*Mathematical Structures in Computer Science* (2023), **33**, pp. 913–957
doi:10.1017/S0960129523000324

CAMBRIDGE UNIVERSITY PRESS

PAPER

# Dilations and information flow axioms in categorical probability

Tobias Fritz[1], Tomáš Gonda[1], Nicholas Gauguin Houghton-Larsen[2] Antonio Lorenzin[1], Paolo Perrone[3] and Dario Stein[4]

[1]University of Innsbruck, Innsbruck, Austria, [2]Copenhagen, Denmark, [3]University of Oxford, Oxford, UK and [4]Radboud Universiteit Nijmegen, Nijmegen, Netherlands

**Corresponding author:** Tobias Fritz; Email: tobias.fritz@uibk.ac.at

(Received 26 December 2022; revised 18 August 2023; accepted 11 September 2023; first published online 25 October 2023)

**Abstract**

We study the positivity and causality axioms for Markov categories as properties of dilations and information flow and also develop variations thereof for arbitrary semicartesian monoidal categories. These help us show that being a positive Markov category is merely an additional *property* of a symmetric monoidal category (rather than extra structure). We also characterize the positivity of representable Markov categories and prove that *causality implies positivity*, but not conversely. Finally, we note that positivity fails for quasi-Borel spaces and interpret this failure as a privacy property of probabilistic name generation.

**Keywords:** Categorical probability; Markov category; Semicartesian category; Information flow; Quasi-Borel space

## 1. Introduction

Markov categories are a categorical approach to the foundations of probability and statistics. Recent developments of this framework have resulted in purely categorical proofs of various classical theorems, including theorems on sufficient statistics (Fritz 2020), 0/1-laws (Fritz and Rischel 2020), comparison of statistical experiments (Fritz et al. 2023), the de Finetti theorem (Fritz et al. 2021; Moss and Perrone 2022), development of multinomial and hypergeometric distributions (Jacobs 2021), ergodic systems (Moss and Perrone 2023), and the $d$-separation criterion for Bayesian networks (Fritz and Klingler 2023). The Markov categories framework has also found use in probabilistic programming theory (Stein 2021; Stein and Staton 2021) and cognitive science (St. Clere Smithe *https : //arxiv.org/abs/*2109.04461).

Many of these developments do not apply to arbitrary Markov categories, they require additional conditions, such as the *existence of conditionals*, the *causality axiom* or the *positivity axiom* (see Section 1.2 for more details). The fact that these axioms hold in measure-theoretic probability is the only measure-theoretic input that is needed for developments like this. The string-diagrammatic nature of these conditions also suggests that one can think of them as conditions on information flow; hence, we propose to call them information flow axioms.

The purpose of the present paper is to conduct a more detailed study of these axioms. Previously it was known that the existence of conditionals implies both the causality and the positivity axioms (Fritz 2020, Proposition 11.34 and Lemma 11.24). The converse is not true. For example, the Markov category Stoch of all measurable spaces and Markov kernels satisfies causality and positivity but does not have conditionals (Fritz 2020, Example 11.3). The relation

between causality and positivity remained an open question. Our main result here is that causality implies positivity, but not conversely (Theorem 2.24). For both axioms, we also prove various reformulations which elucidate their meaning further.

Besides adding further clarity to the intuition behind the axioms, these reformulations can also help in deciding whether a given Markov category satisfies them. As a case in point, we consider the Markov category $\mathsf{QBStoch}$, which is the Kleisli category of the probability monad on the category $\mathsf{Qbs}$ of quasi-Borel spaces. We find that $\mathsf{QBStoch}$ violates positivity, and that it does so in an interesting way. As was recently discovered, $\mathsf{Qbs}$ validates the privacy equation (Sabok et al. 2021), which originally describes a phenomenon of fresh name generation in theoretical computer science (Stark 1996). As we show in Proposition 3.5, the privacy equation and positivity are incompatible.

A secondary theme of this paper is the idea of developing categorical probability in terms of semicartesian monoidal categories only, which have a weaker structure than Markov categories. This is achieved using the concept of dilations. A dilation of a Markov kernel is another Markov kernel with an additional output such that marginalizing over the latter recovers the original Markov kernel. In the case of kernels with trivial input, this amounts to an extension of the original probability space. While already many of our investigations of the positivity axiom for Markov categories are phrased in terms of dilations, the concept of dilation comes to shine in the purely semicartesian setting. There, we use them to define concepts that mimic those of almost sure equality and of deterministic morphisms in Markov categories. We also note that the structure of a positive Markov category can be recovered from its structure as a semicartesian category, and we provide a characterization of positive Markov categories in semicartesian and dilational terms.

### *1.1 Summary of the paper*

- Section 1.2 starts with the definitions of semicartesian categories and Markov categories, sketches the most important examples for this paper and recalls the definition of dilation.
- Section 2 presents a detailed study of the positivity axiom. After a reformulation of positivity as deterministic marginal independence (DMI) in Definition 2.4 and Theorem 2.8, we derive a characterization of positivity for representable Markov categories in Proposition 2.14. We then turn to the causality axiom and state its equivalence with parametrized equality strengthening in Definition 2.16 and Proposition 2.17. Theorem 2.24 shows that causality implies positivity; an intricate counterexample for the converse is given by a Markov category with semiring-valued Markov kernels as morphisms for a carefully crafted semiring in Proposition 2.25.
- Section 3 recalls the main features of quasi-Borel spaces before presenting the privacy equation as Theorem 3.2. Proposition 3.3 then uses our earlier reformulation of positivity to show that the Markov category of quasi-Borel spaces violates positivity. Proposition 3.5 generalizes this to arbitrary categorical models of name generation by first observing that every such model defines a Markov category.
- Section 4 treats aspects of categorical probability, and in particular the positivity theme, in purely semicartesian categories. To this end, Definition 4.1 introduces dilational equality. It coincides with almost sure equality in Markov categories if and only if the said Markov category satisfies causality (Proposition 4.3). In Definition 4.4, we associate a category of dilations to every morphism. Its initial objects are dubbed initial dilations (Definition 4.7). This provides yet further characterizations of positivity for Markov categories as Proposition 4.12 and Corollary 4.16. Finally, Corollary 4.20 characterizes positive Markov categories in terms of their structure as semicartesian categories alone, and Theorem 4.19 achieves the same for a slightly more general class of Markov categories.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

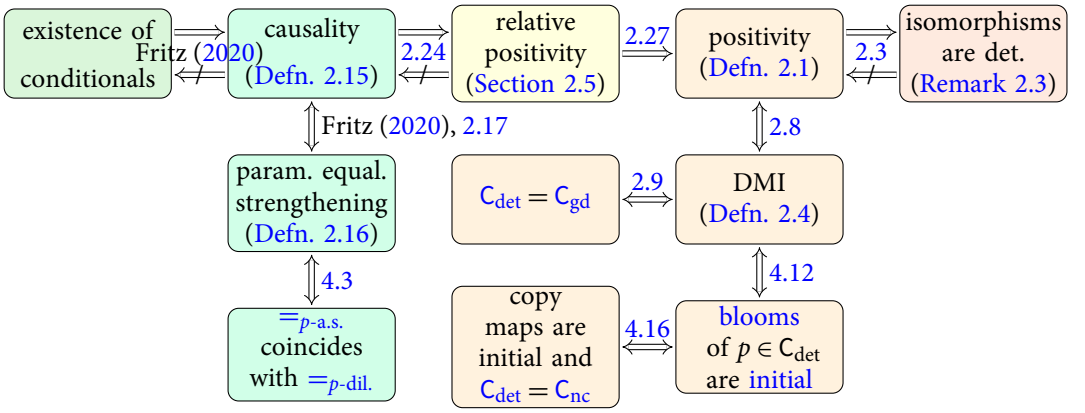



**Figure 1.** Implications between various information flow axioms considered in this paper, with pointers to theorem numbers on the arrows.

Fig. 1 summarizes various information flow axioms considered in this paper together with their relations.

#### *1.1.1 Prerequisites for reading*

We assume that the reader has some basic familiarity with symmetric monoidal categories and string diagrams (Baez and Stay 2011; Piedeleu and Zanasi *https://arxiv.org/abs/2305.08768*). Some prior exposure to Markov categories (Cho and Jacobs 2019; Fritz 2020) will be helpful, but is not strictly necessary. On the probability theory side, basic knowledge of discrete probability theory suffices, as measure-theoretic probability does not play a central role in this paper. Somewhat of an exception is Section 3; although we briefly recall the most relevant theoretical background, prior exposure to the theory of quasi-Borel spaces (Heunen et al. 2017) is helpful to understand it.

### *1.2 Semicartesian categories, Markov categories, and dilations*

All categories that are of interest to us in this paper are symmetric monoidal categories with the following extra property, which for categorical probability implements the normalization of probability, and can be thought of as saying that there is a unique way to forget information.

**Definition 1.1.** *A symmetric monoidal category* $\mathsf{C}$ *is semicartesian if the monoidal unit* $I$ *is terminal.*

We usually abbreviate the lengthy phrase "semicartesian symmetric monoidal category" to "semicartesian category," leaving in particular the symmetry implicit since other cases will not be considered. Although the concept of semicartesian category is standard, we do not know where it was considered first.

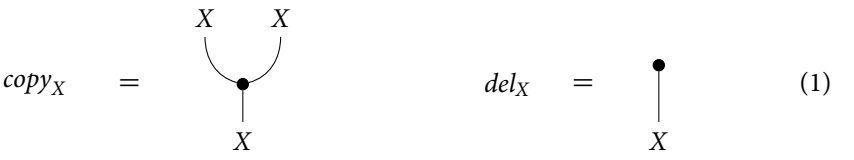
Besides Markov categories, that we turn to next, interesting examples of semicartesian categories occur in quantum probability, where for example the category with finite-dimensional Hilbert spaces as objects and quantum channels as morphisms is a widely studied example (see Houghton-Larsen 2021 and references therein).

**Markov categories.** Roughly speaking, a Markov category (Fritz 2020) is a semicartesian category where all objects are commutative comonoids. Here is the precise definition.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

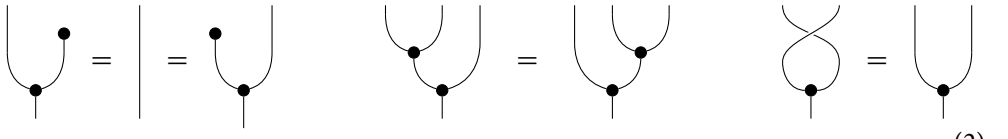
**Definition 1.2.** *A Markov category is a symmetric monoidal category* C *where:*

- *Each object $X$ is equipped with "copy" and "discard" maps*



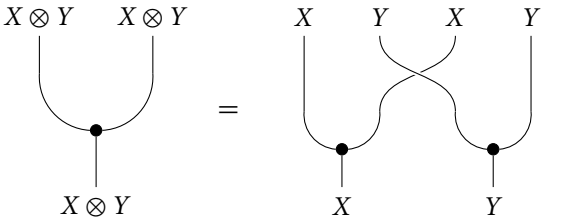

(1)

*satisfying the identities of a commutative comonoid:*



(2)

- *The copy maps are compatible with the monoidal structure in the following way:*

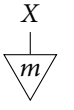



(3)

- C *is semicartesian.*

This definition is motivated by the fact that statements in probability theory often refer to the same variable multiple times, which explains why the copy morphisms are relevant. In a semicartesian category, we still have the discarding maps del, for which we use the same symbol in string diagrams. For further background on Markov categories, we refer the reader to Fritz (2020). For a history of the concept, see Fritz and Liang (2023, Introduction and Remark 2.2)

A state in a Markov category, or more generally in a semicartesian category, is a morphism from the monoidal unit, i.e., in the form $m\colon I \to X$. We denote it by a triangle,




States are the abstract categorical generalization of probability measures.

**Example 1.3.** (Probabilistic Markov categories). Here are some Markov categories of interest in probability theory and which are used in this work.

- The category FinStoch has as objects finite sets, and as morphisms stochastic matrices with their usual composition. We denote the entries of a matrix $m\colon X \to Y$ by $m(y|x)$, which we can interpret as a discrete transition probability.
- The category Stoch has as objects measurable sets, and as morphisms Markov kernels with their usual composition. Given such a kernel $k\colon X \to Y$, we denote its value on a point $x \in X$ and on a measurable subset $B \subseteq Y$ by $k(B|x)$, and again we can interpret it as the probability of obtaining an outcome in $B$ given input $x$.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

- The category BorelStoch is the full subcategory of Stoch whose objects are *standard Borel spaces*, i.e., measurable spaces whose $\sigma$-algebra can be written as the Borel $\sigma$-algebra of a complete separable metric space (often called a *Polish space*).
- The category QBStoch has as objects quasi-Borel spaces and as morphisms kernels between them. We refer to Section 3 and Heunen et al. (2017), Sabok et al. (2021) for details.

In all these examples, states are probability measures. In the case of FinStoch, they are discrete and finitely supported.

**Example 1.4.** (Semiring-valued kernels). One can generalize the Markov category FinStoch to one in which transition "probabilities" are valued in an arbitrary commutative semiring $R$ instead of the semiring of non-negative real numbers. Furthermore, by requiring that, for each $x \in X$, the transition probability $m(y|x) \in R$ is nonzero for finitely many $y \in Y$ only, we can extend its objects to include all sets rather than merely the finite ones. In this way, we obtain a Markov category $\mathsf{Kl}(D_R)$, which is equivalently the Kleisli category of the $R$-distribution monad $D_R$ on the category of sets and functions. See Fritz et al. (2023, Example 3.3) or Coumans and Jacobs (2013, Section 5.1) for more details. For example, taking $R := \mathbb{R}_+$ to be the nonnegative reals results in the category $\mathsf{Kl}(D_{\mathbb{R}_+})$, within which FinStoch is the full subcategory on finite sets.

We return to Markov categories of semiring-valued kernels throughout the article. In particular, we identify properties of the semiring $R$ that characterize when they are positive and representable (Proposition 2.12) as well as when they satisfy the causality axiom (Proposition 2.20). We use these to show that the Markov category $\mathsf{Kl}(D_R)$ is causal (Proposition 2.23) whenever $R$ is a bounded distributive lattice and to construct a Markov category that is positive but not causal (Proposition 2.25).

**Example 1.5.** ($\mathbb{R}$-valued kernels). For some counterexamples in this paper, we also need the category $\mathsf{FinStoch}_{\pm}$, which can be defined as the full subcategory of $\mathsf{Kl}(D_{\mathbb{R}})$ on finite sets. More explicitly:

- Just as in FinStoch, objects are finite sets;
- Just as in FinStoch, a morphism $f\colon X \to Y$ is a $Y$-by-$X$ matrix whose columns sum to one,

$$\sum_{y \in Y} f(y|x) = 1. \tag{4}$$

  However, unlike in FinStoch, we do not require that the entries $f(y|x)$ are nonnegative.
- The copy and discard structures are the same as in FinStoch, which is embedded as a full subcategory.

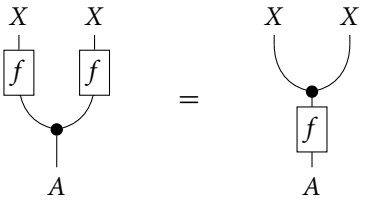
**Example 1.6.** (Cartesian categories). Every cartesian monoidal category is a Markov category with the copy morphisms given by the diagonal maps $X \to X \times X$. The categories Set and FinSet are examples.

The main theme of this work is the positivity axiom, which we recall in Definition 2.1. It has played an important role in the proofs of theorems on sufficient statistics in Markov categories in Fritz (2020), and variants of it have also been used in the quantum context (Parzygnat 2020).

In this work, we also make use of the following additional concepts for Markov categories, for which we also refer to earlier sources for more detail.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

**Definition 1.7.** (Fritz 2020, Definition 10.1). *A morphism $f\colon A \to X$ in a Markov category is deterministic if it commutes with copying in the following way:*

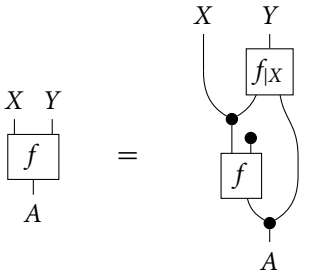



(5)

*We denote by $\mathsf{C}_{det}$ the wide subcategory of $\mathsf{C}$ consisting of deterministic morphisms.*

For example, a state in Stoch is deterministic if and only if it is a probability measure that assigns either 0 or 1 to each measurable subset. In BorelStoch, these are exactly the Dirac delta measures. A cartesian monoidal category is exactly a Markov category where each morphism is deterministic.

**Definition 1.8.** (Fritz 2020, Definition 11.5). *Given a morphism $f\colon A \to X \otimes Y$ in a Markov category, a conditional of $f$* given *$X$ is a morphism $f_{|X}\colon X \otimes A \to Y$ such that*

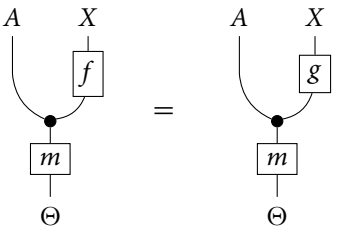



(6)

*holds. We say that a Markov category has conditionals if every morphism admits a conditional.*

For example, if $f$ is a state $I \to X \otimes Y$ in Stoch, the conditional $f_{|X}$ recovers exactly the notion of *regular conditional probability* (Fritz 2020, Example 11.3), and for general $f$, a conditional is the same thing with an additional measurable dependence on an additional parameter. BorelStoch has conditionals, but Stoch does not since there are probability measures on product spaces without regular conditional probabilities (Faden 1985). We suggest (Fritz 2020, Section 11) for additional context.

