Introduction to Quantum Logical Information Theory

David Ellerman
University of California at Riverside

July 15, 2017

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 as the distinctions 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., *qubits 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. The conclusion is that quantum logical entropy is the simple and natural notion of information for a quantum information theory focusing on the distinguishing of quantum states.

Contents

1 Duality of Subsets and Partitions

The logical foundations for classical and quantum 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*." [19, p. 85] That suggests that the Boolean logic of subsets should have a dual logic of partitions ([9], [10]).

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) $f^{-1}(Y)$ 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.

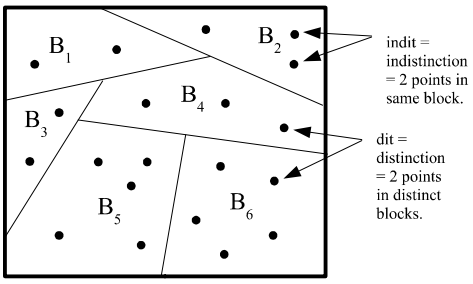

Figure 1: Distinctions and indistinctions of a partition.

The element-distinction duality can be illustrated by noting that the very notion of a function $f : X \to Y$ can be defined in a dual manner using those concepts. A *function* $f : X \to Y$ is a binary relation $R \subseteq X \times Y$ that transmits elements and dually reflects distinctions–where:

- a binary relation $R \subseteq X \times Y$ *transmits elements* if for each element $x \in X$, there is an ordered pair $(x, y) \in R$ for some $y \in Y$, and
- a binary relation $R \subseteq X \times Y$ *reflects distinctions* if for any pairs (x, y) and (x', y') in R, if $y \neq y'$, then $x \neq x'$.

f: X→Y	Subsets	Partitions
General case	Image f(X) = Subset of codomain Y	Inverse-image $f^{-1}(Y)$ = Partition on domain X
Definition of a function	Binary relation f ⊆ X×Y that transmits elements	+Binary relation f ⊆ X×Y that reflects distinctions
Basic duality	Elements of Subset	Distinctions of Partition

Table 1: Duality of Subsets and Partitions &
Duality of Elements and Distinctions ("Its" & "Dits")

The Boolean logic of subsets is usually treated in modern texts solely in terms of the special case of "propositional logic." For instance, Given a formula $\Phi(\pi, \sigma, ...)$ composed with Boolean operations (e.g., $\vee$, $\wedge$, $\Longrightarrow$, $\emptyset$) on the atomic variables π, σ, ..., then a *Boolean tautology* would be defined as a formula such that no matter what subsets of the nonempty universe U are substituted for the atomic variables, then the whole formula evaluates (using the corresponding set operations) to the universe U, the top of the power-set Boolean algebra $\wp(U)$. It is then a theorem (not a definition) that the same set of valid formulas is obtained if one only considers the one-element universe $U = 1$, in which case it is convenient to interpret the variables and formulas as being propositions. Most modern texts just start with this propositional special case and *define* a valid formula as a truth table tautology, i.e., as a formula such that no matter what subsets $\emptyset, 1$ of the universe 1 are substituted for the atomic variables, the whole formula will evaluate to the universe 1. This neglect of the general Boolean logic of subsets in favor of the propositional special case is one of the reasons for the long delay in developing the dual logic of partitions–since propositions, unlike subsets, don't have a category-theoretic dual [9].

The algebra associated with the subsets $S \subseteq U$ is the power-set Boolean algebra $\wp(U)$ of subsets of U with the partial order as the inclusion of elements. The corresponding algebra of partitions π on U is the *partition algebra* $\prod(U)$ defined as follows:

- the *partial order* $\sigma \preceq \pi$ of partitions $\sigma = \{C_1, ..., C_J\}$ and $\pi = \{B_1, ..., B_I\}$ holds when π *refines* σ in the sense that for every block $B_i \in \pi$ there is a block $C_j \in \sigma$ such that $B_i \subseteq C_j$, or, equivalently, using the element-distinction ('its' & 'dits') pairing, the partial order is the inclusion of distinctions: $\sigma \preceq \pi$ if and only if (iff) $\text{dit}(\sigma) \subseteq \text{dit}(\pi)$;

