The Mean and the Variance as Dual Concepts in a Fundamental Duality

David Ellerman

Independent Researcher, 1000 Ljubljana, Slovenia; david@ellerman.org

Abstract: A basic duality arises throughout the mathematical and natural sciences. Traditionally, logic is thought to be based on the Boolean logic of subsets, but the development of category theory in the mid-twentieth century shows the duality between subsets and partitions (or equivalence relations). Hence, there is an equally fundamental dual logic of partitions. At a more basic or granular level, the elements of a subset are dual to the distinctions (pairs of elements in different blocks) of a partition. The quantitative version of subset logic is probability theory (as developed by Boole), and the quantitative version of the logic of partitions is information theory re-founded on the notion of logical entropy. The subset side of the duality uses a one-sample (or one-element) approach, e.g., the mean of a random variable; the partition side uses a two-sample (or pair-of-elements) approach. This paper gives a new derivation of the variance (and covariance) based on the two-sample approach, which positions the variance on the partition and information theory side of the duality and thus dual to the mean.

Keywords: fundamental subsets–partitions duality; logic of subsets; logic of partitions; logical entropy; variance; covariance; mean

MSC: 03A10; 03B60; 06B99; 62A99; 94A17

1. Introduction: A Basic Duality in the Exact Sciences

The (new) logic of partitions (or equivalence relations) ([1–3]) is category-theoretically dual to the usual Boolean logic of subsets. This initiated a series of developments, some new and some reformulations of older ideas, that showed how the duality extended throughout the mathematical sciences [4] (including quantum mechanics and beyond into the life sciences). The dual logics have quantitative versions. The quantitative version of subset logic is probability theory, which is why George Boole's book was entitled "*An Investigation of the Laws of Thought on which are founded the Mathematical Theories of Logic and Probabilities*" [5]. Gian-Carlo Rota conjectured on numerous occasions [6] that subsets are to probability as partitions are to information:

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

Accordingly, partitions quantitatively defined the notion of *logical entropy*. Ordinarily, information and coding theory is based on Shannon entropy, which is not a measure (in the sense of measure theory) and has various anomalies, e.g., negative mutual information of some sets of three random variables (r.v.s), which are pair-wise independent but not mutually independent. Logical entropy quantifies the notion of information-as-distinctions and is a measure—specifically, a probability measure—and thus always non-negative.

There is a certain methodology used in developing this duality. That is, the notions associated with the subset logic are defined in terms of *single* elements, and the notions associated with partition logic are defined in terms of *pairs* of elements (or even ordered pairs). The simplest notion in mathematical statistics is the *mean* of a random variable (r.v), which is associated with the single samplings of a random variable. The partition-motivated notion of a pair of samples of an r.v. has been found to be the logical basis for variance (and similarly covariance). This is not, however, the usual definition of variance (or covariance) in textbooks. This definition of the variance seems to be so little known that it has been announced as a new discovery [7], although that derivation of variance is much older [8] (p. 42) and was known as the variance formula that does not use the mean.

Our purpose here is to apply this derivation of variance and covariance in the context of the fundamental category-theoretic duality, starting with the dual logics of subsets and partitions. Moreover, this treatment of variance and covariance places those notions firmly on the information-theoretic side of the duality. Indeed, the mean and the variance are dual concepts in terms of the duality. Hence, we begin with the duality at the logical level.

2. Methods

2.1. The Duality of Subsets and Partitions

Today, Boolean logic is almost always presented as "proposition logic", which is a special case of the logic of subsets. Category theory started in 1945 [9]. It provides the full mathematical treatment of the aforementioned basic duality as the "turn-around-the-arrows" duality. A subset or generally a subobject may be called a "part", and "The dual notion (obtained by reversing the arrows) of the "part" is the notion of the *partition*" [10] (p. 85).

Propositions do not have category–theory duals, so the idea of a dual logic of partitions was missed when Boolean logic was confined to propositions. Richard Dedekind and Ernst Schröder defined the lattice operations of join and meet in the nineteenth century. A "logic" of certain types of equivalence relations was defined by Gian-Carlo Rota and colleagues [11] in the twentieth century, but it was without any notion of implication for partitions—which is necessary to be considered a logic (as opposed to just a lattice). Indeed, no new operations on partitions or equivalence relations were developed in the 20th century.

> 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 ∨ and meet ∧ operations. [12] (p. 445)

2.2. The Two Lattices of Subsets and of Partitions

A set $\pi = \{B_1, \ldots, B_m\}$ of non-empty subsets (blocks or cells) of a universe set $U = \{u_1, \ldots, u_n\}$ ($|U| \geq 2$) is called a *partition*. An *equivalence relation* on U is a reflexive, symmetric, and transitive binary relation. A partition and an equivalence relation are equivalent notions, just looked at from different perspectives. The classes of equivalent elements of U for an equivalence relation are the blocks of the corresponding partition.

The category-theoretic (CT) reverse-the-arrows duality of subsets and partitions is brought about by their CT characterizations: A *subset* of U is the image of a function $X \to U$, and a *partition* on U is the co-image (i.e., inverse-image) of a function $U \to X$.

An ordered pair of elements in different blocks of a partition is called a "distinction", or "dit", and an ordered pair of equivalent elements, i.e., elements in the same block, is

called an "indistinction", or *indit*, of the partition. The set of distinctions, or *ditset*, is $\text{dit}(\pi) \subseteq U \times U$, and its complement in $U \times U$ is the set of *indistinctions or indits* $\text{indit}(\pi)$, which is just the equivalence relation version of the partition π.

Things are simple in the Boolean lattice of subsets since the join is just the union of subsets, the meet is the intersection of subsets, and the partial order is the inclusion of subsets. The top of the lattice is the universe set U, and the bottom is the empty set $\varnothing$. To form a Boolean logic or Boolean algebra, the implication or conditional operation on subsets of U is denoted $S \supset T$ and is defined as $S^c \cup T$ (where S^c is the complement of S). It is important to notice that when the implication $S \supset T$ is equal to the top U, then the partial order holds, i.e., $S \subset T$.

