

Probability monads with submonads of deterministic states – Extended version

Sean Moss
University of Oxford
Department of Computer Science
Oxford, UK
sean.moss@cs.ox.ac.uk

Paolo Perrone
University of Oxford
Department of Computer Science
Oxford, UK
paolo.perrone@cs.ox.ac.uk

Abstract

Probability theory can be studied synthetically as the computational effect embodied by a commutative monad. In the recently proposed Markov categories, one works with an abstraction of the Kleisli category and then defines deterministic morphisms equationally in terms of copying and discarding. The resulting difference between ‘pure’ and ‘deterministic’ leads us to investigate the ‘sober’ objects for a probability monad, for which the two concepts coincide. We propose natural conditions on a probability monad which allow us to identify the sober objects and define an idempotent sobrification functor. Our framework applies to many examples of interest, including the Giry monad on measurable spaces, and allows us to sharpen a previously given version of de Finetti’s theorem for Markov categories.

This is an extended version of the paper accepted for the Logic In Computer Science (LICS) conference 2022. In this document we include more mathematical details, including all the proofs, of the statements and constructions given in the published version.

About citing this work. All the definitions, propositions, and theorems appearing in the published version also appear here, with the same numbering as in the published version. There is one result here, Lemma 3.18, not present in the published version. The numbering of particular equations is however inevitably different between the two versions. Because of this, if future readers need to refer to any of the equations contained here, we recommend them to refer to the corresponding definition or theorem instead.

1 Introduction

This paper is about different models of ‘abstract’ or ‘synthetic’ probability theories. Such theories talk about both *deterministic quantities* and *random quantities* (i.e. *random variables*). The difference is analogous to the distinction between *values* and *computations* in the semantics of programming languages. Indeed, particular non-standard models of synthetic probability have proved useful in applications to probabilistic programming:

- Quasi-Borel spaces ([19], [30]) model Kock’s synthetic measure theory [22]. Unlike traditional foundations

for probability in measurable spaces, they are well-suited to higher-order data.

- While naive handling of conditional probabilities can lead to paradoxes [20], it was shown in [33] that in more restrictive models of probability ‘exact conditioning’ can be given a consistent meaning. Fritz’s Markov categories [6] were used to formulate the result.

A useful way to present a model of probability is with a *commutative monad* [22]. This gives the link to the semantics of more general programming effects, since commutative monads are a special case of the *strong monads* used by Moggi for the semantics of call-by-value languages [25]. A potentially strange aspect of monadic semantics is that it effectively gives both values and computations separately, with no need for one to be a subset of the other. In terms of probability, we are led to ask whether the deterministic quantities can be characterized as certain well-behaved random quantities. Two useful criteria, relevant to both probability and more general computation are the following, described informally.

1. *Discardable*: A computation that has negligible effect if its result is not used, so may be safely discarded.
2. *Copyable*: A computation which can be run once and have its output used twice instead of being run twice.

Unlike monadic semantics, in a Markov category only the random quantities are explicitly given. All the quantities in a Markov category are discardable, and then deterministic quantities are *defined* to be those that are also copyable.

Having deterministic quantities characterized equationally has enabled many interesting results from traditional probability theory to be expressed and proved synthetically. However, it is often the case that we wish to preserve some connection with the ‘pure’ quantities, i.e. those morphisms in the original category hosting a monad for probability. This is because that category is where we have explicit descriptions of the objects. The goal of this paper is to elucidate such a connection for several examples of interest. More specifically, we address whether probability monads of interest admit ‘sobrification’ submonads, i.e. a universal way of replacing each object with one for which pure quantity coincides with deterministic quantity.

In the remainder of this introduction, we provide some informal explanation of the background material on probability monads and a high-level picture of the development in this paper. Our main contribution rests in packaging up properties of certain probability monads into the *observationality* and *S-observationality* conditions, showing that these are indeed satisfied by examples and interest, and showing that these conditions abstractly imply the desired results on sobriification.

1.1 Probability monads

The canonical mathematical model of probability theory is in the category Meas of measurable spaces. A measurable space is a pair (X, Σ_X) where X is a set and $\Sigma_X \subseteq \mathcal{P}(X)$ is a collection of subsets of X containing \emptyset and X and closed under countable unions and complements. Morphisms $(X, \Sigma_X) \rightarrow (Y, \Sigma_Y)$, or ‘measurable functions’, are functions $f : X \rightarrow Y$ such that $f^{-1}(E) \in \Sigma_X$ for every $E \in \Sigma_Y$. It is necessary to consider measurable spaces rather than just sets because there are foundational problems with constructing probability distributions that assign probability to *every* subset of a given set X . Instead, we equip X with a family Σ_X of ‘measurable sets’, and require a *probability measure* to be a function $m : \Sigma_X \rightarrow [0, 1]$ satisfying $m(X) = 1$, $m(\emptyset) = 0$ and countable additivity.

Interestingly, the set $PX = P(X, \Sigma_X)$ of probability measures on a measurable space (X, Σ_X) can itself be considered as a measurable space. In fact, P is the functor part of the *Giry monad* on Meas (see Example 2.2). This allows us to consider ‘probability measures on the set of probability measures’. The Giry monad is a strong monad, and indeed a commutative monad, so following Moggi [25] we can use it to interpret a first-order call-by-value language.

The basic setting in this paper is that of a category \mathbb{C} with finite products and a monad T on it. The idea is that a monad includes an assignment $A \mapsto TA$, where TA is the object of ‘distributions’ or ‘measures’ on A . Hence, for any pair of objects $A, X \in \text{ob } \mathbb{C}$, we can consider morphisms $A \rightarrow X$ to be ‘pure’ functions and morphisms $A \rightarrow TX$ to be ‘stochastic functions’. In the language of computation, the latter would be a computation that produces an X .

1.2 Thinkable morphisms

Besides the aforementioned properties of copyability and discardability, there is an additional property that makes a monadic computation look ‘ordinary’: *thinkability*. A *think* is a computation that has been ‘frozen’. In probability, the ‘think’ of a state x is a ‘Dirac delta’ probability distribution δ_x which assigns probability 1 to x and 0 to everything else. In other words, sampling from δ_x almost surely returns x . Informally, a program $M : A$ is *thinkable* if it satisfies the equation

$$\lambda_.M = \text{let } x \leftarrow M \text{ in } \lambda_.x : 1 \rightarrow A.$$

In terms of probability, we think of M as a distribution, the difference between the two sides is that the left-hand side returns a think that samples anew from M every time it is run, but the right hand side creates a think by sampling from M once and for all then wrapping the result.

From the point of view of probability theory, thinkable morphisms are those that commute with forming Dirac deltas. This can be seen as related to determinism, since any stochastic map which ‘spreads’ the mass of a measure, from a single point to several ones, cannot possibly commute with forming deltas. (See also [8, Remark 3.11] for additional context.)

It therefore seems that thinkability is yet another property that sets pure computations apart from the other ones, and so it is interesting to study the relationship between purity, thinkability, copyability, and discardability [14, 15]. We recap these properties and their relationships in Section 3. We show that in nice situations thinkable morphisms are encoded by a submonad of the original monad, often idempotent, which one can think of as measuring the extent to which a generic thinkable morphism fails to be pure.

1.3 Observations and de Finetti’s theorem

In the practice of probabilistic programming, as well as in statistics, one is sometimes given two random variables and has to test whether they follow the same distribution or not. The problem is not an easy one. In the best case one can draw independent samples from them, and compare the (random) sequences obtained by the repeated draws. This idea is also reflected by the famous *de Finetti theorem* [5], which roughly says that random probability measures correspond bijectively to exchangeable random sequences. In this work we make mathematically precise, in terms of monads, the intuition that *random variables are tested by repeated draws*, or more generally, that in some contexts, effectful computations can be compared by taking repeated independent runs. In terms of a program $M : 1 \rightarrow (1 \rightarrow A)$, we test with the program contexts

$$C_n[M] := \text{let } z \leftarrow M() \text{ in let } x_1 \leftarrow z() \text{ in } \dots \\ \text{let } x_n \leftarrow z() \text{ in return } (x_1, \dots, x_n)$$

where $C_n[M] : A^n$, for $n \in \mathbb{N}$. To see why repeated sampling is necessary, consider the following programs

$$M_1 := \lambda_.\lambda_.\text{or}(\text{true}, \text{false})$$

$$M_2 := \lambda_.\text{or}(\lambda_.\text{true}, \lambda_.\text{false})$$

of type $1 \rightarrow (1 \rightarrow \text{bool})$ where *or* is a non-deterministic choice. Then $C_1[M_1]$ and $C_1[M_2]$ can each evaluate to both true and false, but $C_2[M_1]$ can evaluate to only two of the four possibilities and $C_2[M_2]$ can evaluate to any of them.

We call monads that exhibit this property *observational*. While this property is often considered typical of probabilistic contexts, somewhat surprisingly it can also hold for monads which are not strictly about probability. For example,

the lower Vietoris monad (or Hoare powerdomain) is observational (Theorem 9.4), and it generally does not encode randomness, but rather, nondeterminism. For observational monads, as we prove in Theorem 7.1, every thinkable morphism is deterministic.

1.4 Outline

In Section 2 we recall the basic concepts of monad theory, with a view on the probability case. We in particular look at the *equalizing requirement*, which is one of the less known concepts which is of great relevance for this work, and we sketch how a monad interacts with the products of a monoidal category, in order to form ‘joint states’.

In Section 3 we look in detail at the structures that select the different versions of effectful (in particular, random) and noneffectful computation. We recall (Theorem 3.14) that every pure morphism is thinkable and every thinkable morphism is copyable and discardable, but not the other way around (as counterexamples show).

In Section 4 we construct for every monad T a submonad D whose Kleisli morphisms are the thinkable morphisms of T . In Section 5 we define sober objects as those for which every thinkable morphism is pure, and look at the relationship with the submonad D . We show that in several cases, such as for the Giry monad, the submonad D is idempotent, and that sober objects generalize sober topological spaces.

In Section 6 we define observational monads, those for which computations can be tested by repeated independent runs. We look at the connection with ground types, the ones that are directly observable (e.g. by the user or experimenter), and we give technical conditions to show observability of monads. In Section 7 we prove what could be considered the main result of this work (Theorem 7.1), namely that for observational monads, deterministic morphisms are the same as thinkable morphisms. Therefore, in that case the submonad D equivalently encodes deterministic morphisms.

In Section 8 we connect the notion of observability with de Finetti’s theorem. In particular, we show that for observational monad one can sharpen the known version of the synthetic de Finetti theorem for Markov categories [7].

Finally, in Section 9 we prove observability for three apparently very different monads: the Giry monad, the lower Vietoris monad of nondeterminism, and the monad of name generation. We give concluding remarks in Section 10.

2 Background on monads

Recall [24] that a *monad* on a category \mathbb{C} is a triple $\mathbb{T} = (T, \eta, \mu)$ (subsequently also just denoted by T) where

1. T is a functor $\mathbb{C} \rightarrow \mathbb{C}$,
2. η is a natural transformation $\text{id} \Rightarrow T$,
3. μ is a natural transformation $TT \Rightarrow T$,

satisfying $\mu_A \circ \eta_{TA} = 1_{TA} = \mu_A \circ T\eta_A$ and $\mu_A \circ \mu_{TA} = \mu_A \circ T\mu_A$ for all $A \in \mathbb{C}$.