- the minimum or bottom partition is the *indiscrete partition* (or blob) $\mathbf{0} = \{U\}$ with one block consisting of all of U;

- the maximum or top partition is the *discrete partition* $\mathbf{1} = \{\{u\}\}_{u \in U}$ consisting of singleton blocks;

- the *join* $\pi \vee \sigma$ is the partition whose blocks are the non-empty intersections $B_i \cap C_j$ of blocks of π and blocks of σ, or, equivalently, using the element-distinction pairing, $\text{dit}(\pi \vee \sigma) = \text{dit}(\pi) \cup \text{dit}(\sigma)$;

- the *meet* $\pi \wedge \sigma$ is the partition whose blocks are the equivalence classes for the equivalence relation generated by: $u \sim u'$ if $u \in B_i \in \pi$, $u' \in C_j \in \sigma$, and $B_i \cap C_j \neq \emptyset$; and

- $\sigma \Rightarrow \pi$ is the *implication partition* whose blocks are: (1) the singletons $\{u\}$ for $u \in B_i \in \pi$ if there is a $C_j \in \sigma$ such that $B_i \subseteq C_j$, or (2) just $B_i \in \pi$ if there is no $C_j \in \sigma$ with $B_i \subseteq C_j$, so that trivially: $\sigma \Rightarrow \pi = \mathbf{1}$ iff $\sigma \preceq \pi$.

The same formulas $\Phi(\pi, \sigma, ...)$ can be interpreted as subset formulas or partition formulas. A *partition tautology* is analogously defined as a formula such that no matter what partitions on any U ($|U| \geq 2$) are substituted for the variables, the whole formula will evaluate by the partition operations to the discrete partition $\mathbf{1}$, the top of the partition algebra $\prod(U)$. For instance, *modus ponens*, $(\sigma \wedge (\sigma \Rightarrow \pi)) \Rightarrow \pi$, is a partition tautology. There is no U, analogous to $U = 1$, such that a formula is a partition tautology if and only if it always evaluates to $\mathbf{1}$ for partitions on U [9, Proposition 1.18].

There is a better way to connect subsets and partitions to propositions by considering a generic element u and a generic distinction (u, u') (with $u \neq u'$ understood). If a formula $\Phi(\pi, \sigma, ...)$ is construed as a subset formula, then "u is an element of $\Phi(\pi, \sigma, ...)$" [i.e., $u \in \Phi(\pi, \sigma, ...)$] is the corresponding proposition that is always true when $\Phi(\pi, \sigma, ...)$ is a Boolean tautology. If the formula $\Phi(\pi, \sigma, ...)$ is construed as a partition formula, then the corresponding proposition "(u, u') is a

distinction of $\Phi(\pi, \sigma, ...)$" is always true if $\Phi(\pi, \sigma, ...)$ is a partition tautology.[1] These results are summarized in Table 2.

Table 2	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 2: Dual Logics: Boolean subset logic of subsets and partition logic.

2 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 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 $\mathrm{dit}(\pi) \subseteq U \times U$, so the parallel quantitative development of the logic of partitions would start with the size or number of distinctions $|\mathrm{dit}(\pi)|$ in a partition π on U, which could then be normalized to $\frac{|\mathrm{dit}(\pi)|}{|U \times U|}$ and given 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 [22] (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" [18, p. 30] or symbolically:

$$\text{information} : \text{partitions} :: \text{probability} : \text{subsets}.$$

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? [22, 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".

[1] See [9] for how to use this propositional connection to develop a consistent and complete system of semantic tableaus for partition tautologies.

3 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. [17, 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 $U \times U$:

$$h(\pi) = \frac{|\operatorname{dit}(\pi)|}{|U \times U|} = \sum_{(u_j, u_k) \in \operatorname{dit}(\pi)} \frac{1}{|U|} \frac{1}{|U|}$$
Logical entropy of π (equiprobable case).