Partitions on U also form a lattice, the *partition lattice* $\Pi(U)$. Given $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_{m'}\}$ partitions on U, the *join* $\pi \vee \sigma$ is the partition on U whose blocks are the non-empty intersections $B_j \cap C_{j'}$ for $j = 1, \ldots, m$ and $j' = 1, \ldots, m'$. The union of the ditsets, i.e., $\text{dit}(\pi \vee \sigma) = \text{dit}(\pi) \cup \text{dit}(\sigma)$, is the ditset of the join. Since the join $S \cup T$ in the lattice of subsets is just the union of the elements of the two subsets $S, T \subseteq U$, we see that the distinctions or "Dits" of a partition play the same role as the elements or "Its" of a subset in the granular duality between subsets and partitions.

The intersection of two equivalence relations is always an equivalence relation, so there is a the smallest equivalence relation containing the union $\text{indit}(\pi) \cup \text{indit}(\sigma)$, and the meet $\pi \wedge \sigma$ is the corresponding partition.

The newly defined *implication* operation on partitions, denoted $\sigma \Rightarrow \pi$, is the partition that is like π except when a block $B_j \in \pi$ is contained in some block $C_{j'} \in \sigma$, i.e., $B_j \subseteq C_{j'}$, and then B_j is replaced by its discretization, i.e., by the singleton blocks of its elements.
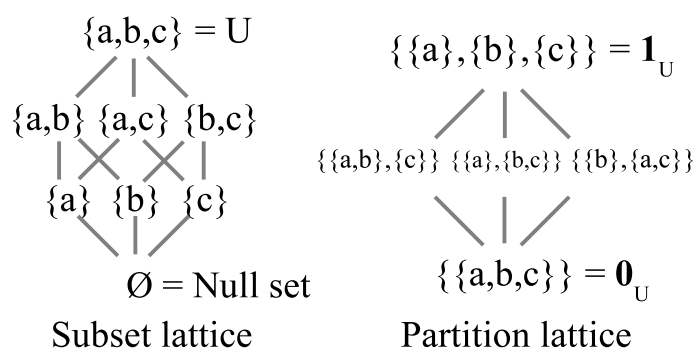
The partitions on a set have a partial order called *refinement* where π *refines* σ, written $\sigma \precsim \pi$, if for every block $B_j \in \pi$, there is a block $C_{j'}$ of σ such that $B_j \subseteq C_{j'}$. Intuitively, π can be obtained from σ by chopping up some blocks of σ. Some older texts ([13,14]) define the "lattice of partitions" with the reverse partial order (which reverses the join and meet), which Gian-Carlo Rota called "unrefinement" or "reverse refinement" [6] (p. 30). The partial order on subsets is the inclusion of elements, and the refinement of partitions is just the inclusion of distinctions, i.e., $\sigma \precsim \pi$ if and only if $\text{dit}(\sigma) \subseteq \text{dit}(\pi)$. This is one way the dual connection between "elements" and "distinctions" shows itself. The top of the lattice of partitions is the *discrete partition* $\mathbf{1}_U = \{\{u\}\}_{u \in U}$, where all blocks are singletons and the bottom is the *indiscrete partition* $\mathbf{0}_U = \{U\}$, where the only block is U. The addition of the implication operation turns the (19th-century) notion of a partition lattice into the (21st-century) notion of a partition algebra.

When the respective implications are equal to the top element, then and only then the partial order relation holds, i.e.,

$$S \supset T = U \text{ iff } S \subseteq T$$
$$\sigma \Rightarrow \pi = \mathbf{1}_U \text{ iff } \sigma \precsim \pi.$$

This cements the full dual relationship between the two lattices illustrated in Figure 1 for $U = \{a, b, c\}$.

Table 1 summarizes the duality between the subset lattice $\wp(U)$ and the partition lattice $\Pi(U)$.

$$\{a,b,c\} = U$$

$$\{a,b\} \; \{a,c\} \; \{b,c\}$$

$$\{a\} \; \{b\} \; \{c\}$$

$$\varnothing = \text{Null set}$$

Subset lattice

$$\{\{a\},\{b\},\{c\}\} = \mathbf{1}_U$$

$$\{\{a,b\},\{c\}\} \; \{\{a\},\{b,c\}\} \; \{\{b\},\{a,c\}\}$$

$$\{\{a,b,c\}\} = \mathbf{0}_U$$

Partition lattice

Figure 1. Lattices of subsets and of partitions.

Table 1. Elements–distinctions duality between the two dual lattices.

Dualities	**Boolean Lattice of Subsets**	**Lattice of Partitions**
"Its" or "Dits"	Elements of subsets	Distinctions of partitions
Partial order	$S \subseteq T$	$\mathrm{dit}(\sigma) \subseteq \mathrm{dit}(\pi)$
Join	$S \vee T = S \cup T$	$\mathrm{dit}(\pi \vee \sigma) = \mathrm{dit}(\pi) \cup \mathrm{dit}(\sigma)$
Top	Subset U with all elements	Partition $\mathbf{1}_U$ with all distinctions
Bottom	Subset $\varnothing$ with no elements	Partition $\mathbf{0}_U$ with no distinctions

Hence, we can extend Rota's duality statement to that between elements of subsets and distinctions of partitions:

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

One aspect of the dual concepts of elements and distinctions (that we will employ later), is that elements are a one-variable concept while distinctions are pairs of elements. The basic questions about elements are, for instance, whether or not a predicate applies to it, which is a question about existence, i.e., whether or not it exists in the subset of elements with that property. The basic questions about distinctions are about, given two elements, whether or not they are distinct in the partition or whether or not they are equivalent or not in the corresponding equivalence relation. As we will see, this duality between single-element concepts and pair-concepts comes out in statistics.

2.3. Fundamental Status of the Two Lattices

The notions of a subset of a set and a partition on a set are defined without any additional structure on the sets. Hence, the lattices, algebras, and logics of those two notions have a certain fundamental status. Given a topology or a partial order on a set, then other lattices can be defined in terms of that additional structure. But the notions of subsets and partitions require no such additional structure, and thus implying their fundamentality. Moreover, subsets and partitions are category-theoretically dual concepts.

