

Theorem A1.1 (*Conservation of Entanglement Entropy*). For a bipartite system AB in a pure density operator state ρ_{AB} associated with the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, the entanglement entropy S_A of subsystem A is invariant under transformations of the form $U_A \otimes U_B$, where U_A and U_B are unitary operators that act on $\mathcal{H}_A$ and $\mathcal{H}_B$.

> *Proof*. Under the action of $U_A \otimes U_B$, ρ_A becomes $\rho_A' = U_A \rho_A U_A^\dagger$, and the entanglement entropy becomes $S(\rho_A') = -\mathrm{Tr}[U_A \rho_A U_A^\dagger \ln (U_A \rho_A U_A^\dagger)] = -\mathrm{Tr}(\rho_A \ln \rho_A) = S(\rho_A)$, expanding the logarithm and using cyclicity of the trace.

Conservation of entanglement entropy entails that the entanglement of a bipartite pure state, as measured by its entanglement entropy, cannot be affected by unitary operations on either, or both, of the states of its subsystems.

Theorem A1.2 (*No-Signalling*). For a bipartite system AB in a pure density operator state ρ_{AB} associated with the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, the reduced density operator state ρ_A for subsystem A is invariant under transformations of the form $I_A \otimes U_B$, where I_A is the identify on $\mathcal{H}_A$, and U_B is a unitary operator that acts on $\mathcal{H}_B$.

> *Proof*. Under the action of $I_A \otimes U_B$, ρ_A becomes $\rho_A' = I_A \rho_A I_A^\dagger = \rho_A$.

The statistics that govern measurement outcomes on A are encoded in the reduced density operator ρ_A, so Theorem A2 entails that a unitary operation on B (as represented by the operator $I_A \otimes U_B$) cannot affect the statistics of measurement outcomes performed on A. In other words, doing something to one of the subsystems of a composite system in an entangled state cannot be detected in measurements on the other subsystem, and hence cannot be used to signal the other subsystem.

Appendix A2: Entanglement Correlations

Recall from Section 4.2 that the observables represented by the terms in a product operator exhibit a correlation in a quantum entangled state that is not due to a classical mixture. Let's call such a non-classical correlation, a *quantum entanglement correlation*. The task now is to further understand the sense in which a quantum entanglement correlation is non-classical.

A2.1. Correlation

First let's recall what it means for two observables of a physical system in a state represented by a vector $|\psi\rangle$ to be correlated. From Def. 8, the observables represented by A and B are correlated just when there is a vector state $|\psi\rangle$ such that $\langle\psi|(A \otimes B)|\psi\rangle \neq \langle\psi|A|\psi\rangle\langle\psi|B|\psi\rangle$. This says that the expectation value $\langle\psi|(A \otimes B)|\psi\rangle$ of the bipartite operator $A \otimes B$ in the vector state $|\psi\rangle$ (i.e., the average value of the observable represented by $A \otimes B$ in the vector state $|\psi\rangle$) is not equal to the product of the expectation values of A and B separately. And this means (via the Born Rule) that

$$\mathrm{Pr}_\psi(a, b|A, B) \neq \mathrm{Pr}_\psi(a|A)\mathrm{Pr}_\psi(b|B) \tag{A2.1}$$

which says that the joint probability of getting the values a, b of A and B in the vector state $|\psi\rangle$ is not equal to the product of the probabilities of getting a for A, and b for B, separately. This means that the observables A and B are *statistically dependent* (which is another way of saying that A and B are correlated).

A2.2. Entanglement Correlation

An entanglement correlation is supposed to be a non-classical correlation between two observables of the subsystems of a composite system in an entangled state. One way to flesh this out is to characterize a classical correlation as due to either a direct cause, or a common cause (or both), and then demonstrate that an entanglement correlation can be both direct cause-violating and common cause-violating.

Let's characterize a correlation that is not due to a direct cause in the following way:

Def. A2.1 (*Direct cause-violating correlation*). The observables represented by A and B exhibit a **direct cause-violating correlation** just when they are correlated and the distance between them (the distance between the regions on which A and B have support) exceeds an appropriate bound on causal signal propagation.

This means that the A and B observables are so far apart that their correlation cannot be due to a causal signal that propagates between them. So their correlation cannot be due to a direct cause.