This is just the product probability measure of the equiprobable or uniform probability distribution on U applied to the information set or ditset $\operatorname{dit}(\pi)$. The inditset of π is $\operatorname{indit}(\pi) = \cup_{i=1}^{I} (B_i \times B_i)$ so where $p(B_i) = \frac{|B_i|}{|U|}$ in the equiprobable case, we have:

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

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 $S \subseteq U \times U$ the value of:

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

The *logical entropy* of π in general is the product-probability measure of its ditset:

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

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 π–just as $p(S)$ is the one-draw probability of getting an element of S.

4 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. Given partitions $\pi = \{B_1, ..., B_I\}, \sigma = \{C_1, ..., C_J\}$ on U, the *joint information set* is the union of the ditsets which is the *ditset for their join* is: $\operatorname{dit}(\pi) \cup \operatorname{dit}(\sigma) = \operatorname{dit}(\pi \vee \sigma) \subseteq U \times U$.

Given probabilities $p = \{p_1, ..., p_n\}$, the *joint logical entropy* is:

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

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

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

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

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

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.

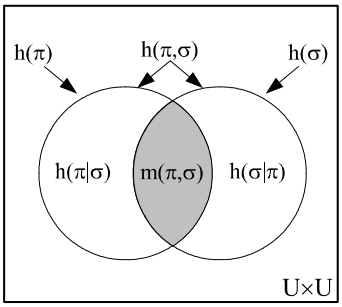

Figure 2: Logical entropies Venn diagram

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

5 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 from the brow' of Claude Shannon [23] already satisfying the standard Venn diagram relationships. 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 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:

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

The normalized number of distinctions or 'dit-count' of the discrete partition $\mathbf{1}$ is $1 - \frac{1}{n} = 1 - p(B_i)$.

The general case $\pi = \{B_1, ..., B_m\}$ is the average of the dit-counts $1 - p(B_i)$:

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

In the canonical case of 2^n equiprobable elements, the minimum number of binary partitions ("yes-or-no questions") or "bits" it takes to uniquely determine or *encode* each distinct element or block is n, so the Shannon-Hartley entropy [14] is:

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

The general case is the average of the bit-counts $\log_2\left(\frac{1}{p(B_i)}\right)$:

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

The *Dit-Bit Transform* essentially replaces the dit-counts by the bit-counts. First one expresses any logical entropy concept (simple, joint, conditional, or mutual) as an average of dit-counts $1 - p(B_i)$, and then substitutes the bit-count $\log\left(\frac{1}{p(B_i)}\right) = -\log(p(B_i))$ to obtain the corresponding formula as defined by Shannon. Table 3 gives examples of the dit-bit transform.

Table 3	The Dit-Bit Transform: $1-p(B_i) \rightsquigarrow \log(1/p(B_i))$
Entropy	$h(\pi) = \Sigma_i\, p(B_i)(1-p(B_i))$ $H(\pi) = \Sigma_i\, p(B_i)(\log(1/p(B_i)))$
Joint Entropy	$h(\pi,\sigma) = \Sigma_{i,j} p(B_i\cap C_j)(1-p(B_i\cap C_j))$ $H(\pi,\sigma) = \Sigma_{i,j} p(B_i\cap C_j)\log(1/p(B_i\cap C_j))$
Conditional Entropy	$h(\pi\|\sigma) = \Sigma_{i,j} p(B_i\cap C_j)(1-p(B_i\cap C_j)) - \Sigma_i p(C_j)(1-p(C_j))$ $H(\pi\|\sigma) = \Sigma_{i,j} p(B_i\cap C_j)\log(1/p(B_i\cap C_j)) - \Sigma_i p(C_j)\log(1/p(C_j))$
Mutual Information	$m(\pi,\sigma) = \Sigma_{i,j} p(B_i\cap C_j)[(1-p(B_i)) + (1-p(C_j)) - (1-p(B_i\cap C_j))]$ $I(\pi,\sigma) = \Sigma_{i,j} p(B_i\cap C_j)[(\log(1/p(B_i)) + (\log(1/p(C_j)) - (\log(1/p(B_i\cap C_j))]$

Table 3: Summary of the dit-bit transform

For instance,

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

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

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

The dit-bit transform is linear in the sense of preserving plus and minus, so the Shannon formulas satisfy the same Venn diagram formulas in spite of not being a measure (in the sense of measure theory):

