Research Statement

Tobias Fritz

March 22, 2026

Preface. This document is written for a mixed audience of computer scientists, mathematicians, and physicists. Since my research is topically broad, the statement is longer than typical, but its sections can be read independently.

Two recurring features characterize my research:

- **Interdisciplinarity**: I often develop mathematical frameworks for problems motivated by physics, computer science, or engineering. These problems are usually not well known to mathematicians, but they tend to exhibit interesting mathematical structures whose clarification and distillation I view as a central task.
- **Pioneering nature**: Several of the tools and methods I have developed have since seen broad adoption. Once the core ideas are in place, I often move on to new topics.

Contents

1 Operator algebras and quantum information theory

Towards the end of my PhD, I realized that an open problem in quantum information theory known as **Tsirelson's problem** is closely related to **Connes' embedding problem** in operator algebras [1], in the sense that the latter conjecture implies the former. This connection between the two problems was proven independently and concurrently by a team of six established quantum information theory researchers [2]. Ozawa subsequently proved the converse direction, thus establishing that the two conjectures are in fact equivalent [3].

I also had the hope that these conjectures could be disproven by methods from computability theory. This resulted in some partial results joint with Netzer and Thom [4]. The connection between Tsirelson's problem and Connes' embedding problem as well as the idea of using computability theory to disprove these conjectures is indeed what was realized in the 2020 breakthrough solution $\mathbf{MIP}^* = \mathbf{RE}$ by Ji et al. [5].[1] That being said, their undecidability argument is more sophisticated in its complexity-theoretic aspects than what I had been able to anticipate. The idea behind my uncomputability idea was recently confirmed by Goldbring and Sinclair [7], who showed that it follows from [5].

Goldbring's recent surveys [8, 9] provide a good overview of these developments and the significance of my contributions. Other overviews include the theses [10, 11]. These ideas have since become part of the research area of **nonlocal games** at the intersection of quantum information theory and operator algebras.[2] The strong results developed by others in this area have enable me to derive the undecidability of propositional quantum logic as a corollary [12].

[1] Recently, this result has also been shown to imply a negative solution to the Aldous–Lyons conjecture on random graphs [6].

[2] See e.g. recent workshops like Quantum information, quantum groups and operator algebras (Cambridge, 2024), Operator Algebras and Quantum Information (Waterloo, 2025), Workshop on Quantum Information and Operator Systems (Oslo, 2025), Noncommutative Harmonic Analysis and Quantum Information, (Oberwolfach, 2026), and Operator Algebras and Quantum Information (Mittag-Leffler Institute, 2026).

2 Characterizations of entropy

Entropy is a fundamental but mysterious concept in many areas of science. In order to clarify its meaning, it is natural to ask for a characterization of entropy in terms of simple axioms. My work with Baez and Leinster on characterizations of information measures [13, 14] provides the perhaps simplest **axiomatic characterizations of Shannon entropy and relative entropy (Kullback-Leibler divergence)** known to date. This simplicity is achieved through the use of categorical language, which has revealed that entropic quantities can meaningfully be considered as functors:

Theorem 2.1 ([13]). *Let* FinProb *be the category of finite probability spaces* (X, p) *and measure-preserving maps* $f : (X, p) \to (Y, q)$. *Suppose* F *is any map sending morphisms in* FinProb *to numbers in* $[0, \infty)$ *and obeying these three axioms:*

(a) Functoriality:

$$F(f \circ g) = F(f) + F(g)$$

whenever f, g *are composable morphisms.*

(b) Convex linearity:

$$F(\lambda f \oplus (1 - \lambda)g) = \lambda F(f) + (1 - \lambda)F(g)$$

for all morphisms f, g *and scalars* $\lambda \in [0, 1]$.

(c) Continuity: F *is continuous.*

Then there exists a constant $c \geq 0$ *such that for any morphism* $f : (X, p) \to (Y, q)$ *in* FinProb*,*

$$F(f) = c(H(p) - H(q)),$$

where $H(p)$ *is the Shannon entropy of* p*. Conversely, for any constant* $c \geq 0$*, this formula determines a map* F *obeying conditions (a)–(c).*

While this was proven by a reduction to an earlier characterization of Shannon entropy due to Faddeev, our characterization of relative entropy in [14] had to be proven from scratch. These two papers have resulted in a substantial number of follow-up works by other authors who connect category theory and information theory, including [15–30].

3 Causal inference and Bell's theorem

Bell's theorem in the foundations of quantum mechanics is a mathematical result which shows that certain predictions of quantum mechanics cannot be reproduced by any theory that is based on local hidden variables. Therefore, in order to reproduce the predictions of quantum mechanics by means of standard probability theory, one needs to allow for nonlocal interactions, which is unattractive since such interactions have not been observed in nature.

In [31, 32][3], I developed a new mathematical framework for Bell's theorem, which shows that it is an instance of a type of problem known as **causal inference with hidden variables**. Based on this, I derived new impossibility results analogous to Bell's theorem but with other causal structures. These results clarify that one does not need a notion of "agent" or "free will" in order to apply Bell-type theorems to rule out local hidden variable theories, in contrast to what is sometimes claimed in the literature. My main example of this lives on the "triangle scenario" causal structure. The probability distribution underlying this example has become known as the "Fritz distribution" [33–40]. See also e.g. [41–44] for further work on the triangle scenario and [45–47] for recent experimental tests.

In [48], and together with my physics colleagues Elie Wolfe and Rob Spekkens, I built on these insights into causal inference in order to develop the **inflation technique**, which is the first generally applicable method for solving causal inference problems with hidden variables. Navascués and Wolfe later showed that this method provides necessary and sufficient conditions [49]. Among many other follow-up works, there is now a generalization to quantum causal inference [50] as well as a Python package[4] for both our classical and the more recent quantum version of inflation [51].

[3] [31] was selected for the journal's *Highlights of 2012*.

[4] Installable via `pip install inflation`.

While the inflation technique has so far mainly been used by physicists working on quantum information theory and quantum foundations, the personal interactions I have had with the other participants in the research programme Causal inference: From theory to practice and back again indicate growing awareness in the causal inference community.

4 Resource theories and Vergleichsstellensätze for semirings

Over the past ~10 years, an active research area emerging out of quantum information theory has been that of **resource theories**. The basic idea behind this research is to develop a common mathematical framework for the analysis of resource efficiency for problems involving entanglement transformations, quantum thermodynamics, or magic states in quantum computation, to name a few examples [52, 53]. Moreover, in my view there is no reason to restrict this research to quantum information and quantum computation: the ultimate goal should be to develop a general theory of resource efficiency that applies to a wide range of problems in physics, engineering, computer science and economics. This should have classical results like Shannon's channel coding theorem and Carnot's theorem on the efficiency of heat engines as special cases, and provide an improved conceptual understanding of these theorems as well as a toolbox that allows for easier discovery of new results of this type.

Therefore my own contributions to this area focus primarily on the development of a mathematical toolbox that applies quite generally to problems of resource efficiency. My initial works in this direction consist of a general mathematical framework for resource considerations based on category theory [54][5] and the use of separation theorems from functional analysis to prove optimality of bounds on resource efficiency [55]. As a first application, this has facilitated the first comprehensive treatment of work and heat in quantum thermodynamics from the resource-theoretic perspective [56].

After that, I have been working on the development of deeper mathematical results on **preordered semirings**, motivated by certain intended applications that I will sketch below. This is based on the observation that many resource theories not only have a notion of combining resources, which tends to behave multiplicatively, but also a more mysterious notion of addition given by some kind of direct sum over which one has a distributive law. Hence one arrives at the algebraic structure of a preordered semiring as a natural setting for the study of resource theories, where an ordering relation $x \geq y$ means that x can be converted into y. This turns out to be a very fruitful perspective which allows for the development of deeper mathematical results in the form of separation theorems with powerful consequences. Historically, the first such result was Strassen's separation theorem for Archimedean preordered semirings [57], which he applied to give a theoretical formula for the complexity of matrix multiplication.[6]

Strassen's proof of his separation theorem was by a reduction to a separation theorem for ordered rings, namely the classical **Positivstellensatz** of Krivine–Kadison–Dubois. Following the tradition of naming such separation theorems by analogy with Hilbert's Nullstellensatz, I have coined the term **Vergleichsstellensatz** for the semiring versions, based on the fact that these kinds of results are about asymptotic comparisons between semiring elements[7]. Here is my first Vergleichsstellensatz for preordered semirings [58], which generalizes Strassen's result to a much wider class of preordered semirings.

Theorem. *Let S be a preordered semiring with $1 \geq 0$ and a power universal element u, and let $x, y \in S$ nonzero. Then the following are equivalent:*

(a) $\phi(x) \leq \phi(y)$ for all monotone homomorphisms $\phi : S \to \mathbb{R}_+$ or $\phi : S \to \mathbb{TR}_+$.

(b) For every $\varepsilon > 0$, we have

$$x^n \leq u^{\lfloor \varepsilon n \rfloor} y^n$$

for all $n \gg 1$.

Moreover, suppose that $\phi(x) < \phi(y)$ for all such ϕ. Then also the following hold:

*(c) **Asymptotic comparison I:** There is $k \in \mathbb{N}$ such that*

$$u^k x^n \leq u^k y^n \qquad \forall n \gg 1.$$

[5] Selected for *Computing Reviews 21st Annual Best of Computing*.

[6] I call it "theoretical" because actually computing this formula requires understanding the spectrum of the associated preordered semiring of tensors, which is a very difficult problem.

[7] German *vergleichen* = to compare.

(d) ***Asymptotic comparison II:*** *If y is power universal as well, then*

$$x^n \leq y^n \qquad \forall n \gg 1.$$

(e) ***Catalytic comparison:*** *There is nonzero $a \in S$ such that*

$$ax \leq ay.$$

Moreover, there is $k \in \mathbb{N}$ such that $a \coloneqq u^k \sum_{j=0}^n x^j y^{n-j}$ does the job for any $n \gg 1$.

