Science & Philosophy Volume 14(1), 2026

From Rota to the Logic of Partitions and the Notion of Logical Entropy

David Ellerman*

Abstract

This paper traces an intellectual journey that starts with some unfinished work of Gian-Carlo Rota on making a logic of equivalence relations or partitions. Rota understood the category-theoretic duality between subsets and partitions which implied there should be a logic of partitions dual to the usual Boolean logic of subsets. Rota also conjectured that subsets are to probability as partitions are to information. And just as probability starts quantitatively with the size of a subset, so he saw that information should start with some notion of size of a partition. In the duality between subsets and partitions, the elements of a subset are dual to the distinctions (or dits) of a partition, i.e., the ordered pairs of elements in different blocks of the partition. Since logical probability starts with the normalized number of elements in a subset, the corresponding logical notion of information is the normalized number of distinctions in a partition so that is the definition of *logical entropy*. Finally, logical entropy is contrasted with the usual Shannon notion of entropy.

Keywords: Gian-Carlo Rota, logic of subsets, dual logic of partitions, logical probability, dual notion of logical entropy.[1]

*University of Ljubljana, Ljubljana, Slovenia; david@ellerman.org; https://orcid.org/0000-0002-5718-618X

[1]Received on April 12, 2026. Accepted on May21, 2026. Published on June 30, 2026.
DOI: 10.23756/sp.v14i1.1738; eISSN 2282-7765.

1 Introduction

This paper is a result of my collaboration with Gian-Carlo Rota in the mid-80s. Rota died suddenly at the young age of 66 in 1999. Those who had worked with him picked up unfinished threads of his work to see if we could developed them further. My choice, which turned out to be more fruitful than I could imagine, was his joint paper with two young colleagues on the logic of a certain type of equivalence relations Finberg et al. [1996]. An *equivalence relation* on a universe set $U = \{u_1, ..., u_n\}$ is a reflexive, symmetric, and transitive binary relation on the set. Each point in the set has an equivalence class of other points equivalent to it and the equivalence classes form a partition on the set, i.e., a set of non-empty subsets ("blocks") of U that are disjoint and cover all of U. Thus the notions of a partition and an equivalence relation are essentially the same mathematical concept (Rota would say "cryptomorphic") viewed from different viewpoints. The viewpoint taken here will focus on the notion of partitions. In spite of the label "logic", that paper used only the lattice operations of join and meet and didn't have an operation of implication on partitions. The lattices of partitions are so varied that the only formulas true in all partition lattices are general lattice identities Whitman [1946] which is why the Rota paper focused on a specific type ("commuting") of equivalence relation.

Hence the first step of extending Rota's work on this thread had two challenges: to develop the operation of implication for partitions and to develop the general logic of arbitrary partitions on a universe set. This step lead to developing Rota's conjecture that subsets are to probability as partitions are to information. Since probability starts with the quantitative version of subsets, the quantitative version of partitions would define a logical notion of information, logical entropy– which can then be compared and contrasted with the usual Shannon notion of entropy.

2 The logic of partitions

2.1 The subset-partition duality

"Logic" is usually taken to be the Boolean logic of subsets, usually presented only in the special case of propositional logic, and its sublogics and extensions. Rota knew that it was important to focus on general subsets, rather than just subsets of 1 (1 and $\emptyset$ representing truth and falsehood of propositions) since subset and partitions are dual concepts in category theory. Given a set function $f : X \to Y$, its *image* $f(X)$ is a subset of Y and its dual *coimage* (inverse image) is a partition $f^{-1} = \{f^{-1}(y)\}_{y \in f(X)}$ on X. By paying attention to this

duality, the logic of partitions can be developed in parallel to the usual Boolean logic of subsets.

2.2 The Boolean lattice and algebra of subsets

The powerset $\wp(U)$ is a Boolean lattice with the join as set union, the meet as set intersection, and the partial order as set inclusion. The lattice has a top (maximal element) U and a bottom (minimal element) $\emptyset$. The Boolean lattice become a Boolean algebra by the addition of the implication or conditional operation: for $S, T \in \wp(U)$, $S \supset T := S^c \cup T$ (where $S^c = U - S$ is the complement of S). The key relationship of the set implication is that when it equals the top, then the partial order of inclusion holds:

$$S \supset T = U \text{ iff (if and only if) } S \subseteq T.$$

2.3 The lattice and algebra of partitions

A *partition* π on a set $U = \{u_1, ..., u_n\}$ is a set $\pi = \{B_1, ..., B_m\}$ of non-empty subsets of U called *blocks* that are disjoint and their union if all of U (for our expository purposes, we stick to finite universe sets). A *distinction* or *dit* of π is an ordered pair of elements of U that are in different blocks of π, and the set of all distinctions of π is the ditset $\mathrm{dit}(\pi) \subseteq U \times U$. An *indistinction* or *indit* is an ordered pair of elements of U in the same block of π, and the set of all indistinctions of π is the indit set $\mathrm{indit}(\pi) \subseteq U \times U$. The ditset and indit set are complementary subsets of $U \times U$, i.e., $\mathrm{dit}(\pi)^c = \mathrm{indit}(\pi)$. The indit set $\mathrm{indit}(\pi)$ is the equivalence relation version of the partition π. The binary relations that are the ditsets of a partition might be called *partition relations* but they are also know as *apartness relations*.