The two dual logics provide a modern model for the old Aristotelian duality of *substance* versus *form* (as in information) [15]. We can tell two abstract "creation stories" by moving up in the two lattices from the bottom to the top. We see that the existence of substance increases in the subset lattice while form stays constant (always classical in the sense of being fully distinct), and the reverse happens in the partition lattice.

For each lattice where $U = \{a, b, c\}$, we start at the bottom and move towards the top in Figure 2.
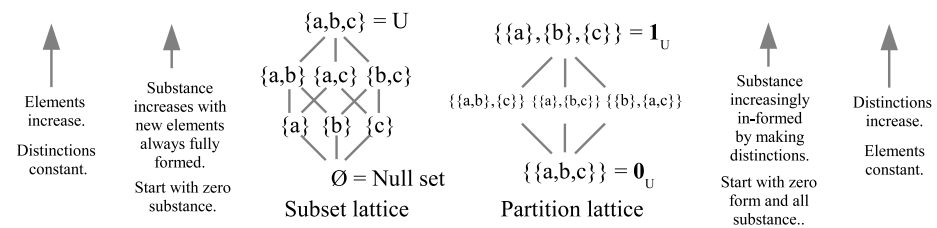
Figure 2. Moving up the subset and partition lattices.

There are no substances or "Its" (elements) at the bottom $\varnothing$ of the subset lattice, and as one moves up, new "Its" or elements appear but are always fully formed, i.e., no indefiniteness, until one reaches the full universe U. At the bottom $\mathbf{0}_U$ of the partition lattice, there are no dits (distinctions), i.e., $\mathrm{dit}(\mathbf{0}_U) = \varnothing$, and all the substance already appears but with no form (i.e., no "Dits" in $\mathbf{0}_U$, just as no "Its" in $\varnothing$). As one moves up the lattice of partitions, form (as in in-form-ation) is created by making new dits or distinctions until reaching the partition $\mathbf{1}_U$ that makes all possible distinctions, i.e., $\mathrm{dit}(\mathbf{1}_U) = U \times U - \Delta$, where $\Delta = \{(u_i, u_i) | u_i \in U\}$ is the *diagonal* of all the self-pairs of elements of U, since an element cannot be distinguished from itself.

The progress from the bottom to the top of the two lattices can be described as two creation stories.

- *Subset creation story*: "In the Beginning was the Void", and then elements were created, fully propertied and distinguished from one another, until finally reaching all the elements of the universe set U.

- *Partition creation story*: "In the Beginning was Undifferentiated Substance (e.g., "Form-less Chaos"), and then there was a "Big Bang" where the substance is was objectively informed by the making of distinctions (i.e., symmetry-breaking) until the final result was the fully distinguished (i.e., the equivalence classes are singletons) elements of the universe U.

2.4. Logical Entropy

2.4.1. A Little History of Information-as-Distinctions

When the logic of partitions was developed [3],it could imitate Boole's development of logical probability as the quantitative version of subsets [5]. Logical probability was the number of elements in a subset $\frac{|S|}{|U|}$ normalized by the cardinality of U. The dual to the notion of an element of a subset is a distinction of a partition.

Gregory Bateson, an eclectic anthropologist, said that information is "differences that make a difference." [16] (p. 99). Charles Bennett, a founder of quantum information theory, described information as "the notion of distinguishability abstracted away from what we are distinguishing, or from the carrier of information…" [17] (p. 155).

But the notion of information as the quantification of distinctions or differences goes back almost four centuries. In *The Information: A History, A Theory, A Flood* by James Gleick, he noted the focus on distinctions or differences in the work of John Wilkins, a 17th-century polymath and founder of the Royal Society. In the year before Newton was born (1641), Wilkins published one of the earliest books on cryptography, *Mercury or the Secret and Swift Messenger*. Moreover, he pointed out the fundamental role of differences and noted that any (finite) set of different things could be encoded by words in a binary code.

> For in the general we must note, that whatever is capable of a competent difference, perceptible to any sense, may be a sufficient means whereby to express the cogitations. It is more convenient, indeed, that these differences should be of as great variety as the letters of the alphabet; but it is sufficient if

they be but twofold, because two alone may, with somewhat more labour and
time, be well enough contrived to express all the rest. [18] (p. 67)

Wilkins explains that a five-letter binary code would be sufficient to code the letters of the
alphabet since $2^5 = 32$.

Thus any two letters or numbers, suppose A. B. being transposed through five
places, will yield thirty-two differences, and so consequently will superabun-
dantly serve for the four and twenty letters,.... [18] (pp. 67-68)

Gleick dates modern information theory from Claude Shannon's work published in
1948 [19]:

Any difference meant a binary choice. Any binary choice began the expressing
of cogitations. Here, in this arcane and anonymous treatise of 1641, the essential
idea of information theory poked to the surface of human thought, saw its
shadow, and disappeared again for [three] hundred years. [20] (p. 161) [an old
Pennsylvania Dutch superstition is that if on the second of February each year,
a groundhog emerges from its den and sees its shadow, then it stays in its den
for another six weeks].

2.4.2. The Mathematics of Logical Entropy

We have seen how probability theory starts with the notion of the logical probability
of a subset (or event) as the normalized number of elements in the subset S:

$$\Pr(S) = \frac{|S|}{|U|}.$$

We have also seen the duality between subsets and partitions expressed as

$$\frac{\text{Subsets}}{\text{Probability}} \approx \frac{\text{Partitions}}{\text{Information}} \text{ and } \frac{\text{Elements}}{\text{Subsets}} \approx \frac{\text{Distinctions}}{\text{Partitions}}$$

which also means that

$$\frac{\text{Elements}}{\text{Probability}} \approx \frac{\text{Distinctions}}{\text{Information}}.$$

