Article

LOGICAL ENTROPY: INTRODUCTION TO CLASSICAL AND QUANTUM LOGICAL INFORMATION THEORY

David Ellerman [1]

[1] University of California Riverside; david@ellerman.org

Abstract: Logical information theory is the quantitative version of the logic of partitions just as logical probability theory is the quantitative version of the dual Boolean logic of subsets. The resulting notion of information is about distinctions, differences, and distinguishability, and is formalized using the distinctions ('dits') of a partition (a pair of points distinguished by the partition). All the definitions of simple, joint, conditional, and mutual entropy of Shannon information theory are derived by a uniform transformation from the corresponding definitions at the logical level. The purpose of this paper is to give the direct generalization to quantum logical information theory that similarly focuses on the pairs of eigenstates distinguished by an observable, i.e., qudits of an observable. The fundamental theorem for quantum logical entropy and measurement establishes a direct quantitative connection between the increase in quantum logical entropy due to a projective measurement and the eigenstates (cohered together in the pure superposition state being measured) that are distinguished by the measurement (decohered in the post-measurement mixed state). Both the classical and quantum versions of logical entropy have simple interpretations as "two-draw" probabilities for distinctions. The conclusion is that quantum logical entropy is the simple and natural notion of information for quantum information theory focusing on the distinguishing of quantum states.

Keywords: logical entropy; partition logic; qudits of an observable

1. Introduction

The formula for 'classical' logical entropy goes back to the early twentieth century [17]. It is the derivation of the formula from basic logic that is new and accounts for the name. The ordinary Boolean logic of subsets has a dual logic of partitions [12] since partitions (= equivalence relations = quotient sets) are category-theoretically dual to subsets. Just as the quantitative version of subset logic is the notion of logical finite probability, so the quantitative version of partition logic is logical information theory using the notion of logical entropy [13]. This paper generalizes that 'classical' (i.e., non-quantum) logical information theory to the quantum version. The classical logical information theory is briefly developed before turning to the quantum version.[1]

2. Duality of Subsets and Partitions

The foundations for classical and quantum logical information theory are built on the logic of partitions–which is dual (in the category-theoretic sense) to the usual Boolean logic of subsets. F. William Lawvere called a subset or, in general, a subobject a "part" and then noted: "The dual notion (obtained by reversing the arrows) of 'part' is the notion of *partition*." [24, p. 85] That suggests that the Boolean logic of subsets has a dual logic of partitions ([11], [12]).

[1] Applications of logical entropy have already been developed in several special mathematical settings; see [26] and the references cited therein.

This duality can be most simply illustrated using a set function $f : X \to Y$. The image $f(X)$ is a *subset* of the codomain Y and the inverse-image or coimage[2] $f^{-1}(Y)$ is a *partition* on the domain X–where a *partition* $\pi = \{B_1, ..., B_I\}$ on a set U is a set of subsets or blocks B_i that are mutually disjoint and jointly exhaustive ($\cup_i B_i = U$). But the duality runs deeper than between subsets and partitions. The dual to the notion of an "element" (an 'it') of a subset is the notion of a "distinction" (a 'dit') of a partition, where $(u, u') \in U \times U$ is a *distinction* or *dit* of π if the two elements are in different blocks. Let $\text{dit}(\pi) \subseteq U \times U$ be the set of distinctions or *ditset* of π. Similarly an *indistinction* or *indit* of π is a pair $(u, u') \in U \times U$ in the same block of π. Let $\text{indit}(\pi) \subseteq U \times U$ be the set of indistinctions or *inditset* of π. Then $\text{indit}(\pi)$ is the equivalence relation associated with π and $\text{dit}(\pi) = U \times U - \text{indit}(\pi)$ is the complementary binary relation that might be called a *partition relation* or an *apartness relation*. The notions of a distinction and indistinction of a partition are illustrated in Figure 1.

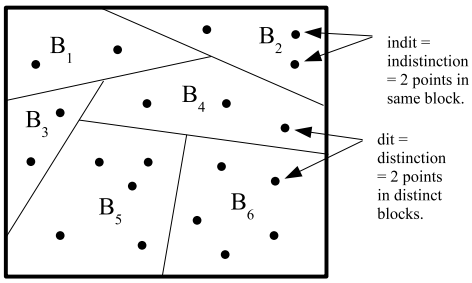

Figure 1: Distinctions and indistinctions of a partition.

The relationships between Boolean subset logic and partition logic are summarized in Table 1 which illustrates the dual relationship between the elements ('its') of a subset and the distinctions ('dits') of a partition.

	Subset Logic	Partition Logic
Logic of...	Subsets S ⊆ U	Partitions π on U
Elements (its or dits)	*Elements* u of a subset S	*Distinctions* (u,u') of a partition π
All elements	Universe set U (all elements)	Discrete partition **1** (all dits)
No elements	Empty set ∅ (no elements)	Indiscrete partition **0** (no dits)
Partial order on...	S ⊆ T = Inclusion of elements	Refinement σ ⩽ π of partitions = inclusion of distinctions: dit(σ) ⊆ dit(π)
Formula variables	Subsets of U	Partitions on U
Logical operations ∪, ∩, ⇒,...	Operations on subsets	Operations on partitions
Propositional interp. of Φ(π,σ,...)	Subset Φ(π,σ,...) contains an element u.	Partition Φ(π,σ,...) makes a distinction (u,u').
Valid formula Φ(π,σ,...)	Φ(π,σ,...) = U for any subsets π,σ,... of any U (|U| ≥ 1), i.e., contains all elements u.	Φ(π,σ,...) = **1** for any partitions π,σ,... on any U (|U| ≥ 2), i.e., makes all distinctions (u,u').

Table 1: Dual Logics: Boolean subset logic of subsets and partition logic.

3. From the logic of partitions to logical information theory

George Boole [4] developed the quantitative version of his logic of subsets by starting with the size or number of elements $|S|$ in a subset $S \subseteq U$, which could then be normalized to $\frac{|S|}{|U|}$ and given

[2] In category theory, the duality between subobject-type constructions (e.g., limits) and quotient-object-type constructions (e.g., colimits) is often indicated by adding the prefix "co-" to the latter. Hence the usual Boolean logic of "images" has the dual logic of "coimages."

the probabilistic interpretation as the probability that a randomly drawn element from U would be an element of S.

The algebra of partitions π on U is isomorphically represented by the algebra of ditsets $\operatorname{dit}(\pi) \subseteq U \times U$, so the parallel quantitative development of the logic of partitions would start with the size or number of distinctions $|\operatorname{dit}(\pi)|$ in a partition π on U, which could then be normalized to $\frac{|\operatorname{dit}(\pi)|}{|U \times U|}$. It has the probabilistic interpretation as the probability that two randomly drawn elements from U (with replacement) would be a distinction of π.

In Gian-Carlo Rota's Fubini Lectures [31] (and in his lectures at MIT), he remarked in view of duality between partitions and subsets that, quantitatively, the "lattice of partitions plays for information the role that the Boolean algebra of subsets plays for size or probability" [22, p. 30] or symbolically:

$$\frac{\text{Logical Probability Theory}}{\text{Boolean Logic of Subsets}} = \frac{\text{Logical Information Theory}}{\text{Logic of Partitions}}.$$

Since "Probability is a measure on the Boolean algebra of events" that gives quantitatively the "intuitive idea of the size of a set", we may ask by "analogy" for some measure to capture a property for a partition like "what size is to a set." Rota goes on to ask:

> How shall we be led to such a property? We have already an inkling of what it should be: it should be a measure of information provided by a random variable. Is there a candidate for the measure of the amount of information? [31, p. 67]

We have just seen that the parallel development suggests the normalized number of distinctions of a partition as "the measure of the amount of information".

4. The logical theory of information

Andrei Kolmogorov has suggested that information theory should start with sets, not probabilities.

> Information theory must precede probability theory, and not be based on it. By the very essence of this discipline, the foundations of information theory have a finite combinatorial character. [21, p. 39]

The notion of information-as-distinctions does start with the *set of distinctions*, the *information set*, of a partition $\pi = \{B_1, ..., B_I\}$ on a finite set U where that set of distinctions (dits) is:

$$\operatorname{dit}(\pi) = \{(u, u') : \exists B_i, B_{i'} \in \pi, B_i \neq B_{i'}, u \in B_i, u' \in B_{i'}\}.$$

The normalized size of a subset is the logical probability of the event, and the normalized size of the ditset of a partition is, in the sense of measure theory, "the measure of the amount of information" in a partition. Thus we define the *logical entropy* of a partition $\pi = \{B_1, ..., B_I\}$, denoted $h(\pi)$, as the size of the ditset $\operatorname{dit}(\pi) \subseteq U \times U$ normalized by the size of $U \times U$:

$$h(\pi) = \frac{|\operatorname{dit}(\pi)|}{|U \times U|} = \frac{|U \times U| - \sum_{i=1}^{I} |B_i \times B_i|}{|U \times U|} = 1 - \sum_{i=1}^{I} \left(\frac{|B_i|}{|U|}\right)^2 = 1 - \sum_{i=1}^{I} \Pr(B_i)^2.$$

In two independent draws from U, the probability of getting a distinction of π is the probability of not getting an indistinction.

Given any probability measure $p : U \to [0, 1]$ on $U = \{u_1, ..., u_n\}$ which defines $p_i = p(u_i)$ for $i = 1, ..., n$, the *product measure* $p \times p : U \times U \to [0, 1]$ has for any binary relation $R \subseteq U \times U$ the value of:

$$p \times p(R) = \sum_{(u_i, u_j) \in R} p(u_i)\, p(u_j) = \sum_{(u_i, u_j) \in R} p_i p_j.$$

The *logical entropy* of π in general is the product-probability measure of its ditset $\operatorname{dit}(\pi) \subseteq U \times U$, where $\Pr(B) = \sum_{u \in B} p(u)$:

$$h(\pi) = p \times p(\operatorname{dit}(\pi)) = \sum_{(u_i,u_j) \in \operatorname{dit}(\pi)} p_i p_j = 1 - \sum_{B \in \pi} \Pr(B)^2.$$

There are two stages in the development of logical information. Before the introduction of any probabilities, the information set of a partition π on U is its ditset $\operatorname{dit}(\pi)$. Then given a probability measure $p : U \to [0,1]$ on U, the logical entropy of the partition is just the product measure on the ditset, i.e., $h(\pi) = p \times p(\operatorname{dit}(\pi))$. The standard interpretation of $h(\pi)$ is the two-draw probability of getting a distinction of the partition π–just as $\Pr(S)$ is the one-draw probability of getting an element of the subset-event S.

5. Compound logical entropies

The compound notions of logical entropy are also developed in two stages, first as sets and then, given a probability distribution, as two-draw probabilities. After observing the similarity between the formulas holding for the compound Shannon entropies and the Venn diagram formulas that hold for any measure (in the sense of measure theory), the information theorist, Lorne L. Campbell, remarked in 1965 that the similarity:

> suggests the possibility that $H(\alpha)$ and $H(\beta)$ are measures of sets, that $H(\alpha,\beta)$ is the measure of their union, that $I(\alpha,\beta)$ is the measure of their intersection, and that $H(\alpha|\beta)$ is the measure of their difference. The possibility that $I(\alpha,\beta)$ is the entropy of the "intersection" of two partitions is particularly interesting. This "intersection," if it existed, would presumably contain the information common to the partitions α and β. [6, p. 113]

Yet, there is no such interpretation of the Shannon entropies as measures of sets, but the logical entropies precisely fulfill Campbell's suggestion (with the "intersection" of two partitions being the intersection of their ditsets). Moreover, there is a uniform requantifying transformation (see next section) that obtains all the Shannon definitions from the logical definitions and explains how the Shannon entropies can satisfy the Venn diagram formulas (e.g., as a mnemonic) while not being defined by a measure on sets.

Given partitions $\pi = \{B_1, ..., B_I\}$ and $\sigma = \{C_1, ..., C_J\}$ on U, the *joint information set* is the union of the ditsets which is also the *ditset for their join* is: $\operatorname{dit}(\pi) \cup \operatorname{dit}(\sigma) = \operatorname{dit}(\pi \vee \sigma) \subseteq U \times U$. Given probabilities $p = \{p_1, ..., p_n\}$ on U, the *joint logical entropy* is the product probability measure on the union of ditsets:

$$h(\pi,\sigma) = h(\pi \vee \sigma) = p \times p(\operatorname{dit}(\pi) \cup \operatorname{dit}(\sigma)) = 1 - \sum_{i,j} \Pr(B_i \cap C_j)^2.$$

The information set for the *conditional logical entropy* $h(\pi|\sigma)$ is the difference of ditsets, and thus that logical entropy is:

$$h(\pi|\sigma) = p \times p(\operatorname{dit}(\pi) - \operatorname{dit}(\sigma)) = h(\pi,\sigma) - h(\sigma).$$

The information set for the *logical mutual information* $m(\pi,\sigma)$ is the intersection of ditsets, so that logical entropy is:

$$m(\pi,\sigma) = p \times p(\operatorname{dit}(\pi) \cap \operatorname{dit}(\sigma)) = h(\pi,\sigma) - h(\pi|\sigma) - h(\sigma|\pi) = h(\pi) + h(\sigma) - h(\pi,\sigma).$$

Since all the logical entropies are the values of a measure $p \times p : U \times U \to [0,1]$ on subsets of $U \times U$, they automatically satisfy the usual Venn diagram relationships as in Figure 2.

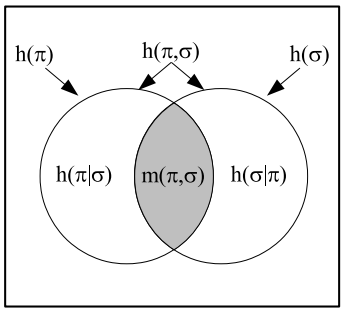

Figure 2: Venn diagram for logical entropies as values of a probability measure $p \times p$ on $U \times U$.

