HYPERNORMALISATION IN AN ABSTRACT SETTING

RICHARD GARNER

Abstract. Jacobs' *hypernormalisation* is a construction on finitely supported discrete probability distributions, obtained by generalising certain patterns occurring in quantitative information theory. In this paper, we generalise Jacobs' notion in turn, by describing a notion of hypernormalisation in the abstract setting of a symmetric monoidal category endowed with a linear exponential monad—a structure arising in the categorical semantics of linear logic.

We show that Jacobs' hypernormalisation arises in this fashion from the finitely supported probability measure monad on the category of sets, which can be seen as a linear exponential monad with respect to a non-standard monoidal structure on sets which we term the *convex monoidal structure*. We give the construction of this monoidal structure in terms of a quantum-algebraic notion known as a *tricocycloid*. Besides the motivating example, and its generalisations to the continuous context, we give a range of other instances of our abstract hypernormalisation, which swap out the side-effect of probabilistic choice for other important side-effects such as non-deterministic choice, ranked choice, and input from a stream of values.

1. Introduction

Hypernormalisation was introduced by Jacobs in [16] in order to provide, among other things, a smooth category-theoretic formulation of certain concepts [23] of quantitative information flow. The main theoretical contribution of this paper is to analyse, in turn, Jacobs' notion of hypernormalisation, showing how it arises naturally out of well-studied category-theoretic concepts, and how it generalises to other settings. The main application of this paper uses our framework in order to relate bisimilarity and trace equivalence for a range of coalgebraic generative systems. As is well-known, for such systems, bisimulation equivalence is a finer relation than trace equivalence, so that multiple different behaviours (i.e., states up to bisimilarity) may have the same underlying trace. We will describe a process of "normalisation-by-trace-evaluation" which normalises any given behaviour to a maximally-efficient one with the same trace.

To motivate our development, it will be useful to take a step back, and first describe the ideas of quantitative information flow which motivated [16]. These ideas are concerned with the following question: given a probabilistic process P which, when run on a *private* input returns a *public* output, how can we measure

Date: 2nd March 2023.

2000 *Mathematics Subject Classification.* Primary:

The support of Australian Research Council grants DP160101519, FT160100393 and DP190102432 is gratefully acknowledged. Thanks also to the referees, whose constructive and robust analyses have improved the paper immensely.

the leakage inherent in P, that is, the extent to which knowledge of the output allows an adversary to infer knowledge about the input?

The natural way of answering this question is information-theoretic: we model the input and output types as finite sets A and B, and P as a discrete channel, that is, a function assigning to each input in A a discrete probability distribution over possible outputs in B. Using this, any prior distribution π on A yields a joint distribution on $A \times B$ with associated marginals on A (viz. π) and B; we can now define the leakage of P as the "mutual information", i.e., the difference between the sum of the marginals' entropies and the entropy of the joint distribution; said another way, the leakage of P is the difference between the entropy of the prior distribution on A, and the expected entropy (over all possible observed outputs) of the posterior distribution on A.

Now, if Q is another process of input type A—but possibly different output type—then we can compare their drops in entropy to ascertain whether P or Q leaks more information about A. *Prima facie*, the answer to this depends not only on a choice of prior π for A, but also on the flavour of entropy chosen for the calculations—which has to do with questions like: must a successful attack guess the input precisely, or would partial knowledge suffice? A key contribution of [23] is to exhibit a leakage-ordering which is *robust*, in the sense of being largely independent of these choices—and moreover, depends not on the channels P and Q themselves, but only on their so-called *abstract channel* denotations.

According to [23], an *abstract channel* of type A is a function $\mathcal{D}A \to \mathcal{D}(\mathcal{D}A)$ from the set $\mathcal{D}A$ of probability distributions on A to the set of finitely supported probability distributions on $\mathcal{D}A$; said another way, it is a discrete time-homogeneous Markov chain with set of states $\mathcal{D}A$. In the case of the abstract channel P^r associated to a channel P from A to B, this Markov chain encodes the probabilities that a given prior distribution will update to a given posterior distribution on account of an observation in B. Crucially, however, the identity of these observations is suppressed, which is what makes the abstract channel "abstract".

Now in [23], the status of probabilistic channels as a kind of monadic computation [27] is clearly acknowledged; indeed, as is well known, the operation $A \mapsto \mathcal{D}A$ underlies the finitely supported discrete distribution monad D on the category of sets. However, the construction in [23] of an abstract channel P^r from a channel P does not exploit this fact. This is where Jacobs' [16] enters the picture; one of its objectives is to explain the construction $P \mapsto P^r$ via the calculus of monadic computation, so providing a framework for generalisation beyond the finite discrete case.

In Jacobs' analysis, the abstract channel P^r associated to a discrete channel $P\colon A \to \mathcal{D}B$ is found as a composite

$$\mathcal{D}A \xrightarrow{\tilde{P}} \mathcal{D}(A \times B) \xrightarrow{\mathcal{N}} \mathcal{D}(\mathcal{D}A \times B) \xrightarrow{\mathcal{D}(\pi_1)} \mathcal{D}(\mathcal{D}A)\ , \tag{1.1}$$

whose three terms we now explain. The last map is the action of the monad $\mathcal{D}$ on arrows by pushforward—which in this case takes a joint distribution on $\mathcal{D}A \times B$ to the associated marginal distribution on $\mathcal{D}A$. The first map, by contrast, takes a prior distribution π on A to the associated joint distribution on $A \times B$ given by $a, b \mapsto \pi(a) \cdot (Pa)(b)$. In categorical terms, this $\tilde{P}$ arises as the composite

$$\mathcal{D}A \xrightarrow{\mathcal{D}(1,P)} \mathcal{D}(A \times \mathcal{D}B) \xrightarrow{\mathsf{str}} \mathcal{D}\mathcal{D}(A \times B) \xrightarrow{\mu} \mathcal{D}(A \times B) \ , \tag{1.2}$$

where μ is the monad multiplication, and str is its cartesian *strength*.

The remaining part of (1.1) is the map $\mathcal{N}$, which is Jacobs' hypernormalisation. It can be described as follows. Given a joint distribution $\omega \in \mathcal{D}(A \times B)$, we have for each $b \in B$ the marginal probability $\omega(b) = \Sigma_a \omega(a, b)$; and when this is non-zero, we have also the conditional distribution $\omega_{A|b}$ on A with $\omega_{A|b}(a) = \omega(a, b)/\omega(b)$. Now $\mathcal{N}(\omega) \in \mathcal{D}(\mathcal{D}A \times B)$ is the distribution which takes the value $(\omega_{A|b}, b)$ with probability $\omega(b)$, for all $b \in B$ with $\omega(b) > 0$ (note that it is important for this that B is a *finite* set). In particular, $\mathcal{N}$ encodes the process of normalising a non-zero sub-probability distribution $\omega(-, b)$ on A to the probability distribution $\omega_{A|b}$—while avoiding the impossibility of normalising $\omega(-, b)$ when it is everywhere-zero. This explains the name *hypernormalisation* chosen for this map.

In [16], Jacobs introduces hypernormalisation by an element-based definition, and verifies by hand a number of desirable equational properties—the implication being that, to generalise away from the finite discrete setting, it would suffice to define a corresponding map $\mathcal{N}$, and to verify the corresponding properties. Our objective here is to replace this *axiomatic* approach with a *synthetic* one: rather than defining hypernormalisation on a case-by-case basis, we will show how it arises naturally from a certain well-known categorical framework. This will, in particular, allow us in a principled way to generalise hypernormalisation (and so also channel-abstraction) to diverse other settings.

The framework in question is that of a symmetric monoidal category endowed with a *linear exponential monad*. A linear exponential monad T is one for which the symmetric monoidal structure of the base category lifts to the category of T-algebras and there becomes finite coproduct. Linear exponential monads originate in the categorical semantics of linear logic [1], but also have applications in studying abstract differentiation in mathematics and computer science [5]. A key observation of this paper is that linear exponential monads *always* have an associated notion of hypernormalisation, satisfying all the equational axioms one may hope for.

The motivating example fits into this framework via the discrete distribution monad D on the category of sets. This turns out to be a linear exponential monad, but with respect to a non-standard monoidal structure on $\mathcal{S}\mathrm{et}$ which we term the *convex monoidal structure*. The convex monoidal structure has the empty set as unit and binary tensor given by

$$A \star B = A + \big((0, 1) \times A \times B\big) + B \tag{1.3}$$

where $(0, 1)$ denotes the open interval; while its associativity constraints are controlled by a map $v\colon (0, 1) \times (0, 1) \to (0, 1) \times (0, 1)$ which encodes a particular change of coordinates for points of the topological 2-simplex.

The unfamiliar aspect here is the monoidal structure (1.3); but it turns out that this can, in turn, be understood via another established piece of category theory. A *tricocycloid* [33] in a symmetric monoidal category is an object H endowed with an invertible map $v\colon H \otimes H \to H \otimes H$ satisfying suitable axioms;

and by a general construction of *loc. cit.*, any tricocycloid H in $\mathcal{C}$ gives a new monoidal structure on $\mathcal{C}$ defined by $A \star B = A + H \otimes A \otimes B + B$.

Typical examples of tricocycloids arise in the k-linear context from Hopf algebras [34] and multiplier Hopf algebras [37]; but their relevance here stems from the fact that $(0, 1)$ is a tricocycloid in the cartesian monoidal category of sets, so that we can derive the convex monoidal structure via the general construction described above. We refer to $(0, 1)$ endowed with the map v as the *convex tricocycloid*; it is a basic combinatorial object which lies at the heart of probability theory.

Beyond recapturing our motivating example, we also give a range of other examples of hypernormalisation arising from other linear exponential monads. One obvious direction of generalisation we pursue exhibits various *continuous* probability monads as linear exponential monads. However, and more interestingly, we also give a range of *non*-probabilistic examples. From the well-known perspective of [27, 29], monads encode computational effects, and the discrete distribution monad D in particular encodes (finite) probabilistic choice. We will see that the monads encoding other computational effects, including non-deterministic choice, ranked choice, logical choice over tests valued in a Boolean algebra, and input from a stream of B-values, also admit hypernormalisation.