Hence, the logical notion of information in a partition should be the normalized
number of distinctions in the partition. Following Shannon's labeling of his quantification
of information as "entropy", we will call the notion based on quantifying the logic of
partitions as "logical entropy" ([21–23]). For a partition $\pi = \{B_1, \ldots, B_m\}$, the *logical
entropy* $h(\pi)$ of π is the normalized number of distinctions:

$$h(\pi) = \frac{|\text{dit}(\pi)|}{|U \times U|} = \frac{|U \times U - \text{indit}(\pi)|}{|U \times U|} = 1 - \frac{\left|\cup_{j=1}^m B_j \times B_j\right|}{|U \times U|} = 1 - \sum_{j=1}^m \left(\frac{|B_j|}{|U|}\right)^2$$

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

where the product version of the formula follows from

$$1 = \left(\sum_{j=1}^m \Pr(B_j)\right)^2 = \sum_{j=1}^m \Pr(B_j)^2 + 2\sum_{j<k} \Pr(B_j)\Pr(B_k).$$

As in the case of logical probability, we are, for the moment, assuming equiprobable
points in U. Moreover, it should be noted that for probability, we are dealing with single
samples, and for information, we are dealing with pairs, indeed, ordered pairs of elements
drawn from U, so that each pair of distinct blocks B_j and B_k are counted twice in the
sum $\sum_{j \neq k} \Pr(B_j)\Pr(B_k)$. This yields an immediate simple interpretation of logical entropy,
namely, that $h(\pi)$ is the probability that in two independent samples from U, one obtains a
distinction of π.

The formulas generalize immediately to the general case of finite probability the-
ory, where the points $\{u_1, \ldots, u_n\}$ of U have the respective point probabilities of $p =
(p_1, \ldots, p_n)$. Then, $\Pr(S) = \sum_{u_i \in S} p_i$ and

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

when the given data is just the probability distribution p on U, then,

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

which is the logical entropy of the discrete partition $\mathbf{1}_U$, i.e., $h(p) = h(\mathbf{1}_U)$. In all the cases, the logical entropy is the two-sample probability of obtaining a distinction. The lowest value of logical entropy is 0 for the indiscrete partition $\mathbf{0}_U$ or for p, where some $p_i = 1$. The maximum value for logical entropy is for the discrete partition with equiprobable points:

$$h(\mathbf{1}_U) = 1 - \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 = 1 - \frac{n}{n^2} = 1 - \frac{1}{n}$$

which is interpreted simply as the probability that the first draw is not repeated in the second draw.

Diagrammatically, the logical entropy is illustrated using a box diagram with a unit square. For instance, for $p = (1/2.1/4.1/4)$, the box diagram is given in Figure 3, where each of the halves is

$$h(p)/2 = \sum_{i < k} p_i p_k = p_1 p_2 + p_1 p_3 + p_2 p_3 = \frac{1}{2}\frac{1}{4} + \frac{1}{2}\frac{1}{4} + \frac{1}{4}\frac{1}{4} = \frac{5}{16}.$$
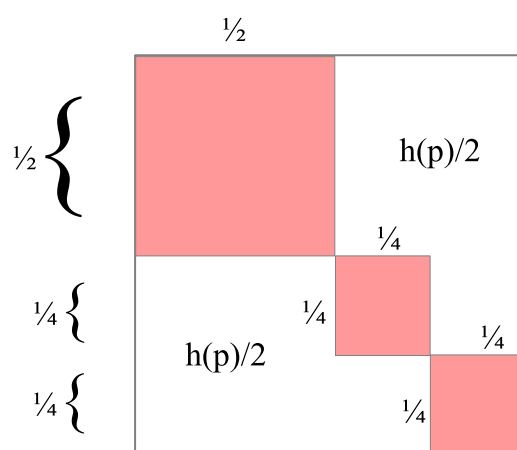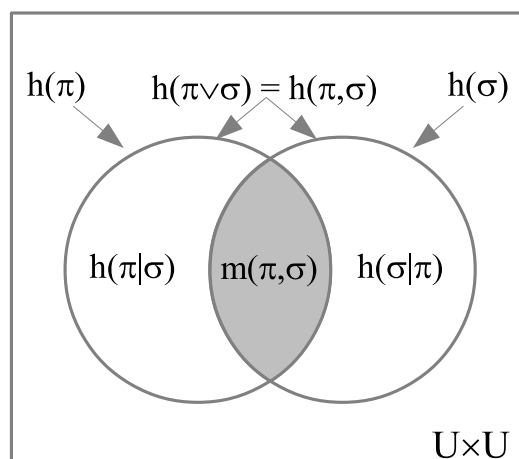

Figure 3. Logical entropy box diagram.

Logical entropy can be obtained as the value on a subset of a probability measure. The probability measure is the product measure $p \times p$ on the set $U \times U$, and the subset giving the logical entropy is $\text{dit}(\pi)$, i.e.,

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

Since $\text{dit}(\pi \vee \sigma) = \text{dit}(\pi) \cup \text{dit}(\sigma)$, $h(\pi \vee \sigma) = p \times p(\text{dit}(\pi \vee \sigma))$, which is sometimes written as $h(\pi, \sigma)$ as the *joint entropy*. Even though there is no structure on U, a natural closure operation on subsets $S \subseteq U \times U$, is defined as the reflexive–symmetric–transitive (RST) closure of S, which is the smallest indit set or equivalence relation containing S. Its complementary ditset will define a partition. Although they are the complements of a set under a closure operation, ditsets are not like the open sets in a topological space, and the RST closure is not a topological closure operation. The arbitrary union of ditsets is a ditset, but the intersection of ditsets is not necessarily a ditset (unlike open sets). But that intersection is nevertheless a subset of $U \times U$, so its probability measure defines the *mutual information*: $m(\pi, \sigma) = p \times p(\text{dit}(\pi) \cap \text{dit}(\sigma))$. Similarly, the *difference information* is $h(\pi | \sigma) = p \times p(\text{dit}(\pi) - \text{dit}(\sigma))$, which is interpreted as the information in π that is not in σ. Then, we have the usual Venn diagram relationships illustrated in Figure 4:

$$h(\pi, \sigma) = h(\pi|\sigma) + m(\pi, \sigma) + h(\sigma|\pi).$$

Figure 4. Venn diagram relationships for logical entropy.