What about a correlation that is not due to a common cause? Intuitively, a common cause of a correlation between observables A and B is a random variable λ in their pasts that is the cause of their correlation. Under a standard notion, this requires, in part, that A and B be *conditionally statistically independent* with respect to λ:

$$\Pr_\psi(a, b|A, B,\lambda) = \Pr_\psi(a|A,\lambda)\Pr_\psi(b|B,\lambda) \tag{A2.2}$$

This means that λ screens off A from B so that, with respect to λ, they appear statistically independent.[9] In addition to (A2.2), the standard notion of a common cause also requires a few more conditions on λ, A and B.[10] But these aren't important for our purposes. What's important for us is that the standard notion considers (A2.2) to be a necessary condition for λ to be a common cause of a correlation between A and B. So if A and B are correlated, and there is no random variable λ such that (A2.2) holds, then their correlation cannot be due to a common cause (according to the standard notion). This suggests the following way to characterize a correlation that is not due to a common cause:

---

[9] Example: Let $\lambda =$ a drop in atmospheric pressure, $A =$ a storm, and $B =$ a drop in the mercury level in a barometer. B is relevant to A in the absence of λ, but irrelevant in the presence of λ (and similarly, A is relevant to B in the absence of λ, but irrelevant in the presence of λ).

[10] See, e.g., Hitchcock and Redei (2020). These conditions are as follows: (i) $\sim\lambda$ screens A off from B: $\Pr_\psi(a, b|A, B,\sim\lambda) = \Pr_\psi(a|A,\sim\lambda)\Pr_\psi(b|B,\sim\lambda)$; (ii) A is more probable given λ than in its absence: $\Pr_\psi(a|A, \lambda) > \Pr_\psi(a|A, \sim\lambda)$; (iii) B is more probable given λ, than in its absence: $\Pr_\psi(b|B, \lambda) > \Pr_\psi(b|B, \sim\lambda)$.

Def. A2.2 (*Common cause-violating correlation*). The observables represented by A and B exhibit a **common cause-violating correlation** just when they are correlated and there is no random variable λ such that they are *conditionally statistically independent* with respect to λ.

To show that there can be correlations between observables with respect to an entangled vector state that cannot be due to a common cause, we have to show that there is no random variable λ with respect to which these observables are conditionally statistically independent. We'll now prove the following, slightly more specific, claim:

Claim A2.1. There are pair-wise correlations in the entangled vector state $|\Psi^-\rangle = \sqrt{½}\{|01\rangle - |10\rangle\}$ between four spin-½ observables such that a particular sum of their expectation values violates an inequality that it must satisfy if the correlated observables are conditionally statistically independent.

In other words,

(a) If these correlated observables are conditionally statistically independent, then a particular sum of their expectation values must satisfy an inequality.

(b) This sum does not satisfy the inequality.

If we can prove Claims (a) and (b), then these correlations are common cause-violating. The inequality in Claim (a) is called the "CHSH" inequality (after Clauser, Horne, Shimony and Holt 1969). Claim (a) is given specifically by:

Claim A2.2 (*CHSH inequality*). Let A_x, B_y, x, $y \in \{0, 1\}$ be four spin-½ operators that act on 2-dim vector spaces $\mathcal{H}_A$, $\mathcal{H}_B$, respectively, with values a, $b \in \{-1, +1\}$, and let $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. If A_x, B_y are conditionally statistically independent, then

$$S \equiv \langle A_0 \otimes B_0\rangle_\psi + \langle A_0 \otimes B_1\rangle_\psi + \langle A_1 \otimes B_0\rangle_\psi - \langle A_1 \otimes B_1\rangle_\psi \leq 2$$

To prove this, recall first that (A2.2) expresses what it means for A_x and B_y to be conditionally statistically independent with respect to λ. In general, the random variable λ may vary from measurement to measurement, and hence should be characterized by a probability distribution $q(\lambda)$, so that $\Pr_\psi(a, b|A_x, B_y) = \int d\lambda\, q(\lambda) \Pr_\psi(a, b|A_x, B_y, \lambda)$. Conditional statistical independence with respect to λ is then given by

$$\Pr_\psi(a, b|A_x, B_y) = \int d\lambda\, q(\lambda) \Pr_\psi(a|\, A_x, \lambda) \Pr_\psi(b|\, B_y, \lambda) \qquad \text{(A2.3)}$$

Note, too, that the expectation values of the four product operators $A_x \otimes B_y$ in the state $|\psi\rangle$ are given by

$$\langle A_x \otimes B_y\rangle_\psi = \sum_{a,b} ab \Pr_\psi(a, b|A_x, B_y) \qquad \text{(A2.4)}$$

So, for example,

$$\langle A_0 \otimes B_0\rangle_\psi = \Pr_\psi(-1,-1|A_0, B_0) - \Pr_\psi(-1,+1|A_0, B_0) - \Pr_\psi(+1,-1|A_0, B_0) + \Pr_\psi(+1,+1|A_0, B_0)$$