We conclude this introduction with a brief overview of the contents of the paper. We begin Section 2 by recalling Jacobs' notion of hypernormalisation; as in (1.1), this is a certain map associated to the monad D for finitely supported probability distributions on the category of sets. The algebras for this monad are *abstract convex spaces*—the variety of algebras generated by the quasivariety of convex subsets of affine spaces (cf. [28]), and our first contribution is to explain how hypernormalisation can be understood in terms of finite coproducts in the category of abstract convex spaces. The key is the (well-known) observation that the binary coproduct of abstract convex spaces A and B is given by $A \star B$ as in (1.3), endowed with a suitable convex structure.

In fact, the results just described do not *quite* recapture hypernormalisation. In Section 3, we rectify this, and in doing so arrive at the key idea of this paper: that an appropriate general setting for hypernormalisation is a symmetric monoidal category endowed with a linear exponential monad. In this setting, we define a notion of hypernormalisation, and show that it inherits almost all of the good equational properties of hypernormalisation noted in [16]; we also show that the qualifier "almost" can be removed so long as the symmetric monoidal structure on the base category is *co-affine*—meaning that the unit object is initial—and the linear exponential monad T is *affine*—meaning that $T1 \cong 1$. We also show that, when $\mathcal{C}$ is a category with finite products and distributive finite coproducts, and T has a cartesian strength, we have a perfect analogue of the channel-to-abstract-channel construction following Jacobs' pattern (1.1).

In Section 4, we exhibit the motivating example of hypernormalisation as an instance of our general setting by showing that the convex monoidal structure (1.3) is indeed a symmetric monoidal structure on $\mathcal{S}et$ for which the discrete distribution monad D is a linear exponential monad. As discussed, we do this by first constructing the convex tricocycloid, and applying the general construction of [33]. Along the way, we fill out some aspects of the theory of tricocycloids, in particular relating them to *operads* in the sense of [22].

In Sections 5 and 6, we turn to examples of hypernormalisations beyond the motivating one. In Section 5, we exhibit the structure required to obtain hypernormalisation maps for three non-discrete probability monads: the expectation monad on sets [17]; the monad of Radon probability measures on compact Hausdorff spaces [25]; and the Kantorovich monad [36] on 1-bounded metric spaces. Then in Section 6, we turn to combinatorial examples, including the monad for (total finite) non-deterministic choice P_f^+; the monad for ranked choice, i.e., the non-empty list monad; a monad for "logical distributions" valued in a Boolean algebra B; and the monad for input from an alphabet B, i.e., the monad of B-ary branching trees. In particular, in this last example, hypernormalisation implements a normalisation algorithm for continuous functions defined on a coinductive type of streams.

2. Hypernormalisation and convex coproducts

2.1. **Hypernormalisation.** In this section, we first recall from [16] the notion of hypernormalisation for finitely supported discrete probability distributions, and then explain its relation to coproducts in the category of *abstract convex spaces*—the category of algebras for the discrete distribution monad.

Definition 1. A *finitely supported sub-probability distribution* on a set A is a function $\omega\colon A \to [0,1]$ such that $\mathrm{supp}(\omega)$ is finite and $\omega(A) \leqslant 1$. We call ω a *probability distribution* if $\omega(A) = 1$.

Here, we write $\mathrm{supp}(\omega)$ for the set $\{a \in A : \omega(a) > 0\}$ and, for any $B \subseteq A$, write $\omega(B)$ for $\sum_{b\in B}\omega(b)$. It will often be convenient to write a sub-probability distribution ω on A as a formal convex combination

$$\sum_{a\in\mathrm{supp}(\omega)} \omega(a)\cdot a \tag{2.1}$$

of elements of A; so, for example, $\omega\colon \{a,b,c,d\} \to [0,1]$ with $\omega(a) = \omega(c) = \frac{1}{3}$, $\omega(d) = \frac{1}{6}$ and $\omega(b) = 0$ could also be written as $\frac{1}{3}\cdot a + \frac{1}{3}\cdot c + \frac{1}{6}\cdot d$.

Definition 2. If ω is a sub-probability distribution on A such that $\omega(A) > 0$, then its *normalisation* is the probability distribution $\overline{\omega}$ with $\overline{\omega}(a) = \omega(a)/\omega(A)$.

Of course, if $\omega\colon A \to [0,1]$ is everywhere-zero, then we cannot normalise it. One way of understanding Jacobs' hypernormalisation [16] is as a principled way of avoiding this singularity. In the definition, and henceforth, we write $\mathcal{D}A$ for the set of probability distributions on a set A.

Definition 3. Let A be a set and $n \in \mathbb{N}$. The *n-ary hypernormalisation* function

$$\mathcal{N}\colon \mathcal{D}(\underbrace{A+\cdots+A}_{n}) \longrightarrow \mathcal{D}(\underbrace{\mathcal{D}A+\cdots+\mathcal{D}A}_{n})$$

is given as follows. For $1 \leqslant i \leqslant n$ and $a \in A$, we write (a,i) for the image of a under the ith coproduct injection $A \to A+\cdots+A$. Each $\omega \in \mathcal{D}(A+\cdots+A)$

yields n sub-probability distributions ω_i on A with $\omega_i(a) = \omega(a,i)$, and we take

$$\mathcal{N}(\omega) = \sum_{\substack{1 \leqslant i \leqslant n \\ \omega_i(A) > 0}} \omega_i(A) \cdot (\overline{\omega_i}, i) \ . \tag{2.2}$$

In other words $\mathcal{N}(\omega)$ "normalises the *non-zero* sub-probability distributions among $\omega_1, \dots, \omega_n$ and records the total weights". Note that this agrees with the description of hypernormalisation given in the introduction: the only difference is that there, we wrote B for the finite set $\{1, \dots, n\}$, and wrote $A \times B$ for $A + \dots + A$, so that hypernormalisation became a map $\mathcal{D}(A \times B) \to \mathcal{D}(\mathcal{D}A \times B)$.

As suggested in Section 8 of [16], one may generalise the hypernormalisation maps by replacing the n copies of A with n possibly distinct sets, yielding maps

$$\mathcal{N} \colon \mathcal{D}(A_1 + \dots + A_n) \to \mathcal{D}(\mathcal{D}A_1 + \dots + \mathcal{D}A_n) \tag{2.3}$$

defined in an entirely analogous manner to before. In this paper it will be this asymmetric version of hypernormalisation that we use. In fact, the key features of hypernormalisation are fully alive in the $n = 2$ case, and so in large part we will concentrate on the binary hypernormalisation maps

$$\mathcal{N} \colon \mathcal{D}(A + B) \to \mathcal{D}(\mathcal{D}A + \mathcal{D}B) \ . \tag{2.4}$$

Of particular note is the case of (2.4) where B is a singleton set $1 = \{*\}$, so that we have a map $\mathcal{N} \colon \mathcal{D}(A+1) \to \mathcal{D}(\mathcal{D}A + \mathcal{D}1) \cong \mathcal{D}(\mathcal{D}A + 1)$. An element of $\mathcal{D}(A+1)$ can be identified with a sub-probability distribution ω on A, with the one additional point $*$ necessarily being given the weight $1 - \omega(A)$; likewise, an element of $\mathcal{D}(\mathcal{D}A + 1)$ can be identified with a sub-probability distribution on $\mathcal{D}A$. Under these identifications, the action of $\mathcal{N}$ can be described as follows:

- If ω is the zero sub-probability distribution on A, then $\mathcal{N}(\omega)$ is the zero sub-probability distribution on $\mathcal{D}A$;
- Otherwise, $\mathcal{N}(\omega)$ is the sub-probability distribution on $\mathcal{D}A$ which assigns the weight $\omega(A)$ to the single point $\overline{\omega}$.

2.2. **Convex coproducts.** As explained in the introduction, hypernormalisation is closely bound up with the *discrete distribution monad* D on the category of sets. We now recall this monad, and explain how hypernormalisation is related to coproducts in the category of D-algebras.

Definition 4. The functor $\mathcal{D} \colon \mathcal{S}\mathrm{et} \to \mathcal{S}\mathrm{et}$ takes A to $\mathcal{D}A$ on objects; while on maps, $\mathcal{D}f \colon \mathcal{D}A \to \mathcal{D}B$ sends $\omega \in \mathcal{D}A$ to the *pushforward* $f_*(\omega) \in \mathcal{D}B$ given by

$$f_*(\omega)(b) = \omega(f^{-1}(b)) \ . \tag{2.5}$$

The unit $\eta \colon 1_{\mathcal{S}\mathrm{et}} \Rightarrow \mathcal{D}$ and multiplication $\mu \colon \mathcal{D}\mathcal{D} \Rightarrow \mathcal{D}$ of the discrete distribution monad D have respective components at a set A given by

$$\begin{aligned} \eta_A \colon A &\to \mathcal{D}A \\ a &\mapsto 1 \cdot a \end{aligned} \qquad\qquad \begin{aligned} \mu_A \colon \mathcal{D}\mathcal{D}A &\to \mathcal{D}A \\ \sum_{1 \leqslant i \leqslant n} \lambda_i \cdot \omega_i &\mapsto \big(a \mapsto \sum_{1 \leqslant i \leqslant n} \lambda_i \omega_i(a) \big) \ . \end{aligned}$$

(Note that, in giving μ_A, we have to the left a *formal* convex combination of elements of $\mathcal{D}A$, and to the right, an *actual* convex combination in $[0,1]$.)

In [16], the discrete distribution monad is discussed in terms of its Kleisli category; our interest here is in the algebras of this monad, which are sometimes known as *abstract convex spaces.*

Definition 5. An *abstract convex space* is a set A endowed with an operation

$$\begin{aligned}(0,1)\times A\times A&\to A\\ (r,a,b)&\mapsto r(a,b)\end{aligned} \tag{2.6}$$

satisfying the following axioms for all $a,b,c\in A$ and $r,s\in(0,1)$:

(i) $r(a,a)=a$;
(ii) $r(a,b)=r^*(b,a)$ (recall we write r^* for $1-r$);
(iii) $r(s(a,b),c)=(rs)(a,\frac{r\cdot s^*}{(rs)^*}(b,c))$.