Monads are used in denotational semantics to model a distinction between *pure values* and *effectful computations* [25]. If morphisms $A \rightarrow X$ are ordinary ‘ X -valued functions’, then morphisms $A \rightarrow TX$ are ‘ X -producing computations’. For the purposes of this work, we are mainly interested in those monads involving probability, i.e. for which a morphism $A \rightarrow TX$ can be seen as a *stochastic map* from A to X , involving random chance.

Example 2.1. Consider the category Set whose objects are sets and morphisms are functions. The *distribution monad* on Set is the monad T with $TX = \{\pi \in [0, 1]^X : \sum_{x \in X} \pi(x) = 1\}$, $\eta_X(x) = \lambda x'. \llbracket x = x_0 \rrbracket$ (i.e. it forms a “delta at x ”) and

$$\mu_X(\rho)(x) = \sum_{\pi \in TX} \rho(\pi) \times \pi(x).$$

Example 2.2. Consider the category Meas whose objects are measurable sets and morphisms are measurable functions. The *Giry monad* P [18] consists of

- The functor P assigning to each measurable space X the set PX of probability measures over X , equipped with the coarsest σ -algebra which makes the evaluation of measures measurable;
- The natural transformation of components $\eta : X \rightarrow PX$ assigning to each point $x \in X$ the “Dirac delta” measure δ_x , such that for every measurable $A \subseteq X$, $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise.
- The natural transformation of components $\mu : PPX \rightarrow PX$ which, analogously to the case of the distribution monad, assigns to each measure $\rho \in PPX$ the “mixture” measure $\mu(\rho)$, such that for every measurable $A \subseteq X$,

$$\mu(\rho)(A) = \int_{PPX} p(A) \mu(d\rho).$$

On Meas we can analogously define the monad M of *subprobability* measures, where now instead of having $p(X) = 1$ we only require $0 \leq p(X) \leq 1$.

For more information, we refer to [18]. For an introduction to the concepts, see [27, Chapter 1], [10, Section 6] and [26].

In this view of a monad, $\eta_A : A \rightarrow TA$ is the computation which, given $a \in A$, just returns a . The rest of the monad data tells us how to sequence computations together. The *Kleisli category* $\text{Kl}(T)$ is the category with the same objects as \mathbb{C} but homsets the $\text{Kl}(T)(A, B)$ is in bijection with $\mathbb{C}(A, TB)$. We explicitly denote this bijection with

$$(-)^\# : \text{Kl}(T)(A, B) \rightarrow \mathbb{C}(A, TB)$$

$$(-)^b : \mathbb{C}(A, TB) \rightarrow \text{Kl}(T)(A, B).$$

The composition in $\text{Kl}(T)(A, B)$, denoted by \circ , is given by

$$(g \circ f)^\# = \mu \circ T(g^\#) \circ f^\#$$

with the identity maps $(1_A)^\# = \eta_A$. Since the morphisms of $\text{Kl}(T)$ can easily be confused with those of \mathbb{C} , in diagrams we use an ordinary arrow $f : A \rightarrow TB$ when considering morphisms in \mathbb{C} and a wavy arrow $f^\flat : A \rightsquigarrow B$ for morphisms in $\text{Kl}(T)$. The bijection above gives an adjunction between \mathbb{C} and $\text{Kl}(T)$ called the *Kleisli adjunction*.

Example 2.3. For the distribution monad on Set , a Kleisli morphism is a function $X \rightarrow TY$, i.e. an X -indexed family of probability distributions on Y , or a function on X whose output is random, also called a *finite probability kernel*. We can denote its entries by $k(y|x)$, interpreted as the conditional probability of obtaining output y from the input x .

The Kleisli composition of $k : X \rightarrow TY$ and $h : Y \rightarrow TZ$ returns the kernel $h \circ k : X \rightarrow TZ$ given by

$$h \circ k(z|x) = \sum_{y \in Y} h(z|y) h(y|x).$$

also known as the *Chapman-Kolmogorov* composition of probability kernels.

Example 2.4. We denote the Kleisli categories of P and M by Stoch and SubStoch , respectively. Given measurable spaces X and Y , a morphism $k : X \rightsquigarrow Y$ of Stoch is a *Markov kernel* (or *stochastic map*) k from X to Y , i.e. either a measurable function $X \rightarrow PY$, or equivalently a map

$$\begin{aligned} X \times \Sigma_Y &\xrightarrow{k} [0, 1] \\ (x, B) &\longmapsto k(B|x) \end{aligned}$$

such that

- for each measurable subset $B \subseteq Y$, the assignment $x \mapsto k(B|x)$ is measurable;
- for each $x \in X$, the assignment $B \mapsto k(B|x)$ is a sub-probability measure (i.e. $k(Y|x) = 1$).

The Kleisli composition is the continuous analogue of the Chapman-Kolmogorov formula, it is given by the integral

$$h \circ k(C|x) = \int_Y h(C|y) k(dy|x).$$

The morphisms of SubStoch , called *substochastic maps*, are defined similarly, except that $0 \leq k(Y|x) \leq 1$.

In some sense, the unit of the monad allows to transport morphisms of the base category into the Kleisli category.:

Definition 2.5. A morphism $f : X \rightsquigarrow Y$ of $\text{Kl}(T)$ is called *pure* if its counterpart $f^\# : X \rightarrow TY$ is in the form

$$X \xrightarrow{g} Y \xrightarrow{\eta} TY$$

for some $g : X \rightarrow Y$ of \mathbb{C} . We call f *uniquely pure* if $f^\#$ can be written as $\eta \circ g$ for a unique g .

Pure morphisms are, in some sense, those that “come from the base category”, or “do not really use the monad”.

Example 2.6. For the Giry monad of probability measures, a pure morphism $f : X \rightsquigarrow Y$ is a kernel assigning to each point of x a Dirac measure on Y , in a measurable way. So, in some sense, it is simply a measurable function $X \rightarrow Y$. Note that, if the σ -algebra of Y does not separate points, different measurable functions $X \rightarrow Y$ might define the same kernel, and so in general a pure morphism $f : X \rightsquigarrow Y$ is not uniquely pure. An example of that will given in Example 2.8.

2.1 Unit fork and equalizing requirement

Recall that a *fork* is a diagram

$$A \xrightarrow{f} B \begin{array}{c} \xrightarrow{g} \\ \xrightarrow{h} \end{array} C$$

such that $g \circ f = h \circ f$ (but not necessarily $g = h$). One can view an equalizer as a universal fork.

Given a monad T , for each object X , the unit of the monad forms the following fork,

$$X \xrightarrow{\eta} TX \begin{array}{c} \xrightarrow{\eta} \\ \xrightarrow{T\eta} \end{array} TTX \quad (1)$$

which is indeed a fork by naturality of η . We call the diagram (1) the *unit fork* at X .

The monads for which the unit fork is an equalizer for every X are said to satisfy the *equalizing requirement* [25, Section 4].

Example 2.7. The distribution monad on Set satisfies the equalizing requirement. Indeed, given $p \in TX$, the two distributions $\eta(p)$ and $T\eta(p) \in TTX$ are respectively, a delta peaked at p , and a convex combination of deltas at points x with coefficients $p(x)$. These are equal if and only if p is itself a delta at some point x .

Example 2.8. (This example comes from [6, Example 10.5].) The Giry monad on Meas does not satisfy the equalizing requirement. Let X be the 2-point space $\{x, x'\}$, equipped with the codiscrete sigma-algebra (i.e. the only measurable sets are the empty set and X itself). Then the measures δ_x and $\delta_{x'}$ are equal, even if the points x and x' are not. As the unit η (i.e. δ) is not injective, it cannot be an equalizer.

Remark 2.9. The unit fork for objects in the form TX is always an equalizer, in fact a split one.

$$TX \xrightarrow{\eta} TTX \begin{array}{c} \xrightarrow{\eta} \\ \xrightarrow{T\eta} \end{array} TTTX$$

$\xleftarrow{\mu} \quad \xleftarrow{T\mu}$

2.2 Monads on monoidal categories

A *symmetric monoidal category* (SMC) is a category equipped with a “tensor product”, i.e. a binary functor $\otimes : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$, and a “unit” object I , together with isomorphisms

$$X \otimes I \cong X \cong I \otimes X, \quad (X \otimes Y) \otimes Z \cong X \otimes (Y \otimes Z),$$

$$\tau_{X,Y} : X \otimes Y \cong Y \otimes X$$

satisfying appropriate coherence conditions and $\tau_{X,Y} \circ \tau_{Y,X} = 1$. By a coherence theorem we can calculate as though the first three are actually *identities* – see [24] for precise definitions and details. The interpretation is that given objects A, B, C, D and morphisms $f : A \rightarrow B$ and $g : X \rightarrow Y$, we can form new objects $A \otimes X$ and $B \otimes Y$ of “joint states”, and the morphism $f \otimes g : A \otimes X \rightarrow B \otimes Y$, which as a process it consists of “executing f and g independently, in parallel”.

The interaction between monads and the monoidal structure of a category are of interest for probability theory: in general the probability of a product is not the product of the probabilities, and this difference encodes correlation and other statistical interaction. See [11] for more on this.

So let (\mathbb{C}, \otimes, I) be an SMC. A *monoidal monad* is a monad T together with a natural transformation of components $\nabla : TA \otimes TB \rightarrow T(A \otimes B)$ and a morphism $I \rightarrow TI$ satisfying associativity, unitality, and compatibility with the monad structure. The monad is *symmetric monoidal* if the map ∇ is compatible with permutation of the factors. (See for example [11, Appendix A] for the detailed definition.)

A more general interaction between a monad and the monoidal structure of a category is the notion of *strength*, widely used in theoretical computer science at least since Moggi [25]. A strength consists of a natural transformation of components $\sigma : X \otimes TY \rightarrow T(X \otimes Y)$, satisfying suitable consistency conditions (see the source above for the details). It intuitively turns a computation in Y paired with a value of X into a computation in X and Y , which is in some sense “trivial in X ”. A strength can always be obtained from a monoidal structure using the unit as $\nabla \circ (\eta \otimes \text{id}) : X \otimes TY \rightarrow T(X \otimes Y)$. Conversely, given a strength, one can obtain a monoidal structure provided that the strength satisfies a particular *commutativity* condition (one speaks of a *commutative monad*). Indeed, a symmetric monoidal structure for a monad is equivalent to a commutative strength, see for example [12, Appendix C] for details. Moreover, in that case $\text{Kl}(T)$ is a monoidal category as well, with the tensor product induced by the one on the base category.

3 Categories of structured objects

By this we mean categories where the objects have structure and the morphisms do not necessarily preserve all of this structure. This paper will essentially be the study of the Kleisli adjunction induced by a commutative monad on a cartesian monoidal category, but we recall here the relation to various abstractions.