**Definition 1.9.** (Fritz 2020, Definition 13.1). *Given morphisms $f, g\colon A \to X$ in a Markov category, and a morphism $m\colon \Theta \to A$, we say that $f$ and $g$ are $m$-almost surely equal, and write $f =_{m\text{-a.s.}} g$, if*




(7)

In Stoch, especially when $m$ is a state, this recovers exactly almost sure equality with respect to a measure (Fritz 2020, Example 13.3).

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

**Definition 1.10.** (Fritz et al. 2023, Definition 3.10). *A Markov category* $\mathsf{C}$ *is representable if the inclusion functor* $\mathsf{C}_{det} \to \mathsf{C}$ *has a right adjoint* $P\colon \mathsf{C} \to \mathsf{C}_{det}$. *In this case, we write*

$$\mathsf{samp}_X\colon PX \to X \tag{8}$$

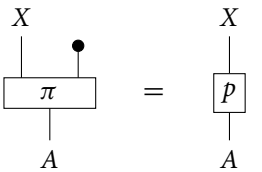
*for the counit of the adjunction, and* $f^\sharp\colon A \to PX$ *for the deterministic counterpart of a morphism* $f\colon A \to X$.

We denote the right adjoint, and also the induced monad on $\mathsf{C}_{\mathrm{det}}$, by $P$ in order to evoke the association with a probability monad. For example, $\mathsf{BorelStoch}$ is representable, since a Markov kernel $A \to X$ is the same thing as a measurable map $A \to PX$, where $PX$ is the measurable space of probability measures on $X$. The counit (8) is the Markov kernel given by $\mathsf{samp}_X(S|\mu) = \mu(S)$ for every measurable $S \in \Sigma_X$, which we interpret as the kernel which outputs a random sample from every distribution $\mu$.

A Markov category $\mathsf{C}$ is representable if and only if it is the Kleisli category of an affine commutative monad on $\mathsf{C}_{\mathrm{det}}$ (Fritz et al. 2023, Section 3.2). Because of this, a representable Markov category can be thought of as adding probabilistic effects (via the monad) to a cartesian category. For example, $\mathsf{BorelStoch}$ is the Kleisli category of the Giry monad on $\mathsf{BorelMeas} = \mathsf{BorelStoch}_{\mathrm{det}}$.

Besides the above, a very important notion for this work is that of *dilations*. While this concept has been established for some time, in particular for quantum states and processes (Chiribella 2014; Chiribella et al. 2010; Selby et al. 2021), dilations have not been systematically studied or applied in the general setting of semicartesian categories before the third author's PhD thesis (Houghton-Larsen 2021).

**Definition 1.11.** *Let* $\mathsf{D}$ *be a semicartesian category (i.e.,* $\mathsf{D}$ *need not be a Markov category). Let* $p\colon A \to X$ *be any morphism in* $\mathsf{D}$. *A dilation of* $p$ *is any morphism* $\pi : A \to X \otimes E$ *for some* $E \in \mathsf{C}$ *satisfying*




(9)

*We call* $E$ *the environment of the dilation.*

Intuitively, a dilation $\pi$ describes a process that coincides with $p$ while potentially leaking information to an "environment" $E$. Conversely, $p$ is obtained from $\pi$ by ignoring the leaked information. Dilations have been studied extensively in Houghton-Larsen (2021) in order to give an abstract account of the so-called *self-testing* of quantum instruments. There, various concepts relating to the structure of dilations were introduced (such as completeness, universality, localizability, purifiability), and these give a flavour of the various types of properties which dilations may or may not enjoy in a given category. Dilations have also been utilized in Selby et al. (2021)[1] to formulate a categorical purification axiom as part of a characterization of quantum theory among other theories of physical processes. We use dilations for a similar purpose in Sections 4.2 and 4.3, namely to characterize positivity in Markov categories, as well as distinguishing positive Markov categories among other semicartesian categories. We study them in detail in Section 4.

We end this background section by relating dilations to convex combinations, which adds some further intuition to the concept of dilation. In Moss and Perrone (2023, Section 3.1), it is shown that for categories of Markov kernels, a convex decomposition of a state can be expressed by means of categorical composition. We now sketch how convex combinations can also be expressed in terms of dilations, in a way that generalizes to all morphisms (not just states). The basic idea is

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

that a dilation $\pi : A \to X \otimes E$ of $p\colon A \to X$ can express explicitly the different terms, indexed by $E$, which appear in a convex decomposition of $p$.

**Example 1.12.** (Convex combinations as dilations). In FinStoch, consider the state $p\colon I \to X$ over $X = \{x, y, z\}$ given by

$$p(x) = 1/6, \qquad p(y) = 3/6, \qquad p(z) = 2/6 \tag{10}$$

We would like to express the distribution $p$ as a convex combination

$$p = \frac{1}{3}\, p_1 + \frac{2}{3}\, p_2 \tag{11}$$

where

$$\begin{aligned} &p_1(x) = 1/2, && p_1(y) = 1/2, && p_1(z) = 0, \\ &p_2(x) = 0, && p_2(y) = 1/2, && p_2(z) = 1/2. \end{aligned} \tag{12}$$

To this end, we can use the environment $E = \{1, 2\}$ and the dilation $\pi : I \to X \times E$ given by

$$\begin{aligned} &\pi(x, 1) = \frac{1}{3}\, p_1(x), && \pi(y, 1), = \frac{1}{3}\, p_1(x), && \pi(z, 1) = \frac{1}{3}\, p_1(x), \\ &\pi(x, 2) = \frac{2}{3}\, p_2(x), && \pi(y, 2), = \frac{2}{3}\, p_2(x), && \pi(z, 2) = \frac{2}{3}\, p_2(x). \end{aligned} \tag{13}$$

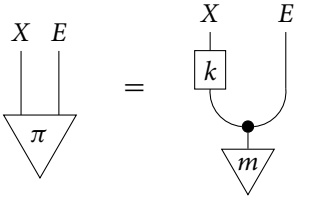
In this way, we can view $\pi$ as equivalent to the convex decomposition from (11). The object $E$ indexes the "components" of $p$ via the conditional distribution $\pi_{|E}\colon E \to X$. Furthermore, the coefficients appearing in Equation (11) correspond to the marginal distribution of $\pi$ on $E$.

In general, if we have a morphism $p\colon A \to X$ for general $A$, viewed as a parametrized state, the same procedure gives a decomposition where both the coefficients and the components in Equation (11) are allowed to depend on a parameter ranging over $A$. While this implements finite convex combinations in FinStoch, in measure-theoretic probability such as BorelStoch, one obtains infinitary convex decompositions in the form of integrals.

Let us now establish a link between dilations and the decompositions of states of Moss and Perrone (2023). Let C be a Markov category, and let $p\colon I \to X$ be a state. In Moss and Perrone (2023), decomposing $p$ means expressing it as a composition

$$I \xrightarrow{\ m\ } E \xrightarrow{\ k\ } X \tag{14}$$

for some object $E$, which indexes the components (and $m$ plays the role of the "coefficients," see Moss and Perrone 2023, Section 3.1 for more). Such a decomposition always gives a dilation given by




(15)

and this dilation represents $p$ as a convex combination in a way that is equivalent to (14). In fact whenever conditionals exist, the two kinds of decomposition are equivalent.

**Proposition 1.13.** *Let* C *be a Markov category with conditionals, and let* $p\colon I \to X$ *be a state. For every object E, the construction of Equation* (15) *establishes a bijective correspondence between:*

- *dilations of p with environment E, and*

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

- *decompositions of $p$ via $E$ up to almost sure equality, i.e., pairs $(m, [k])$ where*
  - *$m\colon I \to E$ is a state;*
  - *$[k]$ is an equivalence class of morphisms $k\colon E \to X$ modulo $m$-a.s. equality;*
  - *$k \circ m = p$ for each $k \in [k]$.*

The proof can be seen as an abstraction of Example 1.12.

*Proof.* Equation (15) shows how to map from decompositions to dilations. The surjectivity of this map is immediate by the existence of conditionals. For injectivity, note that $m$ can be recovered as the marginal on $E$; then, the claim holds by definition of $m$-almost sure equality. □

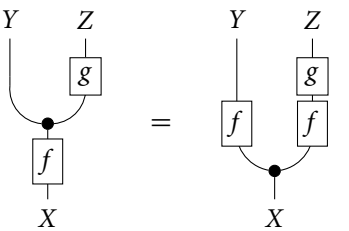
## 2. Conditions for Positivity of Markov Categories

The positivity axiom for Markov categories, as introduced in Fritz (2020), formalizes the idea that any (potentially random) intermediate outcome of a deterministic process is independent of the output given the input. The goal of this section is to present a number of reformulations of the positivity axiom for Markov categories. These underline its significance and shed further light on its intuitive meaning. Along the way, we develop a number of related notions that may be of independent interest.

### *2.1 The positivity axiom*

Here, we recall the definition of a positive Markov category. Readers familiar with this material may proceed directly to Section 2.2.

**Definition 2.1.** *A Markov category* C *is positive if whenever $f\colon X \to Y$ and $g\colon Y \to Z$ are such that $g \circ f$ is deterministic, then we have*




(16)

The positivity property was introduced under this name in Fritz (2020) because the proof that it holds in Stoch relies importantly on the nonnegativity of probabilities. It was also observed there that positivity follows from the existence of conditionals. Moreover, positivity fails in $\mathsf{FinStoch}_{\pm}$ (Fritz 2020, Example 11.27), which provides a nice way to see that $\mathsf{FinStoch}_{\pm}$ does not have conditionals.

**Remark 2.2.** (i) The intuition behind the positivity axiom is that if a composite computation $gf$ is deterministic, then it is possible to calculate the intermediate result (the output of $f$) independently of the result of the entire computation.

(ii) The stronger notion of strict positivity[2] relativizes the positivity axiom with respect to almost sure equality (Fritz 2020, Definition 13.16). This property is relevant for proving properties of sufficient statistics, namely versions of Fisher–Neyman factorization theorem (Fritz 2020, Theorem 14.5) and of Basu's theorem (Fritz 2020, Theorem 15.8). We briefly consider strict positivity, and refer to it as *relative positivity*, in Section 2.5.

(iii) Parzygnat (Parzygnat 2020, Section 4) has considered a quantum analogue of the positivity axiom, but for subcategories of a (quantum) Markov category. It is indeed related to notions

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

of positivity in the quantum setting. In particular, the category of linear unital maps between finite-dimensional C*-algebras includes Schwarz positive maps (and thus also completely positive maps) as a subcategory that satisfies the relevant positivity axiom. Nevertheless, the subcategory of all positive linear maps does not satisfy the categorical notion of positivity (Parzygnat 2020, Example 4.9).

**Remark 2.3.** As a rather weak information flow axiom, one may consider the property that every isomorphism is deterministic. This is not the case in every Markov category (Fritz 2020, Remark 10.10); For example in $\mathsf{FinStoch}_{\pm}$ defined in Example 1.5, every invertible matrix is an isomorphism, but it is only the permutation matrices among them which are also deterministic.

The property that every isomorphism is deterministic is a simple consequence of positivity (Fritz 2020, Remark 11.28). Indeed if $f$ is an isomorphism, then taking $g = f^{-1}$ in Equation (16) and composing with $f$ on the right output shows that $f$ is deterministic.

Conversely, the condition "isomorphisms are deterministic" does not imply positivity. For example, consider the wide subcategory of $\mathsf{FinStoch}_{\pm}$, a morphism of which is

- an arbitrary stochastic matrix (i.e., a morphism of $\mathsf{FinStoch}$), or
- any morphism of $\mathsf{FinStoch}_{\pm}$ whose rank is 1.

Since this class of morphisms is closed under composition, tensor product and contains all the structure morphisms of $\mathsf{FinStoch}_{\pm}$, it follows that it is a Markov category in its own right. To see that every isomorphism is deterministic, note that a stochastic matrix of rank 1 is not invertible unless its domain and codomain are both singletons, in which case its only entry must be 1. To see that positivity fails, consider e.g.

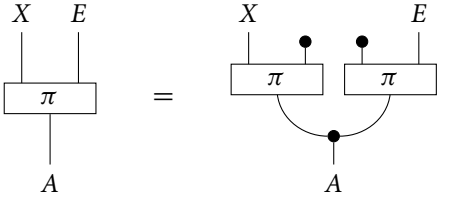
$$f = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \qquad g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \tag{17}$$

for which $g \circ f$ is deterministic, but Equation (16) does not hold.

### *2.2 Deterministic marginal independence is equivalent to positivity*

In probability theory, it is an obvious fact that a deterministic random variable is independent of any other random variable. This fact has previously made a brief appearance in the Markov categories framework in Fritz (2020, Proposition 12.14). Here is the general definition.

**Definition 2.4.** *A Markov category* $\mathsf{C}$ *satisfies deterministic marginal independence (DMI) if for every deterministic morphism* $p\colon A \to X$*, every dilation* $\pi : A \to X \otimes E$ *of* $p$ *displays the conditional independence of* $X$ *and* $E$ *given* $A$*, i.e.*




(18)

Equivalently, DMI says that every morphism $A \to X \otimes E$ that has a deterministic marginal (say, on $X$) must display conditional independence of $X$ and $E$ given $A$.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

Note that the term "deterministic marginal independence" is intended to be understood as "(deterministic marginal) independence," not as "deterministic (marginal independence)."

In words, deterministic marginal independence states that a deterministic output of a process cannot be correlated with another output. The following example already illustrates that this property is also related to the nonnegativity of probabilities.

**Example 2.5.** (DMI for stochastic matrices). FinStoch satisfies deterministic marginal independence. For instance, consider trivial input $A = I$, so that $\pi$ is a joint distribution of $X$ and $E$. Deterministic states in FinStoch are point distributions, so that we have $p = \delta_{x_0}$ for some $x_0 \in X$. The assumption that $\pi$ dilates $p$ means that for $x \in X$,

$$\sum_{e \in E} \pi(x, e) = \begin{cases} 1 & \text{if } x = x_0, \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

holds. By the nonnegativity of probabilities, this implies that $\pi(x, e)$ vanishes for all $e$ whenever $x \neq x_0$. Consequently, the other marginal of $\pi$ is given by

$$\pi_E(e) = \pi(x_0, e), \tag{20}$$

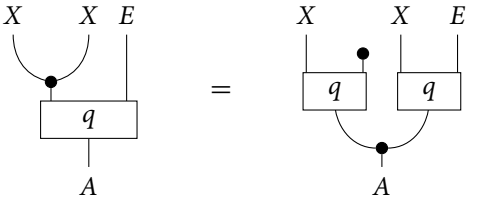
and the whole joint distribution can be written as

$$\pi(x, e) = \delta_{x_0}(x) \cdot \pi(x_0, e) = p(x) \cdot \pi_E(e) \tag{21}$$

which is precisely the desired Equation (18). For more general morphisms, the same holds, where now both $p$ and $\pi$ depend on an additional parameter.

Another closely related notion is the following one.

**Definition 2.6.** *A morphism $q\colon A \to X \otimes E$ is deterministic in $X$ if and only if it satisfies*




(22)

Discarding the output $E$ shows that if $q$ is deterministic in $X$, then its marginal $q_X$ is deterministic. However, the converse does generally not hold, as the following example shows.

**Example 2.7.** (Deterministic marginals with negative probabilities). In $\mathsf{FinStoch}_{\pm}$, a joint distribution $q\colon 1 \to X \otimes E$ is deterministic in $X$ if and only if for all outcomes $x_1, x_2 \in X$ and $e \in E$, we have

$$\delta_{x_1,x_2} \cdot q(x_2, e) = q_X(x_1) \cdot q(x_2, e). \tag{23}$$

In FinStoch, this property is equivalent to the marginal $q_X$ being deterministic; this is an instance of Theorem 2.8 below. In $\mathsf{FinStoch}_{\pm}$, instead, there are joint distributions $q$ which are not deterministic in $X$ although their marginal $q_X$ is deterministic. For example, for $X = \{x, y\}$ and $E = \{a, b\}$, taking the signed distribution with joint probabilities given by

| $q$ | $a$ | $b$ |
|---|---|---|
| $x$ | $1/2$ | $1/2$ |
| $y$ | $1/2$ | $-1/2$ |

(24)

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

has marginal $q_X$ equal to $\delta_x$, which is deterministic. However, setting $x_1 = x_2 = y$ and $e = a$ in (23) results in

$$1 \cdot 1/2 \neq 0 \cdot 1/2. \tag{25}$$

The culprit is that the marginal $q_X$ gives mass zero to $y$, but the joint probability $q$ gives nonzero mass to the point $(y, a)$. This is possible because in this category we are allowing negative probabilities, and some mass cancels out when we form the marginal.