A *map of convex spaces* from A to B is a function $f\colon A\to B$ such that $f(r(a,b))=r(fa,fb)$ for all $a,b\in A$ and $r\in(0,1)$.

If we view the operation (2.6) as an "abstract convex combination" $r(a,b)=r\cdot a+r^*\cdot b$, then the axioms are just what is needed to ensure that this behaves as expected. The two main classes of examples of abstract convex spaces are:

- Convex subsets of vector spaces (so convex spaces in the usual sense) under the usual convex combination operation; and
- Meet-semilattices under the operation $r(a,b)=a\wedge b$ for all $r\in(0,1)$.

If we extend the operation of an abstract convex space to one $[0,1]\times A\times A\to A$ by defining $1(a,b)=a$ and $0(a,b)=b$, then axioms (i)–(iii) are still validated for the new edge cases in $\{0,1\}$ wherever this makes sense (i.e., so long as $rs\neq 1$ in (iii)). This yields the axiomatisation of abstract convex spaces found in [28, §2, Axioms B1–B3]. In particular, these axioms ensure that each of the valid ways of interpreting a formal convex combination

$$\sum_{1\leqslant i\leqslant n} r_i\cdot a_i \quad\in\quad \mathcal{D}A \tag{2.7}$$

as an element of A via repeated application of the operation (2.6) will give the same result, so that we have a well-defined function $\mathcal{D}A\to A$. This function endows A with D-algebra structure, which is the key step in proving:

Lemma 6. *[14, Theorem 4]. The category* $\mathcal{C}\mathrm{onv}$ *of abstract convex spaces and convex maps is isomorphic over* $\mathcal{S}\mathrm{et}$ *to the category* $\mathcal{S}\mathrm{et}^{\mathsf{D}}$ *of* D*-algebras.*

This result justifies us in using expressions of the form (2.7) to denote an element of an abstract convex space A, and we do so without further comment.

The relation betwen abstract convex spaces and hypernormalisation lies in the construction of finite coproducts in $\mathcal{C}\mathrm{onv}$. While coproducts in algebraic categories are usually messy and syntactic, for abstract convex spaces they are quite intuitive. Given $A,B\in\mathcal{C}\mathrm{onv}$, their coproduct must certainly contain copies of A and B; and must also contain a formal convex combination $r\cdot a+r^*\cdot b$ for each $a\in A$, $b\in B$ and $r\in(0,1)$. For a general algebraic theory, this process of iteratively adjoining formal interpretations for operations would continue, but in this case, it stops here:

Lemma 7. *(cf. [18, §2]). If A and B are abstract convex spaces, then their coproduct $A \star B$ in $\mathcal{C}\mathrm{onv}$ is the set $A + ((0,1) \times A \times B) + B$, endowed with the convex combination operator whose most involved case is*

$$r_{A\star B}((s,a,b),(t,a',b')) = (rs + rt^*, (\tfrac{rs}{rs+r^*t})_A(a,a'), (\tfrac{rs^*}{(rs+r^*t)^*})_B(b,b')) \ .$$

This formula was obtained by expanding out the formal convex combination $r \cdot (s \cdot a + s^* \cdot b) + r^* \cdot (t \cdot a' + t^* \cdot b')$, rearranging, and partially evaluating the terms from A and from B. The reader should have no difficulty giving the remaining, simpler, cases (where one or both arguments of $r_{A\star B}$ come from A or B), and in then proving that the resulting object is an abstract convex space.

If we write elements $a \in A$, $(r,a,b) \in (0,1) \times A \times B$ and $b \in B$ in the three summands of $A \star B$ as, respectively,

$$\iota_1(a) \ , \qquad r \cdot a + r^* \cdot b \qquad \text{and} \qquad \iota_2(b) \ ,$$

then the two coproduct injections are given by $\iota_1 \colon A \to A \star B \leftarrow B \colon \iota_2$; and as for the universal property of coproduct, if $f \colon A \to C$ and $g \colon A \to C$ are convex maps, then the unique induced map $\langle f, g\rangle \colon A \star B \to C$ sends $\iota_1(a)$ or $\iota_2(b)$ to $f(a)$ or $g(b)$ respectively, and sends $r \cdot a + r^* \cdot b = r_{A\star B}(a,b)$ to $r \cdot f(a) + r^* \cdot g(b) = r_C(fa, gb)$.

To draw the link with hypernormalisation, consider the free-forgetful adjunction

$$\mathcal{C}\mathrm{onv} \underset{U^{\mathsf{D}}}{\overset{F^{\mathsf{D}}}{\underset{\longrightarrow}{\overset{\longleftarrow}{\bot}}}} \mathcal{S}\mathrm{et} \tag{2.8}$$

associated to the monad D. The left adjoint F^{D} sends the set A to the set $\mathcal{D}A$, seen as an abstract convex space under the convex combination operation given pointwise by the usual one on $[0,1]$. Being a left adjoint, F^{D} preserves coproducts, and so we have for any set A and any $n \in \mathbb{N}$ a bijection of abstract convex spaces

$$\varphi \colon \mathcal{D}(A+B) \to \mathcal{D}A \star \mathcal{D}B \ , \tag{2.9}$$

which, if we spell it out, we see is really just hypernormalisation:

Proposition 8. *The isomorphism (2.9) is given by*

$$\varphi(\omega) \quad = \quad \begin{cases} \iota_1(\omega_1) & \text{if } \omega_1(A) = 1; \\ \iota_2(\omega_2) & \text{if } \omega_2(B) = 1; \\ \omega_1(A) \cdot \overline{\omega_1} + \omega_2(B) \cdot \overline{\omega_2} & \text{otherwise,} \end{cases}$$

where ω_1 and ω_2 are the sub-probability distributions obtained by restricting ω to A and B.

Proof. φ is the extension of the composite function

$$A + B \xrightarrow{\eta+\eta} \mathcal{D}A + \mathcal{D}B \hookrightarrow \mathcal{D}A \star \mathcal{D}B \tag{2.10}$$

to a convex map $\mathcal{D}(A+B) \to \mathcal{D}A \star \mathcal{D}B$. Precomposing (2.10) with $\iota_1 \colon A \to A+B$ yields

$$A \xrightarrow{\eta} \mathcal{D}A \xrightarrow{\iota_1} \mathcal{D}A \star \mathcal{D}B$$

whence φ identifies $\mathcal{D}A \hookrightarrow \mathcal{D}(A+B)$ with the left coproduct summand of $\mathcal{D}A \star \mathcal{D}B$. This proves the first case of the desired formula; the second is similar.

Finally, consider $\omega \in \mathcal{D}(A+B)$ which does not factor through either $\mathcal{D}A$ or $\mathcal{D}B$. Since both sub-probability distributions ω_1 and ω_2 are non-zero, we can form both $\overline{\omega_1} \in \mathcal{D}A$ and $\overline{\omega_2} \in \mathcal{D}B$, and in these terms we now have

$$\omega = \omega_1(A) \cdot (\iota_1)_*(\overline{\omega_1}) + \omega_2(B) \cdot (\iota_2)_*(\overline{\omega_1})$$

in $\mathcal{D}(A+B)$. As φ is a convex map, it follows that $\varphi(\omega) = \omega_1(A) \cdot \overline{\omega_1} + \omega_2(B) \cdot \overline{\omega_2}$ from the two cases already proved. □

3. Abstract hypernormalisation

The map (2.9) of Proposition 8 is related to the hypernormalisation map (2.4), but is not quite the same. In this section, we explain how to derive the latter map from the former one, and isolate the general structure required for this derivation: that of a *linear exponential monad*. Guided by this, we define an abstract notion of hypernormalisation with respect to a linear exponential monad, show that it has the desired equational properties, and explain how it may give rise to a version of the channel-to-abstract channel construction from the introduction.

3.1. **Recapturing hypernormalisation.** Towards bridging the gap between (2.4) and (2.9), we observe that (2.4) is *not* a map of abstract convex spaces, so that to recapture it, we must necessarily leave the category $\mathcal{C}\mathrm{onv}$. We do so in an apparently simple-minded fashion, by considering the category $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$ whose objects are abstract convex spaces and whose maps are *arbitrary* functions.

Now, the binary coproduct $\star$ on $\mathcal{C}\mathrm{onv}$ is part of a symmetric monoidal structure, whose unit is the empty convex space, and whose coherence isomorphisms are induced from the universal properties of finite coproducts. This symmetric monoidal structure *extends* to $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$; by this we mean simply that $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$ has a symmetric monoidal structure with respect to which the inclusion $\mathcal{C}\mathrm{onv} \hookrightarrow \mathcal{C}\mathrm{onv}_{\mathrm{arb}}$ becomes symmetric strict monoidal. This monoidal structure is (necessarily) given on objects as before, while on maps $f \colon A \to A'$ and $g \colon B \to B'$ of $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$, the tensor $f \star g \colon A \star B \to A' \star B'$ is given by

$$f \star g = f + \big((0,1) \times f \times g\big) + g \ , \tag{3.1}$$

i.e., exactly the same formula as the definition of $\star$ on maps in $\mathcal{C}\mathrm{onv}$.

Now suppose we are given abstract convex spaces A and B. Using the extended monoidal structure on $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$, we obtain a function

$$A \star B \xrightarrow{\eta_A \star \eta_B} \mathcal{D}A \star \mathcal{D}B \xrightarrow{\varphi^{-1}} \mathcal{D}(A+B) \tag{3.2}$$

whose second part is the inverse of (2.9) and whose first part is the tensor (3.1) of the (non-convex) functions η_A and η_B. Working through the definitions, we see that this sends elements $\iota_1(a)$ and $\iota_2(b)$ of $A \star B$ to the distributions $1 \cdot a$ and $1 \cdot b$ on $A+B$ concentrated at a single point; while an element $r \cdot a + r^* \cdot b \in A \star B$ is sent to the two-point distribution $r \cdot a + r^* \cdot b$ on $A+B$. Combining this description of (3.2) with Proposition 8, we immediately obtain:

Proposition 9. *The hypernormalisation map* (2.4) *is the composite*

$$\mathcal{D}(A+B) \xrightarrow{\varphi} \mathcal{D}A \star \mathcal{D}B \xrightarrow{\eta_{\mathcal{D}A} \star \eta_{\mathcal{D}B}} \mathcal{D}\mathcal{D}A \star \mathcal{D}\mathcal{D}B \xrightarrow{\varphi^{-1}} \mathcal{D}(\mathcal{D}A + \mathcal{D}B) \ . \tag{3.3}$$

Thus, the hypernormalisation map (2.4) arises inevitably from the isomorphism (2.9) together with the fact that the coproduct monoidal structure on $\mathcal{C}\mathrm{onv}$ extends to $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$. We now give an explanation of why this extension of monoidal structure should exist.