Given another partition $\sigma = \{C_1, ..., C_{m'}\}$, the partition *join* $\pi \vee \sigma$ is the partition whose blocks are all the non-empty intersections $B_j \cap C_{j'}$. The ditset of the join is just the union of the ditsets, $\mathrm{dit}(\pi \vee \sigma) = \mathrm{dit}(\pi) \cup \mathrm{dit}(\sigma)$. By DeMorgan's law, the indit set of the join is the intersection of their equivalence relations, $\mathrm{indit}(\pi \vee \sigma) = \mathrm{indit}(\pi) \cap \mathrm{indit}(\sigma)$ which tells us that the intersection of two equivalence relations is always an equivalence relation. Hence given any subset of $U \times U$, there is always a smallest equivalence relation containing that subset. Taking the subset to be $\mathrm{indit}(\pi) \cup \mathrm{indit}(\sigma)$, the smallest equivalence relation containing it is defined as: $\mathrm{indit}(\pi \wedge \sigma)$, the indit set of the partition *meet* $\pi \wedge \sigma$. The partial order (PO) that turns the set of partitions $\Pi(U)$ on U into a lattice is *refinement*: $\sigma \precsim \pi$ (read π refines σ or σ is refined by π) if for every $B_j \in \pi$, there is a block $C_{j'}$ such that $B_j \subseteq C_{j'}$. Intuitively, π refines σ means that π could be obtained from σ by chopping up some blocks of σ. The refinement

PO is the parallel to the inclusion partial order of the Boolean lattice, and, in fact, refinement is equivalent to the inclusion of ditsets:

$$\sigma \precsim \pi \text{ iff } \operatorname{dit}(\sigma) \subseteq \operatorname{dit}(\pi)$$

This plants the hint that distinctions of a partition are parallel or dual to the elements of a subset.

The top of the lattice of partitions $\Pi(U)$ on U is the *discrete partition* $\mathbf{1}_U = \{\{u_i\}\}_{u_i \in U}$ where all the blocks are singletons, and the bottom is the indiscrete partition $\mathbf{0}_U = \{U\}$ whose only block is all of U. For $U = \{a, b, c\}$,the two dual lattices, the Boolean lattice of subset and the lattice of partitions, are illustrated in Figure 1.

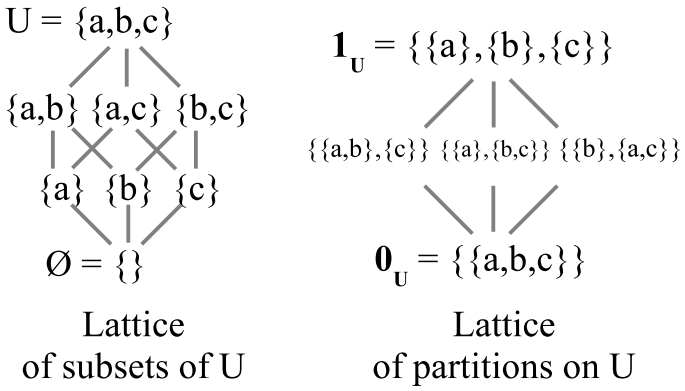

Figure 1: The two dual lattices on $U = \{a, b, c\}$

The lattice of partitions $\Pi(U)$ was known in the 19th century (e.g., Richard Dedekind and Ernst Schröder). But no new operations on partitions, e.g., the implication operation, were defined in the 20th century. In a 2001 paper commemorating Gian-Carlo Rota, the three authors first note the fundamentality of partitions and then acknowledge the sole operations of join and meet.

> Equivalence relations are so ubiquitous in everyday life that we often forget about their proactive existence. Much is still unknown about equivalence relations. Were this situation remedied, the theory of equivalence relations could initiate a chain reaction generating new insights and discoveries in many fields dependent upon it.
>
> This paper springs from a simple acknowledgment: the only operations on the family of equivalence relations fully studied, understood and deployed are the binary join $\vee$ and meet $\wedge$ operations.[Britz et al., 2001, p, 445]

The implication $\sigma \Rightarrow \pi$ operation on partitions can be motivated by requiring that it satisfy the corresponding requirement as in the Boolean algebra, i.e., the implication equals the top iff the partial order holds, i.e., $\sigma \Rightarrow \pi = \mathbf{1}_U$ iff $\sigma \precsim \pi$. The refinement $\sigma \precsim \pi$ means that for every block $B_j \in \pi$, there is a block of σ which contains it. This property of π-blocks can be characterized by a characteristic function. That is, the implication $\sigma \Rightarrow \pi$ can be defined simply like a characteristic or indicator function for that inclusion-property of the π-blocks. Thus if B_j is contained in some $C_{j'} \in \sigma$, then B_j is replaced by blocks of its discrete version, i.e., $\mathbf{1}_{B_j}$, the discrete partition on B_j, and if not, then B_j remains whole as B_j, the block of its indiscrete version $\mathbf{0}_{B_j}$, the indiscrete partition on B_j. Thus $\sigma \Rightarrow \pi$ is motivated by seeing it as the characteristic function taking $1, 0$ values according to whether a block $B_j \in \pi$ is or is not contained in a block of σ. This is how the *partition implication* $\sigma \Rightarrow \pi$ is defined and it automatically satisfies the parallel to $S \supset T = U$ iff $S \subseteq T$:

$$\sigma \Rightarrow \pi = \mathbf{1}_U \text{ iff } \sigma \precsim \pi.$$

The addition of the partition implication operation turns the partition lattice into the *partition algebra* $\Pi(U)$ on U. In this manner, the logic of partitions can be developed that is parallel or dual to the Boolean logic of subsets (Ellerman [2010], Ellerman [2014], Ellerman [2023]).