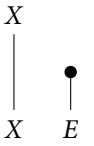
**Theorem 2.8.** *For a Markov category* C*, the following are equivalent:*

*(i)* C *is positive.*
*(ii)* C *satisfies deterministic marginal independence.*
*(iii) For all* $q\colon A \to X \otimes E$*,*

$$q \text{ is deterministic in } X \iff q_X \text{ is deterministic.}$$

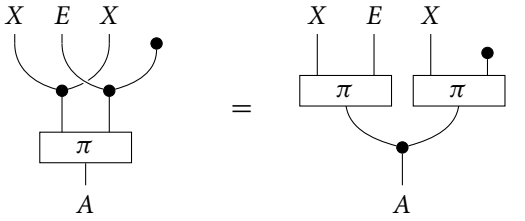
*Proof.*

*(i)* ⇒ *(ii)*: This was proven as Fritz (2020, Proposition 12.14), and we recall the argument here for completeness. If $p$ is deterministic, then in the defining Equation (16) we take a dilation $\pi$ thereof in place of $f$ and the morphism



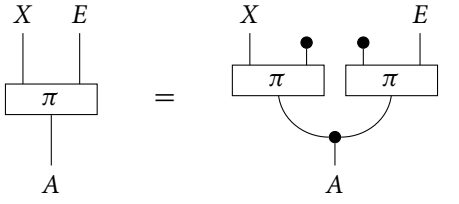

(26)

in place of $g$. Then the composite $g \circ f$ is equal to $p$, which is deterministic by assumption. Equation (16) now reads




(27)

We then get the desired conditional independence by marginalizing over the leftmost output and swapping the other two,

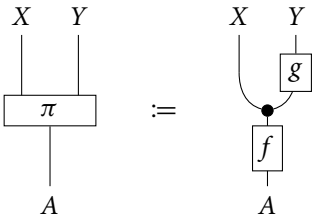



(28)

implicitly using the commutativity of copy.

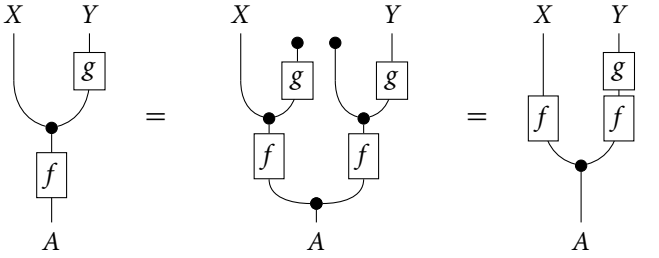
https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

(*ii*) ⇒ (*i*): Consider $f\colon A \to X$ and $g\colon X \to Y$ with a deterministic composite $g \circ f$. Then the morphism




(29)

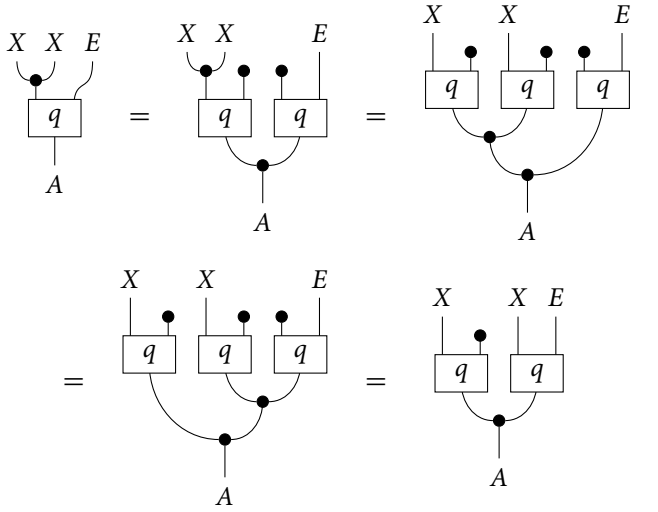
is a dilation of its $Y$-marginal $g \circ f$, which is deterministic by assumption. Therefore deterministic marginal independence applies and gives




(30)

as was to be shown.

(*ii*) ⇒ (*iii*): We already noted that determinism in $X$ of $q$ always implies that its $X$-marginal $q_X$ is deterministic, so we focus on the backward implication, assuming that C satisfies DMI. Consider a morphism $q\colon A \to X \otimes E$ with deterministic marginal $q_X$. By DMI, $q$ displays the conditional independence of $X$ and $E$ given $A$. Using both of these properties entails



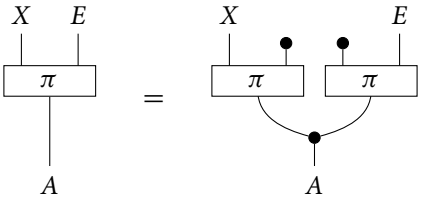

(31)

so that $q$ is indeed deterministic in $X$. The first and last equations hold because of the conditional independence; the second one uses the fact that $q_X$ is deterministic; and the third one holds by associativity of copying.

(*iii*) ⇒ (*ii*): Let $\pi : A \to X \otimes E$ be a dilation of a deterministic morphism $p\colon A \to X$. Then $\pi$ is necessarily deterministic in $X$ by the assumed Property (iii). Marginalizing the middle

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

output in Equation (22) gives




(32)

which is the desired conditional independence.

□

**Remark 2.9.** Let us call a morphism $p\colon A \to X$ globally deterministic if every dilation of $p$ is deterministic in $X$. We can then express Property (iii) as saying

$$\mathsf{C}_{\mathrm{det}} = \mathsf{C}_{\mathrm{gd}} \tag{33}$$

where $\mathsf{C}_{\mathrm{gd}}$ denotes the class of globally deterministic morphisms in $\mathsf{C}$.

As a consequence of Theorem 2.8, Example 2.7 corresponds to the failure of positivity in $\mathsf{FinStoch}_{\pm}$ as noticed in Fritz (2020, Example 11.27).

We can also use Theorem 2.8 to establish the conditions under which Markov categories of semiring-valued stochastic matrices from Example 1.4 are positive. First, let us characterize the deterministic morphisms therein. To that end, recall that a commutative semiring $R$ is entire if $0 \neq 1$ and if $R$ has no zero divisors in the sense that

$$rs = 0 \quad \Longrightarrow \quad r = 0 \text{ or } s = 0 \tag{34}$$

holds.

**Lemma 2.10.** *Let $R$ be an entire commutative semiring. A morphism $f\colon A \to X$ in $\mathsf{Kl}(D_R)$ is deterministic if and only if it can be expressed as*

$$f(x|a) = \delta_{f^\flat(a)}(x) \tag{35}$$

*for a function $f^\flat \in \mathsf{Set}(A, X)$, where $\delta_{x'} \in D_R(X)$ is the delta distribution given by*

$$x \mapsto \begin{cases} 1 & \text{if } x = x', \\ 0 & \text{if } x \neq x'. \end{cases} \tag{36}$$

*Moreover, $f^\flat$ is then uniquely determined by $f$.*

In other words, deterministic morphisms of $\mathsf{Kl}(D_R)$ coincide with its pure morphisms in the sense of Moss and Perrone (2022, Definition 2.5). However, they do not necessarily coincide with pure or dilationally pure morphism in the sense of Selby et al. (2021) and Houghton-Larsen (2021), respectively. Note that the function $f^\flat$ is unique by $1 \neq 0$ in $R$.

*Proof.* Since every morphism of that form is clearly deterministic, it suffices to show the forward implication. Let $f$ be deterministic, meaning that

$$f(x|a)\, f(x'|a) = f(x|a)\, \delta_{x'}(x) \tag{37}$$

holds for all $a \in A$ and all $x, x' \in X$. For every $a$ there is an $x$ with $f(x|a) \neq 0$. Since $R$ has no zero divisors by assumption, Equation (37) which takes the form $f(x|a)\, f(x'|a) = 0$ then implies $f(x'|a) = 0$ for every $x'$ distinct from $x$. Normalization then forces $f(x|a) = 1$, so that $f$ satisfies Equation (35).

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

The uniqueness is clear by $0 \neq 1$ in $R$; equivalently, the canonical Kleisli functor $\mathsf{Set} \to \mathsf{Kl}(D_R)$ is faithful. □

While the proof of Lemma 2.10 is instructive, it is worth noting that this statement also follows from Fritz et al. (2023, Propositions 3.4 and 3.6), which in turn also implies that the Markov category $\mathsf{Kl}(D_R)$ is representable.[3]

**Definition 2.11.** *A semiring $R$ is zerosumfree if it satisfies*

$$r + s = 0 \quad \Longrightarrow \quad r = s = 0 \tag{38}$$

*for all $r, s \in R$.*

**Proposition 2.12.** *Let $R$ be an entire commutative semiring. Then the Markov category $\mathsf{Kl}(D_R)$ from Example 1.4 is positive if and only if $R$ is zerosumfree.*

This generalizes the fact that $\mathsf{FinStoch}_{\pm}$ is not positive and can be taken as further motivation for the term "positivity".

*Proof.* We use the characterization of positive Markov categories as those satisfying deterministic marginal independence.

First, assume that $R$ is zerosumfree and consider a dilation $\pi : A \to X \otimes E$ of a deterministic $p\colon A \to X$. By Lemma 2.10, we have

$$\sum_{e \in E} \pi(x, e|a) = p(x|a) = \delta_{p^\flat(a)}(x). \tag{39}$$

Therefore by zerosumfreeness we have $\pi(x, e|a) = 0$ for every $e \in E$ and every $x \in X$ distinct from $p^\flat(a) \in X$. This means that

$$\pi(x, e|a) = \delta_{p^\flat(a)}(x)\, \tilde{p}(e|a) \tag{40}$$

holds, where $\tilde{p}\colon A \to E$ is the $E$-marginal of $\pi$. But this is precisely the statement of deterministic marginal independence.

Conversely, assume that $R$ is not zerosumfree, i.e., that there exist $r$ and $s$ such that $r + s = 0$ and $r \neq 0$. Consider a morphism $\pi : I \to X \otimes E$ with $X = \{x, x'\}$ and $E = \{e, e'\}$, given by the joint distribution $\pi \in D_R(X \times E)$ with

$$\begin{aligned} \pi(x, e) &= 1, & \pi(x', e) &= r, \\ \pi(x, e') &= 0, & \pi(x', e') &= s. \end{aligned} \tag{41}$$

Then the marginal $p \coloneqq \pi_X$ is deterministic. However, $\pi$ is not equal to the product of its marginals, since the latter is instead given by

$$\begin{aligned} \pi_X(x)\, \pi_E(e) &= 1 + r, & \pi_X(x')\, \pi_E(e) &= 0, \\ \pi_X(x)\, \pi_E(e') &= s, & \pi_X(x')\, \pi_E(e') &= 0. \end{aligned} \tag{42}$$

This means that $\mathsf{Kl}(D_R)$ does not satisfy deterministic marginal independence and thus it is not positive. □

### *2.3 Positivity of representable Markov categories*

We will show next that the following existing notion can be used to detect the positivity of a representable Markov category, in terms of the associated commutative monad on the cartesian monoidal category $\mathsf{C}_{\mathrm{det}}$ of deterministic morphisms.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

**Definition 2.13.** (Jacobs 2016, Definition 1). *A strong monad $(P, \mu, \delta)$ with strength $s$ on a cartesian monoidal category* $\mathsf{D}$ *is strongly affine if for all objects $X$ and $Y$ of* $\mathsf{D}$*, the diagram*

$$\begin{array}{ccc} X \times PY & \xrightarrow{s} & P(X \times Y) \\ \downarrow{\scriptstyle \pi_1} & & \downarrow{\scriptstyle P(\pi_1)} \\ X & \xrightarrow{\delta} & PX \end{array} \tag{43}$$

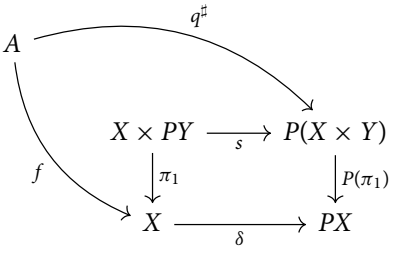
*is a pullback in* $\mathsf{D}$*, where $\pi_1$ is the projection map.*

Taking $X = Y = 1$ shows that a strongly affine monad is in particular affine (satisfies $P1 \cong 1$). Therefore the adverb "strongly" cleverly refers to both a strengthening of affineness and to the fact that the condition involves the strength $s$.

**Proposition 2.14.** *Let* $\mathsf{C}$ *be a representable Markov category with affine commutative monad $P$ on* $\mathsf{C}_{det}$*, so that* $\mathsf{C} = \mathsf{Kl}(P)$*. Then* $\mathsf{C}$ *is positive if and only if $P$ is strongly affine.*

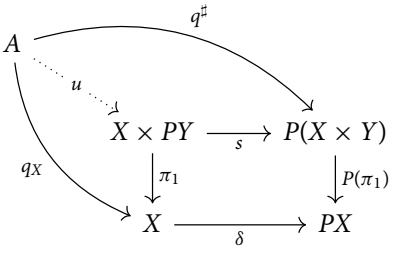
*Proof.* By Theorem 2.8, it suffices to show that $\mathsf{C}$ satisfies deterministic marginal independence if and only if $P$ is strongly affine.

In a representable Markov category, a morphism $p\colon A \to X$ is deterministic if and only if it satisfies $p^\sharp = \delta p$ (Fritz et al. 2023, Proposition 3.12). For the remainder of the proof, we work in $\mathsf{C}_{\mathrm{det}}$ only. Then by the previous statement, a given $q\colon A \to X \times Y$ has deterministic first marginal $q_X$ if and only if there exists $f\colon A \to X$ (namely $q_X$) such that the diagram




(44)

commutes. Now, if the square (43) is a pullback, then $q$ factors (uniquely) through the strength $s$, i.e. there exists a unique morphism $u\colon X \to Y \times PZ$ making the diagram




(45)

commute. By the universal property of the product $X \times PY$ in $\mathsf{C}_{\mathrm{det}}$, the map $u$ is determined by its components. Its $X$-component must be equal to $q_X$ and its $PY$-component must be equal to the marginal $q_Y^\sharp$ in order for the diagram to commute. Indeed, since the diagram

$$\begin{array}{ccc} X \times PY & \xrightarrow{s} & P(X \times Y) \\ & {\scriptstyle \pi_2} \searrow & \downarrow{\scriptstyle P(\pi_2)} \\ & & PY \end{array} \tag{46}$$

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

commutes, we have that $\pi_2 \circ u = P(\pi_2) \circ s \circ u = P(\pi_2) \circ q^\sharp = q^\sharp_Y$. Recall now that the strength $s$ is given by the following composition,

$$X \times PY \xrightarrow{\delta \times \mathrm{id}} PX \times PY \xrightarrow{\nabla} P(X \times Y), \tag{47}$$

where $\nabla$ denotes the lax symmetric monoidal structure morphism of $P$. Therefore, if (43) is a pullback, then we get

$$q^\sharp = s \circ u = \nabla \circ (\delta \times \mathrm{id}) \circ (f, q^\sharp_Y) = \nabla \circ (q^\sharp_X, q^\sharp_Y). \tag{48}$$

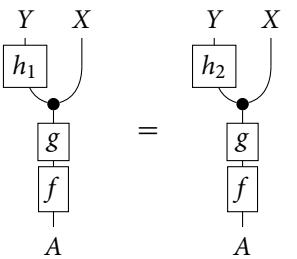
Sampling on both sides produces the desired factorization of Equation (18). Since $q$ was arbitrary, it follows that $\mathsf{C}$ has deterministic marginal independence.

The converse implication follows by the same line of argument upon noting that the above reasoning covers every instance of the universal property of the pullback in $\mathsf{C}_{\mathrm{det}}$. □

### *2.4 Causality and positivity*

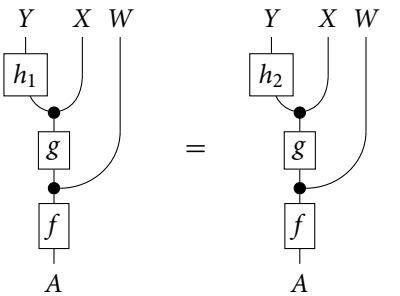
We now turn to another important information flow axiom: causality (Fritz 2020, Definition 11.31).

**Definition 2.15.** (Fritz 2020, Definition 11.31). *A Markov category* $\mathsf{C}$ *is causal if whenever* $f\colon A \to W$, $g\colon W \to X$ *and* $h_1, h_2\colon X \to Y$ *satisfy*




(49)

*then we also have the stronger equation*



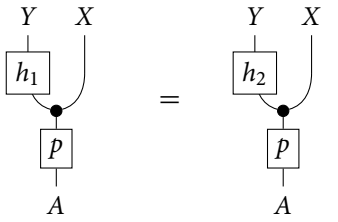

(50)

Intuitively, the axiom states that if a choice between $h_1$ and $h_2$ in the "future" of $g$ does not affect anything that happens from there on, then this choice cannot affect anything that happened in the "past" of $g$ either.

To show that causality is a stronger property than the positivity axiom, it is helpful to have an alternative formulation thereof. The following definition elaborates on Fritz (2020, Remark 11.36).

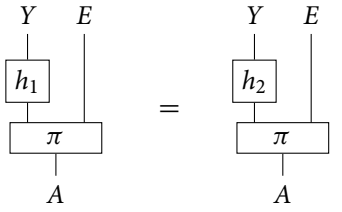
https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

**Definition 2.16.** *A Markov category* C *has parametrized equality strengthening if for any* $h_1, h_2 : X \to Y$ *and any* $p : A \to X$,




(51)

*implies that for every dilation* $\pi$ *of* $p$ *with environment* $E$, *we have*




(52)

Definition 2.16 extends the notion of equality strengthening given by Cho and Jacobs (2019, p. 19), who considered the special case in which $p$ has trivial input, meaning that $A = I$.