To motivate this explanation, observe that the formula (3.1) for the extended tensor product on $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$ works because the underlying set of $A \star B$ depends only on the underlying sets of A and B, and not on their convex structure. So could the symmetric monoidal structure $\star$ on $\mathcal{C}\mathrm{onv}$ be a *lifting* of a symmetric monoidal structure on $\mathcal{S}\mathrm{et}$? In other words, is there a symmetric monoidal structure $(\star, 0)$ on $\mathcal{S}\mathrm{et}$—which as in the introduction we term the *convex monoidal structure*—such that $U^{\mathsf{D}} \colon (\mathcal{C}\mathrm{onv}, \star, 0) \to (\mathcal{S}\mathrm{et}, \star, 0)$ is strict symmetric monoidal?

In Section 4 below, we will see that this is indeed the case; for the moment, let us see how, assuming this fact, we can recover the symmetric monoidal structure of $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$. To do this, we consider the evident factorisation $\mathcal{C}\mathrm{onv} \to \mathcal{C}\mathrm{onv}_{\mathrm{arb}} \to \mathcal{S}\mathrm{et}$ of U^{D} through $\mathcal{C}\mathrm{onv}_{\mathrm{arb}}$, and apply the following result:

Lemma 10. *[30] Let $F \colon \mathcal{E} \to \mathcal{C}$ be a strict symmetric monoidal functor between symmetric monoidal categories, and let*

$$F = \mathcal{E} \xrightarrow{G} \mathcal{D} \xrightarrow{H} \mathcal{C}$$

be a factorisation of the underlying functor F wherein G is bijective on objects and H is fully faithful. There is a unique symmetric monoidal structure on $\mathcal{D}$ making both G and H strict symmetric monoidal.

Proof. Define the unit and the tensor on objects in $\mathcal{D}$ to be those of $\mathcal{E}$, and define the tensor on maps and the coherence morphisms to be those of $\mathcal{C}$. □

At this point, if we still take for granted the existence of the convex monoidal structure $(\star, 0)$ on $\mathcal{S}\mathrm{et}$, then one final category-theoretic transformation will allow us to derive the hypernormalisation maps (2.4) purely in terms of the structure of the discrete distribution monad. We begin by recalling:

Definition 11. A monad T on a symmetric monoidal category $(\mathcal{C}, \otimes, I)$ is *symmetric opmonoidal* if it comes endowed with a map $\upsilon_I \colon TI \to I$ and maps $\upsilon_{XY} \colon T(X \otimes Y) \to TX \otimes TY$ for $X, Y \in \mathcal{C}$, subject to seven coherence axioms; see, for example, [26, Section 7].

The relevance of this definition for us is captured by:

Lemma 12. *[26, Theorem 7.1]. For any monad T on a symmetric monoidal category $(\mathcal{C}, \otimes, I)$, symmetric opmonoidal monad structures on T correspond bijectively to liftings of the symmetric monoidal structure of $\mathcal{C}$ to $\mathcal{C}^{\mathsf{T}}$.*

Proof. Given symmetric opmonoidal structure on T, we define the lifted tensor product of T-algebras by

$$(TX \xrightarrow{x} X) \otimes (TY \xrightarrow{y} Y) \;=\; (T(X \otimes Y) \xrightarrow{\upsilon_{XY}} TX \otimes TY \xrightarrow{x \otimes y} X \otimes Y) \tag{3.4}$$

with as unit the T-algebra $\upsilon_I \colon TI \to I$. Conversely, given a lifted tensor product on T-algebras, we obtain the opmonoidal structure map υ_{XY} as the composite

$$T(X \otimes Y) \xrightarrow{T(\eta_X \otimes \eta_Y)} T(TX \otimes TY) \xrightarrow{\theta} TX \otimes TY$$

where θ is the T-algebra structure of $(\mu_X : TTX \to TX) \otimes (\mu_Y : TTY \to TY)$, and obtain $\upsilon_I : TI \to I$ as the T-algebra structure of the lifted unit. $\square$

Thus, the fact that the coproduct monoidal structure on $\mathcal{C}\mathrm{onv}$ lifts the (assumed) convex monoidal structure $(\star, 0)$ on $\mathcal{S}\mathrm{et}$ can be re-expressed by saying that the discrete distribution monad D on $\mathcal{S}\mathrm{et}$ is symmetric opmonoidal with respect to $(\star, 0)$. Actually, more is true: D is a *linear exponential monad*.

Definition 13. A *linear exponential monad* on a symmetric monoidal category $(\mathcal{C}, \otimes, I)$ is a symmetric opmonoidal monad T on $\mathcal{C}$ such that the lifted symmetric monoidal structure on the category of algebras $\mathcal{C}^{\mathsf{T}}$ is given by finite coproducts. More precisely, we means by this that the lifted unit object (I, ν_I) should be initial in $\mathcal{C}^{\mathsf{T}}$; and that, for any pair of algebras (X, x) and (Y, y), the cospan

$$(X, x) \xrightarrow{\rho_X} (X, x) \otimes (I, \nu_I) \xrightarrow{1 \otimes !} (X, x) \otimes (Y, y) \xleftarrow{! \otimes 1} (I, \nu_I) \otimes (Y, y) \xleftarrow{\lambda_Y} (Y, y)$$

should define a binary coproduct in $\mathcal{C}^{\mathsf{T}}$, where we use ! to denote the unique maps out of the initial object (I, ν_I).

Linear exponential *co*monads originate in the categorical semantics of linear logic [1, Definition 3] where they interpret the exponential modality which allows a resource to be copied freely. Importantly, the co-Kleisli category of a linear exponential comonad on a symmetric monoidal *closed* category is cartesian closed; this is a categorical formulation of the translation of intuitionistic logic into linear logic [12, §5]. Linear exponential comonads also arise in connection with [5]'s differential categories, which are categories endowed with an abstract notion of differentiation, encoded by a comonad which in many cases is linear exponential (note that in this context, the term "monoidal coalgebra modality" is often used rather than "linear exponential comonad").

The dual notion of linear exponential *monad* appears both in linear logic, where it models the de Morgan dual connective ? of !, and in the study of *co*differential categories, of which there are many natural examples; see, for example, [3]. Furthermore, as we will show in the next section, a linear exponential monad is *exactly* the structure one needs for an good abstract notion of hypernormalisation.

The salience of this last observation is not so much that it establishes a deep connection between probabilistic structures and linear logic, but rather that it makes available the well-understood calculus of reasoning for linear exponential (co)monads, as discussed in, for example, [24, Section 7] or [4]. As we will see, this allows to show that our abstract notion of hypernormalisation verifies all the equational axioms one could wish for.

3.2. **Abstract hypernormalisation.** Given a symmetric monoidal category $(\mathcal{C}, \otimes, I)$ with finite coproducts and a linear exponential monad T on $\mathcal{C}$, we continue to write $\varphi : T(A+B) \to TA \otimes TB$ for the map underlying the T-algebra isomorphism $F^{\mathsf{T}}(A + B) \to F^{\mathsf{T}}(A) \otimes F^{\mathsf{T}}(B)$; more generally, we write

$$\varphi : T(A_1 + \cdots + A_n) \to TA_1 \otimes \cdots \otimes TA_n \tag{3.5}$$

for the corresponding n-ary isomorphism. Note that these isomorphisms are natural in maps of $\mathcal{C}$, which is to say that all diagrams of the following form commute:

$$
\begin{array}{ccc}
T(A_1 + \cdots + A_n) & \xrightarrow{\varphi} & TA_1 \otimes \cdots \otimes TA_n \\
{\scriptstyle T(f_1+\cdots+f_n)}\downarrow & & \downarrow{\scriptstyle Tf_1\otimes\cdots\otimes Tf_n} \\
T(B_1 + \cdots + B_n) & \xrightarrow[\varphi]{} & TB_1 \otimes \cdots \otimes TB_n
\end{array} \ . \tag{3.6}
$$

Definition 14. Let $(\mathcal{C}, \otimes, I)$ be a symmetric monoidal category with finite coproducts, and let T be a linear exponential monad on $\mathcal{C}$. The *binary hypernormalisation map* $\mathcal{N}\colon T(A+B) \to T(TA+TB)$ is the composite

$$
T(A+B) \xrightarrow{\varphi} TA \otimes TB \xrightarrow{\eta_{TA}\otimes\eta_{TB}} TTA \otimes TTB \xrightarrow{\varphi^{-1}} T(TA+TB) \ .
$$

More generally, given objects $A_1, \dots, A_n$, the *n-ary hypernormalisation map* $\mathcal{N}\colon T(\Sigma_i A_i) \to T(\Sigma_i TA_i)$ is the composite

$$
T(\sum_i A_i) \xrightarrow{\varphi} \bigotimes_i TA_i \xrightarrow{\otimes_i \eta_{TA_i}} \bigotimes_i TTA_i \xrightarrow{\varphi^{-1}} T(\sum_i TA_i) \ . \tag{3.7}
$$

The leading example, as we confirm in Section 4, is Jacobs' original hypernormalisation, which arises on taking $(\mathcal{C}, \otimes, I)$ to be $(\mathcal{S}\mathrm{et}, \star, 0)$ and $\mathsf{T} = \mathsf{D}$. However, as we will see in Section 5, there are many other interesting examples of abstract hypernormalisation, including continuous probability monads on suitable categories, and examples related to continuous functions on streams.