Here, $\mathbb{TR}_+ = ([0,\infty], \max, \cdot)$ is the **tropical semiring**, and the monotone homomorphisms ϕ form the **test spectrum** of S. As the appearance of tropical points in this spectrum indicates, this Vergleichsstellensatz is genuinely semiring-theoretic, and one thus cannot expect to find a proof by reduction to a Positivstellensatz. Instead, my proof is based on a reduction to the semifield case by formally adjoining multiplicative inverses, where I have shown the following:

Theorem 4.1 ([58]). *(a) Let F be a preordered semifield. Then its preorder $\leq$ is the intersection of all total semifield preorders on F which extend it.*

(b) Let F be a totally preordered semifield. Then F is multiplicatively Archimedean if and only if it order embeds into one of the following preordered semifields:

$$\mathbb{R}_+, \qquad \mathbb{R}_+^{\mathrm{op}}, \qquad \mathbb{TR}_+, \qquad \mathbb{TR}_+^{\mathrm{op}}.$$

The proof of (a) in turn is structurally congruent to standard proofs in real algebra—in particular to Artin's proof of the analogous result for ordered fields—but very different and more demanding in the technical details. For example, the following peculiar polynomial identity plays a key role:

Lemma 4.2. *For $n \in \mathbb{N}$, let $\underline{A} = (A_0, \ldots, A_n)$ and $\underline{B} = (B_0, \ldots, B_n)$ be finite sequences of variables. Then in the semiring $\mathbb{N}[\underline{A}, \underline{B}, X, Y]$, the polynomial*

$$\sum_{k=0}^{n} \left(A_k \sum_{i=0}^{n} B_i X^i + B_k \sum_{i=0}^{n} A_i Y^i \right) \left(\sum_{j=1}^{k} X^{k-j} Y^{j-1} \right).$$

is invariant under $X \leftrightarrow Y$.

In the follow-up paper [59][8], I have proven further generalizations of the above Vergleichsstellensatz, where the assumption $1 \geq 0$ is dropped. This is relevant especially for applications to probability and (classical or quantum) information theory, where the semiring elements are often unnormalized measures or states, but only semiring elements of the same normalization are comparable. The Vergleichsstellensätze of [59] are more technical to state, but they have the same general flavor as the one above. They are arguably my technically most difficult results to date.

The applications worked out so far concern problems where the test spectrum can be explicitly determined, so that the separation results have concrete consequences. Since this is often not easy to do in the setting of resource theories, these existing applications have partly been to questions that are not directly motivated by resource theories. In particular, my first Vergleichsstellensatz has been applied to the semiring of representations of $SU(n)$ ordered under inclusion, where it gives necessary and generically sufficient conditions for asymptotic and catalytic containment of representations [60]. Unexpectedly, (b) has also been used recently as one ingredient in the proof of a characterization of the category of Hilbert spaces and bounded operators [61].

My deeper Vergleichsstellensätze from [59] have also been applied in various ways. My own first application was to a characterization of asymptotic stochastic dominance for random walks, where they specialize to a new relative version of Cramér's large deviation theorem [62]. Together with quantum information theorists from Singapore, I have also applied them to asymptotic matrix majorization [63], where we confirmed an earlier conjectural characterization due to a group of economists [64, Section 6]. My coauthors have since developed further results based on my Vergleichsstellensätze [65, 66], including recent work on information measures [67]. They are currently running an international reading group on my Vergleichsstellensatz papers.

In the future, I would like to extend these Vergleichsstellensätze to preordered semirings with additional structure that allows for the possibility of considering approximate resource conversions, given that approximations play a central role in many results on asymptotic resource efficiency, such as Shannon's channel coding theorem.

[8] Although this is still only a preprint, it has received an enthusiastic referee report and is currently in the second round of review.

5 Categorical probability and statistics with Markov categories

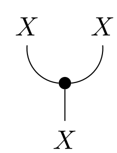
For the past ~7 years, much of my research has focused on the development of a new mathematical framework for probability theory and statistics based on category theory. My main guiding idea behind this endeavour is that working with the measure-theoretic formalism is akin to programming a computer in machine code, while a categorical framework can provide a more usable higher-level language in which much of the same theory can be developed. The advantages of doing so are threefold:

- ▷ Just as in programming, utilizing a higher-level language allows the reasoner to focus on the essential aspects of a problem while disregarding irrelevant implementation details. This should ultimately allow for the treatment of problems of higher complexity than what would be possible in the original framework.
- ▷ Greater generality:[9] The categorical framework can be seen as a general setting for theories of uncertainty which includes probability theory, but similarly applies to other theories of uncertainty as well. For example, there are various flavours of possibility theory, dynamical versions of probability theory involving group actions, and Gaussian probability. By imposing axioms of varying strength, one obtains results that apply to a range of different theories of uncertainty. This generality and flexibility can be exploited by the user in order to help them determine which theory of uncertainty is best suited for their particular problem.
- ▷ As I will argue below, the categorical perspective can also provide new philosophical insights into the nature of probability.

In this section, I will present the theory of Markov categories as providing such a higher-level framework and sketch some of the results that we have obtained in this framework. To follow the technical exposition, familiarity with symmetric monoidal categories will be helpful. Nevertheless, I hope that all readers will be able to get a flavour of what Markov categories are and what one can do with them.

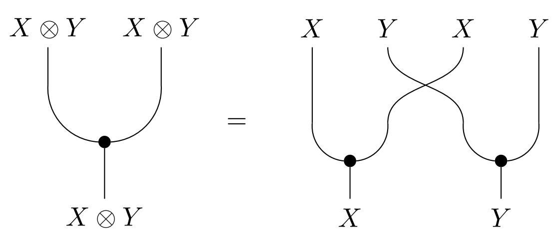
Definition 5.1 ([68])**.** *A **Markov category** C is a symmetric monoidal category, with monoidal unit I which is terminal, and where every object $X \in \mathsf{C}$ is equipped with a distinguished morphism*

copy_X =



(5.1)

which, together with the unique morphism $\mathrm{del}_X \colon X \to I$*, makes X into a commutative comonoid, and such that*



(5.2)

for all $X, Y \in \mathsf{C}$.

Modulo minor variations, Markov categories were rediscovered independently a number of times.[10] However, my paper [68] has apparently been the first to recognize their potential as a higher-level framework for probability and statistics in which one can conduct nontrivial reasoning and prove theorems. This insight has popularized the definition under the name of Markov categories.

Example 5.2. The particular Markov category we use as a model for measure-theoretic probability theory is $\mathsf{BorelStoch}$. In this category, the objects are standard Borel measurable spaces and a morphism $f : X \to Y$

[9] This point is quite similar to how the arithmetic of integers and real numbers has been generalized with enormous benefit to the theory of commutative rings and fields. The latter provides a more general context for the study of number systems, and also offers deep results—such as the fundamental theorem of Galois theory—which can be applied back to number systems and produces conclusions that would not have been obtainable without.

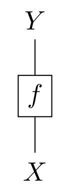
[10] See the introduction to [69] for an account of the history.

is a **Markov kernel**,[11] which is a family of probability measures on Y indexed by the points of X in a measurable way. When X and Y are finite, Markov kernels are equivalent to **stochastic matrices**, which are matrices of nonnegative real numbers

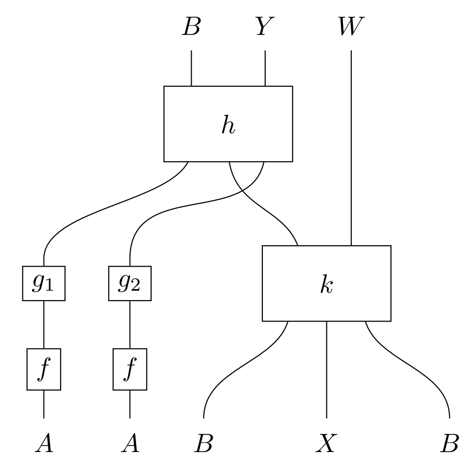
$$(f(y|x))_{x\in X, y\in Y}$$

such that $\sum_{y\in Y} f(y|x) = 1$ for all $x \in X$.

Definition 5.1 and Example 5.2 indicate a paradigm shift in probability theory. In the traditional approach, the primary primitive notions of probability are probability measures and random variables. For us, the primary notion is that of a morphism $f : X \to Y$ in a Markov category[12], drawn like this:



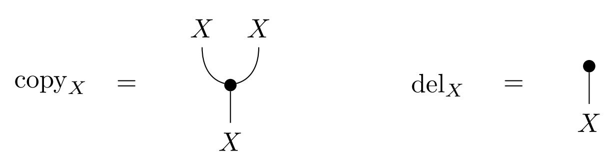
Here, we think of f as some kind of operation that takes an *input* of type X and produces an *output* of type Y. This output should be thought of as possibly uncertain; you may think of a vending machine, which can take one of several coins as input and will produce a random snack (or failure) as output. The structure of a symmetric monoidal category then amounts to the ability to compose an entire network of morphisms into a single morphism, which is drawn as a **string diagram** like this:



(5.3)

The axioms guarantee that these wirings behave as one would expect, e.g. in the sense that they respect simple topological moves. The Markov kernels of BorelStoch provide one possible model for which all of this works. Note that in this case, composing Markov kernels into a network is exactly what one does in the context of causal modelling with Bayesian networks [70, (1.33)]. While this is most commonly formulated for discrete probability, the Markov category framework covers the general measure-theoretic case with the same ease.

I still need to clarify the meaning of the black dots in Definition 5.1. For every object X, there is a distinguished process with one input of type X and two outputs of type X, and another one with input of type X and no output, drawn like this:



We think of these as maps which **copy** and **delete** (or discard) the value that is passed to them. The postulates of Definition 5.1 formalize this intended interpretation. The fact that some form of "copy" is relevant in

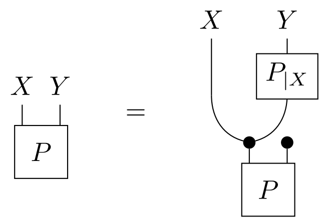
[11]Markov kernels are also known under many different names, such as *stochastic map*, *probabilistic mapping*, *conditional distribution*, *statistical model*, *statistical operation*, *transition probability* and *communication channel*. This proliferation of names for the same concept in different contexts underlines their importance and ubiquity.