3.1 Copy-discard structure

Definition 3.1. Let $(\mathcal{K}, \otimes, I)$ be a symmetric monoidal category. A *comonoid* in \mathcal{K} is a triple (X, c, d) where $X \in \mathcal{K}$, $c : X \rightarrow X \otimes X$, $d : X \rightarrow I$ and these satisfy the following equations.

$$(d \otimes 1_X) \circ c = 1_X = (1_X \otimes d) \circ c \quad (c \otimes 1_X) \circ c = (1_X \otimes c) \circ c$$

A comonoid is *cocommutative* (or *commutative*, for short) if in addition $c = \tau_{X,X} \circ c$. The category of commutative comonoids in \mathcal{K} , written $\text{CComon}(\mathcal{K}, \otimes, I) = \text{CComon}(\mathcal{K})$, has as morphisms $(X, c, d) \rightarrow (X', c', d')$ those maps $f : X \rightarrow X'$ in \mathcal{K} such that $d' \circ f = d$ and $c' \circ f = (f \otimes f) \circ c$.

Definition 3.2. A *copy-delete* or *copy-discard* (CD) category, also called a *garbage-share* (gs) *monoidal category*, is a symmetric monoidal category $(\mathcal{K}, \otimes, I)$ together with a specified section of the function

$$\text{ob CComon}(\mathcal{K}, \otimes, I) \rightarrow \text{ob } \mathcal{K}$$

mapping a commutative comonoid in \mathcal{K} to its underlying object. In other words, each object $X \in \mathcal{K}$ is equipped with maps $\text{copy}_X : X \rightarrow X \otimes X$ and $\text{del}_X : X \rightarrow I$ making it into a commutative comonoid.

CD categories were first defined (in strict form, and under the name “gs-monoidal categories”) in [17], and rediscovered independently several times. See [9, Remark 2.2] (and references therein) for a more detailed history of the subject. While every object of a CD category \mathcal{K} is a comonoid, this does not make \mathcal{K} a subcategory of $\text{CComon}(\mathcal{K})$ since the morphisms of \mathcal{K} do not have to respect the comonoid structures. This allows us to consider subclasses of morphisms in \mathcal{K} which do respect them to various degrees.

Example 3.3. Any cartesian monoidal category is a CD category in an essentially unique way. The copy map is given by the diagonal $X \rightarrow X \times X$, and the discard map is given by the unique map $X \rightarrow 1$.

Example 3.4. More generally, the Kleisli category of a commutative monad T on a cartesian monoidal category \mathbb{C} has a canonical CD structure. The copy and delete structures are inherited by those of \mathbb{C} .

Example 3.5. The Kleisli categories of P and M , which we called *Stoch* and *SubStoch*, have the following copy and discard maps.

$$\begin{array}{ll} X \longrightarrow M(X \times X) & X \longrightarrow M(1) \cong [0, 1] \\ x \longmapsto \delta_{(x,x)} & x \longmapsto 1 \end{array}$$

where instead $P(1) \cong 1$ for the Giry monad. The CD structure of *Stoch* has been studied in detail in [4] and [6].

Example 3.6. A seemingly different example of CD category is a full subcategory of commutative comonoid objects in a symmetric monoidal category with all maps between them, not just the comonoid homomorphisms. Dually, we can also see this as the opposite category to a category of commutative monoid objects, for example rings or algebras.

Definition 3.7. A morphism $f : X \rightarrow Y$ in a CD category is called

- *copyable* if it commutes with the copy map;

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \downarrow \text{copy} & & \downarrow \text{copy} \\ X \otimes X & \xrightarrow{f \otimes f} & Y \otimes Y \end{array}$$

- *discardable* or *normalized* if it commutes with the discard map;

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \searrow \text{del} & & \swarrow \text{del} \\ & I & \end{array}$$

- *deterministic* if it is copyable and discardable.

A *Markov category* [6] is a CD category in which every morphism is normalized.

Example 3.8. In SubStoch , a morphism $k : X \rightsquigarrow Y$ is

- copyable if and only if for every $x \in X$, and for every measurable $B \subseteq Y$,

$$k(B|x) \in \{0, k(Y|x)\};$$

- normalized if and only if for every $x \in X$,

$$k(Y|x) = 1;$$

- deterministic if and only if for every $x \in X$, and for every measurable $B \subseteq Y$,

$$k(B|x) \in \{0, 1\}, \quad k(Y|x) = 1.$$

In other words, deterministic morphisms are the ones that are certain about whether any event (measurable subset $B \subseteq Y$) is going to happen (probability 1), or not (probability 0).

Every morphism of Stoch is normalized, therefore it is a Markov category. A canonical example of a deterministic morphism $1 \rightsquigarrow Y$ of SubStoch (and Stoch) is a Dirac delta $1 \mapsto \delta_y$ at a point $y \in Y$. Not every deterministic morphism is in this form in general:

Example 3.9. Let Y be the unit interval $[0, 1]$, equipped with the countable-cocountable sigma-algebra (i.e. the measurable sets are precisely the countable subsets and their complements). Then the assignment

$$A \mapsto \begin{cases} 1 & \text{if } A \text{ is uncountable} \\ 0 & \text{if } A \text{ is countable} \end{cases}$$

is a deterministic morphism $1 \rightsquigarrow Y$ of SubStoch , and it cannot be written as δ_y for any $y \in Y$.

Note also that measures of the form δ_x do not always correspond to points bijectively, as Example 2.8 shows. One of the main purposes of this paper is, indeed, to study those deterministic kernels which are not (parametrized) Dirac deltas. More generally, to study those deterministic morphisms in a Kleisli category which do not come from morphisms of the base category (i.e. are not pure, according to Definition 2.5).

We conclude this section with a general remark. The following conditions are equivalent for a CD category:

- Every morphism is deterministic;
- The copy and discard maps are natural;
- The category is cartesian monoidal.

In some sense, one can view cartesian monoidal categories as a special case of CD categories where no randomness or nondeterminism is involved.

3.2 Think-force structure

As is a well-known fact in category theory, an adjunction between two categories gives rise to a monad on one and a comonad on the other. Thus when T is a monad on \mathbb{C} , there is a comonad on T^\dagger on $\text{Kl}(T)$. On objects, $T^\dagger X = TX$ and on morphisms $f : A \rightsquigarrow B$ we have

$$(T^\dagger f)^\# = TA \xrightarrow{(f^\#)^\dagger} TB \xrightarrow{\eta} TTB,$$

where $(f^\#)^\dagger$ denotes the Kleisli extension of $f^\#$. It is useful to describe the unit and counit in terms of some slightly richer structure.

Definition 3.10 ([14]). A *think-force category* or *abstract Kleisli category* is a category \mathbb{K} equipped with an endofunctor $L : \mathbb{K} \rightarrow \mathbb{K}$ and two families of maps

$$\text{thunk}_A : A \rightarrow LA \quad \text{force}_A : LA \rightarrow A$$

for $A \in \mathbb{K}$ such that

1. force is a natural transformation $L \Rightarrow \text{id}$,
2. thunk_L is a natural transformation $L \Rightarrow LL$,
3. $L(\text{thunk}_A) \circ \text{thunk}_A = \text{thunk}_{LA} \circ \text{thunk}_A$,
4. $\text{force}_A \circ \text{thunk}_A = 1_A$,
5. $L(\text{force}_A) \circ \text{thunk}_{LA} = 1_{LA}$.

Note thunk is *not* required to be natural in general. It does follow that the endofunctor L underlies a comonad with counit $\text{force} : L \Rightarrow \text{id}$ and comultiplication $\text{thunk}_L : L \Rightarrow LL$.

Example 3.11. $\text{Kl}(T)$ is canonically a think-force category. The endofunctor L is given by the composite $T^\dagger = F \circ G : \text{Kl}(T) \rightarrow \mathbb{C} \rightarrow \text{Kl}(T)$, where (F, G) are the functors of the Kleisli adjunction, and

$$(\text{thunk}_A)^\# = A \xrightarrow{\eta} TA \xrightarrow{\eta} TTA \quad (\text{force}_A)^\# = TA \xrightarrow{1} TA.$$

This suffices to describe the comonad structure of T^\dagger .

In the context of categorical probability, i.e. in [8] and [7], the morphism force is denoted by samp . This can be interpreted as a map taking a probability measure p and returning a random element distributed according to p .

3.3 Thinkable morphisms

In [14] it is shown that every think-force category has the form of Example 3.11. The crucial concept is that of *thinkable* morphism, of which we recall here the basic definitions and constructions.

Definition 3.12. A morphism $f : A \rightarrow B$ in a think-force category is *thinkable* if the following diagram commutes.

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ \text{think}_A \downarrow & & \downarrow \text{think}_B \\ LA & \xrightarrow{Lf} & LB \end{array} \quad (2)$$

The thinkable morphisms form a wide subcategory, $\mathbb{K}_{\text{think}}$. Note that each map think_A is itself thinkable, by one of the axioms in Definition 3.10, as is any map of the form Lf , by naturality of think_L . The identity-on-objects inclusion $\mathbb{K}_{\text{think}} \hookrightarrow \mathbb{K}$ has a right adjoint given by the factoring of $L : \mathbb{K} \rightarrow \mathbb{K}$ through $\mathbb{K}_{\text{think}} \hookrightarrow \mathbb{K}$. At the level of homsets the bijection is

$$\begin{aligned} (A \xrightarrow{f} B) &\mapsto (A \xrightarrow{\text{think}} LA \xrightarrow{Lf} LB) \\ (A \xrightarrow{g} LB) &\mapsto (A \xrightarrow{g} LB \xrightarrow{\text{force}} B). \end{aligned}$$

This adjunction gives rise to a monad on $\mathbb{K}_{\text{think}}$ whose Kleisli category constructed as in Example 3.11 is the original think-force category \mathbb{K} . In [14] it is further observed that in an appropriate sense this is the universal solution to the inverse problem of presenting a think-force category via a monad.

In [8, Remark 3.11] it was remarked that the unit of the monad is natural against morphisms of the base category, but not against generic Kleisli morphisms. From this perspective, we see thinkable morphisms as precisely those against which the unit of the monad is natural.

Thinkable morphisms of a Kleisli category can be characterized in terms of the base category as follows.

Proposition 3.13. *A morphism $f : A \rightsquigarrow B$ of $\text{Kl}(T)$ is thinkable if and only if its counterpart $f^\sharp : A \rightarrow TB$ in \mathbb{C} sits in the following fork.*

$$A \xrightarrow{f^\sharp} TB \begin{array}{c} \xrightarrow{\eta} \\ \xrightarrow{T\eta} \end{array} TTB$$

Proof. One way around square (2) is given by

$$Tf \circ \eta_{TA} \circ \eta_A = \eta_{TB} \circ f$$

and the other one is

$$\mu_{TB} \circ T(\eta_{TB}) \circ T(\eta_B) \circ f = T(\eta_B) \circ f. \quad \square$$

3.4 Relationship between the different classes of maps

Given a commutative monad on a cartesian monoidal category, we get both a copy-discard structure and a think-force structure canonically. The two structures interact in the following way.

Theorem 3.14. *Let T be a commutative monad on a cartesian monoidal category \mathbb{C} . Consider its Kleisli category $\text{Kl}(T)$ together with its canonical think-force and copy-delete structures. We have the following inclusions for morphisms of $\text{Kl}(T)$,*

$$\text{pure} \subseteq \text{thinkable} \subseteq \text{deterministic} \subseteq \text{all}.$$

Before proving the theorem, let's look at some examples.