Before turning to these examples, we investigate the degree to which our abstract hypernormalisation inherits the good equational properties of Jacobs' original definition. In the statement and proof of the following result, we write $\langle f_i \rangle_{i\in I}\colon \sum_i A_i \to B$ to denote a copairing of maps $f_i\colon A_i \to B$ out of a coproduct.

Proposition 15. *Let T be a linear exponential monad on the symmetric monoidal category $(\mathcal{C}, \otimes, I)$. The hypernormalisation maps (3.7) satisfy the conditions expressed by the commutativity of the following diagrams:*

(1) Hypernormalisation has a left inverse:

$$
\begin{array}{ccc}
T(\Sigma_i A_i) & \xrightarrow{\mathcal{N}} & T(\Sigma_i TA_i) \\
\| & & \downarrow{\scriptstyle T\langle T\iota_i\rangle_i} \\
T(\Sigma_i A_i) & \xleftarrow[\mu_{\Sigma_i A_i}]{} & TT(\Sigma_i A_i)
\end{array}
$$

(2) Hypernormalisation is idempotent:

$$
\begin{array}{ccc}
T(\Sigma_i A_i) & \xrightarrow{\mathcal{N}} & T(\Sigma_i TA_i) \\
{\scriptstyle \mathcal{N}}\downarrow & & \downarrow{\scriptstyle \mathcal{N}} \\
T(\Sigma_i TA_i) & \xrightarrow{T(\Sigma_i \eta_{TA_i})} & T(\Sigma_i TTA_i)
\end{array}
$$

(3) Hypernormalisation is natural in maps $f_i \colon A_i \to B_i$:

$$\begin{array}{ccc} T(\Sigma_i A_i) & \xrightarrow{\mathcal{N}} & T(\Sigma_i TA_i) \\ {\scriptstyle T(\Sigma_i f_i)}\downarrow & & \downarrow{\scriptstyle T(\Sigma_i Tf_i)} \\ T(\Sigma_i B_i) & \xrightarrow{\mathcal{N}} & T(\Sigma_i TB_i) \end{array}$$

and in Kleisli maps $f_i \colon A_i \to TB_i$:

$$\begin{array}{ccccc} T(\Sigma_i A_i) & \xrightarrow{\mathcal{N}} & T(\Sigma_i TA_i) & \xrightarrow{T(\Sigma_i Tf_i)} & T(\Sigma_i TTB_i) \\ {\scriptstyle T(\Sigma_i f_i)}\downarrow & & & & \downarrow{\scriptstyle T(\Sigma_i \mu_{B_i})} \\ T(\Sigma_i TB_i) & \xrightarrow{\mathcal{N}} & T(\Sigma_i T^2 B_i) & \xrightarrow{T(\Sigma_i \mu_{B_i})} & T(\Sigma_i TB_i) \end{array} .$$

Proof. To prove (1), we first claim that each diagram as to the left below commutes. This is a general fact about linear exponential monads—see, for example [24, Section 7]—but we include the proof for the sake of self-containedness.

$$(3.8)\qquad \begin{array}{ccc} T(\Sigma_i TA_i) & \xrightarrow{\varphi} & \otimes_i TTA_i \\ {\scriptstyle T\langle T\iota_i\rangle_i}\downarrow & & \\ TT(\Sigma_i A_i) & & \downarrow{\scriptstyle \otimes_i \mu_{A_i}} \\ {\scriptstyle \mu_{\Sigma_i A_i}}\downarrow & & \\ T(\Sigma_i A_i) & \xrightarrow{\varphi} & \otimes_i TA_i \end{array} \qquad \begin{array}{ccccc} TTA_i & & \xrightarrow{\jmath_i} & & \otimes_i TTA_i \\ {\scriptstyle TT\iota_i}\downarrow & \searrow^{\mu_{TA_i}} & & & \\ TT(\Sigma_i A_i) & & TA_i & & \downarrow{\scriptstyle \otimes_i \mu_{A_i}} \\ {\scriptstyle \mu_{\Sigma_i A_i}}\downarrow & \swarrow_{T\iota_i} & & \searrow^{\jmath_i} & \\ T(\Sigma_i A_i) & & \xrightarrow{\varphi} & & \otimes_i TA_i \end{array} .$$

Note that both paths are T-algebra maps $F^{\mathsf{T}}(\Sigma_i TA_i) \to \otimes_i F^{\mathsf{T}} A_i$ with as domain a coproduct of the T-algebras $F^{\mathsf{T}} A_i$. So it suffices to show commutativity on precomposing by a coproduct coprojection $T\iota_i \colon TTA_i \to T(\Sigma_i TA_i)$. This means showing the outside of the diagram to the right above commutes, wherein we write $\jmath_i$ for a coproduct coprojection $X_i \to \otimes_i X_i$ in the category of T-algebras. But the bottom triangle commutes by definition of φ, the left region by naturality of μ and the right region by naturality of the coproduct coprojections $\jmath$.

Now commutativity in (3.8) yields commutativity in the right part of:

$$\begin{array}{ccccccc} T(\Sigma_i A_i) & \xrightarrow{\mathcal{N}} & T(\Sigma_i TA_i) & \xrightarrow{T\langle T\iota_i\rangle_i} & TT(\Sigma_i A_i) & \xrightarrow{\mu_{\Sigma_i A_i}} & T(\Sigma_i A_i) \\ {\scriptstyle \varphi}\downarrow & & \uparrow{\scriptstyle \varphi^{-1}} & & & & \uparrow{\scriptstyle \varphi^{-1}} \\ \otimes_i TA_i & \xrightarrow{\otimes_i \eta_{TA_i}} & \otimes_i TTA_i & & \xrightarrow{\otimes_i \mu_{A_i}} & & \otimes_i TA_i \end{array}$$

whose left part commutes by definition of $\mathcal{N}$. So the outside commutes; now by the monad axioms for T, the lower composite is the identity, whence also the upper one as required for (1).

Turning to (2), we observe that pre-composing by φ^{-1} and post-composing by φ yields the square

$$\begin{array}{ccc} \otimes_i TA_i & \xrightarrow{\otimes_i \eta_{TA_i}} & \otimes_i T^2 A_i \\ {\scriptstyle \otimes_i \eta_{TA_i}}\downarrow & & \downarrow{\scriptstyle \otimes_i \eta_{T^2 A_i}} \\ \otimes_i T^2 A_i & \xrightarrow{\otimes_i T\eta_{TA_i}} & \otimes_i T^3 A_i \end{array}$$

which commutes by functoriality of $\otimes$ and naturality of η. Finally, for (3), commutativity of the first diagram is clear from the naturality (3.6) of the maps $\varphi \colon T(\Sigma_i A_i) \to \otimes_i TA_i$ in the A_i, the functoriality of $\otimes_i$, and the naturality of the unit $\eta \colon 1 \Rightarrow T$. Commutativity of the second diagram follows trivially from the first after postcomposing by $T(\Sigma_i \mu_{B_i})$. □

The preceding conditions generalise ones appearing in [16, Lemma 5]. Our (2) and (3) correspond exactly to its (3) and (5), while our (1) corresponds either to the right diagram of its (2) or to its (4). We have no correlate of the right diagram of part (1) of [16, Lemma 5], since it uses the canonical *strength* of the discrete distribution monad D with respect to the cartesian monoidal structure of $\mathcal{S}et$, and it is not clear what this should be replaced with in general. This leaves only the left diagrams appearing in (1) and (2) of [16, Lemma 5]. Interestingly, while these make sense in our setting, they do not hold without additional assumptions. For the left diagram of (2), this condition is:

Definition 16. A monad T on a category $\mathcal{C}$ with a terminal object 1 is *affine* if the unique map $T1 \to 1$ is invertible (necessarily with inverse $\eta_1 \colon 1 \to T1$).

Proposition 17. *Let T be an affine linear exponential monad on the symmetric monoidal category $(\mathcal{C}, \otimes, I)$ with terminal object 1. The hypernormalisation maps (3.7) satisfy the additional condition that:*

(4) Destroying the output structure destroys hypernormalisation:

$$\begin{array}{ccc} T(\Sigma_i A_i) & \xrightarrow{\quad \mathcal{N} \quad} & T(\Sigma_i TA_i) \\ & {\scriptstyle T(\Sigma_i !)}\searrow \quad \swarrow{\scriptstyle T(\Sigma_i !)} & \\ & T(\Sigma_i 1) & \end{array} .$$

Proof. We may precompose (4) by $\varphi^{-1} \colon \otimes_i TA_i \to T(\Sigma_i A_i)$, postcompose by $\varphi \colon T(\Sigma_i 1) \to \otimes_i T1$, and rewrite using the definition of $\mathcal{N}$ and the naturality (3.6) to obtain the triangle to the left in:

$$\begin{array}{ccc} \otimes_i TA_i & \xrightarrow{\otimes_i \eta_{TA_i}} & \otimes_i TTA_i \\ & {\scriptstyle \otimes_i T!}\searrow \quad \swarrow{\scriptstyle \otimes_i T!} & \\ & \otimes_i T1 & \end{array} \qquad \begin{array}{ccc} \otimes_i TA_i & \xrightarrow{\quad \otimes_i ! \quad} & \otimes_i 1 \\ & {\scriptstyle \otimes_i T!}\searrow \quad \swarrow{\scriptstyle \otimes_i \eta_1} & \\ & \otimes_i T1 & \end{array}$$

whose commutativity is equivalent to that of (4). But by naturality of η, this triangle is equally the triangle on the right above, which commutes since post-composing by the invertible map $\otimes_i ! \colon \otimes_i T1 \to \otimes_i 1$ yields along both sides the map $\otimes_i ! \colon \otimes_i TA_i \to \otimes_i 1$. ☐

Finally, we consider what is necessary for the left diagram of part (1) of [16, Lemma 5] to commute in our setting.

Definition 18. A symmetric monoidal category $(\mathcal{C}, \otimes, I)$ is said to be *co-affine* if its unit object is initial.