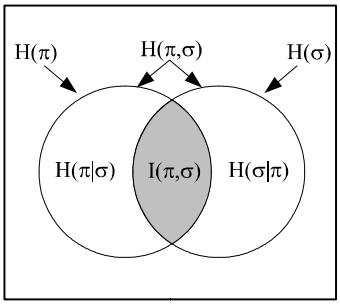

Figure 3: Venn diagram mnemonic for Shannon entropies

6 Logical entropy via density matrices

The transition to quantum logical entropy is facilitated by reformulating the 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 $p(S) = \sum_{u_j \in S} p_j$.

For any event S with $p(S) > 0$, let $|S\rangle = \frac{1}{\sqrt{p(S)}}(\chi_S(u_1)\sqrt{p_1}, ..., \chi_S(u_n)\sqrt{p_n})^t$ (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:

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

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:

$$\rho(\pi) = \textstyle\sum_i p(B_i)\rho(B_i).$$

Then $\rho(\pi)$ represents the *mixed* state, experiment, or lottery where the event B_i occurs with probability $p(B_i)$. The connection with the logical entropy $h(\pi)$ is:

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

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

Example 1 *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 ones) 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 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).

Example 2 (continued) *Now classify or "measure" the elements by the parity (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 where P_{odd} and P_{even} are the projections to the odd or even components:*

$$\begin{aligned} &P_{odd}\rho(\mathbf{0})P_{odd} + P_{even}\rho(\mathbf{0})P_{even} \\ &= \begin{bmatrix} 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 1/6 & 1/6 & 1/6 \end{bmatrix} \\ &= \tfrac{1}{2}\rho(B_{odd}) + \tfrac{1}{2}\rho(B_{even}) = \rho(\pi). \end{aligned}$$

Theorem 3 (Fundamental) *The increase in logical entropy due to a Lüders mixture operation, $h(\rho(\pi)) - h(\rho(\mathbf{0}))$, 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, $\text{tr}\left[\rho^2\right] = \sum_{i,j} |\rho_{ij}|^2$[13, p. 77], we have: $h(\rho(\pi)) - h(\rho(\mathbf{0})) = \left(1 - \text{tr}\left[\rho(\pi)^2\right]\right) - \left(1 - \text{tr}\left[\rho(\mathbf{0})^2\right]\right) = \text{tr}\left[\rho(\mathbf{0})^2\right] - \text{tr}\left[\rho(\pi)^2\right] = \sum_{i,j}\left(|\rho_{ij}(\mathbf{0})|^2 - |\rho_{ij}(\pi)|^2\right)$. ■

The 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.

Example 4 (continued) *In comparison with the matrix $\rho(\mathbf{0})$ of all entries $\frac{1}{6}$, the entries that got zeroed in the Lüders operation $\rho(\mathbf{0}) \rightsquigarrow \rho(\pi)$ correspond to the distinctions created in the transition $\mathbf{0} = \{U\} \rightsquigarrow \pi = \{\{u_1, u_3, u_5\}, \{u_2, u_4, u_6\}\}$, i.e., the odd-numbered faces were distinguished from the even-numbered faces. 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: $h(\pi) = 1 - 2 \times \left(\frac{1}{2}\right)^2 = \frac{1}{2}$ and $h(\mathbf{0}) = 1 - 1^2 = 0$.*

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 is 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.

7 Generalization to Quantum Information Theory: Commuting Observables

The idea of information as being based on 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 basis of eigenvectors of F. The basic idea of a *qubit* is pair of states definitely distinguishable by *some* observable[2]–which is analogous classically to a pair (u, u') of distinct elements of U that are distinguishable by some partition (i.e., $\mathbf{1}$). In general, a qubit[3] can be *relativized to an observable*–just as classically a distinction is a distinction *of a partition* (nor necessarily the discrete partition $\mathbf{1}$). Then there is a set partition $\pi = \{B_i\}_{i=1,...,I}$ on 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 straight forward using the usual ways that set-concepts generalize to vector-space concepts, e. g., subsets → subspaces, set partitions → direct-sum decompositions of subspaces,[4], Cartesian products of sets → tensor

[2] Any nondegenerate self-adjoint operator such as $\sum_{k=1}^{n} kP_{[u_i]}$, 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.