2.4.3. The Relationship with Shannon Entropy

The well-known Shannon entropy [19], $H(p) = \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i}\right)$, is *not* defined as the values of certain subsets of a measure on a set. Shannon defined his notions of joint, mutual, and difference or conditional entropy directly in terms of probabilities and defined them so that the Venn diagrams relationships nevertheless held.

Andrei Kolmogorov objected to Shannon defining information directly in terms of probabilities and instead thought it should be based on a prior combinatorial structure.

> 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. [24] (p. 39)

Kolmogorov had his own ideas, but it might be noticed that the logical entropy definition of information-as-distinctions is based on probability-free sets of a combinatorial nature, namely, ditsets.

The fact that the compound notions of Shannon entropy satisfy the Venn diagram relations, in spite of not being defined as a measure in the sense of measure theory, is explained by the fact that logical entropy is defined as a measure and there is a uniform transformation between the compound logical entropy and Shannon entropy formulas that preserves Venn diagrams. This raises the question of that relationship between the two entropies.

If an outcome has probability close to or equal to 1, then the occurrence of that outcome intuitively gives little or no information. Therefore, it might be said that information is related to the *complement* of 1—but there are two 1-complements, the additive 1-complement $1 - p_i$ and the multiplicative 1-complement $\frac{1}{p_i}$. The additive and multiplicative averages of the respective 1-complements give the two entropies. The logical entropy $\sum_{i=1}^{n} p_i(1 - p_i) = h(p)$ is obtained as the additive probabilistic average of the additive 1-complements. The Shannon entropy in its log-free or anti-log form, $\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i} = \log^{-1}(H(p))$, is obtained as the multiplicative average of the multiplicative 1-complements. The particular log is chosen for the context, e.g., $H(p) = \log_2\left(\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i}\right) = \sum_{i=1}^{n} p_i \log_2\left(\frac{1}{p_i}\right)$ for coding theory or natural logs for statistical mechanics. Since the log of the multiplicative average transforms it to an additive average, the two additive averages can be transformed, one into the other, by the *dit-bit transform* $1 - p_i \rightsquigarrow \log\left(\frac{1}{p_i}\right)$. Once the compound formulas are expressed in terms of the

additive 1-complements, then this non-linear but monotonic dit-bit transform yields the corresponding compound formulas for Shannon entropy [21]. And the dit-bit transform preserves Venn diagrams.

2.4.4. Some History of the Logical Entropy Formula

The derivation of logical entropy as the quantitative version of the new partition logic is new, but the formula itself goes back at least to 1912 in the index of mutability of Corrado Gini [25]. The formula resurfaced in the code-breaking activity in World War II [26,27] where it was the additive 1-complement of Alan Turing's repeat rate $\sum_{i=1}^n p_i^2$, where it became part of the mathematics of cryptography [28]. After the war, Edmund Simpson published the formula as a quantification of biodiversity [29]. Hence, the formula is often known as the Gini–Simpson formula for biodiversity [30]. But Simpson and I. J. Good worked with Turing at Bletchley Park during the Second World War, and, according to Good, "E. H. Simpson and I both obtained the notion [the repeat rate] from Turing" [26] (p. 395). In 1948, when Simpson published the index, he did not acknowledge Turing, "fearing that to acknowledge him would be regarded as a breach of security" [31] (p. 562).

Gini [25] introduced d_{ij} as the "logical distance" between the ith and j^{th} elements where $d_{ij} = 1$ for $i \neq j$ and $d_{ii} = 0$, i.e., $d_{ij} = 1 - \delta_{ij}$, the 1-complement to the Kronecker delta. Since $1 = (p_1 + \ldots + p_n)(p_1 + \ldots + p_n) = \sum_i p_i^2 + \sum_{i \neq j} p_i p_j$, the logical entropy, i.e., Gini's index of mutability, $h(p) = 1 - \sum_i p_i^2 = \sum_{i \neq j} p_i p_j$, is the average logical distance between a pair of independently drawn elements. C. R. (Calyampudi Radhakrishna) Rao in 1982 generalized this by allowing other non-negative distances $d_{ij} = d_{ji}$ for $i \neq j$ (but always $d_{ii} = 0$) between the elements of U so that $Q = \sum_{i,j} d_{ij} p_i p_j$ would be the average distance between a pair of independently drawn elements from U, which was known as *quadratic entropy* [32].

3. Results: The Logical Basis for Variance and Covariance

At the logical level, the fundamental duality starts with the duality between elements ("Its") of a subset and distinctions ("Dits") of a partition. The elements of a subsets are certain *singular* elements from the universe set, and the distinctions of a partition are certain *pairs* (or ordered pairs) of elements of the universe set:

- Given an element, the natural questions are "to be or not to be" in a subset, or existence versus non-existence, the questions of the Boolean logic of subsets.
- Given a pair of elements, the natural questions are identity or difference, distinct or indistinct, or equivalent or in-equivalent, the questions of the logic of partitions (or equivalence relations).

Given $U = \{u_1, \ldots, u_n\}$ with the probability distribution $p = (p_1, \ldots, p_n)$, consider a real-valued random variable (r.v.) $X : U \to \mathbb{R}$. The inverse-image defines a partition $X^{-1} = \{X^{-1}(x_j)\}_{x_j \in X(U)}$ on U, where the values in the image $X(U)$ of the r.v. are $x_1, \ldots, x_m$. Then each value has the probability $\Pr(X = x_j) = \Pr(x_j) = \sum_{u_i \in X^{-1}(x_j)} p_i$.

On the one (single-value) side of the duality, the probability average of the single values is the usual *mean*: $\mu_X = E(X) = \sum_{j=1}^m \Pr(x_j) x_j$. But what is the appropriate notion on the other (pair-of-values) side of the duality?

In the same 1912 book [25] where Gini suggested the index of the mutability $\sum_{i \neq k} p_i p_k$ of a probability distribution, he suggested the *mean difference* $\sum_{j \neq k} \Pr(x_j) \Pr(x_k) |x_j - x_k|$. Maurice Kendall noted that the mean difference "has a certain theoretical attraction, being dependent on the spread of the variate-values among themselves and not on the deviations from some central value" [8] (p. 42). But Kendall went on to note: "It is, however, more difficult to compute than the standard deviation, and the appearance of the absolute values in the defining equations indicates, as for the mean deviation, the appearance of difficulties

in the theory of sampling" [8] (p. 42). At a later date, Kendall summarized his criticism; the mean difference (in comparison to the variance) lacked "ease of calculation, mathematical tractability and sampling simplicity" [33] (p. 223).