At the level of information sets (w/o probabilities), we have the *information algebra* $I(\pi, \sigma)$ which is the Boolean subalgebra of $\wp(U \times U)$ generated by ditsets and their complements.

6. Deriving the Shannon entropies from the logical entropies

Instead of being defined as the values of a measure, the usual notions of simple and compound entropy 'burst forth fully formed from the forehead' of Claude Shannon [33] already satisfying the standard Venn diagram relationships.[3] Since the Shannon entropies are not the values of a measure, many authors have pointed out that these Venn diagram relations for the Shannon entropies can only be taken as "analogies" or "mnemonics" ([6]; [1]). Logical information theory explains this situation since all the Shannon definitions of simple, joint, conditional, and mutual information can be obtained by a uniform requantifying transformation from the corresponding logical definitions, and the transformation preserves the Venn diagram relationships.

This transformation is possible since the logical and Shannon notions of entropy can be seen as two different ways to quantify distinctions–and thus *both* theories are based on the foundational idea of *information-as-distinctions*.

Consider the canonical case of n equiprobable elements, $p_i = \frac{1}{n}$. The logical entropy of $\mathbf{1} = \{B_1, ..., B_n\}$ where $B_i = \{u_i\}$ with $p = \left\{\frac{1}{n}, ..., \frac{1}{n}\right\}$ is:

$$\frac{|U \times U - \Delta|}{|U \times U|} = \frac{n^2 - n}{n^2} = 1 - \frac{1}{n} = 1 - \Pr(B_i).$$

The normalized number of distinctions or 'dit-count' of the discrete partition $\mathbf{1}$ is $1 - \frac{1}{n} = 1 - \Pr(B_i)$. The general case of logical entropy for any $\pi = \{B_1, ..., B_I\}$ is the average of the dit-counts $1 - \Pr(B_i)$ for the canonical cases:

$$h(\pi) = \sum_i \Pr(B_i)(1 - \Pr(B_i)).$$

In the canonical case of 2^n equiprobable elements, the minimum number of binary partitions ("yes-or-no questions" or "bits") whose join is the discrete partition $\mathbf{1} = \{B_1, ..., B_{2^n}\}$ with $\Pr(B_i) = \frac{1}{2^n}$, i.e., that it takes to uniquely *encode* each distinct element, is n, so the Shannon-Hartley entropy [18] is the canonical bit-count:

$$n = \log_2(2^n) = \log_2\left(\frac{1}{1/2^n}\right) = \log_2\left(\frac{1}{\Pr(B_i)}\right).$$

The general case Shannon entropy is the average of these canonical bit-counts $\log_2\left(\frac{1}{\Pr(B_i)}\right)$:

$$H(\pi) = \sum_i \Pr(B_i) \log_2\left(\frac{1}{\Pr(B_i)}\right).$$

[3] One author surmised that "Shannon carefully contrived for this 'accident' to occur" [32, p. 153].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 August 2018 doi:10.20944/preprints201808.0208.v1

Peer-reviewed version available at Entropy 2018, 20, 679; doi:10.3390/e20090679

The *Dit-Bit Transform* essentially replaces the canonical dit-counts by the canonical bit-counts. First express any logical entropy concept (simple, joint, conditional, or mutual) as an average of canonical dit-counts $1-\Pr(B_i)$, and then substitute the canonical bit-count $\log\left(\frac{1}{\Pr(B_i)}\right)$ to obtain the corresponding formula as defined by Shannon. Table 2 gives examples of the dit-bit transform.

	The Dit-Bit Transform: $1-\Pr(B_i) \rightsquigarrow \log(1/\Pr(B_i))$
Entropy	$h(\pi) = \sum_i \Pr(B_i)(1-\Pr(B_i))$ $H(\pi) = \sum_i \Pr(B_i)(\log(1/\Pr(B_i)))$
Joint Entropy	$h(\pi,\sigma) = \sum_{i,j}\Pr(B_i \cap C_j)(1-\Pr(B_i \cap C_j))$ $H(\pi,\sigma) = \sum_{i,j}\Pr(B_i \cap C_j)\log(1/\Pr(B_i \cap C_j))$
Conditional Entropy	$h(\pi\|\sigma) = \sum_{i,j}\Pr(B_i \cap C_j)(1-\Pr(B_i \cap C_j)) - \sum_i \Pr(C_j)(1-\Pr(C_j))$ $H(\pi\|\sigma) = \sum_{i,j}\Pr(B_i \cap C_j)\log(1/\Pr(B_i \cap C_j)) - \sum \Pr(C_j)\log(1/\Pr(C_j))$
Mutual Information	$m(\pi,\sigma) = \sum_{i,j}\Pr(B_i \cap C_j)[(1-\Pr(B_i)) + (1-\Pr(C_j)) - (1-\Pr(B_i \cap C_j))]$ $I(\pi,\sigma) = \sum_{i,j}\Pr(B_i \cap C_j)[(\log(1/\Pr(B_i)) + (\log(1/\Pr(C_j)) - (\log(1/\Pr(B_i \cap C_j))]$

Table 2: Summary of the dit-bit transform

For instance,

$$h(\pi|\sigma) = h(\pi,\sigma) - h(\sigma) = \sum_{i,j}\Pr(B_i \cap C_j)\left[1-\Pr(B_i \cap C_j)\right] - \sum_j \Pr(C_j)\left[1-\Pr(C_j)\right]$$

is the expression for $h(\pi|\sigma)$ as an average over $1-\Pr(B_i \cap C_j)$ and $1-\Pr(C_j)$, so applying the dit-bit transform gives:

$$\sum_{i,j}\Pr(B_i \cap C_j)\log\left(1/\Pr(B_i \cap C_j)\right) - \sum_j \Pr(C_j)\log\left(1/\Pr(C_j)\right) = H(\pi,\sigma) - H(\sigma) = H(\pi|\sigma).$$

The dit-bit transform is linear in the sense of preserving plus and minus, so the Venn diagram formulas, e.g., $h(\pi,\sigma) = h(\sigma) + h(\pi|\sigma)$, which are automatically satisfied by logical entropy since it is a measure, carry over to Shannon entropy, e.g., $H(\pi,\sigma) = H(\sigma) + H(\pi|\sigma)$ as in Figure 3, in spite of it not being a measure (in the sense of measure theory):

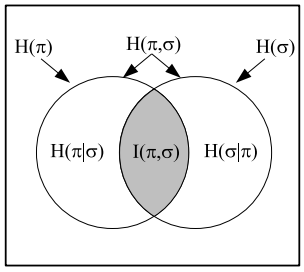

Figure 3: Venn diagram mnemonic for Shannon entropies

7. Shannon information theory for coding and communications

All the compound definitions of the Shannon information theory can be derived from the logical definitions which are a product probability measure on the sets generated by the set operations on ditsets (e.g., union, difference, and intersection). Hence the question of the relation between the two theories naturally arises. The claim here is that the logical theory is the foundational theory that explains what information is at the logical level (information as distinctions) while the Shannon theory is a specialized theory for the purposes of coding and communications theory [33]. In spite of its great success in these fields, the specialized and non-foundational nature of the Shannon theory is emphasized by the way its definitions can breakdown when extended in a number of ways while the logical definitions are stable in all the same cases (see Table 3).

For instance, when the Shannon definition is extended from finite probability distributions to countably infinite probability distributions, it can blow up to infinity as noted in numerous texts,

e.g., [8, p. 48] or [38, p. 30]. But for a countably infinite probability distribution $p = (p_1, p_2, ...)$, we have $\sum_i p_i = 1$ so $h(p) = 1 - \sum_i p_i^2$ is perfectly well-defined. In the process of moving from a discrete distribution p to a continuous distribution $f(x)$, the term $\log\left(\frac{1}{\Delta x_i}\right)$ blows up [27, pp. 34-38] as the mesh size $\Delta x_i \to 0$ so the continuous definition $H(f) = -\int f(x) \log(f(x))\, dx$ is chosen by analogy and "is not in any sense a measure of the randomness of X" [27, p. 38] in addition to possibly having negative values [37, p. 74]. But for a continuous distribution $\int f(x)\, dx = 1$, so the logical entropy $h(f) = 1 - \int f(x)^2\, dx$ is always well-defined. Perhaps the most surprising case is that the conventional mutual information $I(X, Y, Z)$ for three or more random variables can be negative and thus uninterpretable as is noted in many texts, e.g., [1, pp. 128-131], [15, p. 58], [9, p. 52-3], [8, p. 49], and [38, pp.103-4]. In contrast, all the logical entropy notions for any number of random variables have two-draw probability interpretations and thus take values in the half-open interval $[0, 1)$.

Comparisons	Shannon entropy	Logical entropy
Measure (in sense of measure theory)	Is not a measure.	Is a measure.
Countable prob. distribution	$H(p) = \Sigma_i\, p_i(\log(1/p_i))$ can blow up.	$h(p) = 1-\Sigma_i\, p_i^2$ always well defined since $\Sigma_i\, p_i = 1$
Continous limit of discrete	H(p) blows up in continuous limit.	h(p) well-behaved in continuous limit.
Continuous distribution f	$H(f) = -\int f(x)\log f(x)dx$ can be negative.	$h(f) = 1-\int f(x)^2dx$ always non-neg. since $\int f(x)dx=1$
3+ random variables	Mutual information I(X,Y,Z) can be neg.	Mutual info. m(X,Y,Z) is a prob. so always non-neg.

Table 3: Comparisons of generalizations of Shannon and logical entropy

Finally, although not directly related to the comparison with logical entropy, it might be noted that much has been written over the years as if Shannon entropy had the same functional form as Boltzmann's entropy formula $\ln\left(\frac{n!}{n_1!...n_m!}\right)$ in statistical mechanics–even to the point of using phrases like "Shannon-Boltzmann entropy." But the Shannon entropy is obtained as a numerical approximation by using the first two terms in one form of the Stirling series:

$$\ln(n!) = \overbrace{n\ln(n) - n}^{\text{Shannon}} + \tfrac{1}{2}\ln(2\pi n) + \tfrac{1}{(12)n} - \tfrac{1}{(360)n^3} + \tfrac{1}{(1260)n^5} - \tfrac{1}{(1680)n^7} + ...$$

By using more terms, one gets an infinite series of *better and better* approximations to the Boltzmann entropy formula in statistical mechanics (see [25, p. 2] for a similar point).

8. Logical entropy via density matrices

The transition to quantum logical entropy is facilitated by reformulating the classical logical theory in terms of density matrices. Let $U = \{u_1, ..., u_n\}$ be the sample space with the point probabilities $p = (p_1, ..., p_n)$. An event $S \subseteq U$ has the probability $\Pr(S) = \sum_{u_j \in S} p_j$.

For any event S with $\Pr(S) > 0$, let

$$|S\rangle = \tfrac{1}{\sqrt{\Pr(S)}}(\chi_S(u_1)\sqrt{p_1}, ..., \chi_S(u_n)\sqrt{p_n})^t$$

(the superscript t indicates transpose) which is a normalized column vector in $\mathbb{R}^n$ where $\chi_S : U \to \{0, 1\}$ is the characteristic function for S, and let $\langle S|$ be the corresponding row vector. Since $|S\rangle$ is normalized, $\langle S|S\rangle = 1$. Then the *density matrix* representing the event S is the $n \times n$ symmetric real matrix:

$$\rho(S) = |S\rangle\langle S| \text{ so that } (\rho(S))_{j,k} = \begin{cases} \frac{1}{\Pr(S)}\sqrt{p_j p_k} \text{ for } u_j, u_k \in S \\ 0 \text{ otherwise} \end{cases}.$$

Then $\rho(S)^2 = |S\rangle \langle S|S\rangle \langle S| = \rho(S)$ so borrowing language from quantum mechanics, $\rho(S)$ is said to be a *pure state* density matrix.

Given any partition $\pi = \{B_1, ..., B_I\}$ on U, its density matrix is the average of the block density matrices:

$$\rho(\pi) = \sum_i \Pr(B_i)\, \rho(B_i).$$

Then $\rho(\pi)$ represents the *mixed state*, experiment, or lottery where the event B_i occurs with probability $\Pr(B_i)$. A little calculation connects the logical entropy $h(\pi)$ of a partition with the density matrix treatment:

$$h(\pi) = 1 - \sum_{i=1}^{I} \Pr(B_i)^2 = 1 - \text{tr}\left[\rho(\pi)^2\right] = h(\rho(\pi))$$

where $\rho(\pi)^2$ is substituted for $\Pr(B_i)^2$ and the trace is substituted for the summation.

For the throw of a fair die, $U = \{u_1, u_3, u_5, u_2, u_4, u_6\}$ (note the odd faces ordered before the even faces in the matrix rows and columns) where u_j represents the number j coming up, the density matrix $\rho(\mathbf{0})$ is the "pure state" 6×6 matrix with each entry being $\frac{1}{6}$.

$$\rho(\mathbf{0}) = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix} \begin{matrix} u_1 \\ u_3 \\ u_5 \\ u_2 \\ u_4 \\ u_6 \end{matrix}.$$

The nonzero off-diagonal entries represent indistinctions or indits of the partition $\mathbf{0}$, or in quantum terms, "coherences," where all 6 "eigenstates" cohere together in a pure "superposition" state. All pure states have logical entropy of zero, i.e., $h(\mathbf{0}) = 0$ (i.e., no dits) since $\text{tr}[\rho] = 1$ for any density matrix so if $\rho(\mathbf{0})^2 = \rho(\mathbf{0})$, then $\text{tr}\left[\rho(\mathbf{0})^2\right] = \text{tr}[\rho(\mathbf{0})] = 1$ and $h(\mathbf{0}) = 1 - \text{tr}\left[\rho(\mathbf{0})^2\right] = 0$.

The logical operation of *classifying* undistinguished entities (like the six faces of the die before a throw to determine a face up) by a numerical attribute makes distinctions between the entities with different numerical values of the attribute. Classification (also called sorting, fibering, or partitioning [23, Section 6.1]) is the classical operation corresponding to the quantum operation of 'measurement' of a superposition state by an observable to obtain a mixed state.