Example 3.15. Consider the ‘maybe’ monad $TX = X + 1$ on Set . This satisfies the equalizer condition, so pure = thinkable. Every map is copyable, but discardable = pure. Therefore deterministic = thinkable.

Example 3.16. The ‘read-only state’ monad $TX = X \times X$ on Set satisfies the equalizer condition, so pure = thinkable. As observed in [14], every map is copyable and discardable. Thus deterministic \neq thinkable.

Example 3.17. For the Giry monad on Meas , the measure given in Example 3.9 is a deterministic, but not pure morphism. However, as one can verify, applying η and $T\eta$ one gets the same result, and so by Proposition 3.13, the morphism is thinkable. So, pure \neq thinkable. (As we will show in Theorem 9.1, every deterministic morphism for the Giry monad is thinkable.)

Let's now prove the theorem. We make use of the following lemma, which holds in every CD category.

Lemma 3.18 (not present in the published version). *Let*

$$A \xrightarrow{f} B \xrightarrow{g} C$$

be a composable pair of maps in any CD-category.

- (i) *If gf and g are both discardable, then f is discardable.*
- (ii) *If gf and g are both copyable and g is split monic, then f is copyable.*

Proof of Lemma 3.18. For (i), we have

$$\begin{aligned} A &\xrightarrow{f} B \xrightarrow{\text{del}} I \\ &= A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{\text{del}} I \\ &= A \xrightarrow{\text{del}} I. \end{aligned}$$

For (ii), we have

$$\begin{aligned} A &\xrightarrow{f} B \xrightarrow{\text{copy}} B \otimes B \xrightarrow{g \otimes g} C \otimes C \\ &= A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{\text{copy}} C \otimes C \\ &= A \xrightarrow{\text{copy}} A \otimes A \xrightarrow{f \otimes f} B \otimes B \xrightarrow{g \otimes g} C \otimes C \end{aligned}$$

whence the result since $g \otimes g$ is split monic. \square

Proof of Theorem 3.14. Let $f : X \rightsquigarrow Y$ be pure. Then there exists $g : X \rightarrow Y$ in \mathbb{C} such that $f^\sharp = \eta \circ g$, and so the fork of Proposition 3.13 can be decomposed through the unit fork (1) as follows,

$$X \xrightarrow{g} Y \xrightarrow{\eta} TY \begin{array}{c} \xrightarrow{\eta} \\ \xrightarrow{T\eta} \end{array} TTY$$

so f is thinkable.

Let now f be thinkable. Then $\text{think}_Y \circ f = Tf \circ \text{think}_X$ is equal to the composite of two pure (hence deterministic) maps. But as think_Y is deterministic and split monic, Lemma 3.18 says that f is deterministic. \square

4 The submonad of thinkable morphisms

We now want to express the category $\text{Kl}(T)_{\text{think}}$ itself as the Kleisli category of a new monad, a submonad of T . Recall that every pure morphism is thinkable (Theorem 3.14). Therefore we have a functor

$$\begin{aligned} \mathbb{C} &\xrightarrow{i} \text{Kl}(T)_{\text{think}} \\ g &\longmapsto (\eta \circ g)^{\flat}. \end{aligned}$$

which takes $g : X \rightarrow Y$ of \mathbb{C} and gives the pure map $f : X \rightsquigarrow Y$ of $\text{Kl}(T)$ such that, as above, $f^{\sharp} = \eta \circ g$.

Now with the help of Proposition 3.13, let's construct a right-adjoint to this functor i . We will suppose that the parallel pair $(\eta, T\eta)$ has an equalizer. This happens for example if \mathbb{C} has coreflexive equalizers (the common retraction of η and $T\eta$ is $\mu : TTY \rightarrow TY$). Choose an equalizer (DY, θ_Y) .

$$DY \xrightarrow{\theta_Y} TY \xrightarrow[T\eta]{\eta} TTY$$

If $f : X \rightsquigarrow Y$ is thinkable then $\eta \circ f^{\sharp} = T\eta \circ f^{\sharp}$ (and conversely, this is Proposition 3.13). Hence in this case there is a unique arrow $X \rightarrow DY$ making the triangle in the following diagram commute.

$$\begin{array}{ccc} X & & \\ \downarrow & \searrow f^{\sharp} & \\ DY & \xrightarrow{\theta_Y} & TY \xrightarrow[T\eta]{\eta} TTY \end{array}$$

In other words, we have a bijection between thinkable morphisms $X \rightsquigarrow Y$ and arrows $X \rightarrow DY$ of \mathbb{C} :

$$\text{Kl}(T)_{\text{think}}(X, Y) \cong \mathbb{C}(X, DY). \quad (3)$$

Since this bijection is obviously natural in X , we have the required adjunction. As usual, there is a canonical extension of the object assignment D to a functor $\text{Kl}(T)_{\text{think}} \rightarrow \mathbb{C}$. For completeness, we will describe its action of morphisms explicitly. So now let $g : Y \rightsquigarrow Z$ be thinkable. Notice that, by construction, $\theta_Y : DY \rightarrow TY$ forms a fork with the pair $(\eta, T\eta)$, and so its counterpart $(\theta_Y)^{\flat} : DY \rightsquigarrow Y$ of $\text{Kl}(T)$ is thinkable by Proposition 3.13. The composition $g \circ (\theta_Y)^{\flat} : DY \rightsquigarrow Z$ is then thinkable too, and so its counterpart $DY \rightarrow TZ$ in \mathbb{C} fits into the following fork.

$$DY \longrightarrow TZ \xrightarrow[T\eta]{\eta} TTZ$$

Therefore by the universal property of the coequalizer (DZ, θ_Z) , there is a unique morphism $DY \rightarrow DZ$ making the triangle in the following diagram commute.

$$\begin{array}{ccc} DY & & \\ \downarrow & \searrow & \\ DZ & \xrightarrow{\theta_Z} & TZ \xrightarrow[T\eta]{\eta} TTZ \end{array}$$

This morphism $DY \rightarrow DZ$ is Dg . It is automatic but easy to check that D preserves identities and composition, and the θ_Y assemble to form a natural transformation $D \Rightarrow T$.

It is also plain that the bijection (3) is natural in Y , making D a right-adjoint to i . We denote the resulting monad simply by D (instead of Di). Since i is a left-adjoint bijective-on-objects functor, its codomain $\text{Kl}(T)_{\text{think}}$ is isomorphic to $\text{Kl}(D)$. Moreover, as $\text{Kl}(D) \subseteq \text{Kl}(T)$, the monad D is a submonad of T .

As we will see in the next section, D is often idempotent.

5 Sober objects and idempotence

The original notion of *sobriety* (for mathematical objects) refers to a property of topological spaces $(X, \mathcal{O}(X))$. Roughly speaking, a topological space is *sober* if the existence and equality of its points is determined by the frame $\mathcal{O}(X)$ of its open sets. On a sober topological space, a point can be uniquely identified by saying in which open sets it is contained, and conversely, any suitable consistent choice of open sets (called a *completely prime filter*) corresponds to a point of the space. The notion of sober topological space has been linked to an equalizer of a continuation-style monad by several authors [3, 28, 34]. Our definition is less specific, and gives a notion of sobriety which in general depends on the monad.

Categorically, the points of a topological space X correspond to ‘maps in’ $1 \rightarrow X$ and the opens correspond to ‘maps out’ $X \rightarrow \mathbb{S}$ where \mathbb{S} is the Sierpinski space. The latter can be thought of as ‘observable’ [1] or ‘affirmable’ [36] properties of the points of X . They play a similar role to events in probability theory. Indeed, in probability theory one faces a similar challenge: suppose we have a measure that assigns only the values 0 or 1 to each event (measurable set). When is this measure a Dirac delta at a unique point? Whenever this is the case for all measures, we call the space *sober*, by analogy with topological spaces. We can define the concept of sobriety in general, in terms of the monad D , as follows.

Definition 5.1. Let T be a monad on a category \mathbb{C} . An object $X \in \mathbb{C}$ is *sober* for the monad T if its unit fork (1) is an equalizer.

A similar definition was given by Taylor [34, Definition 4.7]. See also [3, Section 3].

Equivalently, X is sober if the unit $e : X \rightarrow DX$ of the monad D is an isomorphism. That is, sober objects for T are exactly the fixed points of the adjunction (3), the one associated to D . Denote by $\text{Sober}(T)$ the full subcategory of \mathbb{C} (equivalently, of $\text{Kl}(D)$) of sober objects. This is also known as the center of the adjunction.

Remark 5.2. The following conditions are equivalent for a monad T on a category \mathbb{C} :

- Every object of \mathbb{C} is sober.

- The monad T satisfies Moggi’s “equalizing requirement” [25, Section 4].
- Every thinkable morphism is uniquely pure.

In particular, every object is sober whenever every deterministic morphism of $\text{Kl}(T)$ is uniquely pure. For the case of Markov categories, this is in particular an instance of *representability*, as defined in [8, Section 3].

Example 5.3. For the distribution monad on Set , every object is sober.

Example 5.4. For the Giry monad restricted to the category of *standard Borel spaces*, every object is sober. See for example [6, Example 10.5] (where the Kleisli category is called BorelStoch).

Moreover, by Remark 2.9, every object in the form TX is sober for every monad T .

In general, not all objects are sober. However, in several cases, given any object we can find a universal “sobrification”. This is the case whenever the adjunction (3) is idempotent. By the general theory of idempotent adjunctions applied to the case of the Kleisli category of D , we have the following statement.

Proposition 5.5. *The following conditions are equivalent.*

- The monad D is idempotent.
- For every X , the object DX is sober.
- The functor $D : \text{Kl}(D) \rightarrow \mathbb{C}$ is fully faithful.
- The inclusion of fixed points $\text{Sober}(T) \hookrightarrow \text{Kl}(D)$ is an equivalence.
- For every X , the counit $\theta^b : DX \rightsquigarrow X$ is an isomorphism.

If any (hence all) of the conditions above holds, we can view D as a “sobrification” functor, analogous to the case of topological spaces, exhibiting sober objects as a reflective subcategory of \mathbb{C} .

Theorem 5.6. *For the following categories \mathbb{C} and monads T , the associated monad D is idempotent:*

- The “Giry” monads of probability and subprobability measures P and M on Meas ;
- The lower Vietoris monad (a.k.a. Hoare powerdomain) on Top .

For more details on the latter see [12, Section 2].

Theorem 5.6 can be proven by the following helpful lemma.

Lemma 5.7. *The monad D is idempotent if and only if for every X , the map $T\theta : TDX \rightarrow TTX$ is monic. In particular, as the map $\theta : DX \rightarrow TX$ is an equalizer, it suffices to show that T maps regular monomorphisms to monomorphisms.*

Proof. First of all, by Proposition 5.5, D is idempotent if and only if for every X , the counit $\theta^b : DX \rightsquigarrow X$ is an isomorphism.