3 Quantitative versions of subset logic and partition logic

The next step in developing Rota's program is to define the quantitative version of the two logics. Rota was well-aware of the category-theoretic duality between subsets and partitions and he used it to make the key analogy: "The lattice of partitions plays for information the role that the Boolean algebra of subsets plays for size or probability." [Kung et al., 2009, p. 30] Rota also attributed the idea to Alfred Rényi [Rota, 2001, p. 65]. This could be symbolized as:

$$\frac{\text{Subsets}}{\text{Probability}} \approx \frac{\text{Partitions}}{\text{Information}}.$$

Probability 'starts' with the probability of a subset S of the set of outcomes U as the (normalized) size of the subset $\Pr(S) = \frac{|S|}{|U|}$. Rota reasoned that 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 "which will capture some property that will turn out to be for [partitions] what size is to a set." [Rota, 2001, p. 67] Since the size of a subset is the number of

elements, we already have a hint that the dual notion to an element of a subset is a distinction of a partition. Table 1 shows that elements and distinctions are indeed parallel or dual notions.

Lattice of subsets $\wp(U)$	Lattice of partitions $\Pi(U)$
Its = Elements of subsets	Dits = Distinctions of partitions
PO Inclusion of subsets $S \subseteq T$	PO $\sigma \precsim \pi$ iff $\operatorname{dit}(\sigma) \subseteq \operatorname{dit}(\pi)$
Join: $S \vee T = S \cup T$	Join: $\operatorname{dit}(\sigma \vee \pi) = \operatorname{dit}(\sigma) \cup \operatorname{dit}(\pi)$
Top: U all elements	Top: $\operatorname{dit}(\mathbf{1}_U)$ with all possible dits
Bottom: $\emptyset$ no elements	Bottom: $\operatorname{dit}(\mathbf{0}_U) = \emptyset$ with no dits

Table 1: Duality of "subset elements" and "partition distinctions":

$$\frac{\text{Elements}}{\text{Subsets}} \approx \frac{\text{Distinctions}}{\text{Partitions}}$$

Since no element in U can be distinct from itself, i.e., cannot be in two different blocks of a partition, the only indits of $\mathbf{1}_U$ are the self-pairs (u_i, u_i), i.e., $\operatorname{indit}(\mathbf{1}_U) = \Delta = \{(u_i, u_i) \,|u_i \in U\}$, i.e., the diagonal, so $\operatorname{dit}(\mathbf{1}_U) = U \times U - \Delta$ is all possible dits. And since all the elements of U are in the one block of $\mathbf{0}_U$, it has no distinctions.

Rota's suggestion to measure information by the " size" of a partition was correct, but he tried to shoehorn the Shannon entropy into that role [Rota, 2001, p. 68] rather than see the duality between elements of a subset and distinctions of a partition illustrated in Table 1. Hence Rota's suggestion can be implemented by taking the measure of information in a partition as the (normalized) number of distinctions in the partition so that is the beginning (equiprobable points in U) definition of *logical entropy*:

$$h(\pi) = \frac{|\operatorname{dit}(\pi)|}{|U \times U|} = \frac{|U \times U| - |\operatorname{indit}(\pi)|}{|U \times U|} = 1 - \frac{\cup_j |B_j \times B_j|}{|U \times U|} = 1 - \sum_j \left(\frac{|B_j|}{|U|}\right)^2.$$

With the points of U having probabilities $p = (p_1, ..., p_n)$, then $\Pr(B_j) = \sum_{u_i \in B_j} p_i$, and:

$$h(\pi) = 1 - \sum_{j=1}^{m} \Pr(B_j)^2 = \sum_{j \neq k} \Pr(B_j) \Pr(B_k)$$

where the last equation holds since: $1 = \left(\sum_{j=1}^{m} \Pr(B_j)\right)^2$. Since the probability distribution p defines the product probability measure $p \times p$ on $U \times U$, we also have:

$$h(\pi) = p \times p(\operatorname{dit}(\pi)).$$

This shows the interpretation of the logical entropy $h(\pi)$ is the probability that in two independent draws from U that one gets a distinction of π.

Since the logical entropy is the value of a (probability) measure, the compound notions of logical entropy are defined by the usual Venn diagram:

- *Joint logical entropy*: $h(\pi \vee \sigma) = p \times p(\operatorname{dit}(\pi) \cup \operatorname{dit}(\sigma))$,
- *Difference logical entropy*: $h(\pi|\sigma) = p \times p(\operatorname{dit}(\pi) - \operatorname{dit}(\sigma))$, and
- *Mutual logical entropy*: $m(\pi, \sigma) = p \times p(\operatorname{dit}(\pi) \cap \operatorname{dit}(\sigma))$,

as illustrated in Figure 2.