Kendall noted that if one tried to improve the mean difference formula by using the square of the difference in values, then that "is nothing but twice the variance" [8] (p. 42) since

$$\sum_{j \neq k} \Pr(x_j) \Pr(x_k)(x_j - x_k)^2 = \sum_{j,k} \Pr(x_j) \Pr(x_k)\left[x_j^2 - 2x_j x_k + x_k^2\right]$$
$$= \sum_{j=1}^{m} \Pr(x_j)x_j^2 - 2\left(\sum_{j-1}^{m} \Pr(x_j)x_j\right)\left(\sum_{k=1}^{m} \Pr(x_k)x_k\right) + \sum_{k=1}^{m} \Pr(x_k)x_k^2$$
$$= 2E(X^2) - 2E(X)^2 = 2Var(X).$$

Kendall further pointed out:

> This interesting relation shows that the variance may in fact be defined as half the mean square of all possible variate differences, that is to say, without reference to deviations from a central value, the mean. [8] (p. 42)

This means that C. R. Rao's quadratic entropy for the distance function $d_{jk} = (x_j - x_k)^2$ is twice the variance. We noted previously that the consideration of pairs could take the form of just pairs $\{x_j, x_k\}$ or ordered pairs (x_j, x_k) (where $j \neq k$). Hence, the variance may be defined using just simple pairs:

$$Var(X) = \sum_{j < k} \Pr(x_j) \Pr(x_k)(x_j - x_k)^2$$

The "interesting relation" also extends to the covariance. Suppose there are two random variables X with values x_i for $i = 1, \ldots, n$ and Y with values y_j for $j = 1, \ldots, m$ with probabilities given by a joint distribution $p(x_i, y_j) : X \times Y \to \mathbb{R}$. The two-samples or two-draws methodology gives two ordered pairs (x_i, y_j) and $\left(x_{i'}, y_{j'}\right)$, so the double-variance formula $\sum_{j \neq k} \Pr(x_j) \Pr(x_k)(x_j - x_k)^2$ generalizes to

$$\sum_{(i,j) \neq (i',j')} p(x_i, y_j) p\left(x_{i'}, y_{j'}\right)(x_i - x_{i'})\left(y_j - y_{j'}\right)$$

which is similarly equal to twice the covariance $Cov(X, Y) = E(XY) - E(X)E(Y)$.

Since $(x_i - x_{i'})\left(y_j - y_{j'}\right) = 0$ if $i = i'$ or $j = j'$, we can sum over all i, j. Abbreviating $p(x_i, y_j) = p_{ij}$, we have

$$\sum_{i,j,i',j'} p_{ij} p_{i'j'}(x_i - x_{i'})\left(y_j - y_{j'}\right) = \sum_{i,j,i',j'} p_{ij} p_{i'j'}\left[x_i y_j - x_i y_{j'} - x_{i'} y_j + x_{i'} y_{j'}\right]$$
$$= \sum_{i,j,i',j'} p_{ij} p_{i'j'} x_i y_j - \sum_{i,j,i',j'} p_{ij} p_{i'j'} x_i y_{j'} - \sum_{i,j,i',j'} p_{ij} p_{i'j'} x_{i'} y_j + \sum_{i,j,i',j'} p_{ij} p_{i'j'} x_{i'} y_{j'}.$$

Then, using

$$\sum_{i,j,i',j'} p_{ij} p_{i'j'} x_i y_j = \sum_{i,j} p_{ij} x_i y_j \sum_{i',j'} p_{i'j'} = \sum_{ij} p_{ij} x_i y_j = E(XY), \text{ and}$$
$$\sum_{i,j,i',j'} p_{ij} p_{i'j'} x_i y_{j'} = \sum_{i,j'} x_i y_{j'} \sum_{i'} \sum_j p_{ij} p_{i'j'} = \sum_{i,j'} x_i y_{j'} \sum_{i'} p_i p_{i'j'}$$
$$= \sum_{i,j/} p_i x_i y_{j'} p_{j'} = \left(\sum_i p_i x_i\right)\left(\sum_{j'} p_{j'} y_{j'}\right) = E(X)E(Y)$$

and similarly for the other cases, we have

$$\sum_{(i,j) \neq (i',j')} p(x_i, y_j) p\left(x_{i'}, y_{j'}\right)(x_i - x_{i'})\left(y_j - y_{j'}\right)$$
$$= E(XY) - E(X)E(Y) - E(Y)E(X) + E(XY) = 2Cov(X, Y).$$

Using the lexicographical ordering of the ordered pairs of indices (i.e., ordering according to first index, or if $i = i'$, then according to the second index), we have

$$\sum_{(i,j) < (i',j')} p(x_i, y_j) p\left(x_{i'}, y_{j'}\right)(x_i - x_{i'})\left(y_j - y_{j'}\right) = Cov(X, Y).$$

4. Discussion

These results shed new light on the variance in terms of the fundamental duality that starts with the two dual logics of subsets and partitions, and extends throughout the exact sciences [4]. The aspects discussed in this paper are given in Table 2.

Table 2. Parts of the fundamental duality discussed here.

Fundamental Duality	Subset or Element Side	Partition or Distinction Side
Its & Dits	Elements of subsets	Distinctions of partitions
Logic	Subset logic $\wp(U)$	Partition logic $\Pi(U)$
"Creation stories"	Ex Nihilo $\varnothing \subseteq U$	Big Bang $1 \cong \mathbf{0}_U \precsim \mathbf{1}_U \cong U$
Quantitative versions	Probability $\sum_{u_i \in S} p_i$	Logical entropy $\sum_{(u_j, u_k) \in \text{dit}(\pi)} p_j p_k$
Sampling	1-draw	2-draw (with replacement)
Random variable X	Mean $\sum_i p_i x_i = E(X)$	Variance $\sum_{j<k} p_j p_k \left(x_j - x_k \right)^2 = Var(X)$

The key point in Table 2 is that in terms of the subsets–partitions duality, the notions of mean and variance are *dual* concepts. Perhaps a better understanding of this would be to consider some of the recent literature developing notions of "logical entropy" motivated ultimately by the notion of a fuzzy set [34] or a subspace and the related algebras, e.g., fuzzy algebras, quantum logics, MV algebras, effect algebras, and D-posets ([35–43]).