Denote now by $e : X \rightarrow DX$ the unit of the monad D , which can be obtained from the universal property of D as an equalizer for the unit fork, as in the following diagram.

$$\begin{array}{ccccc} X & & & & \\ \downarrow e & \searrow \eta & & & \\ DX & \xrightarrow{\theta} & TX & \xrightarrow[T\eta]{\eta} & TTX \end{array}$$

By the triangle identities of the adjunction of D , θ^b is split epi with section given by the map $E := (\eta \circ e)^b : X \rightsquigarrow DX$ induced by unit $e : X \rightarrow DX$. Therefore θ^b is an isomorphism if and only if E is its actual inverse, i.e. if $E \circ \theta^b = \text{id}_{DX}$, which in terms of the category \mathbb{C} reads

$$DX \xrightarrow{\theta} TX \xrightarrow{T_e} TDX = DX \xrightarrow{\eta} TDX. \quad (4)$$

So suppose that (4) holds. Then

$$\begin{aligned} TDX &\xrightarrow{T\theta} TTX \xrightarrow{TT_e} TTDX \xrightarrow{\mu} TDX \\ &= TDX \xrightarrow{T\eta} TTDX \xrightarrow{\mu} TDX \\ &= TDX \xrightarrow{\text{id}} TDX, \end{aligned}$$

so $T\theta$ is split monic.

Conversely, suppose that $T\theta$ is monic. Then

$$\begin{aligned} DX \xrightarrow{\theta} TX \xrightarrow{T_e} TDX \xrightarrow{T\theta} TTX &= DX \xrightarrow{\theta} TX \xrightarrow{T\eta} TTX \\ &= DX \xrightarrow{\theta} TX \xrightarrow{\eta} TTX \\ &= DX \xrightarrow{\eta} TDX \xrightarrow{T\theta} TTX, \end{aligned}$$

which implies (4). \square

Proof of Theorem 5.6. By Lemma 5.7, it suffices to show that the Giry monad maps regular monomorphisms of Meas (i.e. embeddings of measurable spaces) to monomorphisms of Meas (i.e. injective measurable functions). So let $i : X \rightarrow Y$ be an embedding of measurable spaces, that is, an injective function such that the every measurable subset A of X is in the form $f^{-1}(B)$ for some measurable subset B of Y . Let p and q be measures on X , and suppose that $f_*p = f_*q$. Then for every measurable subset $A \subseteq X$ we can find a measurable $B \subseteq Y$ such that

$$p(A) = p(f^{-1}(B)) = f_*p(B) = f_*q(B) = q(f^{-1}(B)) = q(A).$$

Therefore, $p = q$ already on X , and hence $f_* : MX \rightarrow MY$ and its restriction $PX \rightarrow PY$ are injective.

The lower Vietoris monad case is analogous, once one sees closed sets as dual to open sets (as in Section 9.2). \square

The idempotent monad associated to the lower Vietoris monad is not only idempotent, but it is also the sobrification monad of topology, hence the name “sober”.

Theorem 5.8. *The monad D associated to the lower Vietoris monad H on Top is the functor assigning to a topological space X the subset of HX given by the irreducible closed sets.*

The proof of this theorem is given at the end of Section 9.2. The result resembles the known characterizations of sobriety in terms of equalizers [3, 28, 34], but note that HX is not quite a continuation (it behaves more like a subspace of the continuation — but of course Top is not cartesian closed, see also [12, Appendix B]).

6 Observational monads

In higher-order programming languages one typically has a ‘ground type’ such as ‘nat’ or ‘bool’, representing actual data that we can handle as input or output, as well as higher-order types (i.e. function types), or open terms, which are never directly accessible as inputs or outputs and only appear at intermediate stages of computation. Thus it is of interest to consider when two values of a higher-order type can be interchanged in the middle of programs without altering the *observable* behaviour of the computer. From the point of view of probability theory one has the same intuition for *real numbers*, i.e. proving equality of probability measures involves, in the end, proving that certain integrals give the same number. An *observational monad* makes this intuition precise: in the sense proposed here it corresponds to a semantics of open terms which is abstract for a certain kind of observable equivalence.

Let \mathbb{C} be a cartesian monoidal category and T a commutative monad. Since each $X \in \mathbb{C}$ is a commutative comonoid in $\text{Kl}(T)$, for each $n \in \mathbb{N}$ there is a canonical map

$$\text{copy}_n : TX \rightsquigarrow (TX)^{\otimes n}$$

obtained by iterating the copy map (all possibilities are equal by coassociativity).

By post-composing with force $\otimes \dots \otimes \text{force} : (TX)^{\otimes n} \rightsquigarrow X^{\otimes n}$, we define a map

$$\text{samp}_n : TX \rightsquigarrow X^{\otimes n}, \quad (5)$$

the *n*’th *sampling map*. As special cases, $\text{samp}_0 = \text{del}$ and $\text{samp}_1 = \text{force}$. (Note that the map samp appearing in [8] and [7] corresponds to our $\text{samp}_1 = \text{force}$.)

We now want to make the intuition precise that *probability measures can be tested for equality by taking repeated independent samples*.

Definition 6.1. Let T be a commutative monad on a cartesian monoidal category \mathbb{C} . Then T is an *observational monad* if for every object X the family of maps $(\text{samp}_n : TX \rightsquigarrow X^{\otimes n})_{n \in \mathbb{N}}$ is jointly monic in $\text{Kl}(T)$.

Remark 6.2. The notion of observational monad also makes sense for strong monads that are not necessarily commutative, but we will not pursue that line here.

Proposition 6.3 (Remark 6.3 in the published version). *The maps (5) are jointly monic in $\text{Kl}(T)$ iff the maps*

$$TTX \xrightarrow{T(\Delta_n)} T((TX)^n) \xrightarrow{T(\nabla_n)} TT(X^n) \xrightarrow{\mu} T(X^n) \quad (6)$$

are jointly monic in \mathbb{C} .

Intuitively, the map in (6) takes a distribution on distributions on X , samples to get a distribution on X , and then returns the result of n independent samples from that distribution.

Proof. Consider the following diagram in $\text{Kl}(T)$.

$$A \begin{array}{c} \xrightarrow{f} \\ \rightsquigarrow \\ \xrightarrow{g} \end{array} TX \xrightarrow{\text{copy}_n} TX^{\otimes n} \xrightarrow{\text{thunk}^{\otimes n}} X^{\otimes n}$$

The condition that the maps samp_n form a monic family in $\text{Kl}(T)$ means that whenever the composites $\text{thunk}^{\otimes n} \circ \text{copy}_n \circ f$ and $\text{thunk}^{\otimes n} \circ \text{copy}_n \circ g$ in the diagram above are equal for all n , then $f = g$.

Now in terms of the category \mathbb{C} we can rewrite the (Kleisli) composites above as follows.

$$A \begin{array}{c} \xrightarrow{f^\sharp} \\ \rightsquigarrow \\ \xrightarrow{g^\sharp} \end{array} TTX \xrightarrow{T(\Delta_n)} T((TX)^n) \xrightarrow{T(\nabla_n)} TT(X^n) \xrightarrow{\mu} T(X^n)$$

The condition that the maps samp_n form a monic family in $\text{Kl}(T)$ now reads: whenever the composites in this new diagram are equal, then $f = g$, or equivalently $f^\sharp = g^\sharp$. But this means precisely that the maps in (6) are jointly monic in \mathbb{C} . \square

From (6) it is clear that taking one sample $n = 1$ is in general insufficient: that would correspond to hoping that the map $\mu : TTX \rightarrow TX$ alone is monic. This is in general not the case. For the Giry monad, for example, this is far from injective: a probability measure can be in general obtained as a mixture of other measures in several different ways.

The two major consequences of observability will be given in Section 8 and Section 7. In the rest of this section, we give some technical sufficient conditions that one can use in order to prove that a monad is observational, and we introduce the idea of ‘objects which are directly observable’ (such as real numbers for probability).

6.1 Observations via result objects

An observational monad is one for which an ‘observation procedure’ for the powers of X can be transferred to one for TX . Ordinarily, direct observations are only made for *ground types*, e.g. Boolean values or real numbers, and all observations at higher types ultimately implemented in terms of direct ground observations. We call these special types the *result objects*.

Definition 6.4. Let R be an object of \mathbb{C} . We say that the monad T is *R-observational* iff for each object X , the family of morphisms

$$TX \xrightarrow{\text{samp}_n} X^{\otimes n} \xrightarrow{h_1^b \otimes \dots \otimes h_n^b} R^{\otimes n} \quad (7)$$

where $n \in \mathbb{N}$ and $h_1, \dots, h_n : X \rightarrow TR$, is jointly monic. We call R the *result object*.

It is easy to see that if T is R -observational then T is observational. We can also write the R -observationality condition in terms of the category \mathbb{C} , analogously to Proposition 6.3. It reads that the following maps need to be jointly monic for all n .

$$TTX \xrightarrow{T(Th_1, \dots, Th_n)} T((TTR)^n) \xrightarrow{T(\nabla_n \nabla_n)} TTT(R^n) \xrightarrow{\mu\mu} T(R^n) \quad (8)$$

As well as being useful for demonstrating that a given monad is observational, the property of being R -observational allows us to exploit Lemma 5.7. Recall that an object P is \mathcal{M} -injective [24] with respect to a class \mathcal{M} of morphisms iff whenever $(i : A \rightarrow B) \in \mathcal{M}$ and $h : A \rightarrow P$, there exists a (not necessarily unique) map $k : B \rightarrow P$ with $k \circ i = h$.

Proposition 6.5. *Suppose that T is R -observational and that TR is injective with respect to the class of regular monomorphisms. Then the submonad D of thinkable morphisms is idempotent.*

Proof. By Lemma 5.7 it suffices to show that $T\theta_X$ is monic. Since $\eta_{TTX} \circ T\theta_X = TT\theta_X \circ \eta_{TDX}$ it suffices for this to show that $TT\theta_X$ is monic. (Indeed, this is also necessary, since we saw above that if $T\theta_X$ is monic then it is also split monic). The R -observationality condition for DX says that the family of maps (8) is monic, so it suffices to show that

$$\begin{aligned} & \mu\mu \circ T(\nabla_n \nabla_n) \circ T(Th_1, \dots, Th_n) \circ TT\theta_X \\ &= \mu\mu \circ T(\nabla_n \nabla_n) \circ T(T(h_1 \circ \theta_X), \dots, T(h_n \circ \theta_X)) \end{aligned}$$

is monic. By R -observationality, it is sufficient that every map $s : DX \rightarrow TR$ factorize as $h \circ \theta_X$ for some $h : TX \rightarrow TR$, since then this family is equivalent to the one in the R -observationality condition for TX . But this just says that TR is a $\{\theta_X\}$ -injective object. \square

6.2 Result objects with a monoid structure

A particularly useful case of result object is when R is a monoid in \mathbb{C} . Then TR is a monoid in the category of T -algebras, with unit e^{TR} and multiplication m^{TR} given as follows,

$$1 \xrightarrow{\eta} T1 \xrightarrow{Te^R} TR \quad TR \times TR \xrightarrow{\nabla} T(R \times R) \xrightarrow{Tm^R} TR \quad (9)$$

where e^R and m^R are the unit and multiplication of R . For probability monads, we generally take TR to be the unit interval $[0, 1]$, which is a monoid under multiplication (as well as an algebra under integration). In particular,

- For the distribution and Giry monad P , the object $[0, 1]$ can be written as $P\{0, 1\}$ (where the product in $R = \{0, 1\}$ is multiplication);
- For the monad of subprobability measures M , the object $[0, 1]$ can be written as $M1$ (with the trivial monoid structure on $R = 1$).