**Proposition 2.17.** (Fritz 2020, Remark 11.36). *For a Markov category* C, *the following are equivalent:*

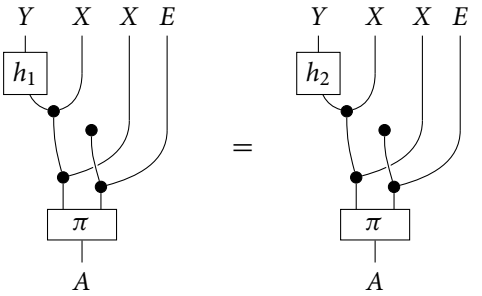
*(i)* C *is causal.*
*(ii)* C *has parametrized equality strengthening.*

*Proof.* The proof amounts to reinterpreting the terms in the respective equalities.

(*i*) ⇒ (*ii*): Consider $h_1, h_2$, and $p$ satisfying Equation (51) and an arbitrary dilation $\pi : A \to X \otimes E$ of $p$. If we define


(53)

then Equation (51) coincides with (49) and applying causality gives

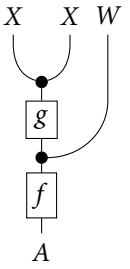



(54)

which, upon marginalizing the two $X$ outputs, gives the required Equation (52).

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

*(ii)* ⇒ *(i)*: Conversely, the morphism




(55)

is a particular dilation of $g \circ f \colon A \to X$ with environment $X \otimes W$, so that applying parametrized equality strengthening to Equation (49) gives Equation (50).

□

Let us use Proposition Proposition 2.17 to characterize causality for Markov categories of semiring-valued stochastic matrices from Example 1.4.

**Definition 2.18.** *An element $r$ of a semiring $R$ is said to have a complement $\bar{r} \in R$ if we have*

$$r + \bar{r} = 1. \tag{56}$$

**Remark 2.19.** A complement need not be unique if it exists. For example, $R$ may satisfy $x + x = x$ and $x^2 = x$ for all $x \in R$, in which case $R$ is equivalently a bounded distributive lattice with join $+$ and meet $\cdot$, and with 0 as bottom and 1 as top element. In this case, 1 is a complement for (and is complemented by) every other element.[4]

**Proposition 2.20.** *Let $R$ be a commutative semiring. The Markov category $\mathsf{Kl}(D_R)$ is causal if and only if, for all $s, t, v, w \in R$ such that $s, t$ and $v + w$ have complements in $R$, we have the following implication:*

$$s(v+w) = t(v+w) \quad \Longrightarrow \quad sv = tv \ \text{ and } \ sw = tw \tag{57}$$

*Proof.* First, let us show that for a semiring satisfying Implication (57), the corresponding Markov category $\mathsf{Kl}(D_R)$ has parametrized equality strengthening. Writing out Equation (51) in components gives

$$h_1(y|x)\, p(x|a) = h_2(y|x)\, p(x|a) \tag{58}$$

which, for any dilation $\pi$ of $p$, reads

$$h_1(y|x) \left[ \sum_{e \in E} \pi(x, e|a) \right] = h_2(y|x) \left[ \sum_{e \in E} \pi(x, e|a) \right]. \tag{59}$$

For any choice of $a \in A$, $e \in E$, $x \in X$, and $y \in Y$, let us define the following elements of $R$,

$$\begin{aligned} &s \coloneqq h_1(y|x), && t \coloneqq h_2(y|x), && v \coloneqq \pi(x, e|a), \\ &\bar{s} \coloneqq \sum_{y' \neq y} h_1(y'|x), && \bar{t} \coloneqq \sum_{y' \neq y} h_2(y'|x), && w \coloneqq \sum_{e' \neq e} \pi(x, e'|a). \end{aligned} \tag{60}$$

Then, by normalization of the respective morphisms, $s$ and $t$ have $\bar{s}$ and $\bar{t}$ as complements, while $v + w$, being equal to $p(x|a)$, has a complement too. Moreover, Equation (59) takes the form of the antecedent of Implication (57). Using this implication then gives $sv = tv$, which reads

$$h_1(y|x)\, \pi(x, e|a) = h_2(y|x)\, \pi(x, e|a). \tag{61}$$

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

This is the componentwise form of Equation (52) that we aimed to show.

Conversely, assume that $\mathsf{Kl}(D_R)$ has parametrized equality strengthening. Let $A = I$ be the singleton and let $X$, $Y$, and $E$ each have cardinality two. We choose morphisms $p$, $\pi$, $h_1$, and $h_2$ with types as above to be given by

$$p \coloneqq \begin{pmatrix} z \\ \bar{z} \end{pmatrix} \qquad h_1 \coloneqq \begin{pmatrix} s & 1 \\ \bar{s} & 0 \end{pmatrix}$$
$$\pi \coloneqq \begin{pmatrix} v & w \\ \bar{z} & 0 \end{pmatrix} \qquad h_2 \coloneqq \begin{pmatrix} t & 1 \\ \bar{t} & 0 \end{pmatrix}. \tag{62}$$

where $z \coloneqq v + w$. In the above matrix notation, $\pi$ ranges over $X$ in its rows and over $E$ in its columns, while $h_1$ and $h_2$ denote stochastic matrices $X \to Y$ with $X$ in columns and $Y$ in rows as usual. With these choices, Equation (58) reads

$$\begin{pmatrix} sz & \bar{s}z \\ \bar{z} & 0 \end{pmatrix} = \begin{pmatrix} tz & \bar{t}z \\ \bar{z} & 0 \end{pmatrix}, \tag{63}$$

as an equality of joint distributions in $D_R(X \times Y)$ with $X$ on the rows and $Y$ on the columns. This equation follows from the assumed antecedent of Implication (57). Applying parametrized equality strengthening to get Equation (61), we then obtain the requisite equations $sv = tv$ and $sw = tw$ by choosing the elements of $X$ and $Y$ that correspond to the upper left corner in Equation (63). □

**Remark 2.21.** Causality implies positivity for semiring-valued kernels. For entire commutative semirings, Propositions 2.20 and 2.12 foreshadow Theorem 2.24, which says that causality implies positivity for arbitrary Markov categories. Indeed, any commutative semiring satisfying Implication (57) is also zerosumfree. To see this, let $s = 1$ and $t = 0$, both of which have complements. Since $v + w = 0$ implies that $v + w$ also has a complement, applying (57) gives $v = 0$ and $w = 0$. Thus, $R$ is then zerosumfree and $\mathsf{Kl}(D_R)$ is a positive Markov category by Proposition 2.12.

**Remark 2.22.** Multiplicative cancellativity implies causality. Another consequence of Proposition 2.20 is that if $R$ is a zerosumfree commutative semiring in which multiplication by nonzero elements can be cancelled, then the Markov category $\mathsf{Kl}(D_R)$ is causal. Indeed, if $v + w = 0$, then by zerosumfreeness we conclude $v = w = 0$, so that Implication (57) is satisfied. Otherwise, the equation $s(v + w) = t(v + w)$ gives us $s = t$ by cancelling $v + w$, and thus the implication also holds in this case. An interesting example where this applies would be the *tropical semiring* $R = [-\infty, +\infty)$ with max as addition and $+$ as multiplication.

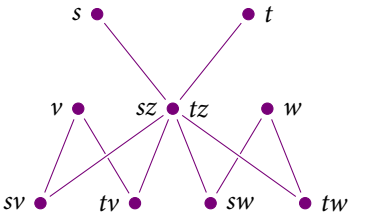
As we show next, the conditions in Proposition 2.20 are also satisfied when $R$ is a bounded distributive lattice (as considered in Remark 2.19). Later on, this will rule such semirings out as potential counterexamples showing that positivity does not imply causality, and we will need to consider more complicated semirings instead (Proposition 2.25).

**Proposition 2.23.** *Let $R$ be a bounded, distributive lattice. Then the Kleisli category $\mathsf{Kl}(D_R)$ is a causal Markov category.*

*Proof.* Let us denote the underlying lattice ordering by $\geq$. We will show that for arbitrary elements $s, t, v, w$ of $R$, Implication (57) holds, from which causality follows by Proposition 2.20. By the

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

antecedent of (57) and the fact that in a lattice we always have $a + b \geq a \geq ab$ and $a + b \geq b \geq ab$, we can infer the order relations depicted in the following Hasse diagram, where $z := v + w$:




(64)

Using transitivity of $\geq$, we extract relations

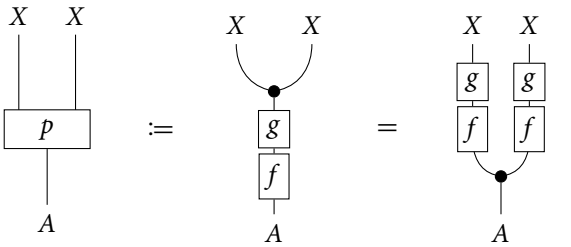
$$s \geq tv, \qquad v \geq tv, \qquad t \geq sv, \qquad v \geq sv. \tag{65}$$

Since $sv$ is the greatest lower bound of $\{s, v\}$, the first two imply $sv \geq tv$, and by similar reasoning the latter two imply $tv \geq sv$. Since $\geq$ is antisymmetric, we finally get $sv = tv$. Analogously, one can obtain the other equation $sw = tw$. By Proposition 2.20, $\mathsf{Kl}(D_R)$ is thus a causal Markov category. □

Returning to the general theory, we now prove a new and surprising implication from causality to positivity.

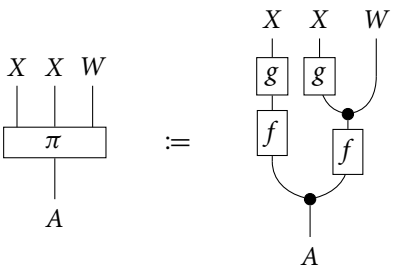
**Theorem 2.24.** *If a Markov category* C *is causal, then* C *is positive. The converse is false.*

*Proof.* Let $f\colon A \to W$ and $g\colon W \to X$ be such that $g \circ f$ is deterministic. Define $p$ to be the morphism




(66)

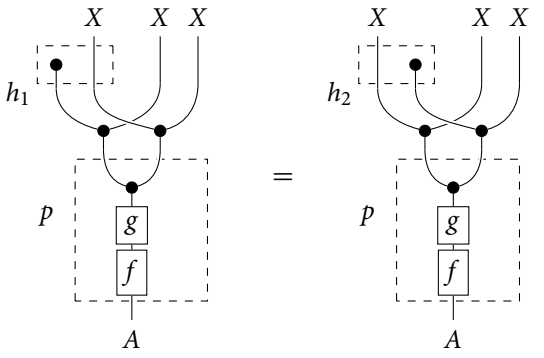
where the two forms of $p$ are equal by the assumption that $g \circ f$ is deterministic. Then there is a dilation $\pi$ of $p$, with environment $W$, given by




(67)

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

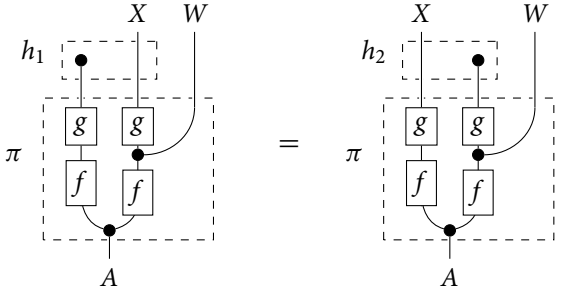
Note that the following equation




(68)

holds by the associativity of copying, where we identify $h_1$ as $\mathrm{del}_X \otimes \mathrm{id}_X$ and $h_2$ as $\mathrm{id}_X \otimes \mathrm{del}_X$.

Since causality is equivalent to parametrized equality strengthening (Proposition Proposition 2.17), we can apply the latter to Equation (68). Replacing $p$ (with its outputs copied) by the dilation $\pi$ from (67) yields




(69)

which is the desired positivity equation up to swapping the outputs (see Definition 2.1).

The fact that positivity does not imply causality in general is shown by constructing a Markov category that is positive but not causal, which we do next in Proposition 2.25. □

In particular, we take inspiration from Propositions 2.12 and 2.20 and look for an entire commutative semiring that is zerosumfree, but does not satisfy Implication (57). By Proposition 2.23, we know that such a semiring cannot be a distributive lattice. Rather, let $\mathcal{I}(\mathbb{Z}[2i])$ be the commutative quantale[5] of ideals of the commutative ring

$$\mathbb{Z}[2i] = \mathbb{Z} \oplus 2i\mathbb{Z} = \{m \oplus 2ik \mid m, k \in \mathbb{Z}\}, \tag{70}$$

which is the subring of the Gaussian integers whose imaginary part is even. The addition and multiplication in $\mathcal{I}(\mathbb{Z}[2i])$ are the ideal addition and multiplication, respectively, with units given by the null ideal $\{0\}$ the whole ring $\mathbb{Z}[2i]$, respectively.

**Proposition 2.25.** *Let $R$ be the semiring of ideals $\mathcal{I}(\mathbb{Z}[2i])$ defined above. Then:*

*(i) $R$ is entire and zerosumfree.*[6]
*(ii) The Markov category $\mathsf{Kl}(D_R)$ is representable and positive.*
*(iii) The Markov category $\mathsf{Kl}(D_R)$ is not causal.*

*Proof.* (i) $R$ is nontrivial as we have $\{0\} \neq \mathbb{Z}[2i]$. To show that it is entire, we thus need to prove that is has no zero divisors. Consider two nonzero ideals $I, J \subseteq \mathbb{Z}[2i]$ with $IJ = \{0\}$. The latter is equivalent to $\alpha\beta = 0$ for all $\alpha \in I$ and $\beta \in J$. Since $\mathbb{Z}[2i]$ itself is entire, this implies $I = J = \{0\}$.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

To show that $R$ is zerosumfree, consider two ideals $I$ and $J$ that sum to zero, i.e., $I + J = \{0\}$. Since the sum of ideals contains each of them, the desired $I = J = \{0\}$ is immediate.

(ii) Representability follows by Fritz et al. (2023, Proposition 3.6) and the previous item. Positivity follows by Proposition 2.12 and the previous item.

(iii) Let us show that Implication (57) fails here. To this end, we choose ideals

$$s = v \coloneqq (2, 4i), \qquad t = w \coloneqq (4, 2i), \tag{71}$$

where $(m, 2ik)$ denotes the set $m\mathbb{Z} \oplus 2ik\mathbb{Z} \subseteq \mathbb{Z}[2i]$. Note that $s$ and $t$ are the principal ideals generated by 2 and $2i$, respectively. We use the $(m, 2ik)$ notation to highlight that these are distinct ideals in $\mathbb{Z}[2i]$, which would not be the case for the Gaussian integers $\mathbb{Z}[i]$. Since every element of $R$ has a complement given by $\mathbb{Z}[2i]$ itself, $s$, $t$, and $(v + w)$ do as well. We can now compute

$$s^2 = t^2 = (4, 8i), \qquad st = (8, 4i), \tag{72}$$

so that we have

$$s(v + w) = s^2 + st = st + t^2 = t(v + w), \tag{73}$$

but nevertheless

$$sv = s^2 = (4, 8i) \neq (8, 4i) = st = tv. \tag{74}$$

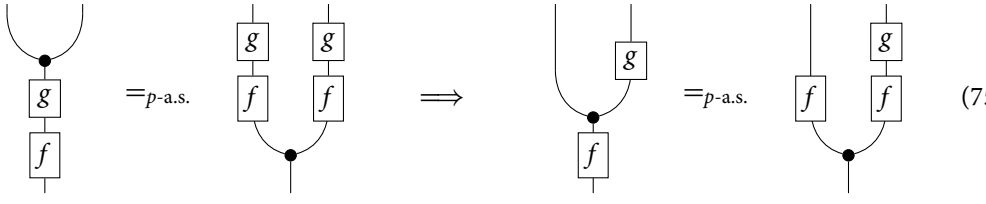
These computations show that Implication (57) fails and thus, by Proposition 2.20, that $\mathsf{Kl}(D_R)$ is not a causal Markov category.

□

**Question 2.26.** Is there a characterization of when a representable Markov category is causal, analogous to Proposition 2.14?