Now classify or "measure" the die-faces by the parity-of-the-face-up (odd or even) partition (observable) $\pi = \{B_{odd}, B_{even}\} = \{\{u_1, u_3, u_5\}, \{u_2, u_4, u_6\}\}$. Mathematically, this is done by the Lüders mixture operation [2, p. 279], i.e., pre- and post-multiplying the density matrix $\rho(\mathbf{0})$ by P_{odd} and by P_{even}, the projection matrices to the odd or even components, and summing the results:

$$P_{odd}\rho(\mathbf{0})\, P_{odd} + P_{even}\rho(\mathbf{0})\, P_{even}$$

$$= \begin{bmatrix} 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

$$= \begin{bmatrix} 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

$$= \tfrac{1}{2}\rho(B_{odd}) + \tfrac{1}{2}\rho(B_{even}) = \rho(\pi).$$

Theorem 1 (Fundamental (Classical)). *The increase in logical entropy, $h(\rho(\pi)) - h(\rho(\mathbf{0}))$, due to a Lüders mixture operation is the sum of amplitudes squared of the non-zero off-diagonal entries of the beginning density matrix that are zeroed in the final density matrix.*

Proof. Since for any density matrix ρ, $\text{tr}\left[\rho^2\right] = \sum_{i,j} |\rho_{ij}|^2$[16, p. 77], we have: $h(\rho(\pi)) - h(\rho(\mathbf{0})) = \left(1 - \text{tr}\left[\rho(\pi)^2\right]\right) - \left(1 - \text{tr}\left[\rho(\mathbf{0})^2\right]\right) = \text{tr}\left[\rho(\mathbf{0})^2\right] - \text{tr}\left[\rho(\pi)^2\right] = \sum_{i,j}\left(|\rho_{ij}(\mathbf{0})|^2 - |\rho_{ij}(\pi)|^2\right)$. If $(u_i, u_{i'}) \in \text{dit}(\pi)$, then and only then are the off-diagonal terms corresponding to u_i and $u_{i'}$ zeroed by the Lüders operation. □

The classical fundamental theorem connects the concept of information-as-distinctions to the process of 'measurement' or classification which uses some attribute (like parity in the example) or 'observable' to make distinctions.

In the comparison with the matrix $\rho(\mathbf{0})$ of all entries $\frac{1}{6}$, the entries that got zeroed in the Lüders operation $\rho(\mathbf{0}) \longrightarrow \rho(\pi)$ correspond to the distinctions created in the transition $\mathbf{0} = \{\{u_1, ..., u_6\}\} \longrightarrow \pi = \{\{u_1, u_3, u_5\}, \{u_2, u_4, u_6\}\}$, i.e., the odd-numbered faces were distinguished from the even-numbered faces by the parity attribute. The increase in logical entropy = sum of the squares of the off-diagonal elements that were zeroed = $h(\pi) - h(\mathbf{0}) = 2 \times 9 \times \left(\frac{1}{6}\right)^2 = \frac{18}{36} = \frac{1}{2}$. The usual calculations of the two logical entropies are: $h(\pi) = 2 \times \left(\frac{1}{2}\right)^2 = \frac{1}{2}$ and $h(\mathbf{0}) = 1 - 1 = 0$.

Since, in quantum mechanics, a projective measurement's effect on a density matrix *is* the Lüders mixture operation, that means that the effects of the measurement are the above-described "making distinctions" by decohering or zeroing certain coherence terms in the density matrix, and the sum of the absolute squares of the coherences that were decohered is the increase in the logical entropy.

9. Quantum Logical Information Theory: Commuting Observables

The idea of information-as-distinctions carries over to quantum mechanics.

> [Information] is the notion of distinguishability abstracted away from what we are distinguishing, or from the carrier of information. ...And we ought to develop a theory of information which generalizes the theory of distinguishability to include these quantum properties... . [3, p. 155]

Let $F : V \to V$ be a self-adjoint operator (observable) on a n-dimensional Hilbert space V with the real eigenvalues $\phi_1, ..., \phi_I$ and let $U = \{u_1, ..., u_n\}$ be an orthonormal (ON) basis of eigenvectors of F. The quantum version of a dit, a *qudit*, is a pair of states definitely distinguishable by *some* observable[4]–which is analogous classically to a pair (u, u') of distinct elements of U that are distinguishable by some partition (i.e., **1**). In general, a *qudit* is *relativized to an observable*–just as classically a distinction is a distinction *of a partition*. Then there is a set partition $\pi = \{B_i\}_{i=1,...,I}$ on the ON basis U so that B_i is a basis for the eigenspace of the eigenvalue ϕ_i and $|B_i|$ is the "multiplicity" (dimension of the eigenspace) of the eigenvalue ϕ_i for $i = 1, ..., I$. Note that the real-valued function $f : U \to \mathbb{R}$ that takes each eigenvector in $u_j \in B_i \subseteq U$ to its eigenvalue ϕ_i so that $f^{-1}(\phi_i) = B_i$ contains all the information in the self-adjoint operator $F : V \to V$ since F can be reconstructed by defining it on the basis U as $Fu_j = f(u_j) u_j$.

The generalization of 'classical' logical entropy to quantum logical entropy is straightforward using the usual ways that set-concepts generalize to vector-space concepts: subsets → subspaces, set partitions → direct-sum decompositions of subspaces[5], Cartesian products of sets → tensor products

---

[4] Any nondegenerate self-adjoint operator such as $\sum_{k=1}^{n} kP_{[u_k]}$, where $P_{[u_k]}$ is the projection to the one-dimensional subspace generated by u_k, will distinguish all the vectors in the orthonormal basis U.

[5] Hence the 'classical' logic of partitions on a set will generalize to the quantum logic of direct-sum decompositions [14] that is the dual to the usual quantum logic of subspaces (since the set duality of subsets and set partitions 'linearizes' to the vector space duality of subspaces and direct-sum decompositions).

of vector spaces, and ordered pairs $(u_k, u_{k'}) \in U \times U \rightarrow$ basis elements $u_k \otimes u_{k'} \in V \otimes V$. The eigenvalue function $f : U \rightarrow \mathbb{R}$ determines a partition $\left\{f^{-1}(\phi_i)\right\}_{i \in I}$ on U and the blocks in that partition generate the eigenspaces of F which form a direct-sum decomposition of V.

Classically, a *dit of the partition* $\left\{f^{-1}(\phi_i)\right\}_{i \in I}$ on U is a pair $(u_k, u_{k'})$ of points in distinct blocks of the partition, i.e., $f(u_k) \neq f(u_{k'})$. Hence a *qudit of* F is a pair $(u_k, u_{k'})$ (interpreted as $u_k \otimes u_{k'}$ in the context of $V \otimes V$) of vectors in the eigenbasis definitely distinguishable by F, i.e., $f(u_k) \neq f(u_{k'})$, distinct F-eigenvalues. Let $G : V \rightarrow V$ be another self-adjoint operator on V which commutes with F so that we may then assume that U is an orthonormal basis of simultaneous eigenvectors of F and G. Let $\left\{\gamma_j\right\}_{j \in J}$ be the set of eigenvalues of G and let $g : U \rightarrow \mathbb{R}$ be the eigenvalue function so a pair $(u_k, u_{k'})$ is a *qudit of* G if $g(u_k) \neq g(u_{k'})$, i.e., if the two eigenvectors have distinct eigenvalues of G.

As in classical logical information theory, information is represented by certain subsets–or, in the quantum case, subspaces–prior to the introduction of any probabilities. Since the transition from classical to quantum logical information theory is straightforward, it will be presented in table form in Table 5 (which does not involve any probabilities)–where the qudits $(u_k, u_{k'})$ are interpreted as $u_k \otimes u_{k'}$.

Table without probs.	'Classical' Logical Info. Theory	Quantum Logical Info. Theory
Universe	$U = \{u_1,...,u_n\}$	Orthonormal basis $\{u_i\}$ Hilbert space V
Attribute/Observable	Real-valued 'random' variables f, g: $U \rightarrow \mathbb{R}$	Commuting self-adjoint operators F, G $\{u_i\}$ O.N. basis of simult. eigenvectors
Values	Image values $\{\phi_i\}_{i \in I}$ of f Image values $\{\gamma_j\}_{j \in J}$ of g	Eigenvalues $\{\phi_i\}_{i \in I}$ of F Eigenvalues $\{\gamma_j\}_{j \in J}$ of G
Partitions / Direct-sum decompositions	Inverse-image $\pi = \{f^{-1}(\phi_i)\}_{i \in I}$ Inverse-image $\sigma = \{g^{-1}(\gamma_j)\}_{j \in J}$	Eigenspace Direct-sum Decomp. F Eigenspace Direct-sum Decomp. G
Distinctions	Dits of π: $(u_k,u_{k'}) \in U^2$, $f(u_k) \neq f(u_{k'})$ Dits of σ: $(u_k,u_{k'}) \in U^2$, $g(u_k) \neq g(u_{k'})$	Qudits of F: $u_k \otimes u_{k'} \in V \otimes V$, $\phi(u_k) \neq \phi(u_{k'})$ Qudits of G: $u_k \otimes u_{k'} \in V \otimes V$, $\gamma(u_k) \neq \gamma(u_{k'})$
Information sets/spaces	$dit(\pi) \subseteq U \times U$ $dit(\sigma) \subseteq U \times U$	[qudit(F)] = subspace gen. by qudits of F [qudit(G)] = subspace gen. by qudits of G
Joint = Conditional = Mutual =	$dit(\pi) \cup dit(\sigma) \subseteq U \times U$ $dit(\pi) - dit(\sigma) \subseteq U \times U$ $dit(\pi) \cap dit(\sigma) \subseteq U \times U$	[qudit(F) ∪ qudit(G] ⊆ V⊗V [qudit(F) – qudit(G)] ⊆ V⊗V [qudit(F) ∩ qudit(G)] ⊆ V⊗V

Table 5: The parallel development of classical and quantum logical information prior to probabilities.

The *information subspace* associated with F is the subspace $[qudit(F)] \subseteq V \otimes V$ generated by the qudits $u_k \otimes u_{k'}$ of F. If $F = \lambda I$ is a scalar multiple of the identity I, then it has no qudits so its information space $[qudit(\lambda I)]$ is the zero subspace. It is an easy implication of the Common Dits Theorem of classical logical information theory ([10, Proposition 1] or [11, Theorem 1.4]) that any two nonzero information spaces $[qudit(F)]$ and $[qudit(G)]$ have a nonzero intersection, i.e., have a nonzero mutual information space. That is, there are always two eigenvectors u_k and $u_{k'}$ that have different eigenvalues both by F and by G.

In a measurement, the observables do not provide the point probabilities; they come from the pure (normalized) state ψ being measured. Let $|\psi\rangle = \sum_{j=1}^{n} \langle u_j|\psi\rangle |u_j\rangle = \sum_{j=1}^{n} \alpha_j |u_j\rangle$ be the resolution of $|\psi\rangle$ in terms of the orthonormal basis $U = \{u_1,...,u_n\}$ of simultaneous eigenvectors for F and G. Then $p_j = \alpha_j \alpha_j^*$ (α_j^* is the complex conjugate of α_j) for $j = 1,...,n$ are the point probabilities on U and the pure state density matrix $\rho(\psi) = |\psi\rangle\langle\psi|$ (where $\langle\psi| = |\psi\rangle^\dagger$ is the conjugate-transpose) has the entries: $\rho_{jk}(\psi) = \alpha_j \alpha_k^*$ so the diagonal entries $\rho_{jj}(\psi) = \alpha_j \alpha_j^* = p_j$ are the point probabilities. Table 6 gives the remaining parallel development with the probabilities provided by the pure state ψ where we write $\rho(\psi)^\dagger \rho(\psi)$ as $\rho(\psi)^2$.

Table with probs.	'Classical' Logical Info. Theory	Quantum Logical Info. Theory
Probability dist.	Pure state density matrix, e.g., $\rho(\mathbf{0}_U)$	Pure state density matrix $\rho(\psi)$
Product prob. dist.	p×p on U×U	$\rho(\psi)\otimes\rho(\psi)$ on V⊗V
Logical entropies	$h(\mathbf{0}_U) = 1 - \mathrm{tr}[\rho(\mathbf{0}_U)^2] = 0$ $h(\pi) = p\times p(\mathrm{dit}(\pi))$ $h(\pi,\sigma) = p\times p(\mathrm{dit}(\pi) \cup \mathrm{dit}(\sigma))$ $h(\pi\|\sigma) = p\times p(\mathrm{dit}(\pi) - \mathrm{dit}(\sigma))$ $m(\pi,\sigma) = p\times p(\mathrm{dit}(\pi) \cap \mathrm{dit}(\sigma))$	$h(\rho(\psi)) = 1 - \mathrm{tr}[\rho(\psi)^2] = 0$ $h(F{:}\psi) = \mathrm{tr}[P_{[\mathrm{qudit}(F)]}\rho(\psi)\otimes\rho(\psi)]$ $h(F,G{:}\psi) = \mathrm{tr}[P_{[\mathrm{qudit}(F)\cup\mathrm{qudit}(G)]}\rho(\psi)\otimes\rho(\psi)]$ $h(F\|G{:}\psi) = \mathrm{tr}[P_{[\mathrm{qudit}(F)-\mathrm{qudit}(G)]}\rho(\psi)\otimes\rho(\psi)]$ $m(F,G{:}\psi) = \mathrm{tr}[P_{[\mathrm{qudit}(F)\cap\mathrm{qudit}(G)]}\rho(\psi)\otimes\rho(\psi)]$
Venn diagram from being prob. measure	$h(\pi,\sigma) = h(\pi\|\sigma) + h(\sigma\|\pi) + m(\pi,\sigma)$ $h(\pi) = h(\pi\|\sigma) + m(\pi,\sigma)$	$h(F,G) = h(F\|G) + h(G\|F) + m(F,G)$ $h(F) = h(F\|G) + m(F,G)$
Interpretation	$h(\pi)$ = two-draw prob. of getting a dit of π, i.e., different f values.	$h(F{:}\psi)$ = prob. in two indep. F meas. of ψ in getting different eigenvalues.
Lüders Mixture	$\rho(\pi) = \Sigma_i P_{B_i}\rho(\mathbf{0}_U)P_{B_i}$ and $h(\pi) = p\times p(\mathrm{dit}(\pi)) = 1 - \mathrm{tr}[\rho(\pi)^2]$	$\rho'(\psi) = \Sigma_i P_{\phi_i}\rho(\psi)P_{\phi_i}$ and $h(F{:}\psi) = 1 - \mathrm{tr}[\rho'(\psi)^2]$
Thm. on L-entropy and measurement.	$h(\pi)$ = sum of squares of terms zeroed in measurement operation: $\rho(\mathbf{0}_U)\to\rho(\pi)$.	$h(F{:}\psi)$ = sum of absol. squares of terms zeroed in the measurement operation: $\rho(\psi)\to\rho'(\psi)$.