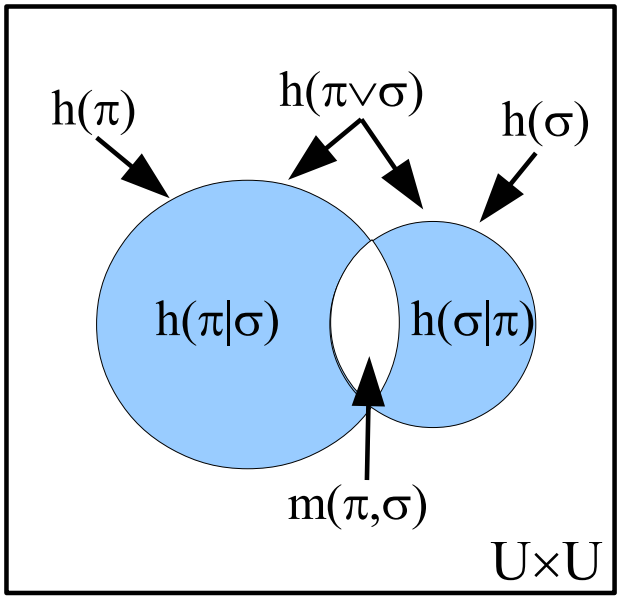

Figure 2: Compound logical entropies as values of subsets of a probability measure on $U \times U$

The logical entropy $h(\mathbf{1}_u)$ could also be taken as just the logical entropy of the probability distribution p:

$$h(p) = 1 - \sum_{i=1}^{n} p_i^2 = \sum_{j \neq k} p_j p_k$$

where the last equation holds since $1 = \left(\sum_{i=1}^{n} p_i\right)^2 = \sum_{i=1}^{n} p_i^2 + \sum_{j \neq k} p_j p_k$.

The logical entropy of a probability distribution can be illustrated in Figure 3 with a box diagram.
For this distribution, the logical entropy is: $h(p) = 1 - \sum p_i^2 = 1 - \left(\frac{1}{4} + \frac{1}{16} + \frac{1}{16}\right) = 1 - \frac{6}{16} = \frac{5}{8}$ with $\frac{5}{16}$ area on each side of the diagonal squares. The derivation of the logical entropy formula as the quantitative version of partitions is new, but the formula goes back to Corrado Gini's index of mutability suggested in 1912 Gini [1912]. Given a probability distribution p on the points of U and a distance function d_{jk} between points u_j and u_k that is non-negative and has

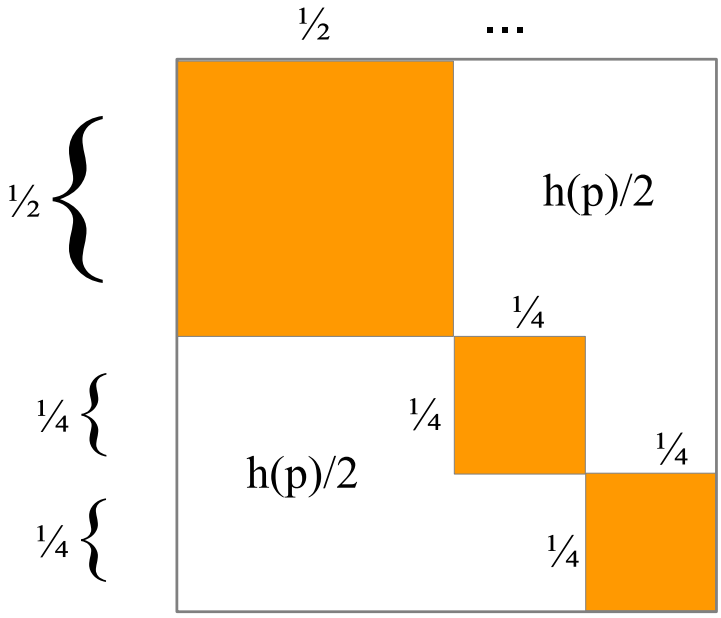

Figure 3: Logical entropy box diagram for $p = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$

$d_{ii} = 0$, C. R. Rao defined the *quadratic entropy* as $\sum_{j,k} p_j p_k d_{jk}$ Rao [2010]. There is a *logical distance function*, $1 - \delta_{jk}$, the 1-complement of the Kronecker delta function, which just represent identity (distance 0) and difference (distance 1). The quadratic entropy of the logical distance function is the logical entropy: $h(p) = \sum_{j,k=1}^{n} p_j p_k (1 - \delta_{jk})$.

To summarize, the duality between subsets and partitions in their quantitative versions gives a duality between probability theory and information theory illustrated in Table 2.