We can now prove Claim A2.2:

<u>*Proof*</u>:

$$\begin{aligned}\langle A_x \otimes B_y\rangle_\psi &= \textstyle\sum_{a,b} ab \Pr_\psi(a,b|A_x, B_y)\\ &= \textstyle\sum_{a,b} ab \int d\lambda\, q(\lambda) \Pr_\psi(a|A_x,\lambda) \Pr_\psi(b|B_y,\lambda) \quad \textit{from (A2.3)}\\ &= \textstyle\int d\lambda\, q(\lambda) \sum_a a \Pr_\psi(a|A_x,\lambda) \sum_b b \Pr_\psi(b|B_y,\lambda)\\ &= \textstyle\int d\lambda\, q(\lambda) \langle A_x\rangle_{\psi,\lambda}\langle B_y\rangle_{\psi,\lambda} \qquad \textit{where, e.g., } \langle A_x\rangle_{\psi,\lambda} \equiv \sum_a a \Pr_\psi(a|A_x,\lambda)\end{aligned}$$

<u>*So*</u>:

$$\begin{aligned}S &= \textstyle\int d\lambda\, q(\lambda) \{\langle A_0\rangle_{\psi,\lambda}\langle B_0\rangle_{\psi,\lambda} + \langle A_0\rangle_{\psi,\lambda}\langle B_1\rangle_{\psi,\lambda} + \langle A_1\rangle_{\psi,\lambda}\langle B_0\rangle_{\psi,\lambda} - \langle A_1\rangle_{\psi,\lambda}\langle B_1\rangle_{\psi,\lambda}\}\\ &= \textstyle\int d\lambda\, q(\lambda) \{\langle A_0\rangle_{\psi,\lambda}[\langle B_0\rangle_{\psi,\lambda} + \langle B_1\rangle_{\psi,\lambda}] + \langle A_1\rangle_{\psi,\lambda}[\langle B_0\rangle_{\psi,\lambda} - \langle B_1\rangle_{\psi,\lambda}]\}\\ &\leq \textstyle\int d\lambda\, q(\lambda) \{|\langle B_0\rangle_{\psi,\lambda} + \langle B_1\rangle_{\psi,\lambda}| + |\langle B_0\rangle_{\psi,\lambda} - \langle B_1\rangle_{\psi,\lambda}|\}\end{aligned}$$

The last line is a result of the fact that the maximum value of $\langle A_x\rangle_{\psi,\lambda}$ is $+1$. This is also the maximum value of $\langle B_y\rangle_{\psi,\lambda}$. And this means that the maximum value of $|\langle B_0\rangle_{\psi,\lambda} + \langle B_1\rangle_{\psi,\lambda}| + |\langle B_0\rangle_{\psi,\lambda} - \langle B_1\rangle_{\psi,\lambda}|$ is 2. Thus $S \leq 2$.

So, if the pair-wise correlations between A_0, A_1, B_0, B_1 satisfy (A2.3), the CHSH inequality must hold. And this means that if the CHSH inequality does not hold, then (A2.3) does not hold; i.e., the pair-wise correlations between A_0, A_1, B_0, B_1 cannot be due to a common cause. The CHSH inequality is one version of a Bell inequality. There are other ways of deriving inequalities for combinations of expectation values, depending on the particular system and the particular observables. All such inequalities go under the general name of "Bell inequalities". They have in common that a violation of the inequality entails a violation of conditional statistical independence.

Now let's prove Claim (b) above. We'll show that a particular choice of A_0, A_1, B_0, B_1 violates the CHSH inequality with respect to the entangled vector state $|\Psi^-\rangle = \sqrt{½}\{|01\rangle - |10\rangle\}$. Our choice will be the following spin-½ operators:

$$A_0 = \hat{x}\cdot\vec{\sigma} \qquad A_1 = \hat{y}\cdot\vec{\sigma}$$
$$B_0 = -\sqrt{½}(\hat{x}+\hat{y})\cdot\vec{\sigma} \qquad B_1 = \sqrt{½}(-\hat{x}+\hat{y})\cdot\vec{\sigma}$$

The vector $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ encodes the Pauli operators which act on 2-dim vectors in the following way (in quantum information theory, these are called X, $-iY$, Z operators):

$$\sigma_x|0\rangle = |1\rangle, \qquad \sigma_x|1\rangle = |0\rangle$$
$$\sigma_y|0\rangle = i|1\rangle, \qquad \sigma_y|1\rangle = -i|0\rangle$$
$$\sigma_z|0\rangle = |0\rangle, \qquad \sigma_z|1\rangle = -|1\rangle$$