### *2.5 Information flow almost surely*

As a brief aside, let us turn to the notion of strict positivity introduced in Fritz (2020, Definition 13.16). As discussed in Section 4.3, the name strict positivity appears unfortunately chosen in retrospect. We thus refer to it as relative positivity in this article. It corresponds to the positivity axiom (Definition 2.1) with both antecedent and consequent relativized to $p$-almost sure equality. That is, we say that $\mathsf{C}$ is a relatively positive Markov category if for all morphisms $f$, $g$, and $p$ of suitable types, we have



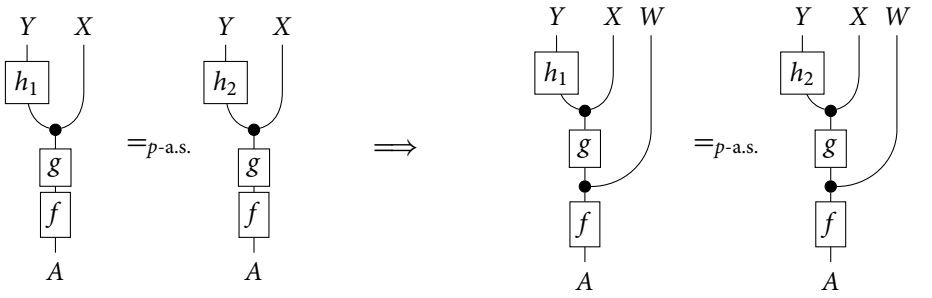

(75)

**Remark 2.27.** Relative positivity implies ordinary positivity by choosing $p = \mathrm{id}$.

In fact, one can formulate similar relative versions of other information flow axioms, such as relative deterministic marginal independence and relative causality. By replacing equalities with $p$-a.s. equalities in the respective proofs, implications in Fig. 1 remain valid also between the relative versions of the axioms, e.g., relative causality implies relative positivity by Theorem 2.24.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

An interesting observation is that the causality axiom is equivalent to its relativized version. That is, the implication




(76)

can be derived by two applications of the causality axiom itself. This lets us show the following strengthening of Theorem 2.24.

**Corollary 2.28.** *If* C *is causal, then* C *is relatively positive.*

*Proof.* By the arguments presented in this section, we have

$$\text{causality} \iff \text{relative causality} \implies \text{relative positivity}.$$

□

It is interesting to note that the first implication cannot be generalized to quantum Markov categories due to Parzygnat (2020, Example 8.28 and Proposition 8.34).[7]

## 3. Quasi-Borel Spaces, the Privacy Equation, and Failure of Positivity

So far, counterexamples to positivity arose in settings constructed for this purpose, such as when we deal with negative probabilities. The purpose of this section is to present *probability theory on function spaces* as a naturally occurring situation in which positivity is violated. The failure of positivity here is not a bug but a *feature*, because it conforms to intuitions about information-hiding and privacy. This failure of positivity is not rooted in the existence of negative probabilities like in $\mathsf{FinStoch}_{\pm}$. Rather, information hiding and a form of destructive interference are central to understanding the failure of positivity in this context.

In short, if $X \sim \nu$ is a random variable sampled from an atomless distribution $\nu$ such as a Gaussian distribution, then we can form the singleton set $A = \{X\}$, which is now a random subset of the real line. As we will show, $A$ is equal in distribution to the empty set, i.e., its law is $\delta_\emptyset$. This means that the distribution of the random pair $(A, X)$ is a state which violates deterministic marginal independence. Indeed its first marginal is deterministic (with value $\emptyset$), but $A$ and $X$ are *not* independent; this can be seen for example because $A \ni X$ holds with probability 1.

In order to make this counterexample precise, we first need to introduce a Markov category capable of expressing random subsets of the real line. This is not possible in Stoch or BorelStoch because the category of (standard Borel) measurable spaces is not cartesian closed (Aumann 1961).

- In Section 3.1, we recall quasi-Borel spaces, which are a model for probabilistic programming with higher order functions and are thus capable of formalizing our example.
- In Section 3.2, we formally define the random singleton distribution as a measure on $2^{\mathbb{R}}$ and show that it equals $\delta_\emptyset$. We then obtain a counterexample to deterministic marginal independence (Proposition 3.3).
- In Section 3.3, we remark that the situation in quasi-Borel spaces has strong connections to fresh name generation in computer science (Sabok et al. 2021). While we do not focus

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

on the models, this matches up well with our analysis of information flow and information leaking and defines another source of interesting Markov categories.

### *3.1 Quasi-Borel spaces*

Quasi-Borel spaces have been introduced in Heunen et al. (2017) as a conservative extension of the category of measurable maps between standard Borel spaces to a cartesian closed category Qbs. This means that one can form function spaces such as $2^{\mathbb{R}}$, the space of all Borel subsets of $\mathbb{R}$, and consider probability distributions on such objects. Quasi-Borel spaces also feature a probability monad $P$ which is commutative, affine, and agrees with the Giry monad on standard Borel spaces. We denote the Markov category obtained as the Kleisli category of $P$ by QBStoch. It serves as an interesting source of counterexamples to information flow axioms, as it can "hide" information flow into objects like $2^{\mathbb{R}}$.

A quasi-Borel space is a pair $(X, M_X)$, where $X$ is a set and $M_X \subseteq X^{\mathbb{R}}$ is a collection of functions $\mathbb{R} \to X$, called random elements, satisfying certain closure properties (Heunen et al. 2017), such as including all constant maps. A morphism of quasi-Borel spaces $(X, M_X) \to (Y, M_Y)$ is a function $f\colon X \to Y$ that preserves random elements. We consider the following quasi-Borel spaces of interest:

- The real line is the quasi-Borel space $\mathbb{R}$ whose random elements are the Borel measurable maps $\mathbb{R} \to \mathbb{R}$.
- The Booleans form the quasi-Borel space 2 with two elements, whose random elements are the Borel measurable maps $\mathbb{R} \to 2$, i.e., Borel subsets of $\mathbb{R}$.
- The exponential $2^{\mathbb{R}}$ in Qbs consists of all Borel measurable maps $\mathbb{R} \to 2$. Its random elements $\mathbb{R} \to 2^{\mathbb{R}}$ are precisely the exponential transposes of Borel measurable maps of type $\mathbb{R} \times \mathbb{R} \to 2$. The evaluation map

$$\mathsf{ev}\colon 2^{\mathbb{R}} \times \mathbb{R} \to 2, \qquad (A, x) \mapsto [\![x \in A]\!] \tag{77}$$

  is a morphism of quasi-Borel spaces, where $[\![x \in A]\!]$ stands for the truth value of the proposition $x \in A$.

Every quasi-Borel space $(X, M_X)$ has an induced $\sigma$-algebra $\Sigma_{M_X}$ given by the largest $\sigma$-algebra which makes all random elements measurable. Equivalently, a subset $A \subseteq X$ is measurable if and only if its characteristic function is a morphism of quasi-Borel spaces of type $(X, M_X) \to 2$.

Random elements $\mathbb{R} \to X$ allow one to push a source of randomness from the quasi-Borel space $\mathbb{R}$ onto the quasi-Borel space $(X, \Sigma_{M_X})$. In this spirit, a probability measure on $(X, M_X)$ is a probability measure on the induced measurable space $(X, \Sigma_{M_X})$ which can be obtained as a pushforward $\alpha_* \mu$, where $\alpha \in M_X$ is a random element and $\mu \in P(\mathbb{R})$ is an ordinary probability measure on $\mathbb{R}$.

**Example 3.1.** The Dirac measure $\delta_x$ on $(X, \Sigma_{M_X})$ is a valid probability measure on the quasi-Borel space $(X, M_X)$ because it can be written as a pushforward of *any* probability measure on $\mathbb{R}$ by the constant random element with image $\{x\}$.

The probability monad $P$ on Qbs assigns, to every quasi-Borel space $(X, M_X)$, the set $P(X)$ of all probability measures on $X$ endowed with a suitable quasi-Borel structure. The unit $\eta_X\colon X \to P(X)$ of the monad is given by the Dirac measure. The monad $P$ is affine and commutative (Heunen et al. 2017), so that its Kleisli category QBStoch is a Markov category.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

### 3.2 Random singleton sets

We can now take advantage of the cartesian closure of quasi-Borel spaces to define random singleton sets. There is a morphism

$$\{-\}\colon \mathbb{R} \to 2^{\mathbb{R}} \tag{78}$$

which sends a number $x \in \mathbb{R}$ to the singleton set $\{x\} \in 2^{\mathbb{R}}$. Note that this morphism is simply the exponential transpose of the equality test $(=)\colon \mathbb{R} \times \mathbb{R} \to 2$. If $\nu \in P(\mathbb{R})$ is a probability measure on the real line, then we can describe the distribution of a *random singleton set* by

$$\mathcal{RS}_\nu = P(\{-\})(\nu) \in P(2^{\mathbb{R}}).$$

Recall that by definition of distributions on a quasi-Borel space, $\mathcal{RS}_\nu$ is the pushforward measure $\{-\}_*\nu$ on the induced measurable space $(2^{\mathbb{R}}, \Sigma_{M_{2^{\mathbb{R}}}})$, defined by

$$\mathcal{RS}_\nu(\mathcal{U}) = \nu\big(\{x \in \mathbb{R} \mid \{x\} \in \mathcal{U}\}\big) \text{ for all } \mathcal{U} \in \Sigma_{M_{2^{\mathbb{R}}}} \tag{79}$$

Using the $\sigma$-algebra $\Sigma_{M_{2^{\mathbb{R}}}}$ is crucial – it ensures, for example, that the set $\{x \in \mathbb{R} \mid \{x\} \in \mathcal{U}\}$ in (79) is measurable. We elaborate on this further in the proof of the next theorem.

**Theorem 3.2.** (Privacy equation Sabok et al. 2021). *For every atomless probability measure $\nu$, the random singleton is equal in distribution to the empty set, i.e.*

$$\mathcal{RS}_\nu = \delta_\emptyset. \tag{80}$$

*Proof idea.* The details of the proof are covered extensively in Sabok et al. (2021, Theorem 4.1). We have to show that for all $\mathcal{U} \in \Sigma_{M_{2^{\mathbb{R}}}}$,

$$\mathcal{RS}_\nu(\mathcal{U}) = \delta_\emptyset(\mathcal{U}) = [\![\emptyset \in \mathcal{U}]\!]. \tag{81}$$

This claim hinges on the fact that the the induced $\sigma$-algebra on $2^{\mathbb{R}}$ is highly restrictive; we have

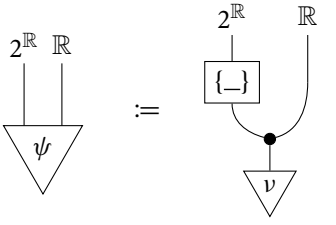
$$\Sigma_{M_{2^{\mathbb{R}}}} = \Big\{\mathcal{U} \mid \forall \alpha\colon \mathbb{R} \times \mathbb{R} \to 2 \text{ measurable}, \ \big\{x \mid \alpha(x, -) \in \mathcal{U}\big\} \in \Sigma_{\mathbb{R}}\Big\}, \tag{82}$$

which is known as Borel-on-Borel in the literature on higher order measurability (e.g. Kechris 1995). All families $\mathcal{U}$ which would be assigned different values by the formulas in (81) turn out to be *not measurable*. As a brief nonexample, consider the family $\mathcal{E} = \{\emptyset\}$. This would clearly differentiate a random singleton from the empty set because we have

$$\mathcal{RS}_\nu(\mathcal{E}) = \nu\big(\{x \in \mathbb{R} \mid \{x\} \in \mathcal{E}\}\big) = 0 \quad \text{and} \quad \delta_\emptyset(\mathcal{E}) = [\![\emptyset \in E]\!] = 1. \tag{83}$$

However, it can be shown that $\mathcal{E} \notin \Sigma_{M_{2^{\mathbb{R}}}}$. In other words, checking if a set is empty is *not* a morphism $2^{\mathbb{R}} \to 2$ in $\mathsf{Qbs}$. □

We call Theorem 3.2 *privacy equation* because the random number $x \in \mathbb{R}$ is "anonymized in distribution" when expressed as a random set $\{x\} \in 2^{\mathbb{R}}$. In particular, we have a dilation




(84)

of $\nu$, with environment $2^{\mathbb{R}}$, which gives no information about the value $x$ in its environment marginal that behaves like the empty set. In this sense, we could view $\psi$ as a "private dilation" – one that leaks no information.

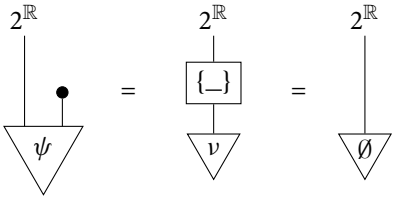
https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

Another way to conceive of private dilations is to require that there is no correlation between the local output and the output leaked into the environment, i.e. that the two outputs are independent. However, the dilation $\psi$ does not factorize, and so it would not be "private" in this sense. In particular, provided access to $x \in \mathbb{R}$, one can distinguish the leaked output from $\emptyset$ by applying the evaluation morphism given in (77).

We employ these curious properties of $\psi$ to show that QBStoch is not a positive Markov category.

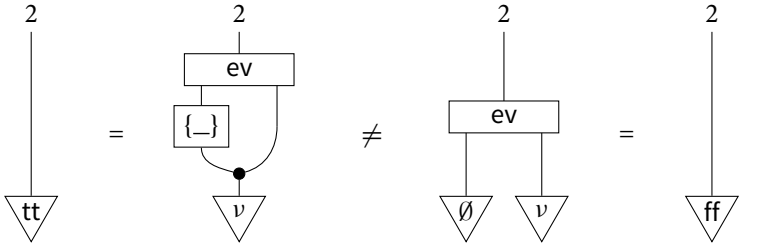
**Proposition 3.3.** *In* QBStoch, *consider the state* $\psi : I \to 2^{\mathbb{R}} \otimes \mathbb{R}$ *defined by Equation* (84). *Then the* $2^{\mathbb{R}}$*-marginal* $\psi_{2^{\mathbb{R}}} : I \to 2^{\mathbb{R}}$ *is deterministic, but* $\psi$ *is not the product of its marginals. Consequently, deterministic marginal independence (Definition 2.4) does not hold in* QBStoch.

*Proof.* By Theorem 3.2, the $2^{\mathbb{R}}$-marginal of $\psi$ equals $\delta_\emptyset$:




(85)

while the second marginal equals $\nu$. Therefore, the product of the marginals is $\delta_\emptyset \otimes \nu$. This is different from $\psi$, as we can witness by postcomposing with the evaluation map




(86)

where $\mathsf{tt}, \mathsf{ff} : 1 \to 2$ are the two boolean truth values. □

By applying Theorems 2.8 and 2.24, we immediately obtain the following consequence, which had already been announced in Sabok et al. (2021), Stein (2021).

**Corollary 3.4.** QBStoch *is neither a positive nor a causal Markov category.*

Note that because QBStoch faithfully contains BorelStoch, positivity *will* hold in the full subcategory of all quasi-Borel spaces that come from standard Borel spaces. The function space $2^{\mathbb{R}}$ is not of that form. This gives a novel, probabilistic reading to Aumann's result that the evaluation morphism ev cannot be made into a measurable map (Aumann 1961). It is, nevertheless, a morphism in Qbs.

### *3.3 Fresh name generation*

Fresh name generation is a classic area of computer science (Pitts and Stark 1993; Stark 1996). A *pure name* is an abstract entity that contains no other information except whether it is equal to other names. Typical examples of names are identifiers such as bound variables: In definitions such as $f(x, y) = xy^2$, the names of the variables $x, y$ do not matter as long as they remain distinct. They could be switched or replaced by say $z, w$ without changing the meaning of the expression.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

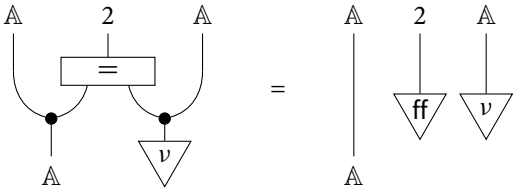
New names are allocated *freshly*, when they are distinct from any other name already in place. In languages such as LISP, this primitive is called `gensym`.

We give a high-level summary of the categorical semantics of name generation and show that it is another instance of information flow which can be modeled using Markov categories. We also show that the information hiding for random functions arises naturally in the context of name generation, which makes it a prototypical example of non-positivity.

A *categorical model of name generation* consists of the following pieces of structure, satisfying further conditions spelled out in Stark (1996, Section 4.1):

- a cartesian closed category $\mathsf{C}$ and a distinguished object $\mathbb{A}$ of names,
- an equality test $(=): \mathbb{A} \times \mathbb{A} \to 2$, where $2 := 1 + 1$ is assumed to exist. (One may think of the coproduct inclusions $\mathsf{tt}, \mathsf{ff}: 1 \to 2$ as represent the boolean truth values true and false.)
- a commutative affine monad $T: \mathsf{C} \to \mathsf{C}$,
- a distinguished state $\nu: I \to T(\mathbb{A})$ which represents picking a fresh name.

Various nondegeneracy axioms are also assumed, such as that the unit $\eta_2: 2 \to T(2)$ is monic. The Kleisli category of $T$ is a Markov category, and the freshness condition for $\nu$ states that testing a fresh name on equality always returns $\mathsf{ff}$. We can write this condition in string diagrams:




(87)

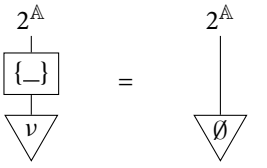
This makes $\nu$ into an abstract version of the atomless measure used in Section 3.2.

An important problem in computer science is then to understand the behavior of higher order functions that generate fresh names locally. The program

$$\texttt{let x} = \texttt{gensym()} \texttt{ in } \lambda \texttt{y}.\,(\texttt{x} = \texttt{y}) \tag{88}$$

generates a fresh name `x` and returns a function $\mathbb{A} \to 2$ which tests its input `y` for equality with `x`. One can show that this function is observably indistinguishable from the function `λy.false` (Stark 1996, Example 8). This is because the name `x` remains *private* or enclosed in the function `λy.(x = y)`, and it can never be extracted programmatically in order to obtain the output `true`.