Table 6: The parallel development of classical and quantum logical entropies for commuting F and G.

The formula $h(\rho) = 1 - \mathrm{tr}\left[\rho^2\right]$ is hardly new. Indeed, $\mathrm{tr}\left[\rho^2\right]$ is usually called the *purity* of the density matrix since a state ρ is *pure* if and only if $\mathrm{tr}\left[\rho^2\right] = 1$ so $h(\rho) = 0$, and otherwise $\mathrm{tr}\left[\rho^2\right] < 1$ so $h(\rho) > 0$ and the state is said to be *mixed*. Hence the complement $1 - \mathrm{tr}\left[\rho^2\right]$ has been called the "mixedness" [20, p. 5] or "impurity" of the state ρ.[6] What is new is not the formula but the whole backstory of partition logic outlined above which gives the logical notion of entropy arising out of partition logic as the normalized counting measure on ditsets–just as logical probability arises out of Boolean subset logic as the normalized counting measure on subsets. The basic idea of information is differences, distinguishability, and distinctions ([10], [13]), so the logical notion of entropy is the measure of the distinctions or dits of a partition and the corresponding quantum version is the measure of the qudits of an observable.

The classical dit-bit transform connecting the logical theory to the Shannon theory also carries over to the quantum version. Writing the quantum logical entropy of a density matrix ρ as $h(\rho) = \mathrm{tr}[\rho(1-\rho)]$, the quantum version of the dit-bit transform $(1-\rho) \longrightarrow -\log(\rho)$ yields the usual Von Neumann entropy $S(\rho) = -\mathrm{tr}[\rho\log(\rho)]$[28, p. 510]. The fundamental theorem connecting logical entropy and the operation of classification-measurement also carries over to the quantum case.

10. Two Theorems about quantum logical entropy

Classically, a pair of elements (u_j, u_k) either 'cohere' together in the same block of a partition on U, i.e., are an indistinction of the partition, or they don't, i.e., are a distinction of the partition. In the quantum case, the nonzero off-diagonal entries $\alpha_j\alpha_k^*$ in the pure state density matrix $\rho(\psi) = |\psi\rangle\langle\psi|$ are called quantum "coherences" ([7, p. 303]; [2, p.177]) because they give the amplitude of the eigenstates $|u_j\rangle$ and $|u_k\rangle$ "cohering" together in the coherent superposition state vector $|\psi\rangle = \sum_{j=1}^n \langle u_j|\psi\rangle |u_j\rangle = \sum_j \alpha_j |u_j\rangle$. The coherences are classically modelled by the nonzero off-diagonal entries $\sqrt{p_j p_k}$ for the indistinctions $(u_j, u_k) \in B_i \times B_i$, i.e., coherences ≈ indistinctions.

For an observable F, let $\phi : U \to \mathbb{R}$ be for F-eigenvalue function assigning the eigenvalue $\phi(u_i) = \phi_i$ for each u_i in the ON basis $U = \{u_1, ..., u_n\}$ of F-eigenvectors. The range of ϕ is the set of F-eigenvalues $\{\phi_1, ..., \phi_I\}$. Let $P_{\phi_i} : V \to V$ be the projection matrix in the U-basis to the eigenspace of ϕ_i. The projective F-measurement of the state ψ transforms the pure state density matrix $\rho(\psi)$ (represented in the ON basis U of F-eigenvectors) to yield the Lüders mixture density matrix $\rho'(\psi) = \sum_{i=1}^I P_{\phi_i}\rho(\psi)P_{\phi_i}$ [2, p. 279]. The off-diagonal elements of $\rho(\psi)$ that are zeroed in $\rho'(\psi)$

[6] It is also called by the misnomer "linear entropy" [5] even though it is obviously quadratic in ρ–so we will not continue that usage. The logical entropy is also the quadratic special case of the Tsallis-Havrda-Charvat entropy ([19], [36]), and the logical special case [10] of C. R. Rao's quadratic entropy [29].

are the coherences (quantum indistinctions or *quindits*) that are turned into 'decoherences' (quantum distinctions or qudits of the observable being measured).

For any observable F and a pure state ψ, a quantum logical entropy was defined as $h(F:\psi) = \text{tr}\left[P_{[qudit(F)]}\rho(\psi)\otimes\rho(\psi)\right]$. That definition was the quantum generalization of the 'classical' logical entropy defined as $h(\pi) = p \times p(\text{dit}(\pi))$. When a projective F-measurement is performed on ψ, the pure state density matrix $\rho(\psi)$ is transformed into the mixed state density matrix by the quantum Lüders mixture operation which then defines the quantum logical entropy $h(\rho'(\psi)) = 1 - \text{tr}\left[\rho'(\psi)^2\right]$. The first test of how the quantum logical entropy notions fit together is showing that these two entropies are the same: $h(F:\psi) = h(\rho'(\psi))$. The proof shows that they are both equal to classical logical entropy of the partition $\pi(F:\psi)$ defined on the ON basis $U = \{u_1,...,u_n\}$ of F-eigenvectors by the F-eigenvalues with the point probabilties $p_j = \alpha_j^*\alpha_j$. That is, the inverse-images $B_i = \phi^{-1}(\phi_i)$ of the eigenvalue function $\phi : U \to \mathbb{R}$ define the eigenvalue partition $\pi(F:\psi) = \{B_1,...,B_I\}$ on the ON basis $U = \{u_1,...,u_n\}$ with the point probabilities $p_j = \alpha_j^*\alpha_j$ provided by the state $|\psi\rangle$ for $j = 1,...,n$. The classical logical entropy of that partition is: $h(\pi(F:\psi)) = 1 - \sum_{i=1}^{I} p(B_i)^2$ where $p(B_i) = \sum_{u_j\in B_i} p_j$.

We first show that $h(F:\psi) = \text{tr}\left[P_{[qudit(F)]}\rho(\psi)\otimes\rho(\psi)\right] = h(\pi(F:\psi))$. Now $qudit(F) = \{u_j\otimes u_k : \phi(u_j)\neq\phi(u_k)\}$ and $[qudit(F)]$ is the subspace of $V\otimes V$ generated by it. The $n\times n$ pure state density matrix $\rho(\psi)$ has the entries $\rho_{jk}(\psi) = \alpha_j\alpha_k^*$, and $\rho(\psi)\otimes\rho(\psi)$ is an $n^2\times n^2$ matrix. The projection matrix $P_{[qudit(F)]}$ is an $n^2\times n^2$ diagonal matrix with the diagonal entries, indexed by $j,k = 1,...,n$: $\left[P_{[qudit(F)]}\right]_{jjkk} = 1$ if $\phi(u_j)\neq\phi(u_k)$ and 0 otherwise. Thus in the product $P_{[qudit(F)]}\rho(\psi)\otimes\rho(\psi)$, the nonzero diagonal elements are the p_jp_k where $\phi(u_j)\neq\phi(u_k)$ and so the trace is $\sum_{j.k=1}^{n}\{p_jp_k : \phi(u_j)\neq\phi(u_k)\}$ which, by definition, is $h(F:\psi)$. Since $\sum_{j=1}^{n} p_j = \sum_{i=1}^{I} p(B_i) = 1$, $\left(\sum_{i=1}^{I} p(B_i)\right)^2 = 1 = \sum_{i=1}^{I} p(B_i)^2 + \sum_{i\neq i'} p(B_i)p(B_{i'})$. By grouping the p_jp_k in the trace according to the blocks of $\pi(F:\psi)$, we have:

$$\begin{aligned} h(F:\psi) = \text{tr}\left[P_{[qudit(F)]}\rho(\psi)\otimes\rho(\psi)\right] &= \sum_{j.k=1}^{n}\{p_jp_k : \phi(u_j)\neq\phi(u_k)\} \\ &= \sum_{i\neq i'}\sum\{p_jp_k : u_j\in B_i, u_k\in B_{i'}\} = \sum_{i\neq i'} p(B_i)p(B_{i'}) \\ &= 1 - \sum_{i=1}^{I} p(B_i)^2 = h(\pi(F:\psi)). \end{aligned}$$

To show that $h(\rho'(\psi)) = 1 - \text{tr}\left[\rho'(\psi)^2\right] = h(\pi(F:\psi))$ for $\rho'(\psi) = \sum_{i=1}^{I} P_{\phi_i}\rho(\psi)P_{\phi_i}$, we need to compute $\text{tr}\left[\rho'(\psi)^2\right]$. An off-diagonal element in $\rho_{jk}(\psi) = \alpha_j\alpha_k^*$ of $\rho(\psi)$ survives (i.e., is not zeroed and has the same value) the Lüders operation if and only if $\phi(u_j) = \phi(u_k)$. Hence the j^{th} diagonal element of $\rho'(\psi)^2$ is

$$\sum_{k=1}^{n}\left\{\alpha_j^*\alpha_k\alpha_j\alpha_k^* : \phi(u_j) = \phi(u_k)\right\} = \sum_{k=1}^{n}\{p_jp_k : \phi(u_j) = \phi(u_k)\} = p_jp(B_i)$$

where $u_j\in B_i$. Then grouping the j^{th} diagonal elements for $u_j\in B_i$ gives $\sum_{u_j\in B_i} p_jp(B_i) = p(B_i)^2$. Hence the whole trace is: $\text{tr}\left[\rho'(\psi)^2\right] = \sum_{i=1}^{I} p(B_i)^2$ and thus:

$$h(\rho'(\psi)) = 1 - \text{tr}\left[\rho'(\psi)^2\right] = 1 - \sum_{i=1}^{I} p(B_i)^2 = h(F:\psi).$$

This completes the proof of the following theorem.

Theorem 2. $h(F:\psi) = h(\pi(F:\psi)) = h(\rho'(\psi))$.

Measurement creates distinctions, i.e., turns coherences into 'decoherences'–which, classically, is the operation of distinguishing elements by classifying them according to some attribute like

classifying the faces of a die by their parity. The fundamental theorem about quantum logical entropy and projective measurement shows how the quantum logical entropy created (starting with $h(\rho(\psi)) = 0$ for the pure state ψ) by the measurement can be computed directly from the coherences of $\rho(\psi)$ that are decohered in $\rho'(\psi)$.

Theorem 3 (Fundamental (Quantum)). *The increase in quantum logical entropy, $h(F:\psi) = h(\rho'(\psi))$, due to the F-measurement of the pure state ψ is the sum of the absolute squares of the nonzero off-diagonal terms [coherences] in $\rho(\psi)$ (represented in an ON basis of F-eigenvectors) that are zeroed ['decohered'] in the post-measurement Lüders mixture density matrix $\rho'(\psi) = \sum_{i=1}^{I} P_{\phi_i}\rho(\psi) P_{\phi_i}$.*

Proof. $h(\rho'(\psi)) - h(\rho(\psi)) = \left(1 - \text{tr}\left[\rho'(\psi)^2\right]\right) - \left(1 - \text{tr}\left[\rho(\psi)^2\right]\right) = \sum_{jk}\left(\left|\rho_{jk}(\psi)\right|^2 - \left|\rho'_{jk}(\psi)\right|^2\right)$. If u_j and u_k are a qudit of F, then and only then are the corresponding off-diagonal terms zeroed by the Lüders mixture operation $\sum_{i=1}^{I} P_{\phi_i}\rho(\psi) P_{\phi_i}$ to obtain $\rho'(\psi)$ from $\rho(\psi)$. □

Density matrices have long been a standard part of the machinery of quantum mechanics. The Fundamental Theorem for logical entropy and measurement shows there is a simple, direct, and quantitative connection between density matrices and logical entropy. The Theorem directly connects the changes in the density matrix due to a measurement (sum of absolute squares of zeroed off-diagonal terms) with the increase in logical entropy due to the F-measurement $h(F:\psi) = \text{tr}\left[P_{[qudit(F)]}\rho(\psi) \otimes \rho(\psi)\right] = h(\rho'(\psi))$ (where $h(\rho(\psi)) = 0$ for the pure state ψ).

This direct quantitative connection between state discrimination and quantum logical entropy reinforces the judgment of Boaz Tamir and Eli Cohen ([34], [35]) that quantum logical entropy is a natural and informative entropy concept for quantum mechanics.

> We find this framework of partitions and distinction most suitable (at least conceptually) for describing the problems of quantum state discrimination, quantum cryptography and in general, for discussing quantum channel capacity. In these problems, we are basically interested in a distance measure between such sets of states, and this is exactly the kind of knowledge provided by logical entropy [Reference to [10]]. [34, p. 1]

Moreover, the quantum logical entropy has a simple "two-draw probability" interpretation, i.e., $h(F:\psi) = h(\rho'(\psi))$ is the probability that two independent F-measurements of ψ will yield distinct F-eigenvalues, i.e., will yield a qudit of F. In contrast, the Von Neumann entropy has no such simple interpretation and there seems to be no such intuitive connection between pre- and post-measurement density matrices and Von Neumann entropy–although Von Neumann entropy also increases in a projective measurement [28, Theorem 11.9, p. 515].