[12]Of course, maps appear because our approach is formulated in terms of category theory. However, the direction of explanation here is: category theory is adequate for our approach *because* we take maps to be primitive, not the other way around.

probability theory becomes apparent when considering the defining equation of conditional probabilities, which in the discrete case is

$$P(x, y) = P(y|x)P(x).$$

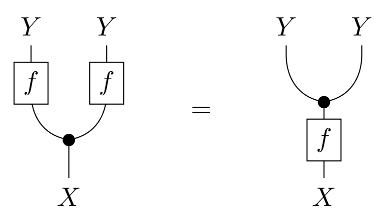
Since x occurs twice on the right, one needs to copy its value in order for the expression to make sense. In terms of string diagrams, this same equation takes the form[13]



The expression on the left has no inputs and two outputs, which makes it into an abstract version of a **probability measure** on the product of measurable spaces X and Y. On the right, we have the marginal on X at the bottom, whose output is copied, and one copy is used as input to $P_{|X}$ in order to determine the second overall output on the top right. Now if $P_{|X}$ exists such that the above equation holds, then we call $P_{|X}$ a **conditional** of P.[14] In BorelStoch, it is well-known that such $P_{|X}$ always exists, because the above equation for $P_{|X}$ turns out to be equivalent to it being a **regular conditional probability** for P.

There are many further notions that have been defined for Markov categories, including almost sure equality, conditional independence, sufficient statistics, etc. Based on these definitions, we also have a range of additional axioms that a given Markov category may or may not satisfy, and which make the Markov category behave more like measure-theoretic probability if they do hold. Let me limit myself to mentioning only one of the definitions.

Definition 5.3. *A morphism $f : X \to Y$ is **deterministic** if it satisfies*



Indeed in BorelStoch, a Markov kernel $f : X \to Y$ is deterministic if and only if all its probabilities are in $\{0, 1\}$, in which case one can identify it with a measurable function $X \to Y$ with no inherent randomness. Here's the intuition behind the general definition: being deterministic means that whenever one feeds the same input to two copies of f independently (as on the left), then these two processes are guaranteed to produce copies of the same output, namely the output produced by a single application of f (as on the right).

I will now present a selection of recent achievements with Markov categories made by myself and my collaborators. These lift classical theorems of probability and statistics to our abstract setting.

The zero–one law of Hewitt and Savage. A *zero–one law* is a result which states that an event with certain properties is deterministic (has probability 0 or 1). In [71], we have proven abstract generalizations of the classical zero–one laws due to Kolmogorov and Hewitt–Savage. The latter is as follows.

Theorem 5.4 (Abstract Hewitt–Savage zero–one law [71])**.** *Suppose that C is a causal Markov category, J an infinite set and $X \in \mathsf{C}$ is such that the Kolmogorov power X^J exists. If $p : A \to X^J$ and deterministic $s : X^J \to T$ are such that*

(a) p displays the conditional independence $\perp_{i \in J} X_i \mid\mid A$, and

(b) Both p and s are invariant under finite permutations of the factors in the power X^J,

then the composite $sp : A \to T$ is deterministic.

[13]Keep in mind that X is an object in our category, i.e. a formal version of a measurable space, *not* a random variable.
[14]Whenever a conditional exists, it is unique up to almost sure equality, even at the general level of our formalism.

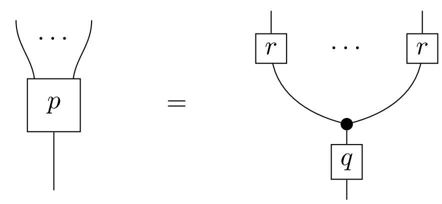
Here, the Kolmogorov power X^J is a J-fold **Kolmogorov product** of X with itself, defined as the cofiltered limit of the finite powers of X, provided that this limit exists and is suitably well-behaved. Instantiating this result in $\mathsf{BorelStoch}$ with J countable, A a singleton, and $T = \{0, 1\}$ recovers the classical Hewitt–Savage zero–one law from [72].[15] Our result generalizes this classical version in three ways:

- ▷ p can be a nontrivial family of probability measures indexed by A, or equivalently a Markov kernel with arbitrary input, rather than merely a single probability measure.
- ▷ s can be a measurable map taking values in any standard Borel space rather than just $\{0, 1\}$.
- ▷ The theorem can also be interpreted in other Markov categories. Depending on which category one considers, the meaning of the statement may end up having a similar flavour or may end up looking radically different.

The first two generalizations are insubstantial, since they can also easily be deduced from the standard formulation mentioned above. The third generalization is not immediate, and its scope is largely unexplored so far.

The de Finetti and Aldous–Hoover theorems. The de Finetti theorem is a classical foundational result which characterizes sequences of random variables, whose distribution is invariant under finite permutations, as mixtures of independent and identically distributed sequences. Here is our categorical generalization.

Theorem 5.5 (Abstract de Finetti theorem [73])**.** *Let C be a Markov category which is a.s.-compatibly representable and has conditionals and countable Kolmogorov products. If a morphism $p : A \to X^{\mathbb{N}}$ is invariant under finite permutations, then it can be written in the form*



(5.4)

for suitable q and r.

Instantiating this result in $\mathsf{BorelStoch}$ and taking A to be the singleton space recovers the classical de Finetti theorem for standard Borel spaces. The appearance of further new proofs of this old theorem in recent literature, even in the simpler special cases of discrete or binary variables [74–76], indicates that there still is demand for a conceptually clean proof like ours. As for Theorem 5.4, our result generalizes the classical one in two ways: by allowing A to be nontrivial (insubstantial), and by considering other Markov categories (significant). For example, Theorem 5.5 has subsequently been used by Forré to obtain a de Finetti theorem for his quasi-measurable spaces [77, Corollary 5.47].

In our more recent work [78], we have proven a deeper categorical result which generalizes the classical **Aldous–Hoover theorem**. This similarly characterizes two-dimensional *arrays* of random variables whose distribution is invariant under finite permutations of rows and columns (separately), such as countably infinite symmetric random bipartite graphs.

Contemporary results in a similar vein are the **hierarchical de Finetti theorems** of Austin and Panchenko [79] and Jung et al. [80]. I expect that our methods will be able to tackle these as well and suggest further generalizations. In a more speculative direction, it could be intriguing to leverage the greater generality of the categorical formalism to explore the known similarities [81] between de Finetti theorems for partial exchangeability and regularity results in extremal combinatorics such as the Szemerédi regularity lemma. Could Markov categories provide a way to formalize these informal analogies, potentially leading to further synergy between extremal combinatorics and the theory of random structures? I imagine—as a dream scenario—to be able to reprove Szemerédi's lemma by instantiating our abstract Aldous–Hoover theorem in a suitable Markov category. If this turns out to be possible, then we gain a formal bridge between probability and extremal combinatorics, which could be fruitful for both fields.

[15] The knowledgeable reader may point out that Hewitt and Savage had already proven a version that applies more generally than just to standard Borel spaces. So far, we have not been able to obtain this more general version from Theorem 5.4, but we also have not tried very hard since the standard Borel case already covers most applications.

Laws of large numbers. Together with the central limit theorem, the weak and strong laws of large numbers are arguably the most fundamental results of probability theory. Here is the version that is relevant for many applications.

Theorem 5.6 (Kolmogorov's strong law)**.** *Let $(X_n)_{n \in \mathbb{N}}$ be an iid (independent and identically distributed) sequence of random variables with finite first moment. Then, almost surely,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}[X_1]. \tag{5.5}$$

Clearly this statement makes use of the assumption that the X_i take values in a vector space, since otherwise we could not even make sense of the sum. But the same flavour of result can even be achieved with a more minimal amount of structure by considering the **empirical distribution** of a sequence of real numbers (X_n), which is encoded in the partially defined Markov kernel[16]

$$\mathsf{es}((-\infty, r]|(x_n)) \coloneqq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,r]}(x_i). \tag{5.6}$$

Here, we consider this to be defined on a given sequence (x_n) if the limit exists uniformly in r, in which case it defines a unique probability measure on $\mathbb{R}$. Then the following classical result is a non-linear version of a strong law.

Theorem 5.7 (Glivenko–Cantelli)**.** *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent and identically distributed real-valued random variables. Then, almost surely, its empirical cumulative distribution function converges uniformly to the underlying cumulative distribution function.*

This result implies Theorem 5.6 relatively easily at least for bounded variables. In recent work, we gave a categorical formulation of this result based on axiomatizing sampling from the empirical distribution as a partial Markov kernel. The relevant theory of partial morphisms, which live in quasi-Markov categories, is developed in a companion paper written by my student Areeb Shah Mohammed [82].

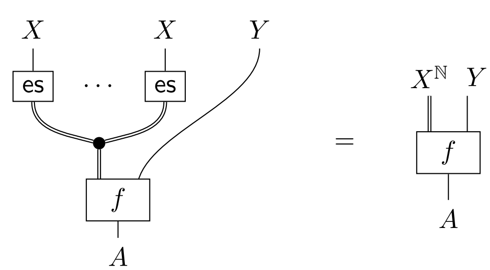
Definition 5.8 ([83])**.** *Let C be a quasi-Markov category with countable Kolmogorov products. Then an **empirical sampling morphism** for an object $X \in \mathsf{C}$ is a morphism*

$$\mathsf{es} : X^{\mathbb{N}} \longrightarrow X$$

satisfying the following conditions:

(a) es *is invariant under finite permutations of its inputs.*

(b) If a morphism $f : A \to X^{\mathbb{N}} \otimes Y$ is invariant under finite permutations in the first factor, then



(5.7)

We have shown that every standard Borel space admits an empirical sampling morphism, where for $X = \mathbb{R}$ one can use (5.6). Informally, this es can be thought of as returning a uniformly random element from the input sequence, whenever this is well-defined, and otherwise being undefined. However, the choice of es for $X = \mathbb{R}$ is highly non-unique, since for example conjugating by a suitable complex measurable automorphism of $\mathbb{R}$ will produce a different es which also satisfies the above axioms.