A_0 represents the spin-½ observable "*spin-along-the-$\hat{x}$-axis*", and A_1 represents "*spin-along-the-$\hat{y}$-axis*". The axes of B_0 and B_1 are at 45°from the $\hat{x}$ and $\hat{y}$ axes.[11] We can now explicitly calculate $S = \langle A_0 \otimes B_0\rangle_{\Psi^-} + \langle A_0 \otimes B_1\rangle_{\Psi^-} + \langle A_1 \otimes B_0\rangle_{\Psi^-} - \langle A_1 \otimes B_1\rangle_{\Psi^-}$.

$$\begin{aligned}
\langle A_0 \otimes B_0\rangle_{\Psi^-} &= ½\{\langle 01| - \langle 10|\}(A_0 \otimes B_0)\{|01\rangle - |10\rangle\} \\
&= ½\{\langle 01| - \langle 10|\}[(\hat{x}\cdot\vec{\sigma})\otimes[-\sqrt{½}(\hat{x}+\hat{y})\cdot\vec{\sigma}]]\{|01\rangle - |10\rangle\} \\
&= ½(-\sqrt{½})\{\langle 01| - \langle 10|\}[\sigma_x \otimes (\sigma_x + \sigma_y)]\{|01\rangle - |10\rangle\} \\
&= ½(-\sqrt{½})\{\langle 01| - \langle 10|\}[(\sigma_x \otimes \sigma_x) + (\sigma_x \otimes \sigma_y)]\{|01\rangle - |10\rangle\} \\
&= ½(-\sqrt{½})\{\langle 01| - \langle 10|\}\{|10\rangle - |01\rangle + [-i|10\rangle - i|01\rangle]\} \\
&= ½(-\sqrt{½})\{(-1-i) + (-1+i)\} = \sqrt{½}
\end{aligned}$$

$$\begin{aligned}
\langle A_0 \otimes B_1\rangle_{\Psi^-} &= ½\{\langle 01| - \langle 10|\}(A_0 \otimes B_1)\{|01\rangle - |10\rangle\} \\
&= ½\{\langle 01| - \langle 10|\}[(\hat{x}\cdot\vec{\sigma})\otimes[\sqrt{½}(-\hat{x}+\hat{y})\cdot\vec{\sigma}]]\{|01\rangle - |10\rangle\} \\
&= ½(\sqrt{½})\{\langle 01| - \langle 10|\}[\sigma_x \otimes (-\sigma_x + \sigma_y)]\{|01\rangle - |10\rangle\} \\
&= ½(\sqrt{½})\{\langle 01| - \langle 10|\}[(\sigma_x \otimes -\sigma_x) + (\sigma_x \otimes \sigma_y)]\{|01\rangle - |10\rangle\} \\
&= ½(\sqrt{½})\{\langle 01| - \langle 10|\}\{-|10\rangle + |01\rangle + [-i|10\rangle - i|01\rangle]\} \\
&= ½(\sqrt{½})\{(1-i) - (-1-i)\} = \sqrt{½}
\end{aligned}$$

$$\begin{aligned}
\langle A_1 \otimes B_0\rangle_{\Psi^-} &= ½\{\langle 01| - \langle 10|\}(A_1 \otimes B_0)\{|01\rangle - |10\rangle\} \\
&= ½\{\langle 01| - \langle 10|\}[(\hat{y}\cdot\vec{\sigma})\otimes[-\sqrt{½}(\hat{x}+\hat{y})\cdot\vec{\sigma}]]\{|01\rangle - |10\rangle\} \\
&= ½(-\sqrt{½})\{\langle 01| - \langle 10|\}[\sigma_y \otimes (\sigma_x + \sigma_y)]\{|01\rangle - |10\rangle\} \\
&= ½(-\sqrt{½})\{\langle 01| - \langle 10|\}[(\sigma_y \otimes \sigma_x) + (\sigma_y \otimes \sigma_y)]\{|01\rangle - |10\rangle\} \\
&= ½(-\sqrt{½})\{\langle 01| - \langle 10|\}\{i|10\rangle + i|01\rangle + [|10\rangle - |01\rangle] \\
&= ½(-\sqrt{½})\{(i-1) - (i+1)\} = \sqrt{½}
\end{aligned}$$

$$\begin{aligned}
\langle A_1 \otimes B_1\rangle_{\Psi^-} &= ½\{\langle 01| - \langle 10|\}(A_0 \otimes B_1)\{|01\rangle - |10\rangle\} \\
&= ½\{\langle 01| - \langle 10|\}[(\hat{y}\cdot\vec{\sigma})\otimes[\sqrt{½}(-\hat{x}+\hat{y})\cdot\vec{\sigma}]]\{|01\rangle - |10\rangle\} \\
&= ½(\sqrt{½})\{\langle 01| - \langle 10|\}[\sigma_y \otimes (-\sigma_x + \sigma_y)]\{|01\rangle - |10\rangle\}
\end{aligned}$$