11. Quantum Logical Information Theory: Non-commuting Observables

11.1. Classical logical information theory with two sets X and Y

The usual ('classical') logical information theory for a probability distribution $\{p(x,y)\}$ on $X \times Y$ (finite) in effect uses the discrete partition on X and Y [13]. For the general case of quantum logical entropy for not-necessarily commuting observables, we need to first briefly develop the classical case with general partitions on X and Y.

Given two finite sets X and Y and real-valued functions $f: X \to \mathbb{R}$ with values $\{\phi_i\}_{i=1}^{I}$ and $g: Y \to \mathbb{R}$ with values $\{\gamma_j\}_{j=1}^{J}$, each function induces a partition on its domain:

$$\pi = \{f^{-1}(\phi_i)\}_{i\in I} = \{B_1, ..., B_I\} \text{ on } X, \text{ and } \sigma = \{g^{-1}(\gamma_j)\}_{j\in J} = \{C_1, ..., C_J\} \text{ on } Y.$$

We need to define logical entropies on $X \times Y$ but first we need to define the ditsets or information sets.

A partition $\pi = \{B_1, ..., B_I\}$ on X and a partition $\sigma = \{C_1, ..., C_J\}$ on Y define a *product partition* $\pi \times \sigma$ on $X \times Y$ whose blocks are $\{B_i \times C_j\}_{i,j}$. Then π induces $\pi \times \mathbf{0}_Y$ on $X \times Y$ (where $\mathbf{0}_Y$ is the indiscrete partition on Y) and σ induces $\mathbf{0}_X \times \sigma$ on $X \times Y$. The corresponding ditsets or information sets are:

- $\operatorname{dit}(\pi \times \mathbf{0}_Y) = \{((x,y),(x',y')) : f(x) \neq f(x')\} \subseteq (X \times Y)^2$;
- $\operatorname{dit}(\mathbf{0}_X \times \sigma) = \{((x,y),(x',y')) : g(y) \neq g(y')\} \subseteq (X \times Y)^2$;
- $\operatorname{dit}(\pi \times \sigma) = \operatorname{dit}(\pi \times \mathbf{0}_Y) \cup \operatorname{dit}(\mathbf{0}_X \times \sigma)$; and so forth.

Given a joint probability distribution $p : X \times Y \to [0,1]$, the product probability distribution is $p \times p : (X \times Y)^2 \to [0,1]$.

All the logical entropies are just the product probabilities of the ditsets and their union, differences, and intersection:

- $h(\pi \times \mathbf{0}_Y) = p \times p(\operatorname{dit}(\pi \times \mathbf{0}_Y))$;
- $h(\mathbf{0}_X \times \sigma) = p \times p(\operatorname{dit}(\mathbf{0}_X \times \sigma))$;
- $h(\pi \times \sigma) = p \times p(\operatorname{dit}(\pi \times \sigma)) = p \times p(\operatorname{dit}(\pi \times \mathbf{0}_Y) \cup \operatorname{dit}(\mathbf{0}_X \times \sigma))$;
- $h(\pi \times \mathbf{0}_Y | \mathbf{0}_X \times \sigma) = p \times p(\operatorname{dit}(\pi \times \mathbf{0}_Y) - \operatorname{dit}(\mathbf{0}_X \times \sigma))$;
- $h(\mathbf{0}_X \times \sigma | \pi \times \mathbf{0}_Y) = p \times p(\operatorname{dit}(\mathbf{0}_X \times \sigma) - \operatorname{dit}(\pi \times \mathbf{0}_Y))$;
- $m(\pi \times \mathbf{0}_Y, \mathbf{0}_X \times \sigma) = p \times p(\operatorname{dit}(\pi \times \mathbf{0}_Y) \cap \operatorname{dit}(\mathbf{0}_X \times \sigma))$.

All the logical entropies have the usual two-draw probability interpretation where the two independent draws from $X \times Y$ are (x,y) and (x',y') and can be interpreted in terms of the f-values and g-values:

- $h(\pi \times \mathbf{0}_Y)$ = probability of getting distinct f-values;
- $h(\mathbf{0}_X \times \sigma)$ = probability of getting distinct g-values;
- $h(\pi \times \sigma)$ = probability of getting distinct f or g values;
- $h(\pi \times \mathbf{0}_Y | \mathbf{0}_X \times \sigma)$ = probability of getting distinct f-values but same g-values;
- $h(\mathbf{0}_X \times \sigma | \pi \times \mathbf{0}_Y)$ = probability of getting distinct g-values but same f-values;
- $m(\pi \times \mathbf{0}_Y, \mathbf{0}_X \times \sigma)$ = probability of getting distinct f and distinct g values.

We have defined all the logical entropies by the general method of the product probabilities on the ditsets. In the first three cases, $h(\pi \times \mathbf{0}_Y)$, $h(\mathbf{0}_X \times \sigma)$, and $h(\pi \times \sigma)$, they were the logical entropies of partitions on $X \times Y$ so they could equivalently be defined using density matrices. The case of $h(\pi \times \sigma)$ illustrates the general case. If $\rho(\pi)$ is the density matrix defined for π on X and $\rho(\sigma)$ the density matrix for σ on Y, then $\rho(\pi \times \sigma) = \rho(\pi) \otimes \rho(\sigma)$ is the density matrix for $\pi \times \sigma$ defined on $X \times Y$, and:

$$h(\pi \times \sigma) = 1 - \operatorname{tr}\left[\rho(\pi \times \sigma)^2\right].$$

The marginal distributions are: $p_X(x) = \sum_y p(x,y)$ and $p_Y(y) = \sum_x p(x,y)$. Since π is a partition on X, there is also the usual logical entropy $h(\pi) = p_X \times p_X(\operatorname{dit}(\pi)) = 1 - \operatorname{tr}\left[\rho(\pi)^2\right] = h(\pi \times \mathbf{0}_Y)$ where $\operatorname{dit}(\pi) \subseteq X \times X$ and similarly for p_Y.

Since the context should be clear, we may henceforth adopt the old notation from the case where π and σ were partitions on the same set U, i.e., $h(\pi) = h(\pi \times \mathbf{0}_Y)$, $h(\sigma) = h(\mathbf{0}_X \times \sigma)$, $h(\pi, \sigma) = h(\pi \times \sigma)$, etc.

Since the logical entropies are the values of a probability measure, all the usual identities hold where the underlying set is now $(X \times Y)^2$ instead of U^2 as illustrated in Figure 4.

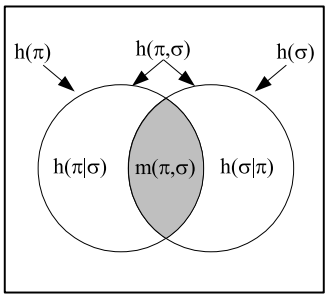

Figure 4: Venn diagram for logical entropies as values of a probability measure $p \times p$ on $(X \times Y)^2$.

The previous treatment of $h(X)$, $h(Y)$, $h(X,Y)$, $h(X|Y)$, $h(Y|X)$, and $m(X,Y)$ in [13] was just the special cases where $\pi = \mathbf{1}_X$ and $\sigma = \mathbf{1}_Y$.

11.2. Quantum logical entropies with non-commuting observables

As before in the case of commuting observables, the quantum case can be developed in close analogy with the previous classical case. Given a finite-dimensional Hilbert space V and not necessarily commuting observables $F, G : V \to V$, let X be an orthonormal basis of V of F-eigenvectors and let Y be an orthonormal basis for V of G-eigenvectors (so $|X| = |Y|$).

Let $f : X \to \mathbb{R}$ be the eigenvalue function for F with values $\{\phi_i\}_{i=1}^{I}$, and let $g : Y \to \mathbb{R}$ be the eigenvalue function for G with values $\{\gamma_j\}_{j=1}^{J}$.

Each eigenvalue function induces a partition on its domain:

$$\pi = \{f^{-1}(\phi_i)\} = \{B_1, ..., B_I\} \text{ on } X, \text{ and } \sigma = \{g^{-1}(\gamma_j)\} = \{C_1, ..., C_J\} \text{ on } Y.$$

We associated with the ordered pair (x,y), the basis element $x \otimes y$ in the basis $\{x \otimes y\}_{x \in X, y \in Y}$ for $V \otimes V$. Then each pair of pairs $((x,y),(x',y'))$ is associated with the basis element $(x \otimes y) \otimes (x' \otimes y')$ in $(V \otimes V) \otimes (V \otimes V) = (V \otimes V)^2$.

Instead of ditsets or information sets, we now have qudit subspaces or information subspaces. For $R \subseteq (V \otimes V)^2$, let $[R]$ be the subspace generated by R. We simplify notation of $qudit(\pi \times \mathbf{0}_Y) = qudit(\pi) = \{(x \otimes y) \otimes (x' \otimes y') : f(x) \neq f(x')\}$, etc.

- $[qudit(\pi)] = [\{(x \otimes y) \otimes (x' \otimes y') : f(x) \neq f(x')\}]$;
- $[qudit(\sigma)] = [\{(x \otimes y) \otimes (x' \otimes y') : g(y) \neq g(y')\}]$;
- $[qudit(\pi,\sigma)] = [qudit(\pi) \cup qudit(\sigma)]$, and so forth.[7]

A normalized state $|\psi\rangle$ on $V \otimes V$ defines a pure state density matrix $\rho(\psi) = |\psi\rangle\langle\psi|$. Let $\alpha_{x,y} = \langle x \otimes y | \psi \rangle$ so if $P_{[x \otimes y]}$ is the projection to the subspace (ray) generated by $x \otimes y$ in $V \otimes V$, then a probability distribution on $X \times Y$ is defined by:

$$p(x,y) = \alpha_{x,y}\alpha^*_{x,y} = \text{tr}\left[P_{[x \otimes y]}\rho(\psi)\right],$$

or more generally, for a subspace $T \subseteq V \otimes V$, a probability distribution is defined on the subspaces by:

$$\Pr(T) = \text{tr}[P_T \rho(\psi)].$$

Then the product probability distribution $p \times p$ on the subspaces of $(V \otimes V)^2$ defines the quantum logical entropies when applied to the information subspaces:

---

[7] It is again an easy implication of the aforementioned Common Dits Theorem that any two nonzero information spaces $[qudit(\pi)]$ and $[qudit(\sigma)]$ have a nonzero intersection so the mutual information space $[qudit(\pi) \cap qudit(\sigma)]$ is not the zero space.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 August 2018 doi:10.20944/preprints201808.0208.v1

Peer-reviewed version available at Entropy 2018, 20, 679; doi:10.3390/e20090679

- $h(F:\psi) = p \times p([qudit(\pi)]) = \text{tr}\left[P_{[qudit(\pi)]}(\rho(\psi) \otimes \rho(\psi))\right]$;
- $h(G:\psi) = p \times p([qudit(\sigma)]) = \text{tr}\left[P_{[qudit(\sigma)]}(\rho(\psi) \otimes \rho(\psi))\right]$;
- $h(F,G:\psi) = p \times p([qudit(\pi) \cup qudit(\sigma)]) = \text{tr}\left[P_{[qudit(\pi) \cup qudit(\sigma)]}(\rho(\psi) \otimes \rho(\psi))\right]$;
- $h(F|G:\psi) = p \times p([qudit(\pi) - qudit(\sigma)]) = \text{tr}\left[P_{[qudit(\pi) - qudit(\sigma)]}(\rho(\psi) \otimes \rho(\psi))\right]$;
- $h(G|F:\psi) = p \times p([qudit(\sigma) - qudit(\pi)]) = \text{tr}\left[P_{[qudit(\sigma) - qudit(\pi)]}(\rho(\psi) \otimes \rho(\psi))\right]$;
- $m(F,G:\psi) = p \times p([qudit(\pi) \cap qudit(\sigma)]) = \text{tr}\left[P_{[qudit(\pi) \cap qudit(\sigma)]}(\rho(\psi) \otimes \rho(\psi))\right]$.

The observable $F : V \to V$ defines an observable $F \otimes I : V \otimes V \to V \otimes V$ with the eigenvectors $x \otimes v$ for any nonzero $v \in V$ and with the same eigenvalues $\phi_1, ..., \phi_I$.[8] Then in two independent measurements of ψ by the observable $F \otimes I$, we have:

$h(F:\psi)$ = probability of getting distinct eigenvalues ϕ_i and $\phi_{i'}$, i.e., of getting a qudit of F.

In a similar manner, $G : V \to V$ defines the observable $I \otimes G : V \otimes V \to V \otimes V$ with the eigenvectors $v \otimes y$ and with the same eigenvalues $\gamma_1, ..., \gamma_J$. Then in two independent measurements of ψ by the observable $I \otimes G$, we have:

$h(G:\psi)$ = probability of getting distinct eigenvalues γ_j and $\gamma_{j'}$.

The two observables $F, G : V \to V$ define an observable $F \otimes G : V \otimes V \to V \otimes V$ with the eigenvectors $x \otimes y$ for $(x,y) \in X \times Y$ and eigenvalues $f(x)g(y) = \phi_i \gamma_j$. To cleanly interpret the compound logical entropies, we assume there is no accidental degeneracy so there are no $\phi_i \gamma_j = \phi_{i'} \gamma_{j'}$ for $i \neq i'$ and $j \neq j'$. Then for two independent measurements of ψ by $F \otimes G$, the compound quantum logical entropies can be interpreted as the following "two-measurement" probabilities:

- $h(F,G:\psi)$ = probability of getting distinct eigenvalues $\phi_i \gamma_j \neq \phi_{i'} \gamma_{j'}$ where $i \neq i'$ or $j \neq j'$;
- $h(F|G:\psi)$ = probability of getting distinct eigenvalues $\phi_i \gamma_j \neq \phi_{i'} \gamma_j$ where $i \neq i'$;
- $h(G|F:\psi)$ = probability of getting distinct eigenvalues $\phi_i \gamma_j \neq \phi_i \gamma_{j'}$ where $j \neq j'$;
- $m(F,G:\psi)$ = probability of getting distinct eigenvalues $\phi_i \gamma_j \neq \phi_{i'} \gamma_{j'}$ where $i \neq i'$ and $j \neq j'$.