In a categorical model of name generation, the function `λy.(x = y)` is modeled using the singleton map $\{-\}: \mathbb{A} \to 2^{\mathbb{A}}$ completely analogously to quasi-Borel spaces. A model of name generation is then said to *satisfy the privacy equation* if




(89)

holds. Given these ingredients, we can follow the reasoning of Proposition 3.3 in an analogous way. This is a striking example how the abstraction of Markov categories enables connections between different areas of mathematics and computer science. That is, we obtain the following statement.

**Proposition 3.5.** *Every categorical model of name generation defines a Markov category. If it is nondegenerate and satisfies the privacy equation, then the Markov category is not positive.*

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

It is relatively involved to construct such a model, but Stark provides one in Stark (1996, Section 6.1) using a categorified version of logical relations. We stress that this model has no probabilistic ingredients whatsoever.

On the other hand, the use of Markov categories lets us apply synthetic probabilistic terminology and intuition to reason about name generation. In fact, choosing names at random is a common strategy to implement `gensym` in practice. By using an atomless measure $\nu$, we formally obtain purely probabilistic semantics for name generation, as described in Sabok et al. (2021):

**Theorem 3.6.** (Sabok et al. 2021, Theorem 3.8). *Quasi-Borel spaces are a nondegenerate categorical model of name generation, where $\mathbb{A}$ is any uncountable standard Borel space and $\nu$ any atomless measure. It furthermore satisfies the privacy equation.*

## 4. Dilations and Positivity Properties in Semicartesian Categories

We now shift the focus from Markov categories to the more general semicartesian categories (Definition 1.1). A substantial theory of information flow can be developed already in semicartesian categories, as shown for example in Houghton-Larsen (2021). This suggests that categorical probability might not need Markov categories after all, but that semicartesian categories could in fact be sufficient. This would be of interest not only as a conceptual clarification on the foundations of probability but also insofar as some of its results may apply to quantum probability.

The primary concept needed for the development of categorical probability in semicartesian terms the notion of dilation, which we have used in the previous sections in the context of Markov categories, but which is meaningful even for semicartesian categories in general. Throughout this section, $\mathsf{D}$ thus refers to a semicartesian category.

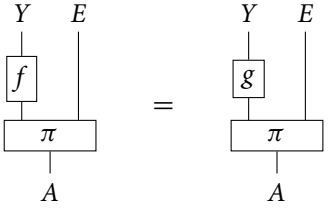
### *4.1 Categories of dilations*

In this part, we study dilations more in detail (recall them from Definition 1.11). The following notion of dilational equality generalizes the definition of strongly almost sure equality (Cho and Jacobs 2019, Definition 5.7), defined originally with respect to a state $p\colon I \to X$. We have already encountered it in Definition 2.16, which defines a Markov category with parametrized equality strengthening as one in which equality almost surely implies dilational equality.

**Definition 4.1.** *Let $p\colon A \to X$ and $f, g\colon X \to Y$ be morphisms in $\mathsf{D}$. We say that $f$ and $g$ are $p$-dilationally equal, written as*

$$f =_{p\text{-dil.}} g,$$

*if for every dilation $\pi$ of $p$, we have*




(90)

Intuitively, $f =_{p\text{-dil.}} g$ means that $f$ and $g$ cannot be distinguished even with access to whatever environment that $p$ may have leaked information to. While this trivially implies $fp = gp$, it is typically a strictly stronger condition:

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

**Example 4.2.** In FinStoch, let $p\colon I \to \{0,1\}$ be the uniform distribution, $f = \mathrm{id}_{\{0,1\}}$, and $g$ the non-identity permutation on $\{0,1\}$. Then clearly $fp = gp$, but using $\pi \coloneqq \mathrm{copy}_{\{0,1\}} \circ p$ witnesses $f \neq_{p\text{-dil.}} g$. This is easy to understand upon noting that the distribution of a fair coin is invariant under exchanging heads and tails, but copying the outcome and switching only one copy while retaining the other clearly changes the distribution.

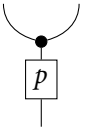
This example suggests that dilational equality in the Markov categories case is closely related to almost sure equality (Fritz 2020, Definition 13.1). We now formalize this relation.

**Proposition 4.3.** *In a Markov category* C*, dilational equality implies equality almost surely. That is, for any* $f, g\colon X \to Y$ *and any* $p\colon A \to X$*,*

$$f =_{p\text{-dil.}} g \quad \Longrightarrow \quad f =_{p\text{-a.s.}} g. \tag{91}$$

*The converse is true if and only if* C *is a causal Markov category.*

*Proof.* Every morphism $p$ has a dilation given by




(92)

Substituting it for $\pi$ in Equation (90) produces exactly the desired $p$-a.s. equality of $f$ and $g$.

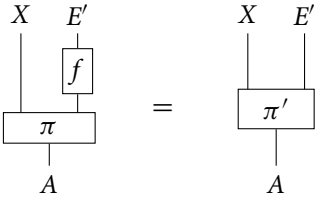
Note that the converse of implication (91) is precisely the property of parametrized equality strengthening from Definition 2.16, which is equivalent to causality as shown in Proposition Proposition 2.17. □

To every morphism in a semicartesian category, we can associate an entire category of dilations — a variant of the one introduced in Houghton-Larsen (2021, Remark 2.2.8).

**Definition 4.4.** *Let* $p\colon A \to X$ *be any morphism in* D*. Then its category of dilations, denoted*

$$\mathsf{Dilations}(p),$$

*has dilations of* $p$ *as objects. Given two dilations of* $p$*, say* $\pi \in \mathsf{D}(A, X \otimes E)$ *and* $\pi' \in \mathsf{D}(A, X \otimes E')$*, morphisms* $\pi \to \pi'$ *in* Dilations$(p)$ *correspond to* $=_{\pi\text{-dil.}}$ *equivalence classes of morphisms* $f \in \mathsf{D}(E, E')$ *such that*

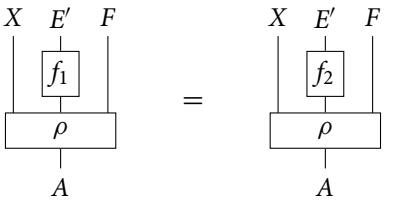



(93)

*holds. Composition of morphisms is defined as composition of representatives in* D*.*

In other words, if $f_1$ and $f_2$ both satisfy Equation (93), then these represent the same morphism of dilations if and only if $f_1 =_{\pi\text{-dil.}} f_2$, which means that for every dilation $\rho$ of $\pi$, we have

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press




(94)

where the third output wire $F$ denotes the additional environment object associated with $\rho$.

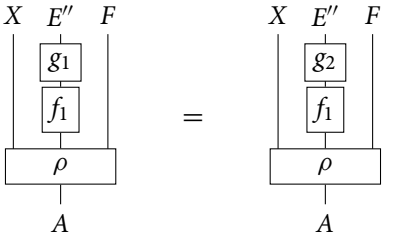
**Lemma 4.5.** *Composition of morphisms in* Dilations($p$) *is well defined.*

*Proof.* Consider $f_1, f_2 \in \mathsf{D}(E, E')$ and $g_1, g_2 \in \mathsf{D}(E', E'')$ as representatives of morphisms $\pi \to \pi'$ and $\pi' \to \pi''$. If $f_1 =_{\pi\text{-dil.}} f_2$, then we have the requisite

$$g_1 \circ f_1 =_{\pi\text{-dil.}} g_1 \circ f_2 \tag{95}$$

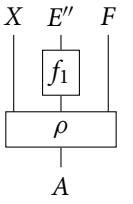
by composing Equation (94) with $g_1$.

On the other hand, given $g_1 =_{\pi'\text{-dil.}} g_2$, we have the requisite




(96)

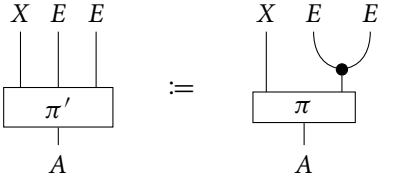
for every dilation $\rho$ of $\pi$ because the morphism




(97)

is itself a dilation of $\pi'$. Thus, the required $g_1 \circ f_1 =_{\pi\text{-dil.}} g_2 \circ f_1$ follows and composition in Dilations($p$) is well-defined. □

**Example 4.6.** (Copying leaked information is irrelevant). If $\pi : A \to X \otimes E$ is a dilation of any morphism $p : A \to X$ in a causal Markov category $\mathsf{C}$, then the dilation given by

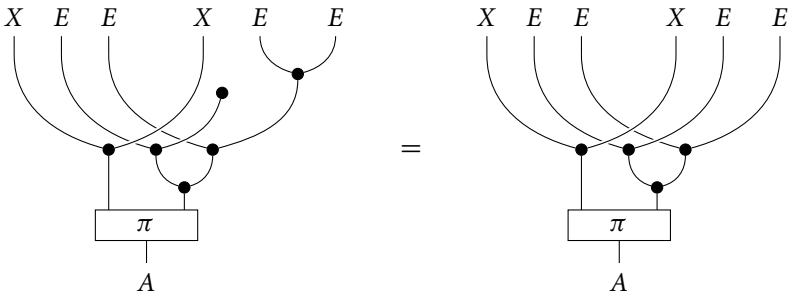



(98)

is isomorphic to $\pi$ in Dilations($p$). To see this, note that the copy morphism itself defines a morphism of dilations $\pi \to \pi'$, while marginalizing either environment output of $\pi'$ defines a morphism $\pi' \to \pi$. The composite $\pi \to \pi' \to \pi$ is trivially equal to $\mathrm{id}_\pi$ already at the level of

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

representatives in C. In the other direction, we use Proposition 4.3 to reduce the claim to proving $\pi'$-almost sure equality. This amounts to the equation




(99)

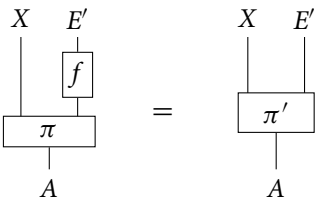
which is a straightforward consequence of the commutative comonoid equations on $E$. Note that for this direction, the composite $\pi' \to \pi \to \pi'$ is generally not equal to $\mathrm{id}_{\pi'}$ on the level of representatives. In fact, in the version of $\mathsf{Dilations}(p)$ where morphisms are *not* identified up to equivalence, the two dilations are generally not isomorphic, since an isomorphism therein in particular constitutes an isomorphism between $X$ and $X \otimes X$, which typically does not exist.

### 4.2 Initial dilations

We introduce and study a variant of the concept of *universal dilations* from Houghton-Larsen (2021, Definition 2.4.1).

**Definition 4.7.** *An initial dilation of a morphism $p$ is an initial object in* $\mathsf{Dilations}(p)$.

Explicitly, a dilation $\pi$ of $p\colon A \to X$ is initial if for every dilation $\pi'$ of $p$ there is a morphism $f$ in $\mathsf{D}$ such that




(100)

holds and moreover such that this $f$ is unique up to $\pi$-dilational equality.

Our notion of initial dilation is intermediate between the notions of universal dilation and of complete dilation from Houghton-Larsen (2021).[8] That is, every universal dilation is initial, and every initial dilation is complete. Indeed, a universal dilation is one for which the $f$ in Equation (100) is unique as a morphism in $\mathsf{D}$ rather than unique up to dilational equality as in the case of initial dilations. On the other hand, a complete dilation is one for which $f$ is merely required to exist with no uniqueness requirement.

**Example 4.8.** The category of finite-dimensional Hilbert spaces and quantum channels has initial dilations in the form of Stinespring dilations. It is shown in Houghton-Larsen (2021, Theorem 2.4.11) that every *minimal* Stinespring dilation is universal and thus initial – but in fact *every* Stinespring dilation is initial. The argument goes as follows. First of all, the relevant morphism $f$ in Equation (100) exists because Stinespring dilations are complete (Houghton-Larsen 2021, Lemma 2.3.8). Moreover, it is unique up to dilational equality because the relation $g =_{\pi\text{-dil.}} h$ for any Stinespring dilation $\pi$ reduces to equality of the composites: $(\mathrm{id} \otimes g) \circ \pi = (\mathrm{id} \otimes h) \circ \pi$. This fact follows because every dilation of such a $\pi$ is given by a tensor product of $\pi$ with a state of the environment (Houghton-Larsen 2021, Corollary 2.3.23).

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

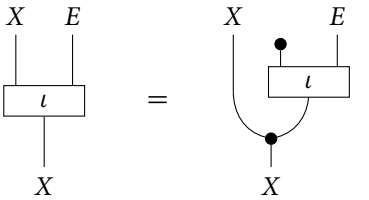
**Example 4.9.** Let $1/\mathsf{Set}$ be the category of pointed sets, and consider $(1/\mathsf{Set})^{\mathrm{op}}$ with cartesian product as the symmetric monoidal structure. This is a semicartesian category. In the remainder of this example, we depict the arrow directions in $1/\mathsf{Set}$ in order to avoid confusion, writing a morphism $f$ from $X$ to $Y$ as $f\colon X \leftarrow Y$. We denote basepoints by the symbol $*$. A dilation of a morphism $p\colon A \leftarrow X$ is a morphism $\pi : A \leftarrow X \times E$ such that $\pi(x, *) = p(x)$ for all $x \in X$.

Writing $X^X$ for the hom-set $(1/\mathsf{Set})(X, X)$ with basepoint the identity map, function evaluation defines a morphism

$$\mathrm{ev}_X\colon X \longleftarrow X \times X^X, \tag{101}$$

which is a dilation of $\mathrm{id}_X$. This dilation is initial, since every dilation $\pi : X \leftarrow X \times E$ of $\mathrm{id}_X$ arises from $\mathrm{ev}_X$ by composing with a unique morphism $X^X \leftarrow E$.

**Example 4.10.** In a Markov category $\mathsf{C}$, a copy morphism $\mathrm{copy}_X$ is a dilation of $\mathrm{id}_X$. If $\mathsf{C}$ is positive, then $\mathrm{copy}_X$ is an initial dilation of $\mathrm{id}_X$. Indeed, any dilation of $\mathrm{id}_X$ can be written as



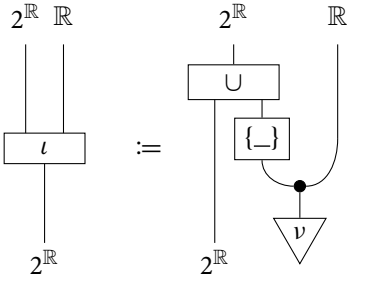

(102)

which follows from the positivity axiom in the form of deterministic marginal independence (Definition 2.4), instantiated with $\pi = \iota$. The uniqueness clause is automatic.

To see that this can fail without positivity, we return to the Markov category of quasi-Borel spaces from Section 3.

**Proposition 4.11.** *In* $\mathsf{QBStoch}$*, the copy map* $\mathrm{copy}_{2^{\mathbb{R}}}$ *is not an initial dilation of the identity.*

*Proof.* In terms of the notation from Section 3, we construct a dilation $\iota\colon 2^{\mathbb{R}} \to 2^{\mathbb{R}} \otimes \mathbb{R}$ of the identity which cannot be obtained from the copy map. To this end, we define

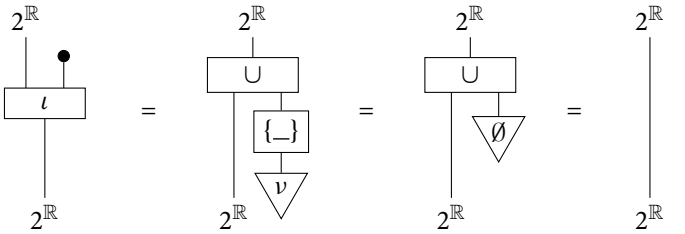



(103)

where $\cup\colon 2^{\mathbb{R}} \otimes 2^{\mathbb{R}} \to 2^{\mathbb{R}}$ takes the union of subsets. That is, $\iota$ modifies its input set $A$ by adding in a random point and records that point in the second output. The map $\cup$ is a morphism of $\mathsf{QBStoch}$ because it can be obtained via cartesian closure and the monad unit from the disjunction map $\vee \in \mathsf{Qbs}(2 \times 2, 2)$.

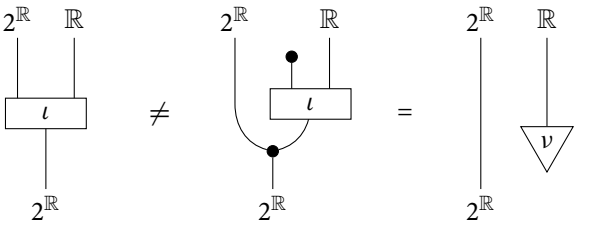
https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

The privacy equation implies that $\iota$ is indeed a dilation of the identity because we have




(104)

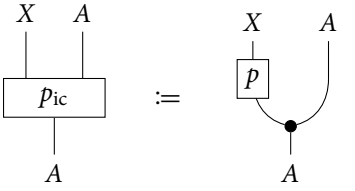
However, $\iota$ cannot be obtained by some dilation morphism from the copy map because we have




(105)

which can be witnessed by postcomposing with the evaluation morphism ev for instance. □

The copy morphism is an example of a specific type of dilation in which only the input is leaked to the environment. More generally, the bloom $p_{\mathrm{ic}}$ (Fullwood and Parzygnat 2021) of a morphism $p\colon A \to X$ is the dilation of $p$ given by[9]




(106)

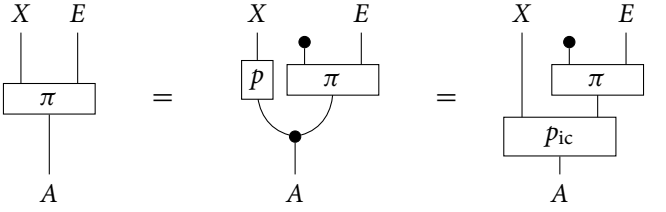
In the following result, we reinterpret positivity as the property that blooms of arbitrary deterministic morphisms are initial dilations.