The striking thing about all these developments is that their origin is on the subset side of the subsets–partitions duality in spite of the formulas called "logical entropy." Subsets linearize to subspaces, so the Birkhoff-von-Neumann-type quantum logics [44] are just the vector (Hilbert) space versions of the logic of subsets. Partitions, in turn, linearize to direct-sum decompositions (DSDs) of vector spaces so the quantum logic on the partition side of the duality would be the logic of DSDs of Hilbert spaces [45].

A subset of U is determined by its characteristic function $\chi : U \to \{0, 1\}$ and a fuzzy subset of U is just the extension to allow for a continuum of membership values $f : U \to [0, 1]$ in the unit interval, so it is clearly on the subset side of the duality.

Logical entropy is properly defined as the quantification of the logic of partitions by the normalized cardinality of the ditset of a partition, where a dit is an ordered pair of elements in different blocks (or equivalence classes). As Kolmogorov emphasized [24], information should be definable without reference to probabilities, e.g., the ditset definition of information-as-distinctions [21]. When a probability distribution is defined on the points of U, then logical entropy as a probability measure is the product measure of the ditset. When the partition is the discrete partition $\mathbf{1}_U$, then the logical entropy is defined using only the probability distribution as $h(p) = \sum_i p_i(1 - p_i)$. That is how the "logical entropy" formulas are defined in the aforementioned algebras all developed from concepts on the subset side of the duality, e.g., fuzzy subsets. Those formulas using probability-like concepts in those algebras may or may not have useful properties but they do not derive from the information (as distinctions) theory side of the subsets–partitions duality.

A fuzzy subset has been described as "non-sharp, non-crisp, and smudged" [34] (p. 7), which are all adjectives used to describe the key non-classical notion in quantum mechanics, namely a superposition state. But it would nevertheless be a mistake to try to model a superposition state as a fuzzy subset. A fuzzy subset is "fuzzy" about which elements of the universe set are in the subset, but it is clear which eigenvectors in an orthonormal basis of an observable are in a superposition state (namely those with non-zero coefficients). The superposition wipes away (or abstracts from) the distinctions between the eigenvectors in

the superposition and those indistinctions, called "quantum coherences" [46] (p. 177) in quantum mechanics, are modeled by an equivalence class in an equivalence relation, i.e., a block in a partition [4].

5. Conclusions

We see the one-draw methodology in the notion of the probability of a subset (the one-draw-from-U probability of obtaining an element of a subset S) and the two-draw methodology in the notion of the logical entropy of a partition (the two-draw-from-U probability of obtaining a distinction of a partition π). Applied to a random variable $X : U \to \mathbb{R}$, the one-draw method gives the mean, and the two-draw method gives the variance.

The quantitative versions of the two logics associates the one-draw method with probability theory and the two-draw method with information theory—which implies that the variance should be seen as an information-theoretic concept, as seen in Table 2. Indeed, if we use the logical distance function $1 - \delta_{jk}$ squared instead of the Euclidean distance squared, then that "logical variance" *is* the logical entropy, i.e., $\sum_{j,k=1}^{n} p_j p_k \left(1 - \delta_{jk} \right)^2 = h(p)$. The usual definition of the variance $Var(X) = E\left[(X - E(X))^2 \right]$ gives no hint of the two-draw approach.

In terms of the element–distinction (or subsets–partitions) duality, the one-draw expectation of the value of an element x_i is its mean, and the two-draw expectation of the value of a distinction $x_i - x_k$ squared is the variance (and similarly for the covariance). This is the new logical basis for the variance that positions it as *dual to the mean* (Table 2) and on the partitions/information theory side of the fundamental subsets–partitions duality.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

Abbreviations

The following abbreviation is used in this manuscript:

RST　　reflexive–symmetric–transitive

References

1. Ellerman, D. The Logic of Partitions: Introduction to the Dual of the Logic of Subsets. *Rev. Symb. Log.* **2010**, *3*, 287–350. [CrossRef]
2. Ellerman, D. An Introduction of Partition Logic. *Log. J. IGPL* **2014**, *22*, 94–125. [CrossRef]
3. Ellerman, D. *The Logic of Partitions: With Two Major Applications. Studies in Logic 101*; College Publications: London, UK, 2023. Available online: https://www.collegepublications.co.uk/logic/?00052 (accessed on 13 June 2025).
4. Ellerman, D. A Fundamental Duality in the Exact Sciences: The Application to Quantum Mechanics. *Foundations* **2024**, *4*, 175–204. [CrossRef]
5. Boole, G. *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*; Macmillan and Co.: Cambridge, UK, 1854.
6. Kung, J.P.S.; Rota, G.-C.; Catherine, H.Y. *Combinatorics: The Rota Way*; Cambridge University Press: New York, NY, USA, 2009.
7. Zhang, Y.; Wu, H.; Cheng, L. Some New Deformation Formulas about Variance and Covariance. In Proceedings of the 2012 International Conference on Modelling, Identification and Control (ICMIC2012), Wuhan, China, 24–26 June 2012; pp. 987–992.
8. Kendall, M.G. *Advanced Theory of Statistics Vol. I*; Charles Griffin & Co.: London, UK, 1945.
9. Eilenberg, S.; Lane, S.M. General Theory of Natural Equivalences. *Trans. Am. Math. Soc.* **1945**, *58*, 231–294. [CrossRef]
10. Lawvere, F.W.; Rosebrugh, R. *Sets for Mathematics*; Cambridge University Press: Cambridge, UK, 2003.