All the quantum logical entropies have been defined by the general method using the information subspaces, but in the first three cases $h(F:\psi)$, $h(G:\psi)$, and $h(F,G:\psi)$, the density matrix method of defining logical entropies could also be used. Then the fundamental theorem could be applied relating the quantum logical entropies to the zeroed entities in the density matrices indicating the eigenstates distinguished by the measurements.

The previous set identities for disjoint unions now become subspace identities for direct sums such as:

$$[qudit(\pi) \cup qudit(\sigma)] = [qudit(\pi) - qudit(\sigma)] \oplus [qudit(\pi) \cap qudit(\sigma)] \oplus [qudit(\sigma) - qudit(\pi)].$$

Hence the probabilities are additive on those subspaces as shown in Figure 5:

$$h(F,G:\psi) = h(F|G:\psi) + m(F,G:\psi) + h(G|F:\psi).$$

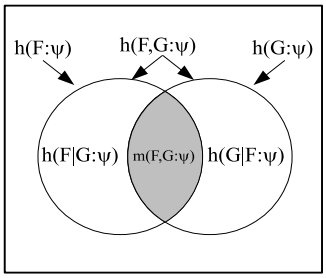

Figure 5: Venn diagram for quantum logical entropies as probabilities on $(V \otimes V)^2$.

[8] The context should suffice to distinguish the identity operator $I : V \to V$ from the index set I for the F-eigenvalues.

11.3. Quantum logical entropies of density operators in general

The extension of the classical logical entropy $h(p) = 1 - \sum_{i=1}^{n} p_i^2$ of a probability distribution $p = (p_1, ..., p_n)$ to the quantum case is $h(\rho) = 1 - \text{tr}\left[\rho^2\right]$ where a density matrix ρ replaces the probability distribution p and the trace replaces the summation. In the previous section, quantum logical entropies were defined in terms of given observables $F, G : V \to V$ (as self-adjoint operators) as well as a state ψ and its density matrix $\rho(\psi)$. An arbitrary density operator ρ, representing a pure or mixed state on V, is also a self-adjoint operator on V so quantum logical entropies can be defined where density operators play the double role of providing the measurement basis (as self-adjoint operators) as well as the state being measured.

Let ρ and τ be two non-commuting density operators on V. Let $X = \{u_i\}_{i=1,...,n}$ be an orthonormal (ON) basis of ρ eigenvectors and let $\{\lambda_i\}_{i=1,...,n}$ be the corresponding eigenvalues which must be non-negative and sum to 1 so they can be interpreted as probabilities. Let $Y = \{v_j\}_{j=1,...,n}$ be an ON basis of eigenvectors for τ and let $\{\mu_j\}_{j=1,...,n}$ be the corresponding eigenvalues which are also non-negative and sum to 1.

Each density operator plays a double role. For instance, ρ acts as the observable to supply the measurement basis of $\{u_i\}_i$ and the eigenvalues $\{\lambda_i\}_i$ as well as being the state to be measured supplying the probabilities $\{\lambda_i\}_i$ for the measurement outcomes. Hence we could define quantum logical entropies as in the previous section. That analysis would analyze the distinctions between probabilities $\lambda_i \neq \lambda_{i'}$ since they are the eigenvalues too. But that analysis would not give the quantum analogue of $h(p) = 1 - \sum_i p_i^2$ which in effect uses the discrete partition on the index set $\{1, ..., n\}$ and pays no attention to when the probabilities of different indices are the same or different. Hence we will now develop the analysis as in the last section but by using the discrete partition $\mathbf{1}_X$ on the set of 'index' states $X = \{u_i\}_i$ and similarly for the discrete partition $\mathbf{1}_Y$ on $Y = \{v_j\}_{j'}$ the ON basis of eigenvectors for τ.

The qudit sets of $(V \otimes V) \otimes (V \otimes V)$ are then defined according to the identity and difference on the index sets and independent of the eigenvalue-probabilities, e.g., $qudit(\mathbf{1}_X) = \left\{(u_i \otimes v_j) \otimes \left(u_{i'} \otimes v_{j'}\right) : i \neq i'\right\}$. Then the qudit subspaces are the subspaces of $(V \otimes V)^2$ generated by the qudit sets of generators:

- $[qudit(\mathbf{1}_X)] = \left[(u_i \otimes v_j) \otimes \left(u_{i'} \otimes v_{j'}\right) : i \neq i'\right]$;
- $[qudit(\mathbf{1}_Y)] = \left[(u_i \otimes v_j) \otimes \left(u_{i'} \otimes v_{j'}\right) : j \neq j'\right]$;
- $[qudit(\mathbf{1}_X, \mathbf{1}_Y)] = [qudit(\mathbf{1}_X) \cup qudit(\mathbf{1}_Y)] = \left[(u_i \otimes v_j) \otimes \left(u_{i'} \otimes v_{j'}\right) : i \neq i' \text{ or } j \neq j'\right]$;
- $[qudit(\mathbf{1}_X|\mathbf{1}_Y)] = [qudit(\mathbf{1}_X) - qudit(\mathbf{1}_Y)] = \left[(u_i \otimes v_j) \otimes \left(u_{i'} \otimes v_{j'}\right) : i \neq i' \text{ and } j = j'\right]$;
- $[qudit(\mathbf{1}_Y|\mathbf{1}_X)] = [qudit(\mathbf{1}_Y) - qudit(\mathbf{1}_X)] = \left[(u_i \otimes v_j) \otimes \left(u_{i'} \otimes v_{j'}\right) : i = i' \text{ and } j \neq j'\right]$; and
- $[qudit(\mathbf{1}_Y \& \mathbf{1}_X)] = [qudit(\mathbf{1}_Y) \cap qudit(\mathbf{1}_X)] = \left[(u_i \otimes v_j) \otimes \left(u_{i'} \otimes v_{j'}\right) : i \neq i' \text{ and } j \neq j'\right]$.

Then as qudit sets: $qudit(\mathbf{1}_X, \mathbf{1}_Y) = qudit(\mathbf{1}_X|\mathbf{1}_Y) \uplus qudit(\mathbf{1}_Y|\mathbf{1}_X) \uplus qudit(\mathbf{1}_Y \& \mathbf{1}_X)$, and the corresponding qudit subspaces stand in the same relation where the disjoint union is replaced by the disjoint sum.

The density operator ρ is represented by the diagonal density matrix ρ_X in its own ON basis X with $(\rho_X)_{ii} = \lambda_i$ and similarly for the diagonal density matrix τ_Y with $(\tau_Y)_{jj} = \mu_j$. The density operators ρ, τ on V define a density operator $\rho \otimes \tau$ on $V \otimes V$ with the ON basis of eigenvectors $\{u_i \otimes v_j\}_{i,j}$ and the eigenvalue-probabilities of $\{\lambda_i \mu_j\}_{i,j}$. The operator $\rho \otimes \tau$ is represented in its ON basis by the diagonal density matrix $\rho_X \otimes \tau_Y$ with diagonal entries $\lambda_i \mu_j$ where $1 = (\lambda_1 + ... + \lambda_n)(\mu_1 + ... + \mu_n) = \sum_{i,j=1}^{n} \lambda_i \mu_j$. The probability measure $p(u_i \otimes v_j) = \lambda_i \mu_j$ on $V \otimes V$ defines the product measure $p \times p$ on $(V \otimes V)^2$ where it can be applied to the qudit subspaces to define the quantum logical entropies as usual.

In the first instance, we have:

$$h(\mathbf{1}_X : \rho \otimes \tau) = p \times p([qudit(\mathbf{1}_X)]) = \sum \left\{ \lambda_i \mu_j \lambda_{i'} \mu_{j'} : i \neq i' \right\}$$
$$= \sum_{i \neq i'} \lambda_i \lambda_{i'} \sum_{j,j'} \mu_j \mu_{j'} = \sum_{i \neq i'} \lambda_i \lambda_{i'} = 1 - \sum_i \lambda_i^2 = 1 - \text{tr}\left[\rho^2\right] = h(\rho)$$

and similarly $h(\mathbf{1}_Y : \rho \otimes \tau) = h(\tau)$. Since all the data is supplied by the two density operators, we can use simplified notation to define the corresponding joint, conditional, and mutual entropies:

- $h(\rho, \tau) = h(\mathbf{1}_X, \mathbf{1}_Y : \rho \otimes \tau) = p \times p([qudit(\mathbf{1}_X) \cup qudit(\mathbf{1}_Y)])$;
- $h(\rho|\tau) = h(\mathbf{1}_X|\mathbf{1}_Y : \rho \otimes \tau) = p \times p([qudit(\mathbf{1}_X) - qudit(\mathbf{1}_Y)])$;
- $h(\tau|\rho) = h(\mathbf{1}_Y|\mathbf{1}_X : \rho \otimes \tau) = p \times p([qudit(\mathbf{1}_Y) - qudit(\mathbf{1}_X)])$; and
- $m(\rho, \tau) = h(\mathbf{1}_Y \& \mathbf{1}_X : \rho \otimes \tau) = p \times p([qudit(\mathbf{1}_Y) \cap qudit(\mathbf{1}_X)])$.

Then the usual Venn diagram relationships hold for the probability measure $p \times p$ on $(V \otimes V)^2$, e.g.,

$$h(\rho, \tau) = h(\rho|\tau) + h(\tau|\rho) + m(\rho, \tau),$$

and probability interpretations are readily available. There are two probability distributions $\lambda = \{\lambda_i\}_i$ and $\mu = \{\mu_j\}_j$ on the sample space $\{1, ..., n\}$. Two pairs (i, j) and (i', j') are drawn with replacement, the first entry in each pair is drawn according to λ and the second entry according to μ. Then $h(\rho, \tau)$ is the probability that $i \neq i'$ or $j \neq j'$ (or both); $h(\rho|\tau)$ is the probability that $i \neq i'$ and $j = j'$; and so forth. Note that this interpretation assumes no special significance to a λ_i and μ_i having the same index since we are drawing a pair of pairs.

In the classical case of two probability distributions $p = (p_1, ..., p_n)$ and $q = (q_1, ..., q_n)$ on the same index set, the *logical cross-entropy* is defined as: $h(p||q) = 1 - \sum_i p_i q_i$, and interpreted as the probability of getting different indices in drawing a single pair, one from p and the other from q. However, this cross-entropy assumes some special significance to p_i and q_i having the same index. But in our current quantum setting, there is no correlation between the two sets of 'index' states $\{u_i\}_{i=1,...,n}$ and $\{v_j\}_{j=1,...,n}$. But when the two density operators commute, $\tau\rho = \rho\tau$, then we can take $\{u_i\}_{i=1,...,n}$ as an ON basis of simultaneous eigenvectors for the two operators with respective eigenvalues λ_i and μ_i for u_i with $i = 1, ..., n$. In that special case, we can meaningfully define the *quantum logical cross-entropy* as $h(\rho||\tau) = 1 - \sum_{i=1}^n \lambda_i \mu_i$, but the general case awaits further analysis below.

12. The logical Hamming distance between two partitions

The development of logical quantum information theory in terms of some given commuting or non-commuting observables gives an analysis of the distinguishability of quantum states using those observables. Without any given observables, there is still a natural logical analysis of the distance between quantum states that generalizes the 'classical' logical distance $h(\pi|\sigma) + h(\sigma|\pi)$ between partitions on a set. In the classical case, we have the logical entropy $h(\pi)$ of a partition where the partition plays the role of the direct-sum decomposition of eigenspaces of an observable in the quantum case. But we also have just the logical entropy $h(p)$ of a probability distribution $p = (p_1, ..., p_n)$ and the related compound notions of logical entropy given another probability distribution $q = (q_1, ..., q_n)$ indexed by the same set.

First we review that classical treatment to motivate the quantum version of the logical distance between states. A binary relation $R \subseteq U \times U$ on $U = \{u_1, ..., u_n\}$ can be represented by an $n \times n$ *incidence matrix* $I(R)$ where

$$I(R)_{ij} = \begin{cases} 1 \text{ if } (u_i, u_j) \in R \\ 0 \text{ if } (u_i, u_j) \notin R. \end{cases}$$

Taking R as the equivalence relation indit(π) associated with a partition $\pi = \{B_1, ..., B_I\}$, the *density matrix* $\rho(\pi)$ *of the partition* π (with equiprobable points) is just the incidence matrix $I(\text{indit}(\pi))$ rescaled to be of trace 1 (i.e., sum of diagonal entries is 1):

$$\rho(\pi) = \frac{1}{|U|} I(\text{indit}(\pi)).$$

From coding theory [27, p. 66], we have the notion of the *Hamming distance between two* 0,1 *vectors or matrices* (of the same dimensions) which is the number of places where they differ. Since logical information theory is about distinctions and differences, it is important to have a classical and quantum logical notion of Hamming distance. The powerset $\wp(U \times U)$ can be viewed as a vector space over $\mathbb{Z}_2$ where the sum of two binary relations $R, R' \subseteq U \times U$ is the symmetric difference, symbolized $R \Delta R' = (R - R') \cup (R' - R) = R \cup R' - R \cap R'$, which is the set of elements (i.e., ordered pairs $(u_i, u_j) \in U \times U$) that are in one set or the other but not both. Thus the Hamming distance $D_H(I(R), I(R'))$ between the incidence matrices of two binary relations is just the cardinality of their symmetric difference: $D_H(I(R), I(R')) = |R \Delta R'|$. Moreover, the size of the symmetric difference does not change if the binary relations are replaced by their complements: $|R \Delta R'| = \left|(U^2 - R) \Delta (U^2 - R')\right|$.