---

[11] The two spin-½ properties associated with the orthogonal directions $\hat{x}$, $\hat{y}$ could be Hardness and Color. The two spin-½ properties associated with the orthogonal directions $-\sqrt{½}(\hat{x}+\hat{y})$, $\sqrt{½}(-\hat{x}+\hat{y})$ are just two other spin-½ properties in directions oriented at 45°from the Hardness and Color directions.

$= ½(\sqrt{½})\{\langle 01| - \langle 10|\}[(\sigma_y \otimes -\sigma_x) + (\sigma_y \otimes \sigma_y)]\{|01\rangle - |10\rangle\}$
$= ½(\sqrt{½})\{\langle 01| - \langle 10|\}\{-i|10\rangle - i|01\rangle + [|10\rangle - |01\rangle]\}$
$= ½(\sqrt{½})\{(-i - 1) - (-i + 1)\} = -\sqrt{½}$

So $S = \langle A_0 \otimes B_0\rangle_\psi + \langle A_0 \otimes B_1\rangle_\psi + \langle A_1 \otimes B_0\rangle_\psi - \langle A_1 \otimes B_1\rangle_\psi = 2\sqrt{2} > 2$. Thus, in the entangled vector state $|\Psi^-\rangle = \sqrt{½}\{|0_A\rangle|1_B\rangle - |1_A\rangle|0_B\rangle\}$, and for our choice of spin-½ observables, we've shown that A_0 and B_0 are correlated, as are A_0 and B_1, and A_1 and B_0, and A_1 and B_1. And these correlations are not conditionally statistically independent, because their expectation values with respect to $|\Psi^-\rangle$ violate the CHSH inequality. So these correlations are common-cause violating! Moreover, if we assume the two entangled subsystems are separated by a distance that exceeds an appropriate bound on causal signal propagation, the correlations between these four observables are direct cause-violating. Thus, to the extent that classical correlations are due either to a direct cause or a common cause (or both), the correlations exhibited by our spin-½ observables in our entangled vector state are non-classical.

Note 1: What's so special about quantum mechanics that certain correlations between observables of a system in an entangled vector state violate a Bell inequality? What's special is the way quantum mechanics represents the expectation values $\langle A_x \otimes B_y\rangle_\psi$ of a bipartite observable in an entangled state; namely, using the machinery of operators and vectors.

Note 2: Many authors interpret the conditional statistical independence condition (A2.2), or (A2.3), as a form of "locality" (sometimes this is called "Bell locality"). They then interpret the violation of a Bell inequality by expectation values of observables of a system in an entangled state as an indication that an entangled state in particular, and quantum mechanics in general, exhibits non-locality. The intuition apparently is that if a direct cause-violating correlation does not admit a common cause, then it involves some form of non-locality. But this is misleading, since the general notion of locality has many meanings, and condition (A2.2), or (A2.3), is easily understood most directly as a common cause condition (as opposed to a "locality" condition). On the other hand, we might use the term "Einstein locality" to refer to the sense of locality that a direct cause-violating correlation violates (it exhibits "spooky action-at-a-distance"); and we might use the term "Bell locality" to refer to the sense of locality that a Bell inequality-violating correlation violates. Then we could say that an entanglement correlation violates both Einstein locality and Bell locality.

References

Brunner, N., D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner (2014) 'Bell Nonlocality', *Reviews of Modern Physics 86*, 419-478

Earman, J. (2015) 'Some Puzzles and Unresolved Issues About Quantum Entanglement', *Erkenntnis 80*, pp. 303-337.

Hitchcock, C. and M. Redei (2020) 'Reichenbach's Principle of Common Cause', *The Stanford Encyclopedia of Philosophy*, /plato.stanford.edu/archives/sum2021/entries/physics-Rpcc/

Nielson, M. and I. Chuang (2010) *Quantum Computation and Quantum Information*, Cambridge University Press, pp. 99-108.

Rieffel, E. and W. Polak (2011) *Quantum Computation: A Gentle Introduction*, MIT Press, pp. 207-225.