	Logical Probability Theory	Logical Information Theory
Outcomes	Elements of S	Distinctions of π
Events	Subsets $S \subseteq U$	Ditsets $\mathrm{dit}(\pi) \subseteq U \times U$
$p_i = \frac{1}{n}$	$\Pr(S) = \frac{\lvert S \rvert}{\lvert U \rvert}$	$h(\pi) = \frac{\lvert \mathrm{dit}(\pi) \rvert}{\lvert U \times U \rvert}$
Probs. p	$\Pr(S) = \sum_{u_i \in S} p_i$	$h(\pi) = \sum_{(u_i,u_k) \in \mathrm{dit}(\pi)} p_i p_k$
Interpretation	1-draw probability of a S-element	2-draw probability of a π-distinction

Table 2: Duality of quantitative subsets and partitions

4 Logical Entropy compared to Shannon Entropy

4.1 Comparison using 'surprise' factors

Intuitively, if $p_i = 1$, then there is no information in the occurrence of u_i, so information is measured by the 1-complement, the 'surprise' factor. But there are two 1-complements, the additive 1-complement of $1 - p_i$ and the multiplicative 1-complement of $1/p_i$.

The additive average of the additive 1-complements is the logical entropy:

$$h(p) = \sum_i p_i(1 - p_i).$$

The multiplicative average of the multiplicative 1-complements is the anti-log of the Shannon entropy:

$$\prod_i \left(\frac{1}{p_i}\right)^{p_i} = \log^{-1}\left(H\left(p\right)\right)$$

so the Shannon entropy is:

$$H(p)) = \log(\prod_i (\tfrac{1}{p_i})^{p_i}) = \sum_i p_i \log(\tfrac{1}{p_i}).$$

Then the log is chosen appropriate to the subject matter, e.g., log to the base 2 for coding theory or natural logs for statistical mechanics.

4.2 Negative Shannon Mutual Information

It is long been held that there is no such thing as "negative information" (at least up to the Trump Era). There is the basic Shannon definition of "entropy" and then there are the compound notions of joint entropy, conditional entropy, and mutual information where the compound notions are all carefully defined so they stand in the usual Venn diagram relationships [Cover and Thomas, 2006, p. 22].

Shannon then showed that if the probability distributions for the random variables X and Y were independent, then the mutual information $I(X;Y)$ was zero. That made a lot of sense, no information in common between independent random variables, and that was a significant "selling point" for that notion of entropy.

For a joint probability distribution $\Pr(X, Y)$, the random variables X and Y are independent if the joint probability is equal to the product of the marginals, i.e., $\Pr(X, Y) = \Pr(X)\Pr(Y)$.

In the case of logical entropy (represented by the lower-case h), there are natural Venn diagrams since it is a probability measure.

In that case, when X and Y are independent, then the logical entropy just mimics the definition using probabilities, $m(X, Y) = h(X)h(Y)$.

All the compound notions of Shannon entropy with one or two random variables are non-negative. But when the Shannon definitions were extended to three or more random variables, then the Shannon mutual information for three variables can be negative. This occurs in the standard examples in most any probability theory texts which notes the independence for three or more random variable is not the same as pairwise independence.

Consider flipping two fair coins with the random variables are $X = 1$ for heads and 0 for tails for the first coin and similarly for Y and the second coin. Then take $Z = 0$ if $X + Y$ is odd and 1 if even. Those three random variables are pairwise independent but mutually dependent, i.e., the values of any two of the variables determines the value of the third variable. In the Venn diagram, the pairwise overlaps have to be zero but the three-way overlap $I(X;Y;Z)$ has to be non-zero, i.e., negative, as in the following Venn diagram.

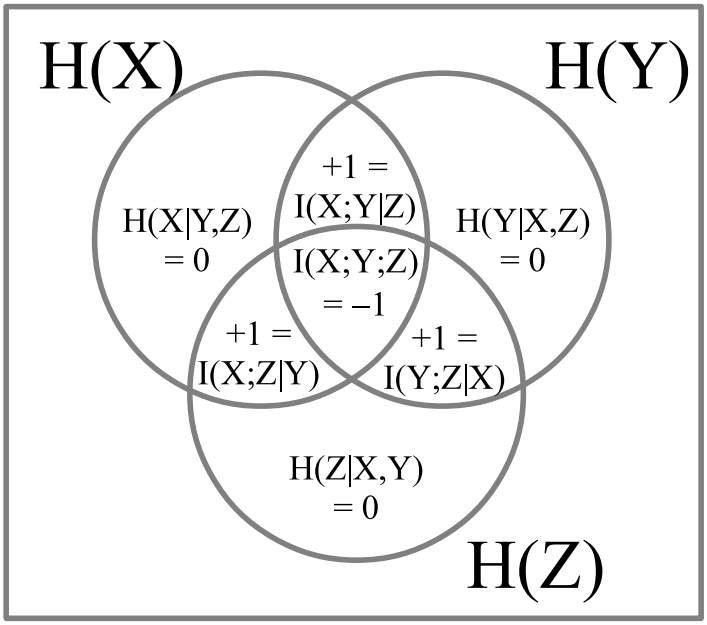

Figure 4: Example of negative Shannon mutual information

The response of Shannon information theorists to this rather anomalous result is interesting to illustrate how Shannon's reputation seems to bias independent judgment of the theory. Shannon himself just avoided the problem by never defining mutual information for more than two random variables. In the most authoritative book on Shannon information theory Cover and Thomas [2006], the other compound Shannon notions are extended to n-variables, but the anomaly for mutual information is only mentioned in a problem (2.25) without any follow-up as if it was not a Problem.