Unfortunately, our developments do not really amount to an abstract *proof* of the Glivenko–Cantelli theorem, since our axiom (5.7) already encodes much of this statement: proving that the axiom holds with

[16] We write lowercase x_n now in order to indicate that this can be an arbitrary sequence of real numbers, not necessarily following any particular distribution.

$X = \mathbb{R}$ and (5.6) is essentially the same as proving Theorem 5.7. Nevertheless, it has been fruitful to take Definition 5.8 as a starting point from which other kinds can be derived. So informally speaking, we are using the law of large numbers as an axiom.

Based on this, we have given new categorical proofs of the de Finetti theorem and the representability of the Markov category.[17] We have also stated and proven a categorical law of large numbers, which has the classical Theorem 5.6 as a special case. Taking these things together, I believe that our paper provides the most concise joint proof of the Glivenko–Cantelli theorem, the strong law of large numbers, and the de Finetti theorem. Our hope is to also prove an abstract generalization of the (pointwise) ergodic theorem in future work.

With these developments in mind, allow me to speculate a bit about the mathematical and philosophical significance of Definition 5.8. Mathematically, I wonder if a choice of empirical sampling morphism on a measurable space could be thought of as an important piece of structure on that space. For example, could it serve as an alternative to the use of topological structure in some measure-theoretic proofs? And to what extent is the existence of empirical sampling morphisms specific to measure-theoretic probability—in which case their existence could be used as an alternative to Kolmogorov's axioms—or do they also exist in other kinds of Markov categories? Philosophically, what does Definition 5.8 tell us about the nature of probability? Can it serve as a foundation of a frequentist interpretation? The reason for why I think that the answer could be positive is as follows. There is a plethora of different notions of when an individual sequence $(x_n)_{n \in \mathbb{N}}$ is considered to be random, formalized in definitions due to von Mises, Martin-Löf, Wald, Schnorr and others [85], and this ambiguity undermines frequentists' arguments in favour of their interpretation. In contrast to that, based on our Definition 5.8 and its mathematical utility, one can now argue that no particular definition is the "right" one, since *any* definition which can be used as the domain of an empirical sampling morphism is reasonable.

Further results. Our existing results on Markov categories are too numerous to go into detail for all of them, but here are some of the other ones:

- My results on sufficient statistics [68] are versions of the classical theorems of Fisher–Neyman, Basu and Bahadur.
- Our results on the informativeness of statistical experiments [84, Theorem 5.13] extend Blackwell's classical result characterizing informativeness of statistical experiments in terms of second-order stochastic dominance to the categorical setting.
- We have a treatment of Bayesian networks in terms of Markov categories, including a categorical definition of d-separation and a complete proof of the d-separation criterion [86]. In contrast to the traditional version, our d-separation criterion has an intuitive formulation in terms of topological connectedness of string diagrams.
- We have developed categorical notions of absolute continuity and support and studied idempotent Markov kernels [87]. Our main result is that every idempotent Markov kernel on a standard Borel space X admits a *splitting*: it can be obtained by first projecting onto another standard Borel space Y and subsequently mapping back into X. This strengthens a characterization of idempotent Markov kernels due to Blackwell [88], and this strengthening is entirely new: no purely measure-theoretic proof exists to date. We expect such splittings to have applications in ergodic theory, where group averaging is often an idempotent Markov kernel (in typical cases only partially defined).
- *Hidden Markov models* and *filters* are an important topic in signal processing, control theory and embodied artificial intelligence. Our categorical treatment of this [89] specializes both to the general nonlinear filter and to the well-known Kalman filter [90].

Categorical quantum probability. Standard probability theory is generalized by quantum probability, which according to quantum physics is the *actual* flavour of probability theory which governs our world. And although the setting of Markov categories is a general framework for theories of uncertainty, this setting is not general enough to facilitate a treatment of quantum probability: the requirement for copy morphisms to exist

[17] Representability is a purely categorical concept playing an important role in the theory of Markov categories itself [84]. For BorelStoch, it amounts to the obvious bijection between Markov kernels $A \to X$ and measurable functions $A \to PX$, where PX is the measurable space of probability measures on a measurable space X.

contradicts the infamous **no-broadcasting theorem** of quantum information theory [91]. Several approaches to overcome this problem have been proposed, including semicartesian probability [92, 93], Parzygnat's quantum Markov categories [94, 95] and the involutive Markov categories by Lorenzin and myself [96].[18] However, I do not want to elaborate on any of this further at present, since I believe that the best setting for categorical quantum probability has not yet been found.

A related problem is the question of what a good definition of **quantum measurable space** would be, which surprisingly seems to be a question that has not been asked before. Of course, it is well-known that von Neumann algebras equipped with faithful normal states are good noncommutative generalizations of probability spaces. However, in Markov categories like BorelStoch, the objects are not probability spaces but rather measurable spaces, and I believe that this is for good reason. We therefore would like to have a category of "quantum measurable spaces" such that its full subcategory of commutative objects is equivalent to BorelStoch. In [98], Lorenzin and I have considered this question and studied a number of candidate definitions revolving around monotone σ-complete C*-algebras, studied in particular by Wright.

Impact. In this section, I have mainly discussed research that I have personally been involved with. But since they were popularized by [68], Markov categories have found broad adoption in a diverse range of topics, including probabilistic programming [99–101], machine learning and AI [102–107], neuroscience [108], cognitive science [109], decision theory [110, 111], causal modeling and inference [112–116], the logical structure of conditional independence [117], information measures [30, 118, 119], ergodic theory [120], the theory of hypergeometric distributions [121], control theory [122] and quantum foundations [123].

Andi Wang—one of the authors of a recent formulation of Markov chain Monte Carlo with Markov categories [124]—has informed me about an ongoing international statistics reading group on Markov categories with ~80 participants.

6 Probability spin-offs

My work on categorical probability has also led to some simple questions in measure-theoretic probability that had apparently not been asked before. The first one of these was based on considerations on the law of large numbers. The particular question which I was pondering was: what can be said about the empirical averages $\frac{1}{n}\sum_{i=1}^n X_i$ when the assumption of finite first moment is removed from Theorem 5.6? For example, assuming nonnegative iid variables, is it possible to upper-bound the probability

$$\mathbf{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i > \frac{\alpha}{m}\sum_{i=1}^{m} X_i\right]$$

for fixed $\alpha > 1$, such that the upper bound tends to 0 as $m, n \to \infty$? In [125], Bellec and I have developed an integer programming approach to problems of this type (optimizing a strict linear inequality over iid distributions on $\mathbb{R}_+$). We have applied our algorithm to the special case

$$\mathbf{P}[X_1 + X_2 + X_3 < 2X_4], \tag{6.1}$$

which is the simplest such problem for which we had not been able to find a solution by hand. Our method has provided a rigorous machine-checkable proof that the optimum lies between 0.400695 and 0.417.[19] Subsequently, work by Google DeepMind and Tao has used this question as one of 67 benchmark problems for the use of AI for structured optimization problems arising from mathematical research [126]. In contrast to what happened with many of the other 66 problems, our human-derived bounds for (6.1) have not been improved on by the AI (yet).

The second spin-off is a result on stochastic dynamics developed jointly with Rivera. It gives a precise sense in which every measure-preserving transformation can be approximated by iterated averaging transformations.

[18] There also is categorical quantum mechanics [97], but this is of quite a different flavour in that the morphisms in the categories typically considered there are operators rather than quantum channels, and the "CP construction" which turns the former kind of categories into the latter works well only for finite-dimensional Hilbert spaces.

[19] To get a feel for the problem, note that if all four variables are constant, then they must be equal and therefore the probability is 0. In the other direction, it feels intuitively clear that the probability cannot get arbitrarily close to 1, but formalizing this already requires some thought. In the paper, we have also considered the simpler case of $\mathbf{P}[X_1 < (X_2 + X_3)/2]$, whose optimization we have phrased as a fun puzzle about a new "Beat the Average" game in a casino.

Theorem 6.1 ([127]). *Let (X, Σ, μ) be a standard Lebesgue probability space and $T : X \to X$ a measure-preserving automorphism. Then for all $f_1, \ldots, f_n \in L^\infty(X)$ and $\varepsilon > 0$, there is a finite sequence of sub-σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_m \subseteq \Sigma$ such that*

$$\|f_i \circ T - \mathbb{E}_m \cdots \mathbb{E}_1 f_i\|_\infty < \varepsilon \qquad \forall i = 1, \ldots, n,$$

where each $\mathbb{E}_j : L^\infty(X) \to L^\infty(X)$ is the conditional expectation operator with respect to $\mathcal{F}_j$.

The proof is based on the following instructive puzzle: given n empty and n full water tanks, all of equal size and shape, suppose that you want to transfer as much water as possible from the full to the empty ones. However, the only tool that you have available is a hose with which you can equilibrate the water levels between any two tanks. Then how much total water can you transfer? Using generating function methods, we have shown that all but a fraction of

$$4^{-n}\binom{2n}{n} = \frac{1}{\sqrt{\pi n}} + O(n^{-3/2})$$

of the water can be transferred, and this is optimal. In particular, for $n \gg 1$ *almost all* of the water can be transferred. This turns out to imply Theorem 6.1 for all transformations T of order 2, from which the general case follows by a result of Ryzhikov [128].

7 Mathematical structures in foundations of physics

I have also recently revived my strong interest in mathematical structures in the foundations of physics through the following works.