So, for example, the convex monoidal structure and the cocartesian monoidal structure on $\mathcal{S}\mathrm{et}$ are co-affine, while the cartesian monoidal structure is not so. The point of this extra condition is that it allows us to prove:

Lemma 19. *Let* T *be a linear exponential monad on the symmetric monoidal co-affine category* $(\mathcal{C}, \otimes, I)$*. Finite coproduct coprojections* $\jmath_i \colon X_i \to \otimes_i X_i$ *in the category of* T*-algebras are natural with respect to* arbitrary *maps of* $\mathcal{C}$*.*

Proof. Since non-empty finite coproducts can be constructed from binary ones, it suffices to prove the binary case. Given T-algebras (X, x) and (Y, y), we know from Definition 13 that the coproduct coprojection $(X, x) \to (X, x) \otimes (Y, y)$ is given by the composite

$$(X, x) \xrightarrow{\rho_X} (X, x) \otimes (I, \nu_I) \xrightarrow{1 \otimes !} (X, x) \otimes (Y, y) \ ,$$

where $! \colon (I, \nu_I) \to (Y, y)$ is the unique map of T-algebras induced by the initiality of (I, ν_I) in $\mathcal{C}^{\mathsf{T}}$. By co-affineness, the underlying map in $\mathcal{C}$ of this composite is

$$X \xrightarrow{\rho_X} X \otimes I \xrightarrow{1 \otimes !} X \otimes Y \ ,$$

where ! is the unique map out of the initial object $I \in \mathcal{C}$. Given this description, the desired naturality with respect to arbitrary maps of $\mathcal{C}$ is now immediate. ☐

Proposition 20. *Let* T *be a linear exponential monad on the symmetric monoidal co-affine category* $(\mathcal{C}, \otimes, I)$*. The hypernormalisation maps* (3.7) *satisfy the additional condition that:*

(5) Normalising trivial input gives trivial output:

$$\begin{array}{ccc} TA_i & \xrightarrow{T\iota_i} & T(\Sigma_i A_i) \\ {\scriptstyle \iota_i}\downarrow & & \downarrow{\scriptstyle \mathcal{N}} \\ \Sigma_i TA_i & \xrightarrow[\eta_{\Sigma_i TA_i}]{} & T(\Sigma_i TA_i) \ . \end{array}$$

Proof. By definition of $\mathcal{N}$ and naturality of η, this is equally to show that the diagram below left commutes. Since φ is the underlying map of the unique comparison between the coproducts $F^{\mathsf{T}}(\Sigma_i A_i)$ and $\otimes_i F^{\mathsf{T}} A_i$ in $\mathcal{C}^{\mathsf{T}}$, it in particular commutes with the coproduct coprojections, so that this diagram is equally the one below right, which commutes by Lemma 19.

$$\begin{array}{ccccc} TA_i & \xrightarrow{T\iota_i} & T(\Sigma_i A_i) & \xrightarrow{\varphi} & \otimes_i TA_i \\ {\scriptstyle \eta_{TA_i}}\downarrow & & & & \downarrow{\scriptstyle \otimes_i \eta_{TA_i}} \\ TTA_i & \xrightarrow{T\iota_i} & T(\Sigma_i TA_i) & \xrightarrow{\varphi} & \otimes_i TTA_i \end{array} \qquad \begin{array}{ccc} TA_i & \xrightarrow{\jmath_i} & \otimes_i TA_i \\ {\scriptstyle \eta_{TA_i}}\downarrow & & \downarrow{\scriptstyle \otimes_i \eta_{TA_i}} \\ TTA_i & \xrightarrow{\jmath_i} & \otimes_i TTA_i \end{array} \qquad \square$$

3.3. **Channel abstraction.** In this short section, we explain how, in our abstract setting, hypernormalisation can be used to build an analogue of the channel-to-abstract-channel construction of [23] described in the introduction. We are motivated in doing this by the examples of hypernormalisation for *continuous* probability monads described in Sections 5.1–5.3 below; thus, in what follows, the reader should keep in mind the interpretation that $\mathcal{C}$ is some category of "spaces", and that T is a monad of "distributions" on $\mathcal{C}$.

Let us recap what the construction should do. The input data, a channel, is simply a map $P\colon A \to TB$, which we think of as giving probabilities that a private input in A will give rise to a public output in B. The output data, the associated abstract channel, is a map $P^r\colon TA \to TTA$, thought of as giving the probabilities that a given prior distribution on A should update to a given posterior distribution on A via conditioning on an observed output in B.

Following Jacobs' lead, we will build P^r from P via a composite (1.1). The main difficulty comes in finding an analogue of the first map $\tilde{P}$ therein. This map, we recall, was itself a composite (1.2), one of whose terms is the canonical cartesian *strength* of the finite discrete distribution monad D. We already remarked in the previous section that it was not clear what to replace this with in general, so here we adopt the most simple-minded approach that is compatible with our examples: we simply assume that, again, our monad T has a cartesian strength.

This is enough to generalise the first map in (1.1); however, there is still a small problem with the second map, which should be a map $T(A \times B) \to T(TA \times B)$ given by hypernormalisation. In the motivating example, this was unproblematic: we could use the fact that B was a finite set $\{1, \dots, n\}$ to express $A \times B$ as an n-fold coproduct $A + \cdots + A$, and then apply n-ary hypernormalisation. In our general context, we can do something similar so long as finite coproducts distribute over finite products in $\mathcal{C}$, and we assume that B is an n-ary coproduct $1 + \cdots + 1$; then by distributivity we have $A \times B \cong A + \cdots + A$ and can proceed as before. The above discussion thus justifies giving:

Definition 21. Let $(\mathcal{C}, \otimes, I)$ be a symmetric monoidal category with finite products and distributive finite coproducts. Let T be a linear exponential monad on $\mathcal{C}$ endowed with a cartesian strength. Given a map $P\colon A \to TB$, where $B = 1 + \cdots + 1$ is an n-fold coproduct of the terminal object, we define the *associated abstract channel* to be the map $P^r\colon TA \to TTA$ given by

$$TA \xrightarrow{\tilde{P}} T(A \times B) \xrightarrow{\mathcal{N}} T(TA \times B) \xrightarrow{T(\pi_1)} TTA$$

wherein the first map is the composite

$$TA \xrightarrow{T(1,P)} T(A \times TB) \xrightarrow{\mathsf{T}(\mathsf{str})} TT(A \times B) \xrightarrow{\mu} T(A \times B) \ ,$$

and the second map is the composite of the n-ary hypernormalisation map $T(A+\cdots+A) \to T(TA+\cdots+TA)$ with isomorphisms $T(A\times B) \cong T(A+\cdots+A)$ and $T(TA+\cdots+TA) \cong T(TA\times B)$.

The restriction we impose on the form of B above is a real one; in our examples, it means that our channel $P\colon A \to TB$ involves a *continuous* space of hidden inputs but only a *finite discrete* space of observable outputs. While this is already progress, allowing an arbitrary observation space B would require something, more general than hypernormalisation, which gave a smooth categorical treatment of disintegration for probability measures on a product space.

4. Tricocycloids and the convex monoidal structure

In this section, we complete our description of the convex monoidal structure $(\star, 0)$ on $\mathcal{S}\mathrm{et}$ with respect to which the discrete distribution monad is linear exponential, and by doing so exhibit Jacobs' hypernormalisation as a particular instance of our abstract hypernormalisation.

The aspects of the convex monoidal structure we have not yet discussed are its unit, associativity and symmetry constraints. Given that it should lift to the coproduct monoidal structure on $\mathcal{C}\mathrm{onv}$, we can read off these constraints from the corresponding ones for coproducts in $\mathcal{C}\mathrm{onv}$. However, there is still work to do: we must show the maps involved can be defined in a way that does not depend on any convex structure, but only on the underlying sets.

We clarify the combinatorics involved in this by using a notion from quantum algebra known as a *tricocycloid*. We begin this section by explaining how tricocycloids give rise to monoidal structures, and how they relate to operads in the sense of [22]; we then exhibit a tricocycloid in $\mathcal{S}\mathrm{et}$ which will allow us to construct the desired convex monoidal structure.

4.1. **Tricocycloids.** Although our applications will primarily be in the category of sets, the construction we are about to give naturally exists in a more general setting. Rather than just the cartesian monoidal category of sets, it starts from a symmetric monoidal category $(\mathcal{C}, \otimes, I)$ with finite *distributive* coproducts—i.e., finite coproducts that are preserved by tensor in each variable. For simplicity, we write $\otimes$ as if it were *strictly* associative, and for brevity, we may denote tensor by mere juxtaposition. We can now ask: given an object $H \in \mathcal{C}$—which in the motivating case will be the set $(0, 1)$—under what circumstances is there a symmetric monoidal structure $(\star, 0)$ on $\mathcal{C}$ with unit the initial object, and tensor

$$A \star B := A \,+\, H \otimes A \otimes B \,+\, B \qquad ? \tag{4.1}$$

First let us consider what is necessary to get a *monoidal* structure. The unit constraints $A \star 0 \to A$ and $0 \star A \to A$ are easy; we have canonical isomorphisms

$$A + HA0 + 0 \xrightarrow{\cong} A + 0 + 0 \xrightarrow{\cong} A \quad 0 + H0B + B \xrightarrow{\cong} 0 + 0 + B \xrightarrow{\cong} B \tag{4.2}$$

using the preservation of the initial object by tensor on each side. The associativity constraint $(A \star B) \star C \to A \star (B \star C)$ is more interesting; it involves a map

$$(A+HAB+B)+H(A+HAB+B)C+C \to A+HA(B+HBC+C)+B+HBC+C$$

which, since tensor preserves binary coproducts in each variable, is equally a map

$$A{+}HAB{+}B{+}HAC{+}HHABC{+}HBC{+}C \;\longrightarrow\; A{+}HAB{+}HAHBC{+}HAC{+}B{+}HBC{+}C\ .$$

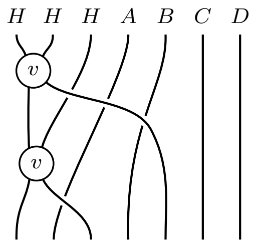
Now the coherence axioms relating the associativity and the unit constraints force this map to take the summands A, B, C, HAC, HAC, HBC of the domain to the corresponding summands of the codomain via identity maps, so leaving only the $HHABC$-summand of the domain unaccounted for. Though we are not forced to, it would be most natural to map this summand to the $HAHBC$-summand of the codomain via a composite

$$HHABC \xrightarrow{v111} HHABC \xrightarrow{1\sigma 11} HAHBC\ , \tag{4.3}$$

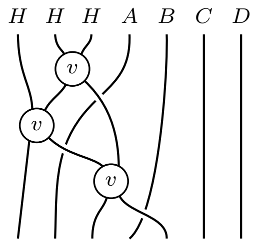
where here $v\colon HH \to HH$ is some fixed invertible map, and σ is the symmetry.