> 2.25 There isn't really a notion of mutual information common to three random variables. Here is one attempt at a definition: Using Venn diagrams, we can see that the mutual information common to three random variables X, Y, and Z can be defined by
>
> $$I(X;Y;Z) = I(X;Y) - I(X;Y|Z).$$
>
> ... Unfortunately, $I(X;Y;Z)$ is not necessarily nonnegative. [Cover and Thomas, 2006, p. 49]

In contrast, all the n-variable compound notions of logical entropy are non-negative. In fact, logical is a measure, indeed a probability measure.

It might also be noted that for certain countable probability distributions, the Shannon entropy is undefined in the sense of going to infinity [Cover and Thomas, 2006, Problem 2.19, p. 48], while the logical entropy for all countable probability distributions is always well-defined, i.e., finite.

4.3 Venn diagrams and Measures

Venn diagrams are creatures of measures in the sense of measure theory (e.g., Halmos [1974]. Shannon carefully defining his compound notions of (two variable) entropy to satisfy the usual Venn diagram relationships. Indeed, "Shannon carefully contrived for this 'accident' to occur" [Rozeboom, 1968, p. 153]. But in spite of Shannon entropy being routinely taken as "the measure" of information, it is actually not a measure in the sense of measure theory. Hence the Venn diagram relations are described as only a "mnemonic" for the relationship between the compound notions. e.g., Campbell [1965], which points out the desirable properties if it was a measure.

In contrast, logical entropy is a measure, even a probability measure, i.e., the logical entropy of a random variable is the probability in two independent samplings of getting different values.

If not a measure, how to account for the Shannon "Venn diagram" mnemonic? There is a nonlinear monotonic transformation, the "dit-bit transform," between the compound two variable logical entropies and the compound Shannon entropies, and that transformation preserves the simple Venn diagram relationships.Ellerman [2021].

4.4 MaxEntropy

Among all finite probability distributions, both Shannon and logical entropy have maximum values on the Laplacian equal distribution. But suppose there are

constraints on the allowed probability distributions so that the equiprobable distribution is not allowed. What then is the distribution closest to the equiprobable one? The maximization of Shannon entropy and logical entropy over the allowed distributions will in general give different results. Which is best?

That depends on how one defines the "closest distance to the equiprobable distribution." If one uses the Kullback-Leibler distance, then the Shannon entropy maximum is the closest distribution by that notion of distance.

There is a whole 'MaxEntropy school' of information theorists, founded by Edwin Jaynes (1922-1998), with many books and annual conferences all using this notion of maximum (Shannon) entropy.

But one might pause at some point and ask if the K-L distance is really a good notion of distance. The K-L distance at least has the distance between a probability distribution and itself is zero, but otherwise it is not even symmetric. That is, the K-L distance from a distribution p to a distribution q is not necessarily as the distance from q to p! And it is not even close to satisfying the triangle inequality.

There is the usual notion of Euclidean distance between probability distributions that is symmetric and satisfies the triangle inequality. Hence one might naturally ask: Is there is a notion of entropy so that maximizing that entropy also minimizes that notion of distance, e.g., the Euclidean distance of a probability distribution to the equiprobable distribution. Yes, it is the logical entropy.

4.5 Entropy in statistical mechanics

Consider a container whose space is divided up into m equally sized cells and the container contains n particles of gas. How can one describe the distribution of the particles between the cells that would characterize "equilibrium" within certain constraints. Ludwig Boltzmann (1844-1906) proposed a notion of entropy, Boltzmann entropy, whose maximum value, he hypothesized, would describe equilibrium. The formula for Boltzmann was a constant times the natural log of the multinomial coefficient. Thus, the formula involved the natural log of n-factorials, i.e., $\ln(n!)$, which is quite intractable for calculations. Hence one should work with numerical approximations, but which ones. There is an infinite series, the Stirling series, for $\ln(n!)$ so an approximation is obtained by taking only a finite number of terms. If one takes the two-term approximation, $\ln(n!) \approx n\ln(n) - n$, from the Stirling series, then the resulting formula obtained by Boltzmann and Josiah Willard Gibbs (1839–1903) is the formula now known as the Shannon entropy which is quite mathematically tractable. Of course, a better approximation would be obtained by considering more terms in the Stirling series but the resulting formulas are too intractable. And there are better approximations (unrelated to the Stirling series) than the two-term one, but the resulting

formulas are also not analytically tractable. It is that sequence of approximation and tractability choices that connects the Boltzmann entropy in statistical mechanics and the Shannon entropy, even though one often finds references to the "Boltzmann-Shannon entropy" as if they were the same notions.

The treatment of statistical mechanics based on logical entropy has not, to my knowledge, been developed. However, the equalizing nature of increasing logical entropy is immediate and trivial—without the sequence of approximation and tractability choices.