Hence given two partitions $\pi = \{B_1, ..., B_I\}$ and $\sigma = \{C_1, ..., C_J\}$ on U, the unnormalized Hamming distance between the two partitions is naturally defined as:[9]

$$D(\pi, \sigma) = D_H(I(\text{indit}(\pi)), I(\text{indit}(\sigma))) = |\text{indit}(\pi) \Delta \text{indit}(\sigma)| = |\text{dit}(\pi) \Delta \text{dit}(\sigma)|,$$

and the *Hamming distance between* π and σ is defined as the normalized $D(\pi, \sigma)$:

$$d(\pi, \sigma) = \frac{D(\pi,\sigma)}{|U \times U|} = \frac{|\text{dit}(\pi) \Delta \text{dit}(\sigma)|}{|U \times U|} = \frac{|\text{dit}(\pi) - \text{dit}(\sigma)|}{|U \times U|} + \frac{|\text{dit}(\sigma) - \text{dit}(\pi)|}{|U \times U|} = h(\pi|\sigma) + h(\sigma|\pi).$$

This motivates the general case of point probabilities $p = (p_1, ..., p_n)$ where we define the *Hamming distance* between the two partitions as the sum of the two logical conditional entropies:

$$d(\pi, \sigma) = h(\pi|\sigma) + h(\sigma|\pi) = 2h(\pi \vee \sigma) - h(\pi) - h(\sigma).$$

To motivate the bridge to the quantum version of the Hamming distance, we need to calculate it using the density matrices $\rho(\pi)$ and $\rho(\sigma)$ of the two partitions. To compute the trace $\text{tr}[\rho(\pi)\rho(\sigma)]$, we compute the diagonal elements in the product $\rho(\pi)\rho(\sigma)$ and add them up: $[\rho(\pi)\rho(\sigma)]_{kk} = \sum_l \rho(\pi)_{kl} \rho(\sigma)_{lk} = \sum_l \sqrt{p_k p_l} \sqrt{p_l p_k}$ where the only nonzero terms are where $u_k, u_l \in B \cap C$ for some $B \in \pi$ and $C \in \sigma$. Thus if $u_k \in B \cap C$, then $[\rho(\pi)\rho(\sigma)]_{kk} = \sum_{u_l \in B \cap C} p_k p_l$. So the diagonal element for u_k is the sum of the $p_k p_l$ for u_l in the same intersection $B \cap C$ as u_k so it is $p_k \Pr(B \cap C)$. Then when we sum over the diagonal elements, then for all the $u_k \in B \cap C$ for any given B, C, we just sum $\sum_{u_k \in B \cap C} p_k \Pr(B \cap C) = \Pr(B \cap C)^2$ so that $\text{tr}[\rho(\pi)\rho(\sigma)] = \sum_{B \in \pi, C \in \sigma} \Pr(B \cap C)^2 = 1 - h(\pi \vee \sigma)$.

Hence if we define the *logical cross-entropy* of π and σ as:

$$h(\pi||\sigma) = 1 - \text{tr}[\rho(\pi)\rho(\sigma)],$$

then for partitions on U with the point probabilities $p = (p_1, ..., p_n)$, the logical cross-entropy $h(\pi||\sigma)$ of two partitions is the same as the logical joint entropy which is also the logical entropy of the join:

$$h(\pi||\sigma) = h(\pi, \sigma) = h(\pi \vee \sigma).$$

Thus we can also express the logical Hamming distance between two partitions entirely in terms of density matrices:

$$d(\pi, \sigma) = 2h(\pi||\sigma) - h(\pi) - h(\sigma) = \text{tr}\left[\rho(\pi)^2\right] + \text{tr}\left[\rho(\sigma)^2\right] - 2\,\text{tr}[\rho(\pi)\rho(\sigma)].$$

---

[9] This is investigated in Rossi [30].

13. The quantum logical Hamming distance

The quantum logical entropy $h(\rho) = 1 - \text{tr}\left[\rho^2\right]$ of a density matrix ρ generalizes the classical $h(p) = 1 - \sum_i p_i^2$ for a probability distribution $p = (p_1, \ldots, p_n)$. As a self-adjoint operator, a density matrix has a spectral decomposition $\rho = \sum_{i=1}^n \lambda_i |u_i\rangle \langle u_i|$ where $\{|u_i\rangle\}_{i=1,\ldots,n}$ is an orthonormal basis for V and where all the eigenvalues λ_i are real, non-negative, and $\sum_{i=1}^n \lambda_i = 1$. Then $h(\rho) = 1 - \sum_i \lambda_i^2$ so $h(\rho)$ can be interpreted as the probability of getting distinct indices $i \neq i'$ in two independent measurements of the state ρ with $\{|u_i\rangle\}$ as the measurement basis. Classically, it is the two-draw probability of getting distinct indices in two independent samples of the probability distribution $\lambda = (\lambda_1, \ldots, \lambda_n)$, just as $h(p)$ is the probability of getting distinct indices in two independent draws on p. For a pure state ρ, there is one $\lambda_i = 1$ with the others zero, and $h(\rho) = 0$ is the probability of getting distinct indices in two independent draws on $\lambda = (0, \ldots, 0, 1, 0, \ldots, 0)$.

In the classical case of the logical entropies, we worked with the ditsets or sets of distinctions of partitions. But everything could also be expressed in terms of the complementary sets of indits or indistinctions of partitions (ordered pairs of elements in the same block of the partition) since: $\text{dit}(\pi) \uplus \text{indit}(\pi) = U \times U$. When we switch to the density matrix treatment of 'classical' partitions, then the focus shifts to the indistinctions. For a partition $\pi = \{B_1, \ldots, B_I\}$, the logical entropy is the sum of the distinction probabilities: $h(\pi) = \sum_{(u_k,u_l) \in \text{dit}(\pi)} p_k p_l$ which in terms of indistinctions is:

$$h(\pi) = 1 - \sum_{(u_k,u_l) \in \text{indit}(\pi)} p_k p_l = 1 - \sum_{i=1}^I \Pr(B_i)^2.$$

When expressed in the density matrix formulation, then $\text{tr}\left[\rho(\pi)^2\right]$ is the sum of the indistinction probabilities:

$$\text{tr}\left[\rho(\pi)^2\right] = \sum_{(u_k,u_l) \in \text{indit}(\pi)} p_k p_l = \sum_{i=1}^I \Pr(B_i)^2.$$

The nonzero entries in $\rho(\pi)$ have the form $\sqrt{p_k p_l}$ for $(u_k, u_l) \in \text{indit}(\pi)$; their squares are the indistinction probabilities. That provides the interpretive bridge to the quantum case.

The quantum analogue of an indistinction probability is the absolute square $|\rho_{kl}|^2$ of a nonzero entry ρ_{kl} in a density matrix ρ and $\text{tr}\left[\rho^2\right] = \sum_{k,l} |\rho_{kl}|^2$ is the sum of those 'indistinction' probabilities. The nonzero entries in the density matrix ρ might be called "*coherences*" so that ρ_{kl} may be interpreted as the amplitudes for the states u_k and u_l to cohere together in the state ρ so $\text{tr}\left[\rho^2\right]$ is the sum of the *coherence probabilities*–just as $\text{tr}\left[\rho(\pi)^2\right] = \sum_{(u_k,u_l) \in \text{indit}(\pi)} p_k p_l$ is the sum of the indistinction probabilities. The quantum logical entropy $h(\rho) = 1 - \text{tr}\left[\rho^2\right]$ may then be interpreted as the sum of the *decoherence probabilities*–just as $h(\rho(\pi)) = h(\pi) = 1 - \sum_{(u_k,u_l) \in \text{indit}(\pi)} p_k p_l$ is the sum of the distinction probabilities.

This motivates the general quantum definition of the joint entropy $h(\pi, \sigma) = h(\pi \vee \sigma) = h(\pi||\sigma)$ which is the:

$$h(\rho||\tau) = 1 - \text{tr}\left[\rho^\dagger \tau\right]$$
quantum logical cross-entropy.

To work out its interpretation, we again take ON eigenvector bases $\{|u_i\rangle\}_{i=1}^n$ for ρ and $\{|v_j\rangle\}_{j=1}^n$ for τ with λ_i and μ_j as the respective eigenvalues, and compute the operation of $\tau^\dagger \rho : V \to V$. Now $|u_i\rangle = \sum_j \langle v_j|u_i\rangle |v_j\rangle$ so $\rho|u_i\rangle = \lambda_i |u_i\rangle = \sum_j \lambda_i \langle v_j|u_i\rangle |v_j\rangle$ and then for $\tau^\dagger = \sum_j |v_j\rangle \langle v_j| \mu_j$, so $\tau^\dagger \rho |u_i\rangle = \sum_j \lambda_i \mu_j \langle v_j|u_i\rangle |v_j\rangle$. Thus $\tau^\dagger \rho$ in the $\{u_i\}_i$ basis would have the diagonal entries $\langle u_i|\tau^\dagger \rho|u_i\rangle = \sum_j \lambda_i \mu_j \langle v_j|u_i\rangle \langle u_i|v_j\rangle$ so the trace is:

$$\text{tr}\left[\tau^\dagger \rho\right] = \sum_i \langle u_i|\tau^\dagger \rho|u_i\rangle = \sum_{i,j} \lambda_i \mu_j \langle v_j|u_i\rangle \langle u_i|v_j\rangle = \text{tr}\left[\rho^\dagger \tau\right]$$

which is symmetrical. The other information we have is the $\sum_i \lambda_i = 1 = \sum_j \mu_j$ and they are non-negative. The classical logical cross-entropy of two probability distributions is $h(p||q) =$

$1-\sum_{i,j} p_i q_j \delta_{ij}$ where two indices i and j are either identical or totally distinct. But in the quantum case, the 'index' states $|u_i\rangle$ and $|v_j\rangle$ have an 'overlap' measured by the inner product $\langle u_i|v_j\rangle$. The trace $\mathrm{tr}\left[\rho^\dagger\tau\right]$ is real since $\langle v_j|u_i\rangle = \langle u_i|v_j\rangle^*$ and $\langle v_j|u_i\rangle\langle u_i|v_j\rangle = |\langle u_i|v_j\rangle|^2 = |\langle v_j|u_i\rangle|^2$ is the probability of getting λ_i when measuring v_j in the u_i basis and the probability of getting μ_j when measuring u_i in the v_j basis. The twofold nature of density matrices as states and as observables then allows $\mathrm{tr}\left[\rho^\dagger\tau\right]$ to be interpreted as the average value of the observable ρ when measuring the state τ or vice-versa.

We may call $\langle v_j|u_i\rangle\langle u_i|v_j\rangle$ the *proportion* or *extent of overlap* for those two index states. Thus $\mathrm{tr}\left[\rho^\dagger\tau\right]$ is the sum of *all* the probability combinations $\lambda_i\mu_j$ weighted by the overlaps $\langle v_j|u_i\rangle\langle u_i|v_j\rangle$ for the index states $|u_i\rangle$ and $|v_j\rangle$. The quantum logical cross-entropy can be written in a number of ways:

$$\begin{aligned} h(\rho||\tau) &= 1-\mathrm{tr}\left[\rho^\dagger\tau\right] = 1-\textstyle\sum_{i,j}\lambda_i\mu_j\langle v_j|u_i\rangle\langle u_i|v_j\rangle \\ &= \mathrm{tr}\left[\tau^\dagger(I-\rho)\right] = \textstyle\sum_{i,j}(1-\lambda_i)\mu_j\langle v_j|u_i\rangle\langle u_i|v_j\rangle \\ &= \mathrm{tr}\left[\rho^\dagger(I-\tau)\right] = \textstyle\sum_{i,j}\lambda_i(1-\mu_j)\langle v_j|u_i\rangle\langle u_i|v_j\rangle. \end{aligned}$$

Classically, the 'index state' $\{i\}$ completely overlaps with $\{j\}$ when $i=j$ and has no overlap with any other $\{i'\}$ from the indices $\{1,\ldots,n\}$ so the 'overlaps' are, as it were, $\langle j|i\rangle\langle i|j\rangle = \delta_{ij}$, the Kronecker delta. Hence the classical analogue formulas are:

$$h(p||q) = 1-\textstyle\sum_{i,j} p_i q_j\delta_{ij} = \sum_{i,j}(1-p_i)q_j\delta_{ij} = \sum_{i,j} p_i(1-q_j)\delta_{ij}.$$

The quantum logical cross-entropy $h(\rho||\tau)$ can be interpreted by considering two measurements, one of ρ with the $\{|u_i\rangle\}_i$ measurement basis and the other of τ with the $\{|v_j\rangle\}_j$ measurement basis. If the outcome of the ρ measurement was u_i with probability λ_i, then the outcome of the τ measurement is different than v_j with probability $1-\mu_j$ but that distinction probability $\lambda_i(1-\mu_j)$ is only relevant to the extent that u_i and v_j are the 'same state' or overlap, and that extent is $\langle v_j|u_i\rangle\langle u_i|v_j\rangle$. Hence the quantum logical cross-entropy is the sum of those two-measurement distinction probabilities weighted by the extent that the states overlap.[10]

When the two density matrices commute, $\rho\tau=\tau\rho$, then (as noted above) we have the essentially classical situation of one set of index states $\{|u_i\rangle\}_i$ which is an orthonormal basis set of simultaneous eigenvectors for both ρ and τ with the respective eigenvalues $\{\lambda_i\}_i$ and $\{\mu_j\}_j$. Then $\langle u_j|u_i\rangle\langle u_i|u_j\rangle = \delta_{ij}$ so $h(\rho||\tau) = \sum_{i,j}\lambda_i(1-\mu_j)\delta_{ij}$ is the probability of getting two distinct index states u_i and u_j for $i\neq j$ in two independent measurements, one of ρ and one of τ in the same measurement basis of $\{|u_i\rangle\}_i$. This interpretation includes the special case when $\tau=\rho$ and $h(\rho||\rho)=h(\rho)$.