- ▷ The paper [129] develops a fully algebraic generalization of pseudo-Riemannian geometry. This is motivated by the observation that when we solve the Einstein field equations, we typically do not fix a manifold to begin with, but rather determine the topology jointly with the metric, and this procedure *feels* more algebraic than geometric. This is based on the simple observation that tensor calculus makes sense over any **algebraifold**, which I defined as an algebra A over a commutative ring k which is such that the A-module of k-linear derivations of A is finitely generated projective.[20,21] Despite its simplicity and several prior approaches to algebraic differential geometry, the algebraifold approach is new.[22] So far, I have developed it up to the point of giving an algebraic treatment of geodesics, which is relevant already because the standard definition of geodesics is point-based and not obviously amenable to an algebraic formulation. There is a diverse range of examples of algebraifolds beyond the $\mathbb{R}$-algebras of the form $C^\infty(M)$, of which I want to highlight three:

  - Every algebraic function field of characteristic zero is an algebraifold over its ground field.
  - Suitable versions of the *Colombeau algebra* of generalized functions on a manifold, which includes Schwartz distributions, are algebraifolds over $\mathbb{R}$.
  - Any smooth submersion $M \to N$ of smooth manifolds makes $C^\infty(M)$ into an algebraifold over $C^\infty(N)$.

  The last example is particularly interesting, as it shows that the algebraifold formalism also specializes to a fibred or parametrized version of differential geometry. This makes me wonder whether there could even be something like a "classifying algebraifold" for solutions to the Einstein equations, where the ground ring would be something like a formal dual of the moduli space of solutions.

- ▷ In physics, the concept of *observable* plays a dual role. On the one hand, observables are quantities that can be measured, at least in an idealized sense; on the other hand, they are the generators of

[20] This condition is importantly distinct from the module of Kähler differentials Ω^1 being finitely generated projective, which is of course not the case for an $\mathbb{R}$-algebra of the form $C^\infty(M)$. See the paper for more discussion of this point, as well as a modified universal property for the module of derivations of an algebraifold.

[21] Readers familiar with algebraic geometry may wonder whether a definition involving sheaves of rings would be more suitable. Although this is indeed the case in algebraic geometry due to the relative scarcity of regular functions on a variety, this reason does not apply to differential geometry, where smooth functions are of ample supply. I therefore do not think that using sheaves is relevant for the purposes of differential geometry.

[22] See the paper for a literature review.

one-parameter groups of transformations. This duality seems to be a feature of all physical theories that we have, including classical mechanics and quantum mechanics. But what is it about these theories that makes this so? More precisely, what is the minimal mathematical structure that the observables must have in order for them to generate one-parameter groups of transformations coherently?

This is the question that I've tried to answer in [130]. My proposal is built on the existing notion of **quandle** Q, which is an algebraic structure "acting on itself": it comes with a binary operation $\triangleright$ satisfying the **self-distributivity** equation

$$x \triangleright (y \triangleright z) = (x \triangleright y) \triangleright (x \triangleright z).$$

Moreover, a quandle is required to satisfy the reflexivity condition $x \triangleright x = x$ and be such that every map $x \triangleright -$ is bijective. A good way to think about this definition is as the fact that every map $x \triangleright - : Q \to Q$ is an automorphism of Q itself (which fixes x). This is already very similar to the situation I've described for observables. The only difference is that an observable should generate a *one-parameter* group of transformations. In order to account for this, I have introduced the notion of **Lie quandle**, which is a manifold[23] together with a family of binary operations $\triangleright_t$ indexed by $t \in \mathbb{R}$, such that the equations

$$x \triangleright_s (x \triangleright_t y) = x \triangleright_{s+t} y, \qquad x \triangleright_0 y = y,$$
$$x \triangleright_s (y \triangleright_t z) = (x \triangleright_s y) \triangleright_t (x \triangleright_s z), \qquad x \triangleright_s x = x$$

for all $x, y, z \in Q$ and $s, t \in \mathbb{R}$ hold, and such that $\triangleright$ is smooth as a map $Q \times \mathbb{R} \times Q \to Q$. Informally, the concept of Lie quandle is a nonlinear generalization of the notion of Lie algebra (over $\mathbb{R}$). Indeed every Lie algebra $\mathfrak{g}$ is a Lie quandle with respect to the adjoint action,

$$x \triangleright_t y := \exp(t \, \mathrm{ad}_x) y.$$

My MSc student Benjamin Walder has worked on reconstructing the vector space structure and Lie bracket on $\mathfrak{g}$ from the Lie quandle structure, and he has shown that this is possible as soon as $\mathfrak{g}$ has trivial centre. I believe that the algebraic and differential structures on a Lie quandle lead to similar synergies as in classical Lie theory, but this largely remains to be explored. As a starting point, I have developed a new understanding of Noether's theorem on the relation between symmetries and conservation laws, extending Baez's perspective on this [131].

It has been argued by Moskovich[24] that self-distributivity is as fundamental as associativity, and that associativity models space and time while self-distributivity models information. I share this idea to some extent and believe that self-distributive structures are currently underappreciated. In fact, self-distributivity appears in physics not just in the context of observables, but *also* in the context of states: taking probabilistic mixtures (convex combinations) of states, as given by

$$\rho \triangleright_t \sigma := e^{-t} \rho + (1 - e^{-t}) \sigma,$$

also satisfies the same self-distributivity and reflexivity equations as in a Lie quandle; the only difference is that we now are restricted to $t \geq 0$. I do not know yet what the deeper meaning of this might be, but it is intriguing.

▷ Most recently [132], I have been working on a vaguely related question in the foundations of quantum theory, namely on why measurements are modelled as tuples of operators indexed by the outcomes.[25] Why could a measurement not be a completely different kind of mathematical object? To make formal sense of this question, I introduced a definition of **generalized measurement theory (GMT)** by analogy with the existing notion of *generalized probabilistic theory (GPT)*. A GMT is simply a functor $\mathcal{M} : \mathsf{FinSet} \to \mathsf{Set}$ with $\mathcal{M}(1) \cong 1$, where we think of $\mathcal{M}(X)$ for a finite set X as the set of measurements on the physical system under consideration which take values in X. Then if $f : X \to Y$ is a function, $\mathcal{M}(f) : \mathcal{M}(X) \to \mathcal{M}(Y)$ is the map which takes a measurement with outcomes in X and produces a measurement with outcomes in Y by post-processing the outcomes according to f. This bare-bones

[23] An attentive reader may wonder this can be done with algebraifolds. I'm not sure yet due to the problem of products described in [129].

[24] See the blog post ldtopology.wordpress.com/2014/07/21/associativity-vs-distributivity/.

[25] More accurately, this is a correct description in the case of discrete outcomes, while for continuous outcomes one needs to use operator-valued measures. I will ignore this technicality for the purposes of this discussion.

structure turns out to be surprisingly powerful, and in particular it can be used to provide an interesting answer to the question above.

Indeed if we consider also the functor $\Delta : \mathsf{FinSet} \to \mathsf{Set}$ which takes a finite set X to the set $\Delta(X)$ of probability distributions on X, then we can define a **probabilistic state** on a GMT $\mathcal{M}$ to be a natural transformation $\mathcal{M} \Rightarrow \Delta$. Based on these notions, I have shown the following: if $\mathcal{M}$ separates probabilistic states, then the elements of every $\mathcal{M}(X)$ are uniquely determined by all possible post-processings along the indicator functions $1_{\{x\}} : X \to \{0,1\}$ of all singletons. In this way, every measurement in $\mathcal{M}(X)$ can be identified with a tuple of elements of $\mathcal{M}(\{0,1\})$ indexed by the outcomes in X.

8 Statement on AI use

Ongoing developments suggest that AI systems are improving at rapid pace. Therefore I find it important to position myself with respect to their use and to evaluate their potential as well as possible risks. My current workflow integrates AI tools as follows:

- I have been using *GitHub Copilot* since 2023 as a writing assistant for all kinds of writing. This includes research papers, lecture notes, emails and this very document. This tool proposes text completions based on the context. I mostly decide what I want to write *before* accepting or declining the suggested completion. With steadily increasing capabilities, I have found that the suggestions match my intentions more and more often. Thanks to this, Copilot has led to significant increases in my writing speed.

- I use *ChatGPT 5.3 Thinking* and *ChatGPT 5.2 Pro* for various professional purposes, but mostly limited to the following: helping me find relevant literature; giving me custom explanations of things that I would like to learn about; and to some extent helping me with simple proofs.

  For proofs, I do not yet find these models intelligent enough to be seriously useful, and this applies even more so to brainstorming or big-picture discussions. I have also tried to use them for designing homework, but have found their suggestions to be generally somewhat dry and uninteresting, while I want to design homework that is engaging and fun for students. It's possible that this can be improved with more prompt engineering.

This workflow is subject to change as capabilities evolve.

References

[1] Tobias Fritz. "Tsirelson's problem and Kirchberg's conjecture". In: *Rev. Math. Phys.* 24.5 (2012). arXiv:1008.1168, pp. 1250012, 67 (cit. on p. 1).

[2] M. Junge, M. Navascues, C. Palazuelos, D. Perez-Garcia, V. B. Scholz, and R. F. Werner. "Connes embedding problem and Tsirelson's problem". In: *J. Math. Phys.* 52.1 (2011). arXiv:1008.1142, pp. 012102, 12 (cit. on p. 1).

[3] Narutaka Ozawa. "About the Connes embedding conjecture: algebraic approaches". In: *Jpn. J. Math.* 8.1 (2013), pp. 147–183 (cit. on p. 1).

[4] Tobias Fritz, Tim Netzer, and Andreas Thom. "Can you compute the operator norm?" In: *Proc. Amer. Math. Soc.* 142.12 (2014). arXiv:1207.0975, pp. 4265–4276 (cit. on p. 1).

[5] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. $MIP^* = RE$. arXiv:2001.04383 (cit. on p. 1).

[6] Lewis Bowen, Michael Chapman, Alexander Lubotzky, and Thomas Vidick. *The Aldous–Lyons Conjecture I: Subgroup Tests*. arXiv:2408.00110. 2024 (cit. on p. 1).

[7] Isaac Goldbring and Thomas Sinclair. *On definability of C*-tensor norms*. arXiv:2509.15086 (cit. on p. 1).