Let now $(S, s : TS \rightarrow S)$ be a T -algebra with monoid structure (for example, in the form $S = TR$ as above, with

R a monoid in \mathbb{C} .) The Eilenberg-Moore adjunction gives a bijection for each object X ,

$$\begin{aligned} \mathbb{C}(X, S) &\xrightarrow{\cong} \text{Alg}(T)(TX, S) \\ f &\longmapsto \varepsilon_f = s \circ Tf. \end{aligned}$$

For the Giry monad, for example, given the function $f : X \rightarrow [0, 1]$, the corresponding morphism $\varepsilon_f : PX \rightarrow [0, 1]$ is the *integral* of f :

$$p \mapsto \int f dp.$$

Because of this correspondence, we can test observationality of T in terms of the maps ε_f . Now, these maps alone are in general not enough to test observationality, but their *products* are. Let's define what we mean by "product". Let's write the unit and multiplication of the monoid S by $e^S : 1 \rightarrow S$ and $m^S : S \times S \rightarrow S$. Denote now by $m_n^S : S^n \rightarrow S$ the maps given by

- $m_0 = e$;
- $m_1 = \text{id}$;
- $m_2 = m$;
- For $m > 2$, m_n is the unique (by associativity) way of multiplying n objects, $S \times \dots \times S \rightarrow S$.

Given now $h_1, \dots, h_n : X \rightarrow S$, define the *pointwise product* $h_1 \cdots h_n$ as the map

$$X \xrightarrow{\Delta_n} X^n \xrightarrow{h_1 \times \dots \times h_n} S^n \xrightarrow{m_n} S. \quad (10)$$

In Meas, if the h_i are functions into $[0, 1]$, this gives the ordinary product of functions $x \mapsto h_1(x) \cdots h_n(x)$.

We can now test observationality by means of the maps ε_f and their products.

Lemma 6.6. *Let R be a monoid in \mathbb{C} , and consider the free algebra $(S, s) = (TR, \mu)$ with its induced monoid structure. The monad T is R -observational (hence observational) if the following maps are jointly monic,*

$$TTX \xrightarrow{T(\varepsilon_{h_1} \cdots \varepsilon_{h_n})} TS \xrightarrow{s} S \quad (11)$$

for all $n \in \mathbb{N}$ and $h_1, \dots, h_n : X \rightarrow S$.

Note that we can write the maps above even more concisely as $\varepsilon_{\varepsilon_{h_1} \cdots \varepsilon_{h_n}}$.

Proof. It suffices to show that the maps of the family (11) can be obtained from the maps of family (8) by postcomposition (because if the former are jointly monic, then surely the latter have to be as well).

Now since $S = TR$ and $s = \mu_R$, we have that $\varepsilon_{h_i} = \mu_R \circ Th_i$, and so we can rewrite (11) as follows,

$$\begin{aligned}
& \mu \circ T(\mu \circ Th_1 \cdots \mu \circ Th_n) \\
&= \mu \circ T(m_n^S) \circ T(\mu \circ Th_1, \dots, \mu \circ Th_n) \\
&= \mu \circ TT(m_n^R) \circ T\nabla_n \circ T(\mu^n) \circ T(Th_1, \dots, Th_n) \\
&= Tm_n^R \circ \mu \circ T\nabla_n \circ T(\mu^n) \circ T(Th_1, \dots, Th_n) \\
&= Tm_n^R \circ \mu \circ T\mu \circ TT\nabla_n \circ T\nabla_n \circ T(Th_1, \dots, Th_n) \\
&= Tm_n^R \circ \mu \circ \mu \circ TT\nabla_n \circ T\nabla_n \circ T(Th_1, \dots, Th_n)
\end{aligned}$$

where we used, in turn,

- the pointwise product formula (10);
- the fact that $m_n^S = Tm_n^R \circ \nabla$, from (9);
- naturality of μ ;
- compatibility of the monoidal structure ∇ of T with μ ;
- the associativity square for μ .

The last line is exactly the composition of Tm_n^R with (8). \square

Lemma 6.6 allows us to prove observability of several monads, including the Girya monad (see Section 9).

7 Determinism in the observational case

The key consequence of a monad's being observational is the following.

Theorem 7.1. *Let T be an observational commutative monad on a cartesian monoidal category \mathbb{C} . Then every deterministic morphism in $\text{Kl}(T)$ is thunkable.*

Proof. Let $f : A \rightsquigarrow B$ be deterministic. It suffices to check that

$$A \xrightarrow{f} B \xrightarrow{\text{thunk}} TB \quad A \xrightarrow{\text{thunk}} TA \xrightarrow{T^b f} TB$$

become equal after postcomposition with each of the samp_n maps from (5). For $n = 1$,

$$\begin{aligned}
& \text{samp}_1 \circ \text{thunk} \circ f \\
&= \text{force} \circ \text{thunk} \circ f \\
&= f \\
&= f \circ \text{force} \circ \text{thunk} \\
&= \text{force} \circ T^b f \circ \text{thunk} \\
&= \text{samp}_1 \circ T^b f \circ \text{thunk}.
\end{aligned}$$

For $n = 0$, we use the facts that f is discardable and thunk and $T^b(f)$ are both pure (and so discardable):

$$\begin{aligned}
& \text{samp}_0 \circ \text{thunk} \circ f \\
&= \text{del} \circ \text{thunk} \circ f \\
&= \text{del} \\
&= \text{del} \circ T^b f \circ \text{thunk} \\
&= \text{samp}_0 \circ T^b f \circ \text{thunk}.
\end{aligned}$$

For $n \geq 2$, we use the facts that f is copyable and that thunk and $T^b(f)$ are both pure (and so copyable) and the already proved $n = 1$ case.

$$\begin{aligned}
& \text{samp}_n \circ \text{thunk} \circ f \\
&= \text{force}^{\otimes n} \circ \text{copy}_n \circ \text{thunk} \circ f \\
&= \text{force}^{\otimes n} \circ \text{thunk}^{\otimes n} \circ \text{copy}_n \circ f \\
&= \text{force}^{\otimes n} \circ \text{thunk}^{\otimes n} \circ f^{\otimes n} \circ \text{copy}_n \\
&= \text{force}^{\otimes n} \circ (T^b f)^{\otimes n} \circ \text{thunk}^{\otimes n} \circ \text{copy}_n \\
&= \text{force}^{\otimes n} \circ \text{copy}_n \circ T^b f \circ \text{thunk} \\
&= \text{samp}_n \circ T^b f \circ \text{thunk}. \quad \square
\end{aligned}$$

Corollary 7.2. *Let T be an observational commutative monad on a cartesian category. The associated submonad D is equivalently characterizing the deterministic morphisms of $\text{Kl}(T)$.*

This will be the case for example for the Girya monad, as we show in Section 9.1.

8 De Finetti's theorem

De Finetti's theorem [5] gives a connection between *random distributions* and *exchangeable sequences*. As we show in this section, the notion of observability can say something about de Finetti's theorem, as infinite sequences can be seen as limits of a finite, arbitrarily large amount of observations.

In general, in a Kleisli category we have monoidal products rather than cartesian ones. Because of that, in order to talk about infinite sequences in X , one cannot take a countable cartesian product of copies of X . One, rather, has to extend monoidal products to the infinite case. This was accomplished in [13] for the case of Markov categories. Here we give the analogous construction for CD categories (which is almost the same). For further context, motivation, and applications we refer to the aforementioned source, as well as to the later [7].

8.1 Kolmogorov products

In order to form Kolmogorov products, let's take a look at the so-called *marginalization maps*. Given objects X and Y in a CD category, their tensor product $X \otimes Y$ can be interpreted as the object of *joint states*. For example, if we are in a Kleisli category, a morphism $I \rightsquigarrow X \otimes Y$ corresponds to an arrow $1 \rightarrow T(X \times Y)$ of the base category, which is a joint probability measure if T is the Girya monad. We can then ap-

ply the map $X \otimes Y \xrightarrow{\text{id} \otimes \text{del}} X \otimes I \cong X$, which intuitively "discards" Y . This maps a joint state into the *marginal* state on X . For the case of the Girya monad, this corresponds to taking the marginal probability.

Since \otimes is a bifunctor, marginalizations are deterministic, and also natural in the sense that the following diagram

commutes.

$$\begin{array}{ccc} X \otimes Y & \xrightarrow{\text{id} \otimes \text{del}} & X \otimes I \\ \text{del} \otimes \text{id} \downarrow & & \downarrow \text{del} \otimes \text{id} \\ I \otimes Y & \xrightarrow{\text{id} \otimes \text{del}} & I \end{array}$$

The same is true for marginalizations of finite sequences $X_1 \otimes \cdots \otimes X_n$.

Definition 8.1. Let I be an infinite set, and let $\{X_i\}$ be an I -indexed collection of objects of a CD category. The *Kolmogorov product* of the family $\{X_i\}$, denoted by X_I , is the (cofiltered) limit of the diagram whose objects are the finite tensor products $\bigotimes_{i \in F} X_i$ over all the finite subsets F of I , and whose morphisms are the marginalizations, if moreover the following two further conditions are satisfied:

- The arrows of the limit cone are deterministic;
- The limit is preserved by the tensor product $X \otimes -$ for each object X .

We call a *Kolmogorov power* a Kolmogorov product where the objects X_i are all isomorphic to a same object X . In that case we denote the product by X^I . In the context of de Finetti, we are interested in *countable Kolmogorov powers*, i.e. where I is, equivalently, the set \mathbb{N} of natural numbers.

Let's now look at the interaction between observational monads and Kolmogorov powers. If the Kolmogorov product $A^{\mathbb{N}}$ exists, the maps samp_n make the following diagram commute for each n and for each marginal projection $\pi : A^{\otimes n} \rightsquigarrow A^{\otimes n-1}$, i.e. $\pi \circ \text{samp}_n = \text{samp}_{n-1}$. Therefore, there exist a unique map $\text{samp}_{\mathbb{N}}$ making the following diagram commute for each n .

$$\begin{array}{ccc} TA & \xrightarrow{\text{samp}_{\mathbb{N}}} & A^{\mathbb{N}} \\ \text{samp}_n \downarrow & & \downarrow \pi \\ A^{\otimes n} & & \end{array}$$

This map is monic if and only if the family $\{\text{samp}_n\}$ is jointly monic. Therefore, if $\text{Kl}(T)$ has countable Kolmogorov products, T is observable if and only if $\text{samp}_{\mathbb{N}}$ is monic for every object A .

8.2 Application to the synthetic de Finetti theorem

In [7] a de Finetti theorem was proved for Markov categories satisfying particular conditions. It reads as follows.

Theorem 8.2 (Theorem 4.4 in [7]). *Consider a Markov category with countable Kolmogorov powers, conditionals, and almost-surely-compatibly representable, with probability monad P^1 . Then a morphism $p : A \rightsquigarrow X^{\mathbb{N}}$ is exchangeable if and only if there is a morphism $\rho : A \rightsquigarrow PX$ such that*

$$p = \text{samp}_{\mathbb{N}} \circ \rho.$$

¹For the precise definition of these properties, we refer to the original source.