At this point, we have all the data of a monoidal structure, satisfying all the axioms except perhaps for the Mac Lane pentagon axiom, which equates the two arrows $((A \star B) \star C) \star D \rightrightarrows A \star (B \star (C \star D))$ constructible from the associativity constraint cells. If we expand out the definitions, we find that this equality is automatic on most summands of the domain; the only non-trivial case to be checked is the equality of the two morphisms $HHHABCD \rightrightarrows HAHBHCD$ given by the respective string diagrams (read from top-to-bottom):

(4.4)



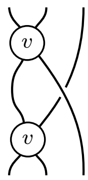
and



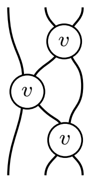
.

By examining the strings, a *sufficient* condition for this equality to hold is the equality of the two maps $HHH \rightrightarrows HHH$ represented by the string diagrams

(4.5)



and



;

and taking $A = B = C = D = I$ in (4.4) shows that this sufficient condition is also *necessary*. In fact, the structure of a map $v\colon HH \to HH$ rendering equal the strings in (4.5) has been studied before:

Definition 22. [33] Let $(\mathcal{C}, \otimes, I)$ be a symmetric monoidal category. A *tricocycloid* in $\mathcal{C}$ comprises an object $H \in \mathcal{C}$ and an invertible map $v\colon H \otimes H \to H \otimes H$ satisfying the equality

$$(v \otimes 1)(1 \otimes \sigma)(v \otimes 1) = (1 \otimes v)(v \otimes 1)(1 \otimes v)\colon H \otimes H \otimes H \to H \otimes H \otimes H\ . \tag{4.6}$$

The preceding argument shows:

Proposition 23. *Let $(\mathcal{C}, \otimes, I)$ be a symmetric monoidal category with finite distributive coproducts, let $H \in \mathcal{C}$ and let $v\colon H \otimes H \to H \otimes H$. The pair (H, v) is a* *tricocycloid if and only if there is a monoidal structure* $(\star, 0)$ *with* $\star$ *as in* (4.1)*, and with unit and associativity constraints as in* (4.2) *and* (4.3)*.*

We can think of the object H underlying a tricocycloid as parametrising "ways of non-trivially combining two things"; the map v then compares two ways in which $H \otimes H$ could parametrise "ways of non-trivially combining three things". This intuition may be clarified in terms of the notion of *operad*. Operads were introduced by May in [22] as a tool for describing certain kinds of topological-algebraic theory arising in homotopy theory, and involve objects of "n-ary operations" for each n, with suitable composition laws. The following notion of *pseudo-operad*, due to Markl, is concerned with the case where the objects of nullary and unary operations are trivial, and so can be omitted.

Definition 24. A *pseudo-operad* [21] in a symmetric monoidal category $\mathcal{C}$ is a sequence $(H_n)_{n\geqslant 2}$ of objects endowed with maps

$$\circ_i \colon H_n \otimes H_m \to H_{n+m-1} \qquad \text{for } n, m \geqslant 1 \text{ and } 1 \leqslant i \leqslant n \tag{4.7}$$

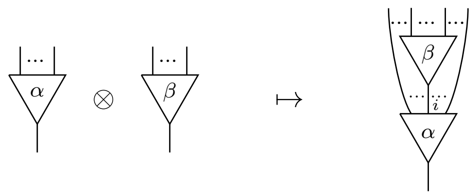
rendering commutative the following diagrams for $n, m, k \geqslant 1$ and $1 \leqslant i < j \leqslant n$:

$$\begin{array}{ccccc} H_n \otimes H_m \otimes H_k & \xrightarrow{1\otimes\sigma} & H_n \otimes H_k \otimes H_m & \xrightarrow{\circ_j \otimes 1} & H_{n+k-1} \otimes H_m \\ {\scriptstyle \circ_i \otimes 1}\downarrow & & & & \downarrow{\scriptstyle \circ_i} \\ H_{n+m-1} \otimes H_k & & \xrightarrow{\circ_{j+m-1}} & & H_{n+m+k-2} \end{array}$$

and the following diagrams for $n, m, k \geqslant 1$ and $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$:

$$\begin{array}{ccc} H_n \otimes H_m \otimes H_k & \xrightarrow{1\otimes\circ_j} & H_n \otimes H_{m+k-1} \\ {\scriptstyle \circ_i \otimes 1}\downarrow & & \downarrow{\scriptstyle \circ_i} \\ H_{n+m-1} \otimes H_k & \xrightarrow{\circ_{j+i-1}} & H_{n+m+k-2} \end{array} .$$

We think of the objects H_n involved in a pseudo-operad as parametrising "ways of non-trivially combining n things"; the maps $\circ_i$ then describe the way of combining $n + m - 1$ things induced by a way of combining n things and a way of combining m things, according to the following schema:



In general, there is no reason to expect the objects H_n parametrising n-ary combinations for $n \geqslant 3$ to be determined by the object H_2 of binary combinations; but when this *is* the case, we get a tricocycloid. This idea dates back to Day [6], is explained in detail in the introduction of [7], and is made precise by:

Lemma 25. *To give a tricocycloid in a symmetric monoidal category* $\mathcal{C}$ *is equally to give a pseudo-operad for which the maps* (4.7) *are all invertible.*

Proof (sketch). From a pseudo-operad H we obtain a tricocycloid with underlying object H_2 and with

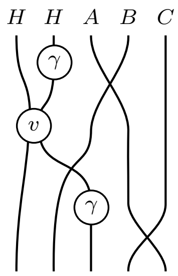
$$v = H_2 \otimes H_2 \xrightarrow{\circ_1} H_3 \xrightarrow{(\circ_2)^{-1}} H_2 \otimes H_2 \ . \tag{4.8}$$

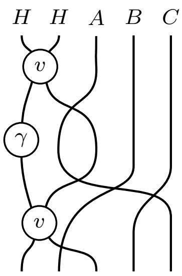
The tricocycloid axiom follows by constructing a suitable commutative diagram relating the various composition operations $H_2 \otimes H_2 \otimes H_2 \to H_4$, and using invertibility of the maps $\circ_i$. Conversely, from a tricocycloid (H, v) we construct a pseudo-operad with $H_n = H^{\otimes(n-1)}$, and with the maps $\circ_i$ given by suitable composites of v which we will not spell out in general; but let us at least say that, in the lowest dimension, we have $\circ_1, \circ_2 \colon H_2 \otimes H_2 \to H_3$ given by $v, \mathrm{id} \colon H \otimes H \to H \otimes H$ ☐

We now describe, following [33, Section 4], the additional structure on a tricocycloid needed to induce a symmetry on the associated monoidal structure. Such a symmetry is given by coherent isomorphisms $\sigma_{AB} \colon A \star B \to B \star A$, i.e., maps $A + HAB + B \to B + HBA + A$, and the coherence axiom relating σ with the unit constraints force the A- and B-summands of the domain to be mapped to the corresponding summands of the codomain. Like before, it is now natural to map the remaining HAB-summand to the HBA-summand via a composite

$$HAB \xrightarrow{\gamma 11} HAB \xrightarrow{1\sigma} HBA \tag{4.9}$$

for some fixed map $\gamma \colon H \to H$. Since a symmetry must satisfy $\sigma_{BA} \circ \sigma_{AB} = 1$, it follows that γ must be an involution (i.e., $\gamma^2 = 1$). As for the hexagon axiom relating the symmetry to the associativity, its only non-trivial case expresses the equality of the maps $HHABC \rightrightarrows HAHBC$ given by the respective diagrams:

(4.10) 

and 

.

Like before, it is necessary and sufficient for this that we should have equality of the diagrams obtained from (4.10) by deleting the A-, B- and C-strings; we encapsulate this requirement in:

Definition 26. Let (H, v) be a tricocycloid in the symmetric monoidal category $(\mathcal{C}, \otimes, I)$. A *symmetry* for H is an involution $\gamma \colon H \to H$ satisfying the equality

$$(1 \otimes \gamma)v(1 \otimes \gamma) = v(\gamma \otimes 1)v \colon H \otimes H \to H \otimes H \ . \tag{4.11}$$

The preceding argument thus shows:

Proposition 27. *Let $(\mathcal{C}, \otimes, I)$ be a symmetric monoidal category with finite distributive coproducts, and let (H, v) be a tricocycloid in $\mathcal{C}$. An involution $\gamma \colon H \to H$ is a symmetry for (H, v) just when the maps $\sigma_{AB} \colon A \star B \to B \star A$ determined by (4.9) endow the associated monoidal structure $(\star, 0)$ on $\mathcal{C}$ with a symmetry.*

A symmetry on a tricocycloid can also be described via the corresponding pseudo-operad. We call a pseudo-operad H in $\mathcal{C}$ *symmetric* if each H_n carries a symmetric group action $\alpha\colon S_n \to \mathcal{C}(H_n, H_n)$, with respect to which composition is equivariant. It is now straightforward to show that giving a symmetric tricocycloid is the same as giving a symmetric pseudo-operad with all maps (4.7) invertible.

4.2. **The convex monoidal structure.** Using the preceding theory, we can obtain the associativity and symmetry constraints of the desired convex monoidal structure $A, B \mapsto A + (0,1) \times A \times B + B$ on $\mathcal{S}\mathrm{et}$ by endowing the set $(0,1)$ with the structure of a symmetric tricocycloid.