Consider a partition where the number of cells or blocks is fixed at m like the cells in a container. The logical entropy is the normalized number of distinctions in the partition where a distinction is an ordered pair of elements in different blocks or cells. Suppose one block has n_1 elements and another block has n_2 elements. Then the number of pairs of different-block elements contributed by those two blocks is just $n_1 n_2$ and the number of ordered pairs is twice that. Suppose the numbers are different with $n_1 < n_2$. If $n_2 = n_1 + 2$, then there are $(n_1 + 2)\, n_1 = n_1^2 + 2n_1$ pairs of elements in different blocks. Now suppose we make the equalizing move of one element from the larger block to the smaller block. Then there are $(n_1 + 1)\,(n_1 + 1) = n_1^2 + 2n_1 + 1$ pairs in different blocks so that increases logical entropy. That simple argument generalizes so that any movement of elements from the larger blocks to the smaller blocks (consistent with constraints if there are any) increases logical entropy. Hence the equilibrium configuration is the one that maximizes logical entropy (subject to any constraints).

4.6 Kolmogorov's criticism

Given a partition on a universe set U, the ditset of the partition is subset of $U \times U$ consisting of all the ordered pairs of elements in different blocks called "distinctions." The notion of a partition on U is equivalent to the notion of an equivalence relation on U where the blocks of the partition are the equivalence classes of the equivalence relation. The equivalence relation as the subset of $U \times U$ is just the set of ordered pairs of elements in the same blocks, i.e., the complement of the ditset. The beginning definition of logical entropy of a partition is just the size of its ditset normalized by the size of $U \times U$. No probabilities are involved. If there is a probability distribution p on U, then there is the product distribution pxp on $U \times U$, and then the logical entropy of the partition is just that product probability of the ditset, i.e., the probability of getting a distinction in two independent draws (i.e., with replacement) from U.

A measure in the sense of measure theory is a certain type of assignment of numbers to subsets of a universe set U, e.g., a counting measure giving the size of the subsets for a finite U. The Shannon entropy is not a measure in the sense

of assigning numbers to certain subsets of U. Instead, the Shannon entropy is defined directly in terms of the probabilities and that was the basis for criticism by Andrei Kolmogorov (1903–1987).

> 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. [Kolmogorov, 1983, p. 39]

Kolmogorov had his own designs, but in the case of logical entropy, the object of "a finite combinatorial character" is the ditset of a partition.

4.7 Shannon entropy as a "measure of uncertainty"

Shannon entropy is often described as a "measure of uncertainty" in, say, a real-valued random variable $X = \{x_1, ..., x_n\}$ where $\Pr(x_i) = p_i$ for $i = 1, ..., n$. But there is already a well-established measure of uncertainty, namely the variance: $Var(X) = E(X^2) - E(X)^2$.

That well-known and well-developed notion of uncertainty can be developed as a special case of logical entropy. Given a probability distribution $p = \{p_1, ..., p_n\}$, the *logical distance* between the indices is: $1 - \delta_{ij}$, i.e., one minus the Kronecker delta. Then in two independent draws from the set of indices $\{1, 2, \ldots, n\}$, the average logical distance (squared) between them is the logical entropy:

$$h(p) = \sum_{i \neq j} p_i p_j (1 - \delta_{ij})^2.$$

For a real-valued random variable, a better distance is Euclidean distance (squared). The logical entropy formula applied to that distance function squared is twice the variance.

$$\sum_{i \neq j} p_i p_j (x_i - x_j)^2 = 2Var(X).$$

If we only consider pairs of distinct indices rather than ordered pairs:

$$\sum_{i<j} p_i p_j (x_i - x_j)^2 = Var(X).$$

Thus the approach through logical entropy yields the usual "measure of uncertainty." The duality between logical probability and logical information theory implies there is a type of duality between the statistical mean and variance Ellerman [2025b].

5 Discussion

The Shannon entropy has its natural applications in coding theory and what Shannon called a "Mathematical Theory of Communications." Shannon [1948] Shannon noted that his theory didn't supply a definition of information: "no concept of information itself was defined." [Shannon, 1993, p. 458] But Shannon considered his entropy to be a quantification of information, and that is correct. There is a non-linear monotonic transformation, the Dit-Bit transform, from logical entropy to Shannon entropy. That transform can be read off of both formulas when expressed as additive probability averages:

$$h(p) = \sum_i p_i(1 - p_i) \text{ and } H(p) = \sum_i p_i \log(\tfrac{1}{p_i}).$$

Thus the Dit-Bit transform takes the term $1 - p_i$ in the logical entropy formula to the term $\log(\frac{1}{p_i})$ in the Shannon formula. The transform is monotonic so the Shannon entropy is indeed a quantification of information-as-distinction–intuitively, it is the minimum average number of *binary* distinctions it takes to distinguish the messages. Moreover, the Dit-Bit transform preserves the Venn diagrams from logical entropy as a measure to Shannon entropy as a mnemonic Campbell [1965]–so it provides an account of how the compound Shannon entropy formulas can satisfy the usual Venn diagrams in spite of not being a measure in the sense of measure theory.

The more controversial question is whether or not Shannon entropy is a satisfactory basis for information theory–as opposed to the theory of communications and coding. We have argued that logical entropy provides that (logical) foundation, and that Shannon entropy is the appropriate transform for the applications in communications and coding theory Ellerman [2021].

For further contrast between the logical entropy and the Shannon entropy, see the special issue of the journal 4Open on logical entropy Manfredi [2022] which includes Ellerman [2022].