We now know that if (and only if) the probability monad P is observational, then $\text{samp}_{\mathbb{N}}$ is monic. Therefore, in the theorem above, we can even conclude that the morphism ρ is unique. This is in particular the case for the Giry monad on standard Borel spaces, and so for the Markov category BorelStoch . In traditional probability theory, the uniqueness of ρ (at least for states, $A = I$) is already known. Our formalism, however, incorporates this statement of uniqueness into the categorical formalism. This opens the road to study de Finetti's theorem and similar statements as categorical universal properties, a path undertaken already (using a different formalism) for example by [32].

9 Examples

Here we give some important examples of observational monads: the Giry and sub-Giry monads and the lower Vietoris monad of nondeterminism. We also give an interesting but unusual example with the name generation monad.

9.1 The Giry monads on measurable spaces

Theorem 9.1. *The Giry monad and the monad M of subprobability measures are observational.*

The proof uses the following version of the celebrated functional monotone class theorem, (a.k.a. the “ π - λ theorem for functions”). See [2, Theorem 2.12.9] for a reference.

Theorem 9.2 (Functional monotone class theorem). *Let X be a measurable space, and denote by $F(X)$ the space of bounded measurable functions $X \rightarrow \mathbb{R}$. Consider a vector subspace $H \subseteq F(X)$ containing the function 1, and such that its positive cone is closed under sequential increasing limits. Consider a subset $K \subseteq H$ closed under pointwise products, and denote by $\sigma(K)$ the σ -algebra generated by the functions in K . Then H contains all $\sigma(K)$ -measurable functions.*

In order to prove Theorem 9.1, let's also recall that given a measurable set X , the (“Giry”) σ -algebra of MX is equivalently generated by the functions

$$\begin{aligned} MX &\xrightarrow{\varepsilon_f} [0, 1] \\ m &\longmapsto \int f dm \end{aligned}$$

for each measurable function $f : X \rightarrow [0, 1]$ (see [18]).

Let's now prove our theorem.

Proof of Theorem 9.1. Let's prove the assert for M , the case of P is analogous. We will use Lemma 6.6 with $S = [0, 1]$ (i.e. result object $R = 1$). So it suffices to show that the maps (11) are jointly monic. Let's unpack the expression (11) for our case. We need to show that on a measurable space X , if p and q are measures on MX , i.e. in MMX , then $p = q$ if and only if p and q agree on the following products,

$$\int_{MX} \varepsilon_{h_1} \cdots \varepsilon_{h_n} dp = \int_{MX} \varepsilon_{h_1} \cdots \varepsilon_{h_n} dq \quad (12)$$

for each finite collection $\{h_1, \dots, h_n\}$ of measurable functions $X \rightarrow [0, 1]$, including the 0-ary product (the function 1, meaning that p and q have the same normalization).

Now consider the collection K of all finite pointwise products of the functions ε_f ,

$$K := \{\varepsilon_{f_1} \cdots \varepsilon_{f_n} : n \in \mathbb{N}, f_i \in F(X), \lambda \in \mathbb{R}\}.$$

This collection generates the sigma algebra of MX , and it is closed under products. Take now the measures p, q on MX , i.e. in MMX . Denote by H the subset of bounded measurable functions $g : MX \rightarrow \mathbb{R}$ such that $\int g dp = \int g dq$. By linearity and monotone continuity of integration, H is a vector subspace of $F(MX)$ and its positive cone is closed under sequential increasing limits. Moreover, it contains 1. Since p and q agree on K (by (12)), $K \subseteq H$, and so we are in the hypothesis of Theorem 9.2. The theorem tells us that every function that is measurable for the σ -algebra generated by K lies in H , i.e. cannot tell p and q apart. But since K generates the whole σ -algebra of MX , this means that $p = q$. \square

Corollary 9.3. *The (sub-)Giry monad admits an idempotent submonad of deterministic states.*

Proof. We know from Theorem 5.6 that the monad D of thinkable morphisms is an idempotent submonad of M (resp. P). Since M (resp. P) is observational, we know by Theorem 7.1 that thinkable and deterministic morphisms coincide. Therefore we can equivalently view D as the monad whose Kleisli morphisms are deterministic (i.e. zero-one) Markov kernels. In particular, the elements of DX are the zero-one measures on X . \square

9.2 The lower Vietoris monad on topological spaces

Theorem 9.4. *The lower Vietoris monad on Top (a.k.a. the Hoare powerdomain) is observational.*

Again as result object we take the terminal object 1, so that $S = H1$ is the Sierpinski space $\{0, 1\}$, equipped with its usual topology (generated by $\{1\}$). This way, a continuous function $f : X \rightarrow \{0, 1\}$ is equivalently and open subset of X (by taking $f^{-1}(1)$).

We use the following statement, [12, Lemma 2.3].

Proposition 9.5. *Let X be a topological space. Let \mathbb{B} be a basis of the topology of X . Let C and D be closed subsets of X , i.e. elements of HX . Then $C = D$ if and only if for every open U in \mathbb{B} , the set C intersects U if and only if D does.*

Moreover, the topology of HX is the weakest topology making the following maps continuous for all open sets U of X ,

$$\begin{array}{ccc} HX & \xrightarrow{\varepsilon_U} & \{0, 1\} \\ C & \longmapsto & \begin{cases} 1 & C \cap U \neq \emptyset; \\ 0 & C \cap U = \emptyset. \end{cases} \end{array} \quad (13)$$

As maps into $\{0, 1\}$ corresponds to open sets, we can view the ε_U as open sets of HX (generating the topology). For more details, see again [12, Section 2].

Let's now prove the theorem. In some sense, the role of the π - λ theorem this time is played by Proposition 9.5.

Proof of Theorem 9.4. We use Lemma 6.6 where S is the Sierpinski space, with monoid structure given by 'meet', and H -algebra structure, i.e. (topological) sup-semilattice structure, given by 'join'. It suffices to show that the maps (11) are jointly monic. Unpacking the expression of Lemma 6.6, and using (13), given closed subsets $C, D \subseteq HX$ we have to show that they are equal if and only if for all n and for all open sets $U_1, \dots, U_n \subseteq X$, the set C intersects the intersection

$$\varepsilon_{U_1} \cap \cdots \cap \varepsilon_{U_n} \quad (14)$$

if and only if D does. Now, as the topology of HX is generated by the ε_U , the ε_U for a subbasis, and hence their intersection form a basis. By Proposition 9.5, then, the sets (14) are indeed enough to test that $C = D$. \square

In particular, Kleisli morphisms for the monad H are deterministic if and only if they are thinkable.

We are now ready to prove Theorem 5.8, i.e. that for the monad H , sober objects are exactly sober topological spaces.

Proof of Theorem 5.8. Let X be a topological space, and let $C \in HX$ be a closed subset of X . We have to prove that C is irreducible if and only if it is thinkable as a morphism $1 \rightarrow HX$ of Top , or equivalently, deterministic. Now first of all, C is discardable as a Kleisli morphism of H if and only if C , as a set, is nonempty. Moreover, C as a morphism is copyable if and only if the closed subsets $C \times C$ and $\text{copy}(C) = \{(c, c) : c \in C\}$ of $X \times X$ are equal. By Proposition 9.5 we can test equalities of closed subsets of $X \times X$ by looking at a basis, and we pick the basis given by the products $U \times V$ of open subsets $U, V \subseteq X$. We have that $C \times C$ intersects $U \times V$ if and only if C intersects U and V separately, and that $\text{copy}(C)$ intersects $U \times V$ if and only if C intersects U and V at the same point, i.e. if $C \cap U \cap V$ is nonempty. In other words,

$$(C \times C) \cap (U \times V) = \text{copy}(C) \cap (U \times V),$$

for all U and V , and so $C \times C$ and $\text{copy}(C)$ are equal, if and only if the mapping on open sets

$$\begin{array}{ccc} O(X) & \longrightarrow & \{0, 1\} \\ U & \longmapsto & \begin{cases} 1 & C \cap U \neq \emptyset; \\ 0 & C \cap U = \emptyset. \end{cases} \end{array}$$

preserves binary intersections. That is, C induces a completely prime filter on the frame $O(X)$, which means precisely that C is irreducible. \square

9.3 Local names

In [29] it was shown that randomness can be used to model fresh name generation, exploiting the fact that repeated uniform sampling from $[0, 1]$ returns distinct values with probability 1. Conversely, the traditional model of local names from [31] is an interesting model of categorical probability. We take $D = [\text{Inj}, \text{Set}]$ and the *name-generation monad*

$$(TX)a = \text{colim}_{b \in \text{Inj}} X(a + b).$$

If a ‘stage of computation’ means the (finite) set of names which are in use so far, an object $X \in [\text{Inj}, \text{Set}]$ consists of a set of values at each stage. A value of TX at stage a consists of a set b of local names that have been generated together with a value of X at the resulting stage $a + b$. The object TX is quotiented so that: 1) the names in b are ‘bound’ or α -convertible; 2) names that are not referenced in the value of X are out of scope and so discarded.

The monad T is commutative and affine (satisfies $T1 \cong 1$). It does not satisfy the equalizing requirement, but it is observational.

Theorem 9.6. *The monad T on $[\text{Inj}, \text{Set}]$ is observational.*

Proof. Let $[b, [c, x]]$ and $[b', [c', x']]$ be representatives of elements of $(TTX)a$ which are equal under all of the samp_n maps. Then, for every $n \in \mathbb{N}$, writing c_1, \dots, c_n and c'_1, \dots, c'_n for tuples of copies of c and c' using fresh names and $x_i = x[c_i/c]$, $x'_i = x'[c'_i/c']$, we have that

$$(b + c_1 + \dots + c_n, (x_1, \dots, x_n))$$

and

$$(b' + c'_1 + \dots + c'_n, (x'_1, \dots, x'_n))$$

are equivalent elements of $(TX)(a)$, meaning that there are injections $f : b + c_1 + \dots + c_n \hookrightarrow d$ and $f' : b' + c'_1 + \dots + c'_n \hookrightarrow d$ such that $X(1_a + f)(x_i) = X(1_a + f')(x'_i) \in X(a + d)$ for each $1 \leq i \leq n$. By choosing n sufficiently large, i.e. bigger than $|b| + |b'|$, we can ensure that for some i the image of c_i under f is disjoint from the image of b' under f' and also that the image of c'_i under f' is disjoint from the image of b under f . Let $d_1 = \text{im}(f|_b) \cup \text{im}(f'|_{b'})$ and $d_2 = d \setminus d_1$. Then $[b, [c, x]]$ is also represented by $[d_1, [d_2, X(1_a + f|_{b + f|_{c_1 + \dots + c_n}})(x_i)]]$ and similarly $[b', [c', x']]$ is also represented by $[d_1, [d_2, X(1_a + f'|_{b' + f'|_{c'_1 + \dots + c'_n}})(x'_i)]]$, but by construction these are actually equal. \square

It is well known that the full subcategory of $[\text{Inj}, \text{Set}]$ whose objects are the pullback-preserving functors is equivalent to the *Schanuel topos*, or topos of *nominal sets* [16]. The following facts are straightforward to establish.