[8] Isaac Goldbring. "The Connes embedding problem: a guided tour". In: *Bull. Amer. Math. Soc.* 59.4 (2022). https://www.math.uci.edu/∼isaac/CEPsurvey.pdf, pp. 503–560 (cit. on p. 1).

[9] Isaac Goldbring. "Undecidability and incompleteness in quantum information theory and operator algebras". In: *Monatsh. Math.* 208 (2025). arXiv:2409.08342, pp. 631–664 (cit. on p. 1).

[10] Viktoria Woergaard Løgstrup and Nadja Enggaard Lausen. "Quantum correlations, C*-algebras and Tsirelson's conjecture". web.math.ku.dk/ rordam/students/Tsirelson.pdf. MA thesis. University of Copenhagen, 2024 (cit. on p. 1).

[11] Yuming Zhao. "Tsirelson's Bound and Beyond: Verifiability and Complexity in Quantum Systems". uwspace.uwaterloo.ca/bitstreams/9229ad7f-0e5d-4add-ab52-0e3a26f1755a/download. PhD thesis. University of Waterloo, 2024 (cit. on p. 1).

[12] Tobias Fritz. "Quantum logic is undecidable". In: *Arch. Math. Logic* 60 (2021). arXiv:1607.05870, pp. 329–341 (cit. on p. 1).

[13] John C. Baez, Tobias Fritz, and Tom Leinster. "A characterization of entropy in terms of information loss". In: *Entropy* 13.11 (2011). arXiv:1106.1791, pp. 1945–1957 (cit. on p. 2).

[14] John C. Baez and Tobias Fritz. "A Bayesian characterization of relative entropy". In: *Theory Appl. Categ.* 29 (2014). arXiv:1402.3067 (cit. on p. 2).

[15] Nicolas Gagné and Prakash Panangaden. "A categorical characterization of relative entropy on standard Borel spaces". In: *The Thirty-third Conference on the Mathematical Foundations of Programming Semantics (MFPS XXXIII)*. Vol. 336. Electron. Notes Theor. Comput. Sci. arXiv:1703.08853, pp. 135–153 (cit. on p. 2).

[16] James Fullwood. "An axiomatic characterization of mutual information". In: *Entropy* 25.4 (2023). arXiv:2108.12647, p. 663 (cit. on p. 2).

[17] James Fullwood. "On a 2-relative entropy". In: *Entropy* 24.1 (2022). arXiv:2112.03582, p. 74 (cit. on p. 2).

[18] Arthur J. Parzygnat. "Towards a functorial description of quantum relative entropy". In: *Geometric science of information*. Vol. 12829. Lecture Notes in Comput. Sci. arXiv:2105.04059. Springer, Cham, 2021, pp. 557–564 (cit. on p. 2).

[19] Arthur Parzygnat. "A functorial characterization of von Neumann entropy". In: *Cah. Topol. Géom. Différ. Catég.* LXIII.1 (2022). arXiv:2009.07125, pp. 89–128 (cit. on p. 2).

[20] Tai-Danae Bradley. "Entropy as a topological operad derivation". In: *Entropy* 23.9 (2021). arXiv:2107.09581, Paper No. 1195, 11 (cit. on p. 2).

[21] Yuri Manin and Matilde Marcolli. "Homotopy Theoretic and Categorical Models of Neural Information Networks". In: *Compositionality* 6 (2024). arXiv:2006.15136 (cit. on p. 2).

[22] Matilde Marcolli. "Motivic information". In: *Boll. Unione Mat. Ital.* 12.1-2 (2019). arXiv:1712.08703, pp. 19–41 (cit. on p. 2).

[23] Matilde Marcolli and Ryan Thorngren. "Thermodynamic semirings". In: *J. Noncommut. Geom.* 8.2 (2014). arXiv:1108.2874, pp. 337–392 (cit. on p. 2).

[24] Matilde Marcolli. "Gamma spaces and information". In: *J. Geom. Phys.* 140 (2019). arXiv:1807.05314, pp. 26–55 (cit. on p. 2).

[25] Cheuk Ting Li. *A Characterization of Entropy as a Universal Monoidal Natural Transformation*. arXiv:2308.05742 (cit. on p. 2).

[26] Kenji Nakahira. "A characterization of von Neumann entropy using functors". In: *Tamagawa University Quantum ICT Research Institute Bulletin* 13.1 (2023). arXiv:2309.10353, pp. 1–5 (cit. on p. 2).

[27] Mee Seong Im and Mikhail Khovanov. *Entropy, cocycles, and their diagrammatics*. arXiv:2409.08462 (cit. on p. 2).

[28] Keenan J. A. Down and Pedro A. M. Mediano. "A Logarithmic Decomposition and a Signed Measure Space for Entropy". In: *IEEE Transactions on Information Theory* 72.1 (2026). arXiv:2409.03732, pp. 17–41 (cit. on p. 2).

[29] Redi Haderi, Cihan Okay, and Walker H. Stern. "The operadic theory of convexity". In: *Appl. Categ. Structures* 33 (2025). arXiv:2403.18102, p. 17 (cit. on p. 2).

[30] Ralph Sarkis and Fabio Zanasi. *Complete Diagrammatic Axiomatisations of Relative Entropy*. arXiv:2603.04530 (cit. on pp. 2, 11).

[31] Tobias Fritz. "Beyond Bell's Theorem: correlation scenarios". In: *New J. Phys.* 14 (2012). arXiv:1206.5115, p. 103001 (cit. on p. 2).

[32] Tobias Fritz. "Beyond Bell's theorem II: scenarios with arbitrary causal structure". In: *Comm. Math. Phys.* 341.2 (2016). arXiv:1404.4812, pp. 391–434 (cit. on p. 2).

[33] Ivan Šupić, Jean-Daniel Bancal, Yu Cai, and Nicolas Brunner. "Genuine network quantum nonlocality and self-testing". In: *Phys. Rev. A* 105 (2022). arXiv:2105.12341, p. 022206 (cit. on p. 2).

[34] Thomas C. Fraser and Elie Wolfe. "Causal Compatibility Inequalities Admitting Quantum Violations in the Triangle Structure". In: *Phys. Rev. A* 98 (2018). arXiv:1709.06242, p. 022113 (cit. on p. 2).

[35] Tamás Kriváchy, Yu Cai, Daniel Cavalcanti, Arash Tavakoli, Nicolas Gisin, and Nicolas Brunner. "A neural network oracle for quantum nonlocality problems in networks". In: *npj Quantum Information* 6 (2020). arXiv:1907.10552, p. 70 (cit. on p. 2).

[36] Emanuele Polino, Davide Poderini, Giovanni Rodari, Iris Agresti, Alessia Suprano, Gonzalo Carvacho, Elie Wolfe, Askery Canabarro, George Moreno, Giorgio Milani, Robert W. Spekkens, Rafael Chaves, and Fabio Sciarrino. "Experimental nonclassicality in a causal network without assuming freedom of choice". In: *Nature Communications* 14.909 (2023). arXiv:2210.07263 (cit. on p. 2).

[37] Pedro Lauand, Bereket Ngussie Bekele, and Elie Wolfe. *Quantum Non-classicality from Causal Data Fusion*. arXiv:2405.19252 (cit. on p. 2).

[38] Ning-Ning Wang, Chao Zhang, Huan Cao, Kai Xu, Bi-Heng Liu, Yun-Feng Huang, Chuan-Feng Li, Guang-Can Guo, Nicolas Gisin, Tamás Kriváchy, and Marc-Olivier Renou. "Experimental genuine quantum nonlocality in the triangle network". In: *Phys. Rev. Lett.* 123 (2022). arXiv:2401.15428, p. 140401 (cit. on p. 2).

[39] Elisa Bäumer, Victor Gitton, Tamás Kriváchy, Nicolas Gisin, and Renato Renner. "Exploring the local landscape in the triangle network". In: *Phys. Rev. A* 111 (2025). arXiv:2405.08939, p. 052453 (cit. on p. 2).

[40] Shashaank Khanna, Matthew Pusey, and Roger Colbeck. *Closing the problem of which causal structures of up to six total nodes have a classical-quantum gap*. arXiv:2512.04058 (cit. on p. 2).

[41] Marc-Olivier Renou, Elisa Bäumer, Sadra Boreiri, Nicolas Brunner, Nicolas Gisin, and Salman Beigi. "Genuine quantum nonlocality in the triangle network". In: *Phys. Rev. Lett.* 123 (2019). arXiv:1905.04902, p. 140401 (cit. on p. 2).

[42] Alejandro Pozas-Kerstjens, Antoine Girardin, Tamás Kriváchy, Armin Tavakoli, and Nicolas Gisin. "Post-quantum nonlocality in the minimal triangle scenario". In: *New J. Phys.* 25 (2023). arXiv:2305.03745, p. 113037 (cit. on p. 2).

[43] Tamás Kriváchy and Martin Kerschbaumer. "Closing the detection loophole in the triangle network with high-dimensional photonic states". In: *Phys. Rev. Lett.* 135 (2025). arXiv:2503.24213, p. 160803 (cit. on p. 2).

[44] Victor Gitton and Renato Renner. *The Elegant Joint Measurement is non-classical in the triangle network*. arXiv:2510.15143. 2025 (cit. on p. 2).

[45] Alessia Suprano, Davide Poderini, Emanuele Polino, Iris Agresti, Gonzalo Carvacho, Askery Canabarro, Elie Wolfe, Rafael Chaves, and Fabio Sciarrino. "Experimental Genuine Tripartite Nonlocality in a Quantum Triangle Network". In: *PRX Quantum* 3 (2022). arXiv:2204.00388, p. 030342 (cit. on p. 2).

[46] Othmane Meskine, Ivan Šupić, Damian Markham, Félicien Appas, Fabien Boitier, Martina Morassi, Aristide Lemaître, Maria Ines Amanti, Florent Baboux, Eleni Diamanti, and Sara Ducci. "Experimental quantum triangle network nonlocality with an AlGaAs multiplexed entangled photon source". In: *PRX Quantum* 6 (2025). arXiv:2410.06860, p. 020313 (cit. on p. 2).