[3] This quantum version of a "dit" might be called a "qudit," but qubit is closer to common usage.

[4] Hence the 'classical' logic of partitions on a set will generalize to the quantum logic of direct-sum decompositions that is the dual to the usual quantum logic of subspaces [12].

products of vector spaces, and ordered pairs $(u_k, u_{k'}) \in U \times U \to u_k \otimes u_{k'} \in V \otimes V$. The eigenvalue function $f : U \to \mathbb{R}$ 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 and they 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 *qubit 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 $G : V \to V$ 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 $\{\gamma_j\}_{j \in J}$ be the set of eigenvalues of G and let $g : U \to \mathbb{R}$ be the eigenvalue function so a pair $(u_k, u_{k'})$ is a *qubit 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 straight forward, it will be presented in table form in Table 4a (which does not involve any probabilities)–where the qubits $(u_k, u_{k'})$ are interpreted as $u_k \otimes u_{k'}$.

Table 4a (w/o 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→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: $(u_k,u_{k'}) \in U^2$, $g(u_k) \neq g(u_{k'})$	Qubits of F: $u_k \otimes u_{k'} \in V \otimes V$, $f(u_k) \neq f(u_{k'})$ Qubits of G: $u_k \otimes u_{k'} \in V \otimes V$, $g(u_k) \neq g(u_{k'})$
Information sets/spaces	dit(π) ⊆ U×U dit(σ) ⊆ U×U	[qubit(F)] = subspace gen. by qubits of F [qubit(G)] = subspace gen. by qubits of G
Joint = Conditional = Mutual =	dit(π) ∪ dit(σ) ⊆ U×U dit(π) – dit(σ) ⊆ U×U dit(π) ∩ dit(σ) ⊆ U×U	[qubit(F) ∪ qubit(G] ⊆ V⊗V [qubit(F) – qubit(G)] ⊆ V⊗V [qubit(F) ∩ qubit(G)] ⊆ V⊗V

Table 4a: The parallel development of 'classical' and quantum logical information prior to probabilities.

The *information subspace* associated with F is the subspace $[qubit(F)] \subseteq V \otimes V$ generated by the qubits $u_k \otimes u_{k'}$ of F. If $F = \lambda I$ is a scalar multiple of the identity I, then it has no qubits so its information space $[qubit(\lambda I)]$ is the zero subspace. It is an easy implication of the Common Dits Theorem of classical logical information theory ([8, Proposition 1] or [9, Theorem 1.4]) that any two nonzero information spaces have a nonzero intersection, i.e., have a nonzero mutual information space.

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 4b gives the remaining parallel development with the probabilities provided by the pure state ψ.

Table 4b (w/ probs.)	'Classical' Logical Info. Theory	Quantum Logical Info. Theory
Probability distribution	Pure state density matrix, e.g., $\rho(\mathbf{0})$	Pure state density matrix $\rho(\psi)$
Product prob. dist.	p×p on U×U	$\rho(\psi)\otimes\rho(\psi)$ on V⊗V
Logical entropies	$h(\mathbf{0}_U) = 1 - \mathrm{tr}[\rho(\mathbf{0})^2] = 0$ $h(\pi) = p\times p(\mathrm{dit}(\pi))$ $h(\pi,\sigma) = p\times p(\mathrm{dit}(\pi) \cup \mathrm{dit}(\sigma))$ $h(\pi\|\sigma) = p\times p(\mathrm{dit}(\pi) - \mathrm{dit}(\sigma))$ $m(\pi,\sigma) = p\times p(\mathrm{dit}(\pi) \cap \mathrm{dit}(\sigma))$	$h(\rho(\psi)) = 1 - \mathrm{tr}[\rho(\psi)^2] = 0$ $h(F{:}\psi) = \mathrm{tr}[P_{[\mathrm{qubit}(F)]}\rho(\psi)\otimes\rho(\psi)]$ $h(F,G{:}\psi) = \mathrm{tr}[P_{[\mathrm{qubit}(F)\cup\mathrm{qubit}(G)]}\rho(\psi)\otimes\rho(\psi)]$ $h(F\|G{:}\psi) = \mathrm{tr}[P_{[\mathrm{qubit}(F)-\mathrm{qubit}(G)]}\rho(\psi)\otimes\rho(\psi)]$ $m(F,G{:}\psi) = \mathrm{tr}[P_{[\mathrm{qubit}(F)\cap\mathrm{qubit}(G)]}\rho(\psi)\otimes\rho(\psi)]$
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	$\rho(\pi) = \Sigma_i P_{B_i}\rho(\mathbf{0})P_{B_i}$ 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 $h(F{:}\psi) = 1 - \mathrm{tr}[\rho'(\psi)^2]$
Fund. Theorem on logical entropy and measurement.	$h(\pi)$ = sum of squares of terms zeroed in measurement operation: $\rho(\mathbf{0})\to\rho(\pi)$.	$h(F{:}\psi)$ = sum of absol. squares of terms zeroed in the measurement operation: $\rho(\psi)\to\rho'(\psi)$.