1. Every object TX is a nominal set.
2. T preserves monos whose codomain is a nominal set.
3. Every nominal set is sober with respect to T .

Corollary 9.7. *The sober objects for the name-generation monad are precisely the nominal sets, and the sobrifcation monad is idempotent.*

10 Conclusion

10.1 Related work

The notion of *sober space* comes from topology, meaning a topological space where the set of points (set of values) is determined by the set of open subsets (observable predicates). The idea that abstractly a ‘sober’ object is one for which the fork (1) is an equalizer has appeared before, e.g. [3, 28, 34]. These works were not focused on probability, but there is some overlap in the examples of interest. In [34], it is shown that the passage $\mathbb{C} \mapsto \text{Kl}(T)_{\text{thunk}}$ ‘freely adjoins sobriety’ to \mathbb{C} . Our concern is a little different: we use thunkable morphisms to transform the *objects* of \mathbb{C} into sober objects in the same category.

The concepts of discardability, copyability, and thunkability have been developed, for example, in [14, 15, 21, 35]. Since most work was on general computational effects, there is another fundamental class of morphisms of interest, the *central morphisms*. For commutative effects as studied in this paper, every morphism is central. Thus in general one needs to consider *symmetric premonoidal categories* [23], rather than symmetric monoidal categories. We note that much of our framework does not rely on monoidality rather than premonoidality, but we leave to future work the investigation of noncommutative examples.

10.2 Summary

We have given natural conditions on a commutative monad, *observationality* (Def. 6) and *S-observational* (Def. 6.4), for which the deterministic computations are precisely the thunkable ones. Under mild conditions we showed that these imply that the monad has an associated, idempotent *sobrifcation* submonad (§4, §5). We showed that these conditions apply to several examples of interest, including the Giry monad on measurable spaces, and the lower Vietoris monad on topological spaces (Theorem 5.6).

Acknowledgements. We would like to thank Tobias Fritz, Tomáš Gonda, and Dario Stein, as well as Sam Staton and all his research group, for the interesting and fruitful conversations and feedback. We also would like to thank the anonymous reviewers for their helpful remarks.

References

- [1] Samson Abramsky. 1991. Domain theory in logical form. *Annals of Pure and Applied Logic* 51, 1 (1991), 1–77. [https://doi.org/10.1016/0168-0072\(91\)90065-T](https://doi.org/10.1016/0168-0072(91)90065-T)
- [2] V. I. Bogachev. 2000. *Measure Theory. Vol. I, II*. Springer.
- [3] Anna Bucalo and Giuseppe Rosolini. 2014. Sobriety for equilogical spaces. *Theoretical Computer Science* 546 (2014), 93–98. <https://doi.org/10.1016/j.tcs.2014.03.002> Models of Interaction: Essays in Honour of Glynn Winskel.
- [4] Kenta Cho and Bart Jacobs. 2019. Disintegration and Bayesian Inversion via String Diagrams. *Math. Structures Comput. Sci.* 29 (2019), 938–971. Issue 7. [arXiv:1709.00322](https://arxiv.org/abs/1709.00322).
- [5] Bruno de Finetti. 1992. *Foresight: Its Logical Laws, Its Subjective Sources*. Springer New York, New York, NY, 134–174. https://doi.org/10.1007/978-1-4612-0919-5_10
- [6] Tobias Fritz. 2020. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Adv. Math.* 370 (2020), 107239. [arXiv:1908.07021](https://arxiv.org/abs/1908.07021).
- [7] Tobias Fritz, Tomáš Gonda, and Paolo Perrone. 2021. De Finetti’s Theorem in Categorical Probability. *Journal of Stochastic Analysis* 2, 4 (2021).
- [8] Tobias Fritz, Tomáš Gonda, Paolo Perrone, and Eigil Fjeldgren Rischel. 2020. Representable Markov Categories and Comparison of Statistical Experiments in Categorical Probability. *arXiv preprint arXiv:2010.07416* (2020). arxiv.org/abs/2010.07416.
- [9] T. Fritz and Wendong Liang. 2022. Free gs-monoidal categories and free Markov categories. [arXiv:2204.02284](https://arxiv.org/abs/2204.02284).
- [10] Tobias Fritz and Paolo Perrone. [n. d.]. Monads, partial evaluations, and rewriting. [arXiv:1810.06037](https://arxiv.org/abs/1810.06037). To appear in *Proceedings of MFPS*.
- [11] Tobias Fritz and Paolo Perrone. 2018. Bimonoidal structure of probability monads. In *Proceedings of the Thirty-Fourth Conference on the Mathematical Foundations of Programming Semantics (MFPS XXXIV) (Electron. Notes Theor. Comput. Sci., Vol. 341)*. Elsevier Sci. B. V., Amsterdam, 121–149. [arXiv:1804.03527](https://arxiv.org/abs/1804.03527).
- [12] Tobias Fritz, Paolo Perrone, and Sharwin Rezagholi. 2019. Probability, valuations, hyperspace: Three monads on Top and the support as a morphism. [arXiv:1910.03752](https://arxiv.org/abs/1910.03752).
- [13] Tobias Fritz and Eigil Fjeldgren Rischel. 2020. The zero-one laws of Kolmogorov and Hewitt–Savage in categorical probability. *Compositionality* 2 (2020), 3. compositionality-journal.org/papers/compositionality-2-3.
- [14] Carsten Führmann. 1999. Direct Models for the Computational Lambda Calculus. *Electr. Notes Theor. Comput. Sci.* 20 (1999), 245–292. [https://doi.org/10.1016/S1571-0661\(04\)80078-1](https://doi.org/10.1016/S1571-0661(04)80078-1)
- [15] Carsten Führmann. 2002. Varieties of Effects. In *Foundations of Software Science and Computation Structures*, Mogens Nielsen and Uffe Engberg (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 144–159.
- [16] Murdoch J. Gabbay and Andrew M. Pitts. 2002. A New Approach to Abstract Syntax with Variable Binding. *Form. Asp. Comput.* 13, 3–5 (jul 2002), 341–363. <https://doi.org/10.1007/s001650200016>
- [17] Fabio Gadducci. 1996. *On the algebraic approach to concurrent term rewriting*. Ph.D. Dissertation, University of Pisa.
- [18] Michèle Giry. 1982. A Categorical Approach to Probability Theory. In *Categorical aspects of topology and analysis*. Lecture Notes in Mathematics, Vol. 915.
- [19] Chris Heunen, Ohad Kammar, Sam Staton, and Hongseok Yang. 2017. A convenient category for higher-order probability theory. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. 1–12. <https://doi.org/10.1109/LICS.2017.8005137>
- [20] Jules Jacobs. 2021. Paradoxes of Probabilistic Programming: And How to Condition on Events of Measure Zero with Infinitesimal Probabilities. *Proc. ACM Program. Lang.* 5, POPL, Article 58 (jan 2021), 26 pages. <https://doi.org/10.1145/3434339>
- [21] Ohad Kammar and Gordon D. Plotkin. 2012. Algebraic Foundations for Effect-Dependent Optimisations. In *POPL ’12: Proceedings of the 39th annual ACM SIGPLAN-SIGACT symposium on Principles of Programming Languages*. 349–360.
- [22] Anders Kock. 2012. Commutative monads as a theory of distributions. *Theory Appl. Categ.* 26 (2012), No. 4, 97–131.
- [23] Paul Blain Levy, John Power, and Hayo Thielecke. 2003. Modelling environments in call-by-value programming languages. *Information and Computation* 185, 2 (2003), 182 – 210. [https://doi.org/10.1016/S0890-5401\(03\)00088-9](https://doi.org/10.1016/S0890-5401(03)00088-9)
- [24] Saunders Mac Lane. 1998. *Categories for the working mathematician* (second ed.). Graduate Texts in Mathematics, Vol. 5. Springer-Verlag, New York. xii+314 pages.
- [25] Eugenio Moggi. 1991. Notions of computation and monads. *Information and Computation* 93, 1 (1991), 55 – 92. [https://doi.org/10.1016/0890-5401\(91\)90052-4](https://doi.org/10.1016/0890-5401(91)90052-4) Selections from 1989 IEEE Symposium on Logic in Computer Science.
- [26] Prakash Panangaden. 2009. *Labelled Markov Processes*. Imperial College Press. <https://doi.org/10.1142/p595> [arXiv:https://www.worldscientific.com/doi/pdf/10.1142/p595](https://arxiv.org/abs/https://www.worldscientific.com/doi/pdf/10.1142/p595)
- [27] Paolo Perrone. 2018. *Categorical Probability and Stochastic Dominance in Metric Spaces*. Ph.D. Dissertation. University of Leipzig. Submitted. Available at <http://personal-homepages.mis.mpg.de/perrone/phdthesis.pdf>.
- [28] Giuseppe Rosolini. 2000. Equilogical spaces and filter spaces. Number 64. 157–175. *Categorical studies in Italy* (Perugia, 1997).
- [29] Marcin Sabok, Sam Staton, Dario Stein, and Michael Wolman. 2021. Probabilistic Programming Semantics for Name Generation. *Proc. ACM Program. Lang.* 5, POPL, Article 11 (jan 2021), 29 pages. <https://doi.org/10.1145/3434292>
- [30] Adam Ścibior, Ohad Kammar, Matthijs Vákár, Sam Staton, Hongseok Yang, Yufei Cai, Klaus Ostermann, Sean K. Moss, Chris Heunen, and Zoubin Ghahramani. 2018. Denotational validation of higher-order Bayesian inference. *PACMPL* 2, POPL (2018), 60:1–60:29. <https://doi.org/10.1145/3158148>
- [31] Ian Stark. 1996. Categorical Models for Local Names. *LISP Symb. Comput.* 9, 1 (1996), 77–107.
- [32] Sam Staton and Bart Jacobs. 2020. De Finetti’s Construction as a Categorical Limit. In *Proceedings of Coalgebraic Methods in Computer Science*, J. Petrisan, D.; Rot (Ed.). Cham : Springer.
- [33] D. Stein and S. Staton. 2021. Compositional Semantics for Probabilistic Programs with Exact Conditioning. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. IEEE Computer Society, Los Alamitos, CA, USA, 1–13. <https://doi.org/10.1109/LICS52264.2021.9470552>
- [34] Paul Taylor. 2002. Sober Spaces and Continuations. *Theory and Applications of Categories* 10, 12 (July 2002), 248–299. [PaulTaylor.EU/ASD/sobsc](https://paul.taylor.eu/ASD/sobsc)
- [35] Hayo Thielecke. 1997. Continuation Semantics and Self-adjointness. *Electronic Notes in Theoretical Computer Science* 6 (1997), 348–364. [https://doi.org/10.1016/S1571-0661\(05\)80149-5](https://doi.org/10.1016/S1571-0661(05)80149-5) MFPS XIII, Mathematical Foundations of Programming Semantics, Thirteenth Annual Conference.
- [36] Steven Vickers. 1989. *Topology via logic*. Cambridge Tracts in Theoretical Computer Science, Vol. 5. Cambridge University Press, Cambridge. xvi+200 pages.