[47] Ning-Ning Wang, Chao Zhang, Huan Cao, Kai Xu, Bi-Heng Liu, Yun-Feng Huang, Chuan-Feng Li, Guang-Can Guo, Nicolas Gisin, Tamás Kriváchy, and Marc-Olivier Renou. *Experimental genuine quantum nonlocality in the triangle network*. arXiv:2401.15428 (cit. on p. 2).

[48] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz. "The inflation technique for causal inference with latent variables". In: *J. Causal Inference* 7.2 (2019). arXiv:1609.00672 (cit. on p. 2).

[49] Miguel Navascués and Elie Wolfe. "The inflation technique completely solves the causal compatibility problem". In: *J. Causal Inference* 8.1 (2020). arXiv:1707.06476, pp. 70–91 (cit. on p. 2).

[50] Elie Wolfe, Alejandro Pozas-Kerstjens, Matan Grinberg, Denis Rosset, Antonio Acín, and Miguel Navascués. "Quantum Inflation: A General Approach to Quantum Causal Compatibility". In: *Phys. Rev. X* 11 (2021). arXiv:1909.10519, p. 021043 (cit. on p. 2).

[51] Emanuel-Cristian Boghiu, Elie Wolfe, and Alejandro Pozas-Kerstjens. "Inflation: a Python library for classical and quantum causal compatibility". In: *Quantum* 7.996 (2023). arXiv:2211.04483 (cit. on p. 2).

[52] Eric Chitambar and Gilad Gour. "Quantum resource theories". In: *Rev. Modern Phys.* 91.2 (2019). arXiv:1806.06107, pp. 025001, 48 (cit. on p. 3).

[53] Bartosz Regula. "Probabilistic transformations of quantum resources". In: *Phys. Rev. Lett.* 128 (2022). arXiv:2109.04481, p. 110505 (cit. on p. 3).

[54] Bob Coecke, Tobias Fritz, and Robert W. Spekkens. "A mathematical theory of resources". In: *Information and Computation* 250 (2016). arXiv:1409.5531., pp. 59–86 (cit. on p. 3).

[55] Tobias Fritz. "Resource convertibility and ordered commutative monoids". In: *Math. Struct. Comp. Sci.* 27.6 (2017). arXiv:1504.03661, pp. 850–938 (cit. on p. 3).

[56] Carlo Sparaciari, Jonathan Oppenheim, and Tobias Fritz. "A Resource Theory for Work and Heat". In: *Phys. Rev. A 96, 052112* (2017). arXiv:1607.01302 (cit. on p. 3).

[57] Volker Strassen. "The asymptotic spectrum of tensors". In: *J. Reine Angew. Math.* 384 (1988), pp. 102–152 (cit. on p. 3).

[58] Tobias Fritz. "Abstract Vergleichsstellensätze for preordered semifields and semirings I". In: *SIAM J. Appl. Algebra Geom.* 7.2 (2023). arXiv:2003.13835, pp. 505–547 (cit. on pp. 3, 4).

[59] Tobias Fritz. *Abstract Vergleichsstellensätze for preordered semifields and semirings II.* arXiv:2112.05949 (cit. on p. 4).

[60] Tobias Fritz. "Asymptotic and catalytic containment of representations of $SU(n)$". In: *Algebr. Comb.* 7.2 (2024). arXiv:2205.10899, pp. 327–336 (cit. on p. 4).

[61] Matthew Di Meglio and Chris Heunen. "Dagger categories and the complex numbers: Axioms for the category of finite-dimensional Hilbert spaces and linear contractions". In: *Appl. Categ. Structures* 33 (2025). arXiv:2401.06584, p. 18 (cit. on p. 4).

[62] Tobias Fritz. "The asymptotic comparison of random walks on topological abelian groups". In: *Bernoulli* 30.4 (2024). arXiv:2004.13655, pp. 2932–2954 (cit. on p. 4).

[63] Muhammad Usman Farooq, Tobias Fritz, Erkka Haapasalo, and Marco Tomamichel. "Asymptotic and catalytic matrix majorization". In: *IEEE Trans. Inform. Theory* 70.5 (2024). arXiv:2301.07353, pp. 3118–3144 (cit. on p. 4).

[64] Xiaosheng Mu, Luciano Pomatto, Philipp Strack, and Omer Tamuz. "From Blackwell Dominance in Large Samples to Rényi Divergences and Back Again". In: *Econometrica* 89.1 (2021). arXiv:1906.02838, pp. 475–506 (cit. on p. 4).

[65] Frits Verhagen, Marco Tomamichel, and Erkka Haapasalo. "Matrix majorization in large samples with varying support restrictions". In: *IEEE Transactions on Information Theory* 71.9 (2025). arXiv:2407.16581, pp. 6517–6545 (cit. on p. 4).

[66] Frits Verhagen, Marco Tomamichel, and Erkka Haapasalo. *Conditions for Large-Sample Majorization of Pairs of Flat States in Terms of α-z Relative Entropies.* arxiv:2507.07520 (cit. on p. 4).

[67] Roberto Rubboli, Erkka Haapasalo, and Marco Tomamichel. *A complete characterisation of conditional entropies.* arXiv:2601.23213 (cit. on p. 4).

[68] Tobias Fritz. "A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics". In: *Adv. Math.* 370 (2020). arXiv:1908.07021, p. 107239 (cit. on pp. 5, 10, 11).

[69] Tobias Fritz and Wendong Liang. "Free gs-monoidal categories and free Markov categories". In: *Appl. Categ. Structures* 31.21 (2023). arXiv:2204.02284 (cit. on p. 5).

[70] Judea Pearl. *Causality.* Second. Models, reasoning, and inference. Cambridge: Cambridge University Press, 2009, pp. xx+464 (cit. on p. 6).

[71] Tobias Fritz and Eigil Fjeldgren Rischel. "Infinite products and zero-one laws in categorical probability". In: *Compositionality* 2 (2020). arXiv:1912.02769, p. 3 (cit. on p. 7).

[72] Paul R. Halmos and L. J. Savage. "Application of the Radon-Nikodym theorem to the theory of sufficient statistics". In: *Ann. Math. Statistics* 20 (1949). projecteuclid.org/euclid.aoms/1177730032, pp. 225–241 (cit. on p. 8).

[73] Tobias Fritz, Tomáš Gonda, and Paolo Perrone. "De Finetti's Theorem in Categorical Probability". In: *J. Stoch. Anal.* 2.4 (2021). arXiv:2105.02639 (cit. on p. 8).

[74] Irfan Alam. "A nonstandard proof of de Finetti's theorem for Bernoulli random variables". In: *J. Stoch. Anal.* 1.4 (2020). arXiv:1912.02784, Art. 15, 18 (cit. on p. 8).

[75] Werner Kirsch. "An elementary proof of de Finetti's theorem". In: *Statistics & Probability Letters* 151 (2019). arXiv:1809.00882, pp. 84–88 (cit. on p. 8).

[76] Lampros Gavalakis and Ioannis Kontoyiannis. "An Information-Theoretic Proof of a Finite de Finetti Theorem". In: *Electron. Commun. Probab.* 26 (2021). arXiv:2104.03882, pp. 1–5 (cit. on p. 8).

[77] Patrick Forré. *Quasi-Measurable Spaces.* arXiv:2109.11631 (cit. on p. 8).

[78] Leihao Chen, Tobias Fritz, Tomáš Gonda, Andreas Klingler, and Antonio Lorenzin. "The Aldous–Hoover Theorem in Categorical Probability". In: *Alg. Stat.* 16 (2025). arXiv:2411.12840, pp. 131–174 (cit. on p. 8).

[79] Tim Austin and Dmitry Panchenko. "A hierarchical version of the de Finetti and Aldous–Hoover representations". In: *Probability Theory and Related Fields* 159.3-4 (2014). arXiv:1301.1259, pp. 809–823 (cit. on p. 8).

[80] Paul Jung, Jiho Lee, Sam Staton, and Hongseok Yang. "A generalization of hierarchical exchangeability on trees to directed acyclic graphs". In: *Ann. H. Lebesgue* 4 (2021). arXiv:1812.06282, pp. 325–368 (cit. on p. 8).

[81] Tim Austin. "On exchangeable random variables and the statistics of large graphs and hypergraphs". In: *Probability Surveys* 5 (2008). arXiv:0801.1698, pp. 80–145 (cit. on p. 8).

[82] Areeb Shah Mohammed. *Partializations of Markov categories.* arXiv:2509.05094 (cit. on p. 9).

[83] Tobias Fritz, Tomáš Gonda, Antonio Lorenzin, Paolo Perrone, and Areeb Shah Mohammed. *Empirical Measures and Strong Laws of Large Numbers in Categorical Probability.* arXiv:2503.21576 (cit. on p. 9).

[84] Tobias Fritz, Tomáš Gonda, Paolo Perrone, and Eigil Fjeldgren Rischel. "Representable Markov Categories and Comparison of Statistical Experiments in Categorical Probability". In: *Theoret. Comp. Sci.* 961.113896 (2023). arXiv:2010.07416 (cit. on p. 10).

[85] Francesca Zaffora Blando. "From Wald to Schnorr: von Mises' definition of randomness in the aftermath of Ville's Theorem". In: *Studies in History and Philosophy of Science* 106 (2024), pp. 196–207 (cit. on p. 10).

[86] Tobias Fritz and Andreas Klingler. "The *d*-separation criterion in Categorical Probability". In: *J. Mach. Learn. Res.* 24.26 (2023). arXiv:2207.05740, pp. 1–49 (cit. on p. 10).

[87] Tobias Fritz, Tomáš Gonda, Antonio Lorenzin, Paolo Perrone, and Dario Stein. *Absolute continuity, supports and idempotent splitting in categorical probability.* arXiv:2308.00651 (cit. on p. 10).