Table 4b: 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" [16, p. 5] or "impurity" of the state ρ.[5] 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 ([8], [11]), 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 qubits of an observable. The 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 $(1-\rho) \rightsquigarrow -\log(\rho)$ yields the usual Von Neumann entropy $S(\rho) = -\mathrm{tr}[\rho\log(\rho)]$[20, p. 510]. The fundamental theorem connecting logical entropy and the operation of classification-measurement also carries over to the quantum case.

8 Fundamental Theorem about logical entropy and measurement

Classically, a pair of elements $(u_k, u_{k'})$ either 'cohere' together in the same block of a partition on U, i.e., are an indistinction of the partition, or they don't and thus 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 $\rho(\psi)$ 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 $|\psi\rangle = \sum_j \langle u_j|\psi\rangle |u_j\rangle = \sum_j \alpha_j |u_j\rangle$. The coherences are classically modelled by the nonzero off-diagonal

[5] It is also called by the misnomer "linear entropy" [5] even though it is obviously a quadratic formula–so we will not continue that usage. The logical entropy is also the quadratic special case of the Tsallis-Havrda-Charvat entropy ([15], [26]), and the logical special case [8] of C. R. Rao's quadratic entropy [21].

entries $\sqrt{p_j p_k}$ for the indistinctions $(u_j, u_k) \in S \times S$, i.e., coherences $\approx$ indistinctions. The off-diagonal elements of $\rho(\psi)$ that are zeroed by the measurement to yield $\rho'(\psi)$ are the coherences (like quantum indistinctions) that are turned into 'decoherences' (like quantum distinctions).

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.

Theorem 5 (Fundamental) *The increase in quantum logical entropy, $h(F:\psi) = h(\rho'(\psi)) - h(\rho(\psi))$, due to the F-measurement of the pure state ψ is the sum of the absolute squares of the nonzero off-diagonal terms in $\rho(\psi)$ that are zeroed in the post-measurement mixed state density matrix $\rho'(\psi) = \sum_i P_{\phi_i} \rho(\psi) P_{\phi_i}$.*

Proof. $h(\rho'(\psi)) - h(\rho(\psi)) = \left(1 - \text{tr}\left[\rho'(\psi)^2\right]\right) - \left(1 - \text{tr}\left[\rho(\pi)^2\right]\right) = \sum_{i,j} \left(|\rho_{i,j}(\psi)|^2 - |\rho'_{i,j}(\psi)|^2\right)$. ■

This Fundamental Theorem for logical entropy and measurement directly connects the basis eigenstates from U that are distinguished by F (i.e., indicated by the zeroed off-diagonal terms) with the increase in logical entropy due to the measurement $h(F:\psi) = h(\rho'(\psi))$ (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 ([24], [25]) 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 [8]]. [24, p. 1]

Moreover, the quantum logical entropy has a simple "two-draw probability" interpretation, i.e., $h(F:\psi)$ is the probability that two independent F-measurements of ψ with yield distinct F-eigenvalues.