11. Finberg, D.; Mainetti, M.; Rota, G.-C. The Logic of Commuting Equivalence Relations. In *Logic and Algebra*; Ursini, A., Agliano, P., Eds.; Marcel Dekker: New York, NY, USA, 1996; pp. 69–96.

12. Britz, T.; Mainetti, M.; Pezzoli, L. Some operations on the family of equivalence relations. In *Algebraic Combinatorics and Computer Science: A Tribute to Gian-Carlo Rota*; Crapo, H., Senato, D., Eds.; Springer: Milano, Italy, 2001; pp. 445–459.

13. Birkhoff, G. *Lattice Theory*; American Mathematical Society: New York, NY, USA, 1948.

14. Grätzer, G. *General Lattice Theory*, 2nd ed.; Birkhäuser Verlag: Boston, MA, USA, 2003.

15. Ainsworth, T. Form vs. Matter. In *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition)*; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2016. Available online: https://plato.stanford.edu/archives/spr2016/entries/form-matter/ (accessed on 13 June 2025).

16. Bateson, G. *Mind and Nature: A Necessary Unity*; Dutton: New York, NY, USA, 1979.

17. Bennett, C.H. Quantum Information: Qubits and Quantum Error Correction. *Int. J. Theor. Phys.* **2003**, *42*, 153–176. [CrossRef]

18. Wilkins, J. Mercury: Or the Secret and Swift Messenger. In *The Mathematical and Philosophical Works of the Right Rev. John Wilkins, Vol. II*; C. Whittingham: London, UK, 1802; pp. 1–87.

19. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]

20. Gleick, J. *The Information: A History, A Theory, A Flood*; Pantheon: New York, NY, USA, 2011.

21. Ellerman, D. *New Foundations for Information Theory: Logical Entropy and Shannon Entropy*; Springer Nature: Cham, Switzerland, 2021. [CrossRef]

22. Manfredi, G. Logical Entropy—Special Issue. *4Open* **2022**, *5*, E1. [CrossRef]

23. Ellerman, D. A New Logical Measure for Quantum Information. *Quantum Inf. Comput.* **2025**, *25*, 81–95. [CrossRef]

24. Kolmogorov, A.N. Combinatorial Foundations of Information Theory and the Calculus of Probabilities. *Russ. Math. Surv.* **1983**, *38*, 29–40. [CrossRef]

25. Gini, C. *Variabilità e Mutabilità*; Tipografia di Paolo Cuppini: Bologna, Italy, 1912.

26. Good, I.J.A.M. Turing's statistical work in World War II. *Biometrika* **1979**, *66*, 393–396. [CrossRef]

27. Rejewski, M. How Polish Mathematicians Deciphered the Enigma. *Ann. Hist. Comput.* **1981**, *3*, 213–234. [CrossRef]

28. Kullback, S. *Statistical Methods in Cryptanalysis*; Aegean Park Press: Walnut Creek, CA, USA, 1976.

29. Simpson, E.H. Measurement of Diversity. *Nature* **1949**, *163*, 688. [CrossRef]

30. Rao, C.R. Gini-Simpson Index of Diversity: A Characterization, Generalization and Applications. *Util. Math. B* **1982**, *21*, 273–282.

31. Good, I.J. Comment on Patil and Taillie: Diversity as a Concept and its Measurement. *J. Am. Stat. Assoc.* **1982**, *77*, 561–563.

32. Rao, C.R. Diversity and Dissimilarity Coefficients: A Unified Approach. *Theor. Popul. Biol.* **1982**, *21*, 24–43. [CrossRef]

33. Kendall, M.G. Review of *Variabilita e Concentrazione*. By Corrado Gini. *J. R. Stat. Soc. Ser. A (Gen.)* **1957**, *120*, 222–223. [CrossRef]

34. Riečan, B.; Neubrunn, T. *Integral, Measure, and Ordering*; Springer Science + Business Media: Bratislava, Slovakia, 1997.

35. Markechová, D.; Riečan, B. Logical Entropy of Fuzzy Dynamical Systems. *Entropy* **2016**, *18*, 157. [CrossRef]

36. Mohammadi, U. The Concept of Logic Entropy on D-Posets. *Algebr. Struct. Their Appl.* **2016**, *3*, 53–61.

37. Ebrahimzadeh, A. Logical Entropy of Quantum Dynamical Systems. *Open Phys.* **2016**, *14*, 1–5. [CrossRef]

38. Markechová, D.; Riečan, B. Logical Entropy and Logical Mutual Information of Experiments in the Intuitionistic Fuzzy Case. *Entropy* **2017**, *19*, 429. [CrossRef]

39. Ebrahimzadeh, A.; Jamalzadeh, J. Conditional Logical Entropy of Fuzzy σ-Algebras. *J. Intell. Fuzzy Syst.* **2017**, *33*, 1019–1026. [CrossRef]

40. Giski, Z.E.; Ebrahimzadeh, A. An Introduction of Logical Entropy on Sequential Effect Algebra. *Indag. Math.* **2017**, *28*, 928–937. [CrossRef]

41. Ebrahimzadeh, A.; Giski, Z.E.; Markechová, D. Logical Entropy of Dynamical Systems—A General Model. *Entropy* **2017**, *5*, 4. [CrossRef]

42. Markechová, D.; Ebrahimzadeh, A.; Giski, Z.E. Logical Entropy of Dynamical Systems. *Adv. Differ. Equ.* **2018**, *70*, 1–17. [CrossRef]

43. Markechová, D.; Mosapour, B.; Ebrahimzadeh, A. Logical Divergence, Logical Entropy, and Logical Mutual Information in Product MV-Algebras. *Entropy* **2018**, *20*, 129. [CrossRef]

44. Birkhoff, G.; Neumann, J. von The Logic of Quantum Mechanics. *Ann. Math.* **1936**, *37*, 823–843. [CrossRef]

45. Ellerman, D. The Quantum Logic of Direct-Sum Decompositions: The Dual to the Quantum Logic of Subspaces. *Log. J. IGPL* **2018**, *26*, 1–13. [CrossRef]

46. Auletta, G.; Fortunato, M.; Parisi, G. *Quantum Mechanics*; Cambridge University Press: Cambridge, UK, 2009.