[88] David Blackwell. "Idempotent Markoff chains". In: *Ann. of Math. (2)* 43 (1942), pp. 560–567 (cit. on p. 10).

[89] Tobias Fritz, Andreas Klingler, Drew McNeely, Areeb Shah Mohammed, and Yuwen Wang. "Hidden Markov Models and the Bayes Filter in Categorical Probability". In: *IEEE Transactions on Information Theory* 71.9 (2025). arXiv:2401.14669, pp. 7052–7075 (cit. on p. 10).

[90] Huazhen Fang, Ning Tian, Yebin Wang, MengChu Zhou, and Mulugeta A. Haile. "Nonlinear Bayesian estimation: from Kalman filtering to a broader horizon". In: *IEEE/CAA J. Autom. Sin.* 5.2 (2018). arXiv:1712.01406, pp. 401–417 (cit. on p. 10).

[91] Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. "Noncommuting mixed states cannot be broadcast". In: *Phys. Rev. Lett.* 76 (1996). arXiv:quant-ph/9511010, pp. 2818–2821 (cit. on p. 11).

[92] Nicholas Gauguin Houghton-Larsen. "A Mathematical Framework for Causally Structured Dilations and its Relation to Quantum Self-Testing". arXiv:2103.02302. PhD thesis. University of Copenhagen, 2021 (cit. on p. 11).

[93] Tobias Fritz, Tomáš Gonda, Nicholas Gauguin Houghton-Larsen, Antonio Lorenzin, Paolo Perrone, and Dario Stein. "Dilations and information flow axioms in categorical probability". In: *Math. Struct. Comp. Sci.* 33.10 (2023). arXiv:2211.02507, pp. 913–957 (cit. on p. 11).

[94] Arthur J. Parzygnat. *Inverses, disintegrations, and Bayesian inversion in quantum Markov categories.* arXiv:2001.08375 (cit. on p. 11).

[95] James Fullwood and Arthur J. Parzygnat. "On quantum states over time". In: *Proc. A.* 478.2264 (2022). arXiv:2202.03607, Paper No. 20220104, 20 (cit. on p. 11).

[96] Tobias Fritz and Antonio Lorenzin. *Involutive Markov categories and the quantum de Finetti theorem.* arXiv:2312.09666 (cit. on p. 11).

[97] Chris Heunen and Jamie Vicary. *Categories for Quantum Theory: An Introduction.* Oxford University Press, 2019 (cit. on p. 11).

[98] Tobias Fritz and Antonio Lorenzin. "Categories of abstract and noncommutative measurable spaces". In: *Adv. Math.* To appear (2026). arXiv:2504.13708 (cit. on p. 11).

[99] Dario Stein and Sam Staton. "Compositional Semantics for Probabilistic Programs with Exact Conditioning". In: *Logic in Computer Science.* arXiv:2101.11351. 2021 (cit. on p. 11).

[100] Dario Stein and Sam Staton. "Probabilistic Programming with Exact Conditions". In: *Journal of the ACM* 71.1 (2024), pp. 1–53 (cit. on p. 11).

[101] Nathanael L. Ackerman, Cameron E. Freer, Younesse Kaddar, Jacek Karwowski, Sean K. Moss, Daniel M. Roy, Sam Staton, and Hongseok Yang. "Probabilistic programming interfaces for random graphs: Markov categories, graphons, and nominal sets". In: *Proceedings of the ACM on Programming Languages.* Vol. 8. POPL. arXiv:2312.17127. Association for Computing Machinery, 2024, pp. 1819–1849 (cit. on p. 11).

[102] Dan Shiebler, Bruno Gavranović, and Paul Wilson. *Category Theory in Machine Learning.* arXiv:2106.07032 (cit. on p. 11).

[103] Benjamin Rodatz, Ian Fan, Tuomas Laakkonen, Neil John Ortega, Thomas Hoffman, and Vincent Wang-Mascianica. *A Pattern Language for Machine Learning Tasks.* arXiv:2407.02424 (cit. on p. 11).

[104] Sean Tull, Robin Lorenz, Stephen Clark, Ilyas Khan, and Bob Coecke. *Towards Compositional Interpretability for XAI.* arXiv:2406.17583 (cit. on p. 11).

[105] Rob Cornish. *Stochastic Neural Network Symmetrisation in Markov Categories.* arXiv:2406.11814 (cit. on p. 11).

[106] Yiyang Jia, Guohong Peng, Zheng Yang, and Tianhao Chen. "Category-Theoretical and Topos-Theoretical Frameworks in Machine Learning: A Survey". In: *Axioms* 14.3 (2025). arXiv:2408.14014, p. 204 (cit. on p. 11).

[107] Sridhar Mahadevan. *A Rose by Any Other Name Would Smell as Sweet: Categorical Homotopy Theory for Large Language Models.* arXiv:2508.10018 (cit. on p. 11).

[108] Sebastian Pardo-Guerra, Johnny Jingze Li, Kalyan Basu, and Gabriel A. Silva. "Neural Networks and Markov Categories". In: *AppliedMath* 5.3 (2025), p. 93 (cit. on p. 11).

[109] Nathaniel Virgo, Martin Biehl, and Simon McGregor. "Interpreting Dynamical Systems as Bayesian Reasoners". In: *Machine Learning and Principles and Practice of Knowledge Discovery in Databases.* arXiv:2112.13523. Springer, 2022 (cit. on p. 11).

[110] Elena Di Lavore and Mario Román. "Evidential Decision Theory via Partial Markov Categories". In: *2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS).* arXiv:2301.12989. IEEE, 2023, pp. 1–14 (cit. on p. 11).

[111] Elena Di Lavore, Mario Román, and Paweł Sobociński. *Partial Markov categories.* arXiv:2502.03477 (cit. on p. 11).

[112] Robin Lorenz and Sean Tull. *Causal models in string diagrams.* arXiv:2304.07638 (cit. on p. 11).

[113] Jun Otsuka and Hiroto Saigo. "Process theory of causality: a category-theoretic perspective". In: *Behaviormetrika* 51 (2024), pp. 21–36 (cit. on p. 11).

[114] Dhurim Cakiqi and Max A. Little. *Algorithmic syntactic causal identification.* arXiv:2403.09580 (cit. on p. 11).

[115] Yimu Yin and Jiji Zhang. "Markov categories, causal theories, and the do-calculus". In: *Studies in Logic* 14.6 (2021). arXiv:2204.04821, pp. 1–24 (cit. on p. 11).

[116] Sridhar Mahadevan. "Universal Causal Inference in a Topos". In: *Advances in Neural Information Processing Systems.* Vol. 38. 2025 (cit. on p. 11).

[117] Tao Gu, Jialu Bao, Justin Hsu, Alexandra Silva, and Fabio Zanasi. "A Categorical Approach to DIBI Models". In: *9th International Conference on Formal Structures for Computation and Deduction (FSCD 2024).* arXiv:2401.05842. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024 (cit. on p. 11).

[118] Paolo Perrone. "Markov Categories and Entropy". In: *IEEE Transactions of Information Theory* 70.3 (2024). arXiv:2212.11719, pp. 1671–1692 (cit. on p. 11).

[119] James Fullwood and Arthur J. Parzygnat. "The information loss of a stochastic map". In: *Entropy* 23.8 (2021). arXiv:2107.01975, p. 1021 (cit. on p. 11).

[120] Sean Moss and Paolo Perrone. "A category-theoretic proof of the ergodic decomposition theorem". In: *Ergodic Theory Dynam. Systems* 43.12 (2023). arXiv:2207.07353, pp. 4166–4192 (cit. on p. 11).

[121] Bart Jacobs. "Multinomial and hypergeometric distributions in Markov categories". In: *Proceedings—37th Conference on Mathematical Foundations of Programming Semantics.* Vol. 351. Electron. Proc. Theor. Comput. Sci. (EPTCS). arXiv:2112.14044. EPTCS, [place of publication not identified], 2021, pp. 98–115 (cit. on p. 11).

[122] Manuel Baltieri, Martin Biehl, Matteo Capucci, and Nathaniel Virgo. *A Bayesian Interpretation of the Internal Model Principle.* arXiv:2503.00511 (cit. on p. 11).

[123] Arash Zaghi. *Natural Emergence of the Born Rule from Relational Quantum Dynamics.* osf.io/preprints/osf/j73ak_v1 (cit. on p. 11).

[124] Rob Cornish and Andi Q. Wang. *A categorical account of the Metropolis-Hastings algorithm.* arXiv:2601.22911 (cit. on p. 11).

[125] Pierre C Bellec and Tobias Fritz. "Optimizing over iid distributions and the Beat the Average game". In: *Amer. Math. Monthly* To appear (2026). arXiv:2412.15179 (cit. on p. 11).

[126] Bogdan Georgiev, Javier Gómez-Serrano, Terence Tao, and Adam Zsolt Wagner. *Mathematical exploration and discovery at scale.* arXiv:2511.02864 (cit. on p. 11).

[127] Tobias Fritz and Nicolás Rivera. *Dynamics from iterated averaging.* arXiv:2508.21416 (cit. on p. 12).

[128] V. V. Ryzhikov. "Representation of Transformations Preserving the Lebesgue Measure in the Form of Periodic Transformations". In: *Mat. Zametki* 38.6 (1985), pp. 860–865 (cit. on p. 12).

[129] Tobias Fritz. "Differential geometry and general relativity with algebraifolds". In: *J. Geom. Phys.* 206 (2024). arXiv:2403.06548 (cit. on pp. 12, 13).

[130] Tobias Fritz. "Self-distributive structures in physics". In: *Int. J. Theor. Phys.* 64 (2025). arXiv:2403.14458, p. 73 (cit. on p. 13).

[131] John C. Baez. "Getting to the Bottom of Noether's Theorem". In: *The Philosophy and Physics of Noether's Theorems.* arXiv:2006.14741. Cambridge University Press, 2022, pp. 66–99 (cit. on p. 13).

[132] Tobias Fritz. *Why measurements are made of effects.* arXiv:2602.18898 (cit. on p. 13).