**Proposition 4.12.** *For a Markov category* C*, the following are equivalent:*

*(i)* C *is positive.*
*(ii) For every deterministic p, its bloom* $p_{\mathrm{ic}}$ *is an initial dilation of p.*

*Proof.*

*(i)* ⇒ *(ii)*: Consider a deterministic morphism $p\colon A \to X$ and a dilation $\pi : A \to X \otimes E$ thereof. By Theorem 2.8, we can apply deterministic marginal independence to $\pi$, which by Equation (18) gives

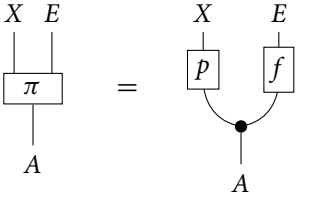



(107)

This makes the bloom $p_{\mathrm{ic}}$ into an initial dilation of $p$, since the uniqueness is automatic by the fact that the $A$-marginal of $p_{\mathrm{ic}}$ is $\mathrm{id}_A$.

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

(*ii*) ⇒ (*i*): We show that C satisfies deterministic marginal independence, which is enough by Theorem 2.8.
So once again let $\pi$ be an arbitrary dilation of a deterministic morphism $p$. Then, by assumption, $\pi$ factors through the bloom of $p$. That is, there exists an $f\colon A \to E$ satisfying




(108)

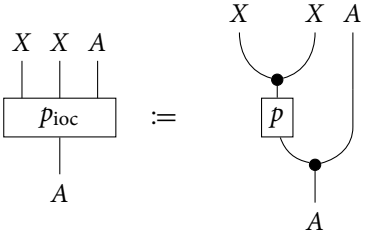
This already constitutes the relevant factorization as in Equation (18).

□

While for the existence of initial dilations for *deterministic* morphisms it suffices to assume positivity, we can give a generic argument for the existence of *all* initial dilations under the stronger assumption that the Markov category in question has conditionals. The following result and proof adapt the third author's argument for the existence of universal dilations in FinStoch (Houghton-Larsen 2021, Theorem 2.4.6).

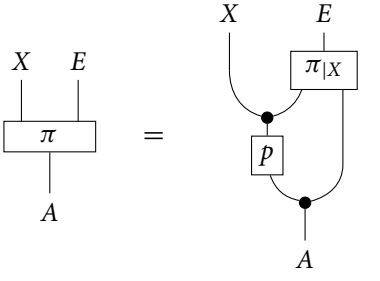
**Proposition 4.13.** *Let* C *be a Markov category with conditionals. Then every morphism in* C *has an initial dilation.*

*Proof.* Let $p\colon A \to X$ be any morphism. We claim that an initial dilation of $p$ is given by the dilation which simply copies both its input and output,




(109)

where the environment is given by $X \otimes A$. To see this, let $\pi : A \to X \otimes E$ be any other dilation of $p$ and let $\pi_{|X}$ be any conditional of $\pi$ with respect to $X$. Then $\pi$ factorizes through $p_{\text{ioc}}$:




(110)

This equation follows directly from the definition of conditionals and the fact that $\pi$ is a dilation of $p$.

On the uniqueness, suppose that we have some other $h\colon X \otimes A \to E$ satisfying (110) in place of $\pi_{|X}$. Then $h$ is itself a conditional of $\pi$ with respect to $X$ and, by the -a.s. uniqueness of conditionals, $h$ is $p_{\text{ioc}}$-a.s. equal to $\pi_{|X}$. Since every Markov category with conditionals is causal (Fritz 2020,

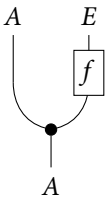
https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

Proposition 11.34), we can use Proposition 4.3 to deduce that $h$ must also be $p_{\mathrm{ioc}}$-dilationally equal to $\pi_{|X}$, thus showing the requisite uniqueness property. □

### *4.3 A dilational characterization of Markov categories*

Here, we give an abstract characterization of positive Markov categories as semicartesian categories subject to additional principles. More precisely, these principles serve to single out the copy morphisms uniquely and ensure their defining properties.

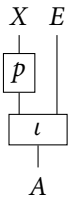
We first introduce the concept of *noncreative* morphisms. It mimics the idea behind deterministic morphisms in a positive Markov category, but its definition applies in the semicartesian case since it does not reference the copy morphisms at all. To motivate this, let us anticipate Lemma 4.15 below: this shows that a morphism $p$ in a positive Markov category is deterministic if and only if every dilation $\pi$ of $p$ factors as in Equation (108). While this factorization of $\pi$ makes explicit reference to $\mathrm{copy}_A$, we also know from Example 4.10 that the copy morphism is an initial dilation of $\mathrm{id}_A$. Thus, the right-hand side of Equation (108) can be expressed as a sequential composition of an arbitrary dilation




(111)

of $\mathrm{id}_A$ with $p \otimes \mathrm{id}_E$. Here is now the general definition.

**Definition 4.14.** *A morphism $p\colon A \to X$ in a semicartesian category is called noncreative if every dilation of $p$ is of the form*




(112)

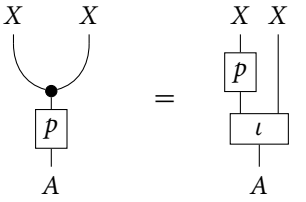
*for some dilation $\iota\colon A \to A \otimes E$ of $\mathrm{id}_A$.*

The term "noncreative" indicates that any information leaked from process $p$ to the environment can be viewed as having leaked already from the input of $p$. For a semicartesian category $\mathsf{D}$, we write $\mathsf{D}_{\mathrm{nc}}$ for the class of noncreative morphisms in $\mathsf{D}$.

**Lemma 4.15.** Determinism, noncreativity and positivity. *Let $\mathsf{C}$ be a Markov category in which the copy morphisms are initial dilations of the identities. Then every noncreative morphism is deterministic and we have $\mathsf{C}_{nc} = \mathsf{C}_{det}$ if and only if $\mathsf{C}$ is positive.*

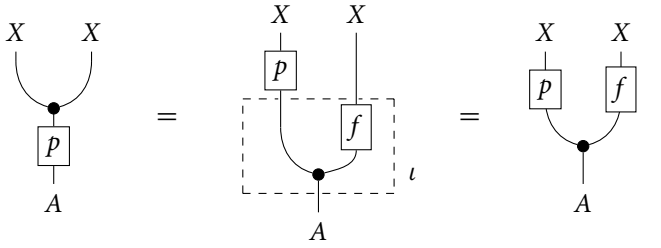
*Proof.* Let $p\colon A \to X$ be a noncreative morphism in $\mathsf{C}$. The morphism $\mathrm{copy}_X \circ p$ is a dilation of $p$, and therefore satisfies

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press




(113)

for some dilation $\iota\colon A \to A \times X$ of $\mathrm{id}_A$ by Definition 4.14. Since $\mathrm{copy}_A$ is an initial dilation of $\mathrm{id}_A$ by assumption, we must have $\iota = (\mathrm{id}_A \otimes f) \circ \mathrm{copy}_A$ for some $f\colon A \to X$. Consequently,

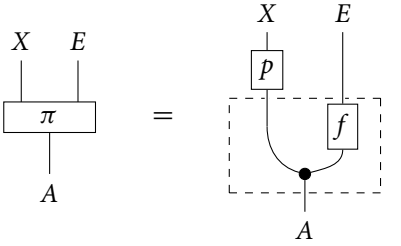



(114)

holds and in particular we must have $f = p$ by marginalizing the left output. The resulting equation precisely asserts that $p$ is deterministic. Hence, the claim $\mathsf{C}_{\mathrm{nc}} \subseteq \mathsf{C}_{\mathrm{det}}$ follows.

It remains to prove that the reverse inclusion $\mathsf{C}_{\mathrm{det}} \subseteq \mathsf{C}_{\mathrm{nc}}$ is equivalent to positivity of $\mathsf{C}$. This can be achieved either by showing that $\mathsf{C}_{\mathrm{det}} \subseteq \mathsf{C}_{\mathrm{nc}}$ is a just a different way to express deterministic marginal independence (Definition 2.4) or by relating it to property (ii) of Proposition 4.12. We take the latter approach. To this end, consider a deterministic morphism $p\colon A \to X$ in $\mathsf{C}$ and an arbitrary dilation $\pi : A \to X \otimes E$ thereof.

Suppose first that $\mathsf{C}$ is positive. By Proposition 4.12, the bloom of $p$ is an initial dilation of $p$, so that we have




(115)

for some $f\colon A \to E$. The fact that the morphism in the dashed rectangle is a dilation of $\mathrm{id}_A$ shows that $p$ is noncreative, so that $\mathsf{C}_{\mathrm{det}} \subseteq \mathsf{C}_{\mathrm{nc}}$ follows.

Conversely, suppose that $\mathsf{C}_{\mathrm{det}} \subseteq \mathsf{C}_{\mathrm{nc}}$ holds. Together with the assumption that $\mathrm{copy}_A$ is an initial dilation of $\mathrm{id}_A$, this implies the factorization of $\pi$ as in Equation (115) for some $f$, which is necessarily the $E$-marginal of $\pi$ and therefore unique. By Proposition 4.12 again, we obtain that $\mathsf{C}$ is a positive Markov category. □

Positivity implies that copy morphisms are initial dilations of the identities (Example 4.10). We therefore obtain another immediate characterization of positivity.

**Corollary 4.16.** *A Markov category* $\mathsf{C}$ *is positive if and only if the copy morphisms are initial dilations of the identities and* $\mathsf{C}_{det} = \mathsf{C}_{nc}$.

Our next goal is to characterize certain classes of Markov categories as semicartesian categories subject to additional axioms. Specifically, these axioms shall serve as a way to reconstruct the copy morphisms and to ensure their required properties. In this way, the copy morphisms emerge as a

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

consequence of dilational axioms rather than being imposed as additional mathematical data on top of a semicartesian category.

**Definition 4.17.** *(i) A morphism $b_X : X \to X \otimes X$ is broadcasting if both of its marginals are the identity:*


(116)

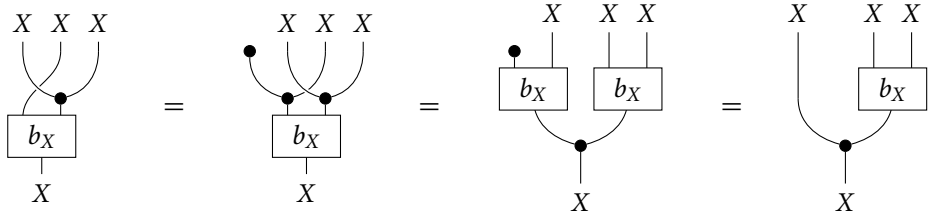
*(ii) An object $X$ admits broadcasting if there is a broadcasting morphism $b_X : X \to X \otimes X$.*

The terminology is inspired by the analogous concept of *broadcasting* in quantum information theory (Barnum et al. 1996). For example, the terminal object $I$ trivially admits broadcasting; and if $X$ and $Y$ admit broadcasting via $b_X$ and $b_Y$, then so does $X \otimes Y$ via the product $b_X \otimes b_Y$ with the middle outputs swapped. Furthermore, every copy morphism in a Markov category is broadcasting by the counitality axiom. In some Markov categories, however, there are other broadcasting morphisms as well. For instance, in $\mathsf{FinStoch}_{\pm}$ (Example 2.7) for $X = \{0, 1\}$, the kernel given by

$$b_X(x_1, x_2|x_0) := (-1)^{x_1+x_2} + \mathrm{copy}_X(x_1, x_2|x_0) \tag{117}$$

is broadcasting, but not equal to the copy map.

On the positive side, in positive Markov categories there are no broadcasting morphisms other than copy,[10] since then for any broadcasting morphism $b_X$, we have




(118)

where the second step is by positivity and the third by the broadcasting property. This shows that $b_X$ is the copy morphism upon marginalizing the third output.

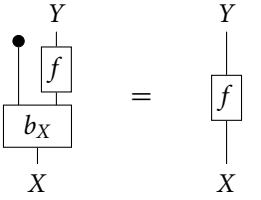
**Proposition 4.18.** *Let* $\mathsf{D}$ *be a semicartesian category. For any object $X$, the following are equivalent:*

*(i) $X$ admits broadcasting.*
*(ii) The discard morphism $\mathrm{del}_X$ is noncreative.*

We therefore obtain a *no-broadcasting theorem* (Barnum et al. 2007): $X$ does not admit broadcasting if and only if $\mathrm{del}_X$ is not noncreative.

*Proof.*

*(i)* ⇒ *(ii)*: Given a broadcasting morphism $b_X$, by equations (116) we have

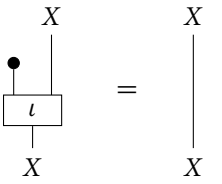



(119)

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

for every $f\colon X \to Y$. In particular, an arbitrary dilation of $\mathrm{del}_X$ is of the form of diagram (112) and thus $\mathrm{del}_X$ is noncreative.

(*ii*) ⇒ (*i*): Consider the dilation $\mathrm{id}_X$ of $\mathrm{del}_X$ with environment $X$. Since $\mathrm{del}_X$ is noncreative, this dilation must be of the form (112). That is, there is a dilation $\iota\colon X \to X \otimes X$ of $\mathrm{id}_X$ whose second output is the environment and which also satisfies




(120)

In particular, $\iota$ is broadcasting.

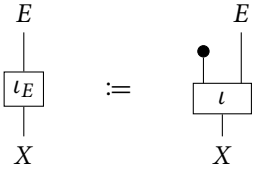
□

In the remainder of this section, we characterize positive Markov categories in purely semicartesian terms.

**Theorem 4.19.** *Let* D *be a semicartesian category. Then the following are equivalent:*

*(i)* D *can be made into a Markov category in which the copy morphisms are initial dilations of the identities.*

*(ii) For every object X, the identity $\mathrm{id}_X$ admits an initial dilation $\iota\colon X \to X \otimes E$ such that the marginal*




(121)

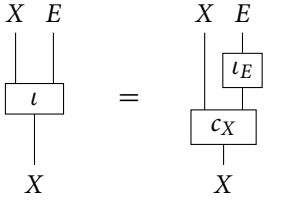
*is noncreative.*

*If these conditions hold, then the Markov category structure in (i) is unique.*

*Proof.*

(*i*) ⇒ (*ii*): If D is a Markov category in which the copy morphisms are initial dilations of the identity, then the requirements of (ii) hold since the identity is noncreative.

(*ii*) ⇒ (*i*): Using the notation of the statement, since $\iota$ is a dilation of the noncreative morphism $\iota_E$ with environment $X$, we have

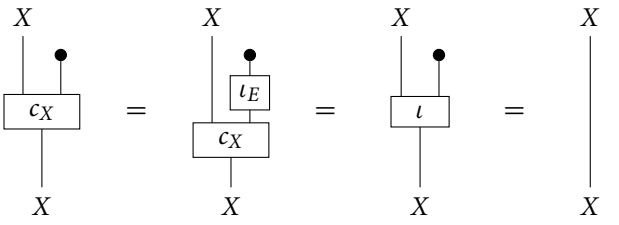



(122)

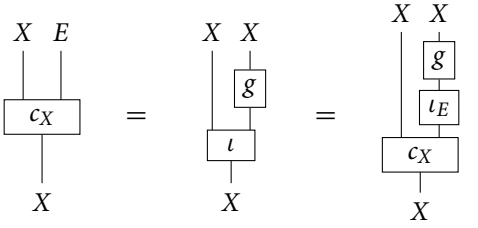
for some $c_X\colon X \to X \otimes X$ whose right-hand marginal is $\mathrm{id}_X$. We argue that this morphism is broadcasting. Indeed, we just noted that the right-hand marginal is the

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

identity, while the left-hand marginal is the identity because $\iota$ is a dilation of the identity:

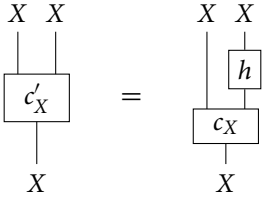



Moreover, since $\iota$ is an initial dilation of the identity, there is $g\colon E \to X$ in $\mathsf{D}$ that corresponds to a morphism of type $\iota \to c_X$ in $\mathsf{Dilations}(\mathrm{id}_X)$, so that



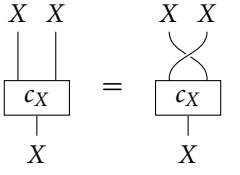

Marginalizing the first output of this equation gives $g \circ \iota_E = \mathrm{id}_X$, while the initiality of $\iota$ implies $\iota_E \circ g =_{\iota\text{-dil.}} \mathrm{id}_E$. As a consequence, $c_X$ is isomorphic to $\iota$ in $\mathsf{Dilations}(\mathrm{id}_X)$ and thus also an initial dilation of $\mathrm{id}_X$.