9 Generalization to Quantum Information Theory: Non-commuting Observables

9.1 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 [11]. 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:

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

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 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: $p_X(x) = \sum_y p(x,y)$ 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.

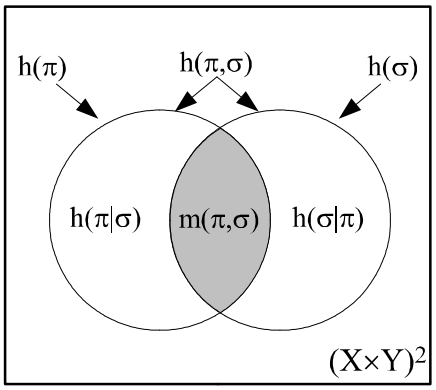

Figure 4: Venn diagram for logical entropies on $(X \times Y)^2$.

The previous treatment of $h(X)$, $h(Y)$, $h(X,Y)$, $h(X|Y)$, $h(Y|X)$, and $m(X,Y)$ in [11] was just the special cases where $\pi = \mathbf{1}_X$ and $\sigma = \mathbf{1}_Y$.

9.2 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 observable $F, G : V \to V$, 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:

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

We associated with 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 dit subspaces or information subspaces. For $S \subseteq (V \otimes V)^2$, let $[S]$ be the subspace generated by S. We simplify notation of $\operatorname{dit}(\pi \times \mathbf{0}_Y) = \operatorname{dit}(\pi) = \{(x \otimes y) \otimes (x' \otimes y') : f(x) \neq f(x')\}$, etc.

- $[\operatorname{dit}(\pi)] = [\{(x \otimes y) \otimes (x' \otimes y') : f(x) \neq f(x')\}]$;
- $[\operatorname{dit}(\sigma)] = [\{(x \otimes y) \otimes (x' \otimes y') : g(y) \neq g(y')\}]$;
- $[\operatorname{dit}(\pi,\sigma)] = [\operatorname{dit}(\pi) \cup \operatorname{dit}(\sigma)]$, and so forth.[6]

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:

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

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

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

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

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:

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

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 $\phi_1, ..., \phi_I$.[7] Then in two independent measurements of ψ by the observable $F \otimes I$, we have:

$$h(F:\psi) = \text{probability of getting distinct eigenvalues } \phi_i \text{ and } \phi_{i'}.$$

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:

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

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:

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

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:

$$[\text{dit}(\pi) \cup \text{dit}(\sigma)] = [\text{dit}(\pi) - \text{dit}(\sigma)] \oplus [\text{dit}(\pi) \cap \text{dit}(\sigma)] \oplus [\text{dit}(\sigma) - \text{dit}(\pi)].$$

Hence the probabilities are additive on those subspaces:

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

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

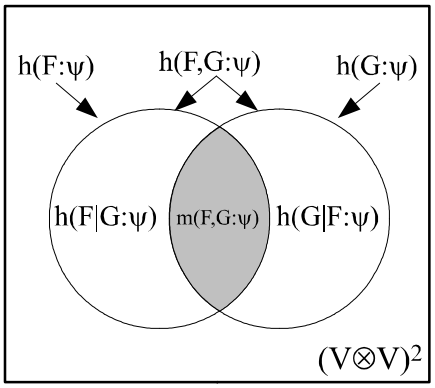

Figure 5: Venn diagram for quantum logical entropies.

10 Concluding Remarks

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–and is thus well-adapted for the theory of coding and communication.

This 'classical' logical information theory generalizes naturally to the quantum case where the distinguishing of points by partitions on an underlying set is replaced by the distinguishing of eigenstates by observables on the ambient Hilbert space. Both the classical and quantum versions of logical entropy have simple interpretations as "two-draw" probabilities. The fundamental theorem for quantum logical entropy and measurement established 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). The conclusion is that quantum logical entropy is the simple and natural notion of information for a quantum information theory focusing on the distinguishing of quantum states.

References

[1] Abramson, Norman 1963. *Information Theory and Coding*. New York: McGraw-Hill.