We saw that classically, the logical Hamming distance between two partitions could be defined as:

$$d(\pi,\sigma) = 2h(\pi||\sigma)-h(\pi)-h(\sigma) = \mathrm{tr}\left[\rho(\pi)^2\right]+\mathrm{tr}\left[\rho(\sigma)^2\right]-2\,\mathrm{tr}\left[\rho(\pi)\rho(\sigma)\right]$$

so this motivates the quantum definition:[11]

$$d(\rho,\tau) = 2h(\rho||\tau)-h(\rho)-h(\tau) = \mathrm{tr}\left[\rho^2\right]+\mathrm{tr}\left[\tau^2\right]-2\,\mathrm{tr}\left[\rho^\dagger\tau\right]$$

quantum logical Hamming distance between two quantum states.

---

[10] The interpretation of $h(\rho)$ and $h(\tau||\rho)$, as well as the later development of the quantum logical conditional entropy $h(\rho|\tau)$ and the quantum Hamming distance $d(\rho,\tau)$, are all based on using the eigenvectors and eigenvalues of density matrices–which Michael Nielsen and Issac Chuang seem to prematurely dismiss as having little or no "special significance." [28, p. 103]

[11] Nielsen and Chuang suggest the idea of a Hamming distance between quantum states–only to then dismiss it. "Unfortunately, the Hamming distance between two objects is simply a matter of labeling, and *a priori* there aren't any labels in the Hilbert space arena of quantum mechanics!" [28, p. 399] They are right that there is no correlation, say, between the vectors in the two ON bases $\{u_i\}_i$ and $\{v_j\}_j$ for V, but the cross-entropy $h(\rho||\tau)$ uses all possible combinations in the terms $\lambda_i(1-\mu_j)\langle v_j|u_i\rangle\langle u_i|v_j\rangle$ and thus the definition of the Hamming distance developed here does not use any arbitrary labeling or correlations.

There is another distance measure between quantum states, namely the *Hilbert-Schmidt norm*, that has been recently investigated in [35] (with an added $\frac{1}{2}$ factor):[12]

$$\underbrace{\mathrm{tr}\left[(\rho-\tau)^2\right]}_{\text{Hilbert-Schmidt norm}}$$

where we write A^2 for $A^\dagger A$. Then the naturalness of this norm as a 'distance' is enhanced by the fact that it is the same as the quantum Hamming distance:

Theorem 4. *Hilbert-Schmidt norm = quantum logical Hamming distance.*

Proof. $\mathrm{tr}\left[(\rho-\tau)^2\right] = \mathrm{tr}\left[\rho^2\right] + \mathrm{tr}\left[\tau^2\right] - 2\,\mathrm{tr}\left[\rho^\dagger\tau\right] = 2h\left(\rho||\tau\right) - h\left(\rho\right) - h\left(\tau\right) = d\left(\rho,\tau\right).$ □

Hence the *information inequality* holds trivially for the quantum logical Hamming distance:

$$d\left(\rho,\tau\right) \geq 0 \text{ with equality iff } \rho = \tau.$$

The Fundamental Theorem can be usefully restated in this broader context of density operators instead of in terms of the density matrix represented in the ON basis for F-eigenvectors. Let ρ be any state to be measured by an observable F and let $\hat{\rho} = \sum_{i=1}^{I} P_{\phi_i}\rho P_{\phi_i}$ be the result of applying the Lüders mixture operation (where the P_{ϕ_i} are the projection operators to the eigenspaces of F). Then a natural question to ask is the Hilbert-Schmidt norm or quantum logical Hamming distance between the pre- and post-measurement states.

Theorem 5 (Fundamental (Quantum))**.** $\mathrm{tr}\left[(\rho-\hat{\rho})^2\right] = h\left(\hat{\rho}\right) - h\left(\rho\right).$

Proof. $\mathrm{tr}\left[(\rho-\hat{\rho})^2\right] = \mathrm{tr}[(\rho^\dagger-\hat{\rho}^\dagger)(\rho-\hat{\rho})] = \mathrm{tr}\left[\rho^2\right] + \mathrm{tr}\left[\hat{\rho}^2\right] - \mathrm{tr}\left[\rho\hat{\rho}\right] - \mathrm{tr}\left[\hat{\rho}\rho\right]$ where $\rho^\dagger = \rho$, $\hat{\rho}^\dagger = \hat{\rho}$, $\mathrm{tr}\left[\rho\hat{\rho}\right] = \mathrm{tr}\left[\hat{\rho}\rho\right]$, and $\mathrm{tr}\left[\rho\hat{\rho}\right] = \mathrm{tr}\left[\rho\sum_{i=1}^{I} P_{\phi_i}\rho P_{\phi_i}\right] = \mathrm{tr}\left[\sum\rho P_{\phi_i}\rho P_{\phi_i}\right]$. Also $\mathrm{tr}\left[\hat{\rho}^2\right] = \mathrm{tr}\left[\left(\sum_i P_{\phi_i}\rho P_{\phi_i}\right)\left(\sum_j P_{\phi_j}\rho P_{\phi_j}\right)\right] = \mathrm{tr}\left[\sum_i P_{\phi_i}\rho P_{\phi_i}^2\rho P_{\phi_i}\right]$ using the orthogonality of the distinct projection operators. Then using the idempotency of the projections and the cyclicity of the trace, $\mathrm{tr}\left[\hat{\rho}^2\right] = \mathrm{tr}\left[\sum_i \rho P_{\phi_i}\rho P_{\phi_i}\right]$ so $\mathrm{tr}\left[\rho\hat{\rho}\right] = \mathrm{tr}\left[\hat{\rho}\rho\right] = \mathrm{tr}\left[\hat{\rho}^2\right]$ and hence $\mathrm{tr}\left[(\rho-\hat{\rho})^2\right] = \mathrm{tr}\left[\rho^2\right] - \mathrm{tr}\left[\hat{\rho}^2\right] = h\left(\hat{\rho}\right) - h\left(\rho\right)$. □

14. Results

Logical information theory arises as the quantitative version of the logic of partitions just as logical probability theory arises as the quantitative version of the dual Boolean logic of subsets. Philosophically, logical information is based on the idea of information-as-distinctions. The Shannon definitions of entropy arise naturally out of the logical definitions by replacing the counting of distinctions by the counting of the minimum number of binary partitions (bits) that are required, on average, to make all the same distinctions, i.e., to uniquely encode the distinguished elements–which is why the Shannon theory is so well-adapted for the theory of coding and communication.

This 'classical' logical information theory may be developed with the data of two partitions on a set with point probabilities. Section 9 gives the generalization to the quantum case where the partitions are provided by two commuting observables, the point set is an ON basis of simultaneous eigenvectors, and the point probabilities are provided by the state to be measured. In Section 10, the fundamental theorem for quantum logical entropy and measurement established a direct quantitative

[12] It is the square of the Euclidean distance between the quantum states and, ignoring the $\frac{1}{2}$ factor, it is the square of the "trace distance" [28, Chapter 9] between the states.

connection between the increase in quantum logical entropy due to a projective measurement and the eigenstates (cohered together in the pure superposition state being measured) that are distinguished by the measurement (decohered in the post-measurement mixed state). This theorem establishes quantum logical entropy as a natural notion for a quantum information theory focusing on distinguishing states.

The classical theory might also start with partitions on two different sets and a probability distribution on the product of the sets. Section 11 gives the quantum generalization of that case with the two sets being two ON bases for two non-commuting observables, and the probabilities are provided by a state to be measured. The classical theory may also be developed just using two probability distributions indexed by the same set, and this is generalized to the quantum case where we are just given two density matrices representing two states in a Hilbert space. Sections 12 and 13 carry over the Hamming distance measure from the classical to the quantum case where it is equal to the Hilbert-Schmidt distance (square of the trace distance). The general Fundamental Theorem relating measurement and logical entropy is that the Hilbert-Schmidt distance (= quantum logical Hamming distance) between any pre-measurement state ρ and the state $\hat{\rho}$ resulting from a projective measurement of the state is the difference in their logical entropies, $h(\hat{\rho}) - h(\rho)$.

15. Discussion

The overall argument is that quantum logical entropy is the simple and natural notion of information-as-distinctions for quantum information theory focusing on the distinguishing of quantum states. There are two related classical theories of information, classical logical information theory (focusing on information-as-distinctions and analyzing classification) and the Shannon theory (focusing on coding and communications theory). Generalizing to the quantum case, there are also two related quantum theories of information, the logical theory (using quantum logical entropy to focus on distinguishing quantum states and analyzing measurement as the quantum version of classification) and the conventional quantum information theory (using von Neumann entropy to develop a quantum version of the Shannon treatment of coding and communications).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Abramson, Norman. *Information Theory and Coding*. McGraw-Hill: New York, 1963.
2. Auletta, Gennaro, Mauro Fortunato, and Giorgio Parisi. *Quantum Mechanics*. Cambridge University Press: Cambridge UK, 2009.
3. Bennett, Charles H. Quantum Information: Qubits and Quantum Error Correction. *International Journal of Theoretical Physics* **2003** *42*, 153-176.
4. Boole, George *An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities*. Macmillan and Co: Cambridge, 1854.
5. Buscemi, Fabrizio, Paolo Bordone, and Andrea Bertoni. Linear Entropy as an Entanglement Measure in Two-Fermion Systems. *ArXiv.org*. **2007** http://arxiv.org/abs/quant-ph/0611223v2.
6. Campbell, L. Lorne. Entropy as a Measure. *IEEE Trans. on Information Theory* **1965**, *IT-11*, 112-114.
7. Cohen-Tannoudji, Claude, Bernard Diu and Franck Laloë. *Quantum Mechanics Vol. 1*. John Wiley and Sons: New York, 2005.
8. Cover, Thomas, and Joy Thomas. *Elements of Information Theory. Second Ed.* John Wiley and Son: Hoboken NJ, 2006.
9. Csiszar, Imre, and Janos Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Academic Press: New York; 1981.
10. Ellerman, David. Counting Distinctions: On the Conceptual Foundations of Shannon's Information Theory. *Synthese* **2009**, *168*, 119-149.
11. Ellerman, David. The Logic of Partitions: Introduction to the Dual of the Logic of Subsets. *Review of Symbolic Logic* **2010**, *3*, 287-350.

12. Ellerman, David. An Introduction of Partition Logic. *Logic Journal of the IGPL* **2014**, *22*, 94-125.
13. Ellerman, David. Logical Information Theory: New Foundations for Information Theory. *Logic Journal of the IGPL* **2017**, *25*, 806-835.
14. Ellerman, David. The Quantum Logic of Direct-Sum Decompositions: The Dual to the Quantum Logic of Subspaces. *Logic Journal of the IGPL* **2018**, *26*, 1-13.
15. Fano, Robert. *Transmission of Information*. MIT Press: Cambridge MA, 1961.
16. Fano, U. Description of States in Quantum Mechanics by Density Matrix and Operator Techniques. *Reviews of Modern Physics* **1957**, *29*, 74-93.
17. Gini, Corrado.*Variabilità e Mutabilità*; Tipografia di Paolo Cuppini: Bologna, 1912.
18. Hartley, Ralph V. Transmission of information. *Bell System Technical Journal* **1928**, *7*, 553-563.
19. Havrda, Jan, and Frantisek Charvat. Quantification Methods of Classification Processes: Concept of Structural Alpha-Entropy. *Kybernetika (Prague)* **1967**, *3*, 30-35.
20. Jaeger, Gregg. *Quantum Information: An Overview*. Springer Science+Business Media: New York, 2007.
21. Kolmogorov, Andrei. Combinatorial Foundations of Information Theory and the Calculus of Probabilities. *Russian Math. Surveys* **1983**, *38*, 29-40.
22. Kung, Joseph P. S., Gian-Carlo Rota, and Catherine H. Yan. *Combinatorics: The Rota Way*. Cambridge University Press: New York, 2009.
23. Lawvere, F. William and Stephen Schanuel. *Conceptual Mathematics: A First Introduction to Categories*. Cambridge University Press: New York, 1997.
24. Lawvere, F. William and Robert Rosebrugh. *Sets for Mathematics*. Cambridge University Press: Cambridge UK, 2003.
25. MacKay, David J.C. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press: Cambridge UK, 2003.
26. Markechová, Dagmar, Batool Mosapour, and Abolfazl Ebrahimzadeh. Logical Divergence, Logical Entropy, and Logical Mutual Information in Product MV-Algebras. *Entropy* **2018**, *20*, 129, doi.org/10.3390/e20020129.
27. McEliece, R. J. *The Theory of Information and Coding: A Mathematical Framework for Communication (Encyclopedia of Mathematics and Its Applications, Vol. 3)*; Addison-Wesley: Reading MA, 1977.
28. Nielsen, M., and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press: Cambridge UK, 2000.
29. Rao, C. R. Diversity and Dissimilarity Coefficients: A Unified Approach. *Theoretical Population Biology* **1982**, *21*, 24-43.
30. Rossi, Giovanni. Partition Distances. *arXiv:1106.4579v1* **2011**.
31. Rota, Gian-Carlo. Twelve Problems in Probability No One Likes to Bring up. In *Algebraic Combinatorics and Computer Science*; Crapo, H., Senato, D., Eds. Springer: Milan, 2001; pp. 57-93.
32. Rozeboom, William W. The Theory of Abstract Partials: An Introduction. *Psychometrika* **1968**, *33*, 133-167.
33. Shannon, Claude E. A Mathematical Theory of Communication. *Bell System Technical Journal* **1948**, *27*, 379-423;623-656.
34. Tamir, Boaz, and Eliahu Cohen. Logical Entropy for Quantum States. *ArXiv.org* **2014**, *Dec.*, http://de.arxiv.org/abs/1412.0616v2.
35. Tamir, Boaz, and Eliahu Cohen. A Holevo-Type Bound for a Hilbert Schmidt Distance Measure. *Journal of Quantum Information Science* **2015**, *5*, 127-133.
36. Tsallis, Constantino. Possible Generalization for Boltzmann-Gibbs Statistics. *J. Stat. Physics* **1988**, *52*, 479-487.
37. Uffink, Jos. *Measures of Uncertainty and the Uncertainty Principle. PhD Thesis*. University of Utrecht, 1990.
38. Yeung, Raymond W. *A First Course in Information Theory*. Springer Science+Business Media: New York, 2002.