Next, we show that every identity morphism $\mathrm{id}_X$ has exactly one broadcasting morphism. To this end, let $c'_X\colon X \to X \otimes X$ be another morphism satisfying equations (116). Then we have




(123)

for some $h\colon X \to X$ by initiality of $c_X$. Marginalizing the left output gives $h = \mathrm{id}_X$, so that $c'_X$ is necessarily equal to $c_X$. This demonstrates the asserted uniqueness. In particular, this immediately implies that $c_X$ is symmetric, i.e., it satisfies



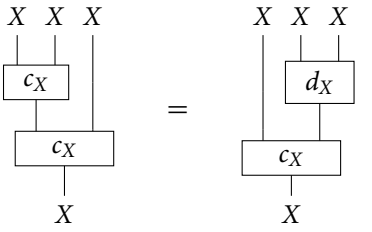

(124)

because swapping the outputs of $c_X$ obviously results in a broadcasting morphism again. Furthermore, $c_X$ is also the only dilation of $\mathrm{id}_X$ which is symmetric in this sense, as being a symmetric dilation implies being broadcasting.

Next, let us show that the unique symmetric dilation equips $X$ with the structure of a commutative comonoid. Commutativity is precisely Equation (124) and counitality

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

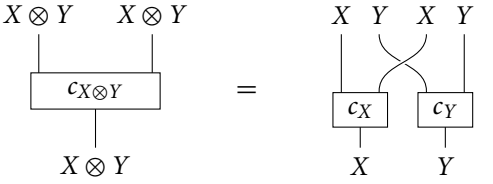
corresponds to Equations (116). For coassociativity, the initiality of $c_X$ implies that there is a $d_X$ of the same type such that




(125)

holds, since the left-hand side of the equation is a dilation of $\mathrm{id}_X$ with environment $X \otimes X$. But then marginalizing the first output shows $d_X = c_X$ since $c_X$ is broadcasting, so that we obtain the required coassociativity equation.

In order to see that $\mathsf{D}$ becomes a Markov category, it is thus enough to prove the multiplicativity equation




(126)

for all objects $X$ and $Y$. As we argued above, symmetric dilations of identities are unique. Therefore, Equation (126) follows upon showing that its right-hand side is a dilation of $\mathrm{id}_{X \otimes Y}$ that is invariant under swapping the two copies of $X \otimes Y$. Since this is a direct consequence of the symmetry of $c_X$ and $c_Y$, we obtain the desired result.

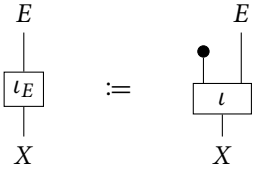
Furthermore, the uniqueness of the broadcasting morphisms implies that the constructed Markov category structure is the only possible one (cf. Fritz 2020, Remark 11.29).

□

With this characterization at hand, also positive Markov categories can now be characterized in semicartesian terms by adding additional requirements to item (ii).

**Corollary 4.20.** (Semicartesian characterization of positive Markov categories). *Let* $\mathsf{D}$ *be a semicartesian category. Then the following are equivalent:*

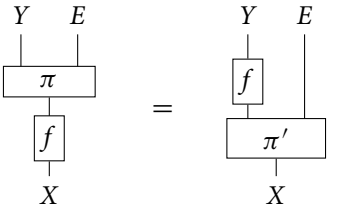
*(i)* $\mathsf{D}$ *can be equipped with copy morphisms making it into a positive Markov category.*
*(ii) For every object* $X$*, the identity* $\mathrm{id}_X$ *admits an initial dilation* $\iota \colon X \to X \otimes E$ *such that the marginal*




(127)

*is noncreative. Moreover, if* $f \colon X \to Y$ *is such that for any dilation* $\pi$ *of the identity morphism* $\mathrm{id}_Y$ *we have*

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

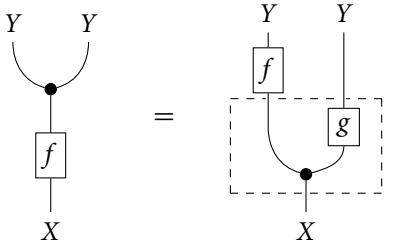



(128)

*for some dilation $\pi'$ of the identity morphism $\mathrm{id}_X$, then $f$ is noncreative.*

*Proof.* By Theorem 4.19, we can assume that $\mathsf{D}$ is a Markov category in which the copy morphisms are initial dilations of the identity. The problem is therefore reduced to showing that, under this assumption, positivity is equivalent to the second part of item (ii).

Consider a morphism $f$ satisfying the condition of Equation (128). For $\pi = \mathrm{copy}_Y$, we have




(129)

where the box on the right-hand side represents an arbitrary dilation of the identity on $X$ since $\mathrm{copy}_X$ is its initial dilation by assumption. Marginalizing the left output gives $g = f$, so that $f$ is deterministic by Equation (129). Conversely, every deterministic morphism satisfies Equation (128), which, once again, follows from copy morphisms being initial dilations of identities. Therefore, the second part of item (ii) can be restated as "every deterministic morphism is noncreative," and this holds if and only if $\mathsf{D}$ is positive by Lemma 4.15. □

**Remark 4.21.** Let us make a couple of comments on Theorem 4.19 and Corollary 4.20.

(i) The conditions stated in items (ii) of both results suggest that being a Markov category of the specified sort is a mere *property* of a symmetric monoidal category rather than extra structure.

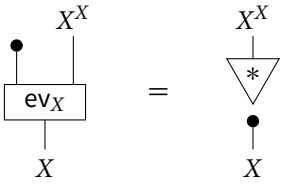
If we restrict to strict monoidal structure for simplicity, then this statement is easy to make precise in terms of the formalism of *stuff, structure, and property* (Baez and Shulman 2010): the obvious forgetful functor from the category of strict Markov categories and Markov functors (Fritz 2020, Definition 10.14) to the category of strict symmetric monoidal categories and strict monoidal functors is full and faithful, and therefore forgets at most property. While the faithfulness is trivial, the fullness amounts to the statement that every strict monoidal functor between positive Markov categories preserves the copy morphisms. This holds because such a functor clearly maps a copy morphism to a broadcasting morphism, which in the positive case must be a copy morphism again.

(ii) Theorem 4.19 also holds substituting all the occurrences of "initial dilation" with different properties – namely that of "complete dilation" (Houghton-Larsen 2021, Definition 2.3.1) and "universal dilation" (Houghton-Larsen 2021, Definition 2.4.1). A complete dilation is like an initial dilation, but for the fact that there could be multiple morphisms $f$ in

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

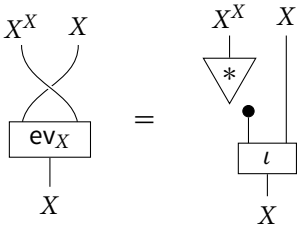
$\mathsf{Dilations}(p)$ that relate it to another dilation of $p$ via Equation (93), while a universal dilation is an initial dilation for which the morphism $f$ is unique as a morphism in $\mathsf{D}$.

(iii) Example 4.9 on pointed sets shows that there are semicartesian categories which have initial dilations, but not initial dilations satisfying item (ii) of Theorem 4.19. Indeed, we have




(130)

i.e. the marginal of the initial dilation given in (101) (the left-hand side) equals the morphism $X \leftarrow X^X$ that evaluates functions at the base point. Since the functions under consideration are all basepoint preserving, this is indeed the map that sends every function in $X^X$ to the basepoint $*$ of $X$, i.e., the right-hand side of Equation (130), where $*\colon I \leftarrow X^X$ is the unique morphism of this type. If the constant morphism from Equation (130) was noncreative, then there would have to be a factorization of the form




(131)

for some dilation $\iota$ of $\mathrm{id}_X$. For any $X$ with at least two elements, this equation cannot be satisfied, because after plugging in a function $f \in X^X$, its right-hand side is independent of $f$ while the left-hand side is not.

(iv) Another important example is the category of finite-dimensional Hilbert spaces and quantum channels. In this category, identities only have trivial dilations (see e.g. Houghton-Larsen 2021, Section 2.2.A), which is a strong form of the no-cloning theorem (cf. Selby and Coecke 2017, Proposition 7.1). In particular, there is no copy morphism and so also item (ii) of Theorem 4.19 cannot hold. The reason why it becomes clearer in light of Proposition 4.18.

**Acknowledgements.** We thank Arthur Parzgynat as well as two anonymous referees for a number of helpful comments on an earlier version. Tobias Fritz and Antonio Lorenzin acknowledge funding by the Austrian Science Fund (FWF) through project P 35992-N. Tomáš Gonda acknowledges the support of the START Prize Y 1261-N of the Austrian Science Fund (FWF). Paolo Perrone acknowledges funding from Sam Staton's grant BLaSt – a Better Language for Statistics of the European Research Council (ERC).

## Notes

**1** The dilations used in Houghton-Larsen (2021) and in Selby et al. (2021) are generally *two sided*, which means that the $g$ in Definition 1.11 may also carry an additional input not shared with $f$. Note, however, that the definitions of two-sided dilations

https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

in Houghton-Larsen (2021) and Selby et al. (2021) are not the same. One-sided dilations as per Definition 1.11 are sufficient for our purposes.

**2** This looks like an unfortunate choice of terminology in hindsight, and we are in favour of changing it in the future.

**3** Because of $\mathsf{Kl}(D_R)_{\mathrm{det}} = \mathsf{Set}$, the relevant right adjoint to the inclusion $\mathsf{Kl}(D_R)_{\mathrm{det}} \hookrightarrow \mathsf{Kl}(D_R)$ is simply the usual Kleisli adjoint of the inclusion $\mathsf{Set} \hookrightarrow \mathsf{Kl}(D_R)$.

**4** Unfortunately our notion of complement does not coincide with the usual notion of lattice complement in this case, since we only require $r \vee \bar{r} = 1$ but not $r \wedge \bar{r} = 0$.

**5** Recall that a quantale is a semiring where addition is given by the join of a complete join-semilattice, and such that multiplication distributes over arbitrary joins. The lattice of ideals of a commutative ring is a commutative quantale under ideal multiplication (Rosenthal 1990, Chapter 4).

**6** As the proof will show, this holds with any integral domain in place of $\mathbb{Z}[2i]$.

**7** We thank Arthur Parzygnat for pointing this out to us.

**8** Strictly speaking, this refers to *one-sided* complete and *one-sided* universal dilations, since in contrast to Houghton-Larsen (2021) we only consider one-sided dilations throughout (see Section 4.3).

**9** We use the subscript "ic" as shorthand for "input-copy".

**10** This is a special case of a general result on broadcasting morphisms in quantum Markov categories (Parzygnat 2020, Theorem 4.17) based on "quantum positivity" (see also Remark 2.2 (iii)).

## References


Aumann, R. J. (1961). Borel structures for function spaces. *Illinois Journal of Mathematics* **5** 614–630.

Baez, J. and Stay, M. (2011). Physics, topology, logic and computation: a Rosetta Stone. In: *New Structures for Physics*, Lecture Notes in Physics, vol. 813, Heidelberg, Springer, 95–172. arXiv:0903.0340.

Baez, J. C. and Shulman, M. (2010). Lectures on $n$-categories and cohomology. In: *Towards Higher Categories*, The IMA Volumes in Mathematics and its Applications, vol. 152, New York, Springer, 1–68. arXiv:math/0608420.

Barnum, H., Barrett, J., Leifer, M. and Wilce, A. (2007). Generalized no-broadcasting theorem. *Physical Review Letters* **99** 240501. arXiv:0707.0620.

Barnum, H., Caves, C. M., Fuchs, C. A., Jozsa, R. and Schumacher, B. (1996). Noncommuting mixed states cannot be broadcast. *Physical Review Letters* **76** 2818–2821. arXiv:9511010.

Chiribella, G. (2014). Dilation of states and processes in operational-probabilistic theories. In: *Proceedings of the 11th Workshop on Quantum Physics and Logic*, vol. 172, EPTCS, 1–14. arXiv:1412.8539.

Chiribella, G., Mauro D'Ariano, G. and Perinotti, P. (2010). Probabilistic theories with purification. *Physical Review A* **81** 062348. arXiv:0908.1583.

Cho, K. and Jacobs, B. (2019). Disintegration and Bayesian inversion via string diagrams. *Mathematical Structures in Computer Science* **29** 938–971.

Coumans, D. and Jacobs, B. (2013). Scalars, monads and categories. In: *Quantum Physics and Linguistics: A Compositional, Diagrammatic Discourse*, Oxford Academic.

Faden, A. M. (1985). The existence of regular conditional probabilities: necessary and sufficient conditions. *Annals of Probability* **13** (1) 288–298.

Fritz, T. (2020). A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics* **370** 107239.

Fritz, T., Gonda, T., Perrone, P. and Rischel, E. F. (2023). Representable Markov categories and comparison of statistical experiments in categorical probability. *Theoretical Computer Science*, 113896.

Fritz, T., Gonda, T. and Perrone, P. (2021). de Finetti's theorem in categorical probability. *Journal of Stochastic Analysis* **2** (4). arXiv:2105.02639.

Fritz, T. and Klingler, A. The $d$-separation criterion in categorical probability. *Journal of Machine Learning Research* **24** (46) 1–49. arXiv:2207.05740.

Fritz, T. and Liang, W. (2023). Free gs-monoidal categories and free Markov categories. *Applied Categorical Structures* **31** (2) Paper No. 21. arXiv:2204.02284.

Fritz, T. and Rischel, E. F. (2020). Infinite products and zero-one laws in categorical probability. *Compositionality* **2** 3.

Fullwood, J. and Parzygnat, A. J. (2021). The information loss of a stochastic map. *Entropy* **23** (8) Paper No. 1021, 27. arXiv:2107.01975.

Heunen, C., Kammar, O., Staton, S. and Yang, H. (2017). A convenient category for higher-order probability theory. In: *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, [Piscataway], NJ, IEEE, 12. arXiv:1701.02547.

Houghton-Larsen, N. G. (2021). *A Mathematical Framework for Causally Structured Dilations and its Relation to Quantum Self-Testing*. Phd thesis, University of Copenhagen. arXiv:2103.02302.

Jacobs, B. (2016). Affine monads and side-effect-freeness. In: *Proceedings of the International Workshop on Coalgebraic Methods in Computer Science*, Lecture Notes in Computer Science, vol. 9608, Springer, 53–72. cs.ru.nl/B.Jacobs/PAPERS/side-effects.pdf.


https://doi.org/10.1017/S0960129523000324 Published online by Cambridge University Press

Jacobs, B. (2021). Multinomial and hypergeometric distributions in Markov categories. In: *Proceedings of the Thirty-Seventh Conference on the Mathematical Foundations of Programming Semantics (MFPS*, Electronic Notes in Theoretical Computer Science, vol. 351, 98–115. arXiv:2112.14044.

Kechris, A. S. (1995). *Classical Descriptive Set Theory*, Graduate Texts in Mathematics, vol. 156, New York, Springer-Verlag.

Moss, S. and Perrone, P. (2022). Probability monads with submonads of deterministic states. In: *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*, 1–13.

Moss, S. and Perrone, P. (2023). A category-theoretic proof of the ergodic decomposition theorem. *Ergodic Theory and Dynamical Systems*, 1–27. arXiv:2207.07353.

Parzygnat, A. J. (2020). Inverses, disintegrations, and Bayesian inversion in quantum Markov categories. arXiv:2001.08375.

Piedeleu, R. and Zanasi, F. An introduction to string diagrams for computer scientists. arXiv:2305.08768.

Pitts, A. and Stark, I. (1993). Observable properties of higher order functions that dynamically create local names, or: What's new? In: *Proceedings of MFCS 1993*.

Rosenthal, K. I. (1990). *Quantales and Their Applications*, Pitman Research Notes in Mathematics Series, vol. 234, Harlow, Longman Scientific & Technical; copublished in the United States with John Wiley & Sons, Inc., New York.

Sabok, M., Staton, S., Stein, D. and Wolman, M. (2021). Probabilistic programming semantics for name generation. In: *Proceedings of the ACM on Programming Languages*, vol. 5, ACM, 1–29. arXiv:2007.08638.

Selby, J. H. and Coecke, B. (2017). Leaks: quantum, classical, intermediate, and more. arxiv.org/abs/1701.07404.

Selby, J. H., Scandolo, C. M. and Coecke, B. (2021). Reconstructing quantum theory from diagrammatic postulates. *Quantum* **5** 445. DOI: 10.22331/q-2021-04-28-445.

St. Clere Smithe, T. Compositional active inference I: Bayesian lenses. statistical games. arXiv:2109.04461.

Stark, I. (1996). Categorical models for local names. *LISP and Symbolic Computation* **9** (1) 77–107.

Stein, D. (2021). *Structural Foundations for Probabilistic Programming Languages*. Phd thesis, University of Oxford. dario-stein.de/thesis.pdf.

Stein, D. and Staton, S. (2021). Compositional semantics for probabilistic programs with exact conditioning. In: *Logic in Computer Science*, IEEE. arXiv:2101.11351.

**Cite this article:** Fritz T, Gonda T, Houghton-Larsen NG, Lorenzin A, Perrone P and Stein D (2023). Dilations and information flow axioms in categorical probability. *Mathematical Structures in Computer Science* **33**, 913–957. https://doi.org/10.1017/S0960129523000324