The extension of Shannon entropy used in quantum mechanics is the von Neumann entropy, and the extension for logical entropy is the quantum logical entropy Ellerman [2025a].

6 Conclusions

The purpose of this paper is to give the results obtained by following out Gian-Carlo Rota's guidance starting with his initial logic of commuting equivalence relations Finberg et al. [1996]. The first result was the general logic of partitions (or equivalence relations) which is dual to the usual Boolean logic of

subsets. Then following Rota's suggestion that subsets are to probability as partitions are to information, we developed the quantitative version of partitions, i.e., logical entropy, dual to the quantitative version of subsets, i.e., logical probability. Then we argued that for the purposes of information theory, i.e., the measure of information-as-distinctions, as opposed to coding and communications theory, the notion of logical entropy was superior to the Shannon entropy.

There is more to the story that runs beyond the scope of this paper. There is a fundamental duality schema that begins with the above mentioned dualities and then connects to and explains category-theoretic duality and then extends to physics itself Ellerman [2024].

References

T. Britz, M. Mainetti, and L. Pezzoli. Some operations on the family of equivalence relations. In H. Crapo and D. Senato, editors, *Algebraic Combinatorics and Computer Science: A Tribute to Gian-Carlo Rota*, pages 445–59. Springer, Milano, 2001.

L. L. Campbell. Entropy as a measure. IT-11:112–114, 1965.

T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, second ed. edition, 2006.

D. Ellerman. The Logic of Partitions: Introduction to the Dual of the Logic of Subsets. *Review of Symbolic Logic*, 3(2 June):287–350, 2010. doi: https://doi.org/10.1017/S1755020310000018.

D. Ellerman. An introduction to partition logic. *Logic Journal of the IGPL*, 22(1):94–125, 2014. doi: https://doi.org/10.1093/jigpal/jzt036. URL `https://doi.org/10.1093/jigpal/jzt036`.

D. Ellerman. *New Foundations for Information Theory: Logical Entropy and Shannon Entropy*. SpringerNature, Cham, Switzerland, 2021. ISBN ISBN 978-3-030-86551-1 ISBN 978-3-030-86552-8 (eBook). URL `https://doi.org/10.1007/978-3-030-86552-8`.

D. Ellerman. Introduction to logical entropy and its relationship to Shannon entropy. *4Open Special Issue: Logical Entropy*, 5(1): 1–33, 2022. doi: https://doi.org/10.1051/fopen/2021004. URL `https://www.4open-sciences.org/component/toc/?task=topic&id=1575`.

D. Ellerman. *The Logic of Partitions: With Two Major Applications*. Studies in Logic 101. College Publications, London, 2023. ISBN 978-1-84890-440-8. URL https://www.collegepublications.co.uk/logic/?00052.

D. Ellerman. A Fundamental Duality in the Exact Sciences: The Application to Quantum Mechanics. *Foundations*, 4(2):175–204, 2024. doi: 10.3390/foundations4020013. URL https://www.mdpi.com/2673-9321/4/2/13.

D. Ellerman. A New Logical Measure for Quantum Information. *Quantum Information and Computation*, 25:81–95, 2025a. ISSN 1533-7146. doi: https://doi.org/10.2478/qic-2025-0005. URL https://sciendo.com/article/10.2478/qic-2025-0005.

D. Ellerman. The Mean and the Variance as Dual Concepts in a Fundamental Duality. *Axioms*, 14(6), 2025b. ISSN 2075-1680. doi: 10.3390/axioms14060466.

D. Finberg, M. Mainetti, and G.-C. Rota. The Logic of Commuting Equivalence Relations. In A. Ursini and P. Agliano, editors, *Logic and Algebra*, pages 69–96. Marcel Dekker, New York, 1996.

C. Gini. *Variabilità e mutabilità*. Tipografia di Paolo Cuppini, Bologna, 1912.

P. R. Halmos. *Measure Theory*. Springer-Verlag, 1974.

A. N. Kolmogorov. Combinatorial foundations of information theory and the calculus of probabilities. 38(4):29–40, 1983.

J. P. S. Kung, G.-C. Rota, and C. H. Yan. *Combinatorics: The Rota Way*. Cambridge University Press, New York, 2009.

G. Manfredi. Logical entropy – special issue. (5):E1, 2022. doi: https://doi.org/10.1051/fopen/2022005. URL www.4open-sciences.org.

C. R. Rao. Quadratic Entropy and Analysis of Diversity. *Sankhyā: The Indian Journal of Statistics*, 72-A(1):70–80, 2010.

G.-C. Rota. Twelve problems in probability no one likes to bring up. In H. Crapo and D. Senato, editors, *Algebraic Combinatorics and Computer Science: A Tribute to Gian-Carlo Rota*, pages 57–93. Springer, 2001.

W. W. Rozeboom. The theory of abstract partials: An introduction. 33(2):133–167, 1968.

C. E. Shannon. A mathematical theory of communication. 27:379–423; 623–56, 1948.

C. E. Shannon. Some topics in information theory. In N. J. A. Sloane and A. D. Wyner, editors, *Claude E. Shannon: Collected Papers*, pages 458–459. IEEE Press, 1993.

P. M. Whitman. Lattices, equivalence relations, and subgroups. 52:507–522, 1946.