[2] Auletta, Gennaro, Mauro Fortunato, and Giorgio Parisi. 2009. *Quantum Mechanics*. Cambridge UK: Cambridge University Press.

[3] Bennett, Charles H. 2003. Quantum Information: Qubits and Quantum Error Correction. *International Journal of Theoretical Physics* 42 (2 February): 153–76.

[4] Boole, George 1854. *An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities*. Cambridge: Macmillan and Co.

[5] Buscemi, Fabrizio, Paolo Bordone, and Andrea Bertoni. 2007. Linear Entropy as an Entanglement Measure in Two-Fermion Systems. *ArXiv.org*. March 2. http://arxiv.org/abs/quant-ph/0611223v2.

[6] Campbell, L. Lorne 1965. Entropy as a Measure. *IEEE Trans. on Information Theory*. IT-11 (January): 112-114.

[7] Cohen-Tannoudji, Claude, Bernard Diu and Franck Laloë 2005. *Quantum Mechanics Vol. 1*. New York: John Wiley & Sons.

[8] Ellerman, David. 2009. Counting Distinctions: On the Conceptual Foundations of Shannon's Information Theory. *Synthese* 168 (1 May): 119–49.

[9] Ellerman, David 2010. The Logic of Partitions: Introduction to the Dual of the Logic of Subsets. *Review of Symbolic Logic*. 3 (2 June): 287-350.

[10] Ellerman, David 2014. An Introduction of Partition Logic. *Logic Journal of the IGPL*. 22, no. 1: 94–125.

[11] Ellerman, David 2017. Logical Information Theory: New Logical Foundations for Information Theory, *Logic Journal of the IGPL* (forthcoming).

[12] Ellerman, David 2017. The Quantum Logic of Direct-Sum Decompositions: The Dual to the Quantum Logic of Subspaces, *Logic Journal of the IGPL* (forthcoming).

[13] Fano, U. 1957. Description of States in Quantum Mechanics by Density Matrix and Operator Techniques. *Reviews of Modern Physics* 29 (1): 74–93.

[14] Hartley, Ralph V. L. 1928. Transmission of information. *Bell System Technical Journal*. 7 (3, July): 535-63.

[15] Havrda, Jan, and Frantisek Charvat. 1967. Quantification Methods of Classification Processes: Concept of Structural α-Entropy. *Kybernetika* (Prague) 3: 30–35.

[16] Jaeger, Gregg. 2007. *Quantum Information: An Overview*. New York: Springer Science+Business Media.

[17] Kolmogorov, Andrei N. 1983. Combinatorial Foundations of Information Theory and the Calculus of Probabilities. *Russian Math. Surveys* 38 (4): 29–40.

[18] Kung, Joseph P. S., Gian-Carlo Rota, and Catherine H. Yan. 2009. *Combinatorics: The Rota Way*. New York: Cambridge University Press.

[19] Lawvere, F. William and Robert Rosebrugh 2003. *Sets for Mathematics*. Cambridge UK: Cambridge University Press.

[20] Nielsen, M., and I. Chuang. 2000. *Quantum Computation and Quantum Information*. Cambridge UK: Cambridge University Press.

[21] Rao, C. R. 1982. Diversity and Dissimilarity Coefficients: A Unified Approach. *Theoretical Population Biology*. 21: 24-43.

[22] Rota, Gian-Carlo. 2001. Twelve Problems in Probability No One Likes to Bring up. In *Algebraic Combinatorics and Computer Science*, edited by Henry Crapo and Domenico Senato, 57–93. Milano: Springer.

[23] Shannon, Claude E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*. 27: 379-423; 623-56.

[24] Tamir, Boaz, and Eliahu Cohen. 2014. Logical Entropy for Quantum States. *ArXiv.org*. December. http://de.arxiv.org/abs/1412.0616v2.

[25] Tamir, Boaz, and Eliahu Cohen. 2015. A Holevo-Type Bound for a Hilbert Schmidt Distance Measure. *Journal of Quantum Information Science* 5: 127–33.

[26] Tsallis, Constantino 1988. Possible Generalization for Boltzmann-Gibbs Statistics. *J. Stat. Physics* 52: 479–87.