This tricycloid is most easily understood by deriving it from a symmetric pseudo-operad. Indeed, for each $n \geqslant 2$ we may consider the set

$$H_n = \{(r_1, \dots, r_n) \in (0,1)^n : r_1 + \cdots + r_n = 1\}\ .$$

We now have maps $\circ_i\colon H_n \times H_m \to H_{n+m-1}$ defined by

$$((r_1, \dots, r_n), (s_1, \dots, s_m)) \mapsto (r_1, \dots, r_{i-1}, r_i s_1, \dots, r_i s_m, r_{i+1}, \dots, r_n) \tag{4.12}$$

and maps $\sigma\colon S_n \times H_n \to H_n$ defined by

$$(g, (r_1, \dots, r_n)) \mapsto (r_{g(1)}, \dots, r_{g(n)}) \tag{4.13}$$

which easily satisfy the axioms for a symmetric pseudo-operad. Moreover, each of the maps $\circ_i$ is invertible, with inverse

$$(t_1, \dots, t_{n+m-1}) \mapsto ((t_1, \dots, t_{i-1}, u, t_{i+m}, \dots, t_{n+m-1}), (\tfrac{t_i}{u}, \dots, \tfrac{t_{i+m-1}}{u})) \tag{4.14}$$

where here $u := \sum_{j=i}^{i+m-1} t_j$. So by Lemma 25, H_2 is a symmetric tricocycloid; transporting this structure along the isomorphism $H_2 \cong (0,1)$ given by $(r,s) \mapsto r$, we conclude that $(0,1)$ is a symmetric tricocycloid. The following result spells the structure out, and gives a direct proof of the symmetric tricocycloid axioms.

Proposition 28. *In the cartesian monoidal category of sets,* $(0,1)$ *is a symmetric tricocycloid, which we term the* convex tricocycloid*, under the operations*

$$\begin{aligned} v\colon (0,1)^2 &\mapsto (0,1)^2 \\ (r,s) &\mapsto (rs, \tfrac{r\cdot s^*}{(rs)^*}) \end{aligned} \qquad\qquad \begin{aligned} \gamma\colon (0,1) &\mapsto (0,1) \\ r &\mapsto r^*\ , \end{aligned}$$

where, as before, we write $r^* = 1 - r$ *for any* $r \in (0,1)$.

Proof. We begin by checking that $\big((0,1), v\big)$ is a tricocycloid. It is easy arithmetic to see that rs and $r \cdot s^*/(rs)^*$ are in $(0,1)$ whenever r and s are, so that v is well-defined. For the tricocycloid axiom, the function $(v \times 1)(1 \times \sigma)(v \times 1)\colon (0,1)^3 \to (0,1)^3$ is given by

$$(r,s,t) \mapsto (rs, \tfrac{r\cdot s^*}{(rs)^*}, t) \mapsto (rs, t, \tfrac{r\cdot s^*}{(rs)^*}) \mapsto (rst, \tfrac{rs\cdot t^*}{(rst)^*}, \tfrac{r\cdot s^*}{(rs)^*})\ ,$$

while the map $(1 \times v)(v \times 1)(1 \times v)\colon (0,1)^3 \to (0,1)^3$ is given by

$$(r,s,t) \mapsto (r, st, \tfrac{s\cdot t^*}{(st)^*}) \mapsto (rst, \tfrac{r(st)^*}{(rst)^*}, \tfrac{s\cdot t^*}{(st)^*}) \mapsto \Big(rst, \tfrac{rs\cdot t^*}{(rst)^*}, \big(\tfrac{r(st)^*}{(rst)^*}\big)\big(\tfrac{s\cdot t^*}{(st)^*}\big)^* \big/ \big(\tfrac{rs\cdot t^*}{(rst)^*}\big)^* \Big).$$

To see that the final terms agree, note first that for any $a, b \in (0,1)$ we have

$$\Big(\tfrac{a\cdot b^*}{(ab)^*}\Big)^* = 1 - \tfrac{a-ab}{1-ab} = \tfrac{1-a}{1-ab} = \tfrac{a^*}{(ab)^*} \tag{4.15}$$

so that the desired equality follows from the calculation

$$\Big(\tfrac{r(st)^*}{(rst)^*}\Big)\Big(\tfrac{s\cdot t^*}{(st)^*}\Big)^*\Big/\Big(\tfrac{rs\cdot t^*}{(rst)^*}\Big)^* \;=\; \Big(\tfrac{r(st)^*}{(rst)^*}\Big)\Big(\tfrac{s^*}{(st)^*}\Big)\Big/\Big(\tfrac{(rs)^*}{(rst)^*}\Big) \;=\; \tfrac{r\cdot s^*}{(rs)^*}\ .$$

We note also that v is invertible, with inverse $v^{-1}(t,u) = ((t^*\cdot u^*)^*, \frac{(t^*\cdot u)^*}{t^*\cdot u^*})$; that $v\circ v^{-1}$ and $v^{-1}\circ v$ are identities follows by a short calculation using (4.15).

We now show that γ provides a symmetry for the tricocycloid $((0,1),v)$. Clearly γ is an involution, so it remains to check the coherence axiom. The map $(1\times\gamma)v(1\times\gamma)\colon (0,1)^2 \to (0,1)^2$ is given by

$$(r,s)\mapsto(r,s^*)\mapsto(r\cdot s^*, \tfrac{r\cdot s^{**}}{(r\cdot s^*)^*})\mapsto\Big(r\cdot s^*, \Big(\tfrac{r\cdot s^{**}}{(r\cdot s^*)^*}\Big)^*\Big)$$

while $v(\gamma\times 1)v\colon (0,1)^2\to(0,1)^2$ is given by

$$(r,s)\mapsto(rs, \tfrac{r\cdot s^*}{(rs)^*})\mapsto((rs)^*, \tfrac{r\cdot s^*}{(rs)^*})\mapsto\Big(r\cdot s^*, (rs)^*\Big(\tfrac{r\cdot s^*}{(rs)^*}\Big)^*\Big/(r\cdot s^*)^*\Big)\ .$$

To check the equality of the second terms, we calculate using (4.15) twice that:

$$(rs)^*\Big(\tfrac{r\cdot s^*}{(rs)^*}\Big)^*\Big/(r\cdot s^*)^* = (rs)^*\tfrac{r^*}{(rs)^*}\Big/(r\cdot s^*)^* = \tfrac{r^*}{(r\cdot s^*)^*} = \Big(\tfrac{r\cdot s^{**}}{(r\cdot s^*)^*}\Big)^*\ . \qquad \square$$

Since the definition of v derives from the coproduct of abstract convex spaces, it is not unreasonable that the same coefficients should appear here as in the convex space axioms. We will see the deeper reason for this in Section 6.

Definition 29. The *convex monoidal structure* is the symmetric monoidal structure $(\star, 0)$ on the category of sets associated to the convex tricocycloid $((0,1),v,\gamma)$.

Working through the details of Lemma 7, the reader should have no difficulty in verifying that the forgetful functor $\mathcal{C}\mathrm{onv}\to\mathcal{S}\mathrm{et}$ is *strict* symmetric monoidal with respect to the coproduct monoidal structure on $\mathcal{C}\mathrm{onv}$ and the convex monoidal structure on $\mathcal{S}\mathrm{et}$. In light of Lemma 6 and Lemma 12, we have thus verified:

Proposition 30. *With respect to the convex monoidal structure on* $\mathcal{S}\mathrm{et}$*, the discrete distribution monad* D *is a linear exponential monad.*

It is now easy to check that the abstract hypernormalisation maps (3.7) reduce in this case to Jacobs' hypernormalisation maps (2.3), as desired.

5. Probabilistic examples

In the following two sections, we describe instances of abstract hypernormalisation which go beyond the motivating case, in which we replace the monad D on $\mathcal{S}\mathrm{et}$ by other monads on (possibly) other categories. It is perhaps worth emphasising that this will only work for rather special monads. Indeed, the key fact which drove our investigations above was Lemma 7, which showed that the underlying set of the coproduct of abstract convex spaces $A\star B$ depended only on the underlying sets of A and B, and not on the corresponding convex structure. In general, if T is a monad on a category $\mathcal{C}$, then in order for it to admit a notion of hypernormalisation, it must be the case that coproduct of T-algebras $\alpha\colon TA\to A$ and $\beta\colon TB\to B$ has the form $\alpha\star\beta\colon T(A\star B)\to A\star B$ wherein the $\mathcal{C}$-object $A\star B$ depends only on the $\mathcal{C}$-objects A and B, but *not* on the T-algebra structures α and β. This is not true, for example, for the exception monad $(-)+1$ on $\mathcal{S}\mathrm{et}$, whose algebras are pointed sets $(A, a\in A)$; indeed, the coproduct of two pointed sets (A, a) and (B, b) cannot be expressed just in terms of the underlying sets A and B, since we have to identify the points a and b.

As such, in order to verify that a monad T gives an example of our framework, we have no choice but to calculate *explicitly* the coproduct of T-algebras, so that we can check that it has the required special property. In this section, we do this for examples involving *continuous* probability monads; the hypernormalisation arising here is suitable for the channel-to-abstract-channel construction of Section 3.3 above, and so for non-discrete generalisations of the theory of [23].

5.1. **The expectation monad.** The next simplest probabilistic monad beyond the finite discrete case is the so-called *expectation monad* E on the category of sets. This was named and investigated in [17], but dates back to [35], where it was described, implicitly, as the monad generated by the chain of adjunctions:

$$\mathcal{KC}\mathrm{onv} \underset{\longrightarrow}{\overset{\longleftarrow}{\bot}} \mathcal{KH} \underset{\longrightarrow}{\overset{\longleftarrow}{\bot}} \mathcal{S}\mathrm{et} \ . \tag{5.1}$$

Here, $\mathcal{KH}$ is the category of compact Hausdorff spaces; while $\mathcal{KC}\mathrm{onv}$ is the category whose objects are compact convex subsets A of locally convex vector spaces, and whose morphisms are continuous affine maps (where affineness is the condition $f(ra + r^*a') = rf(a) + r^*f(a')$.) The two right adjoints in (5.1) are the obvious forgetful functors, while the two left adjoints send, respectively, a set X to its space of ultrafilters βX with the Stone topology, and a compact Hausdorff space Y to its space of Radon probability measures, identified via the Riesz representation theorem with the positive elements of norm 1 in the ordered Banach space of continuous linear functionals $C(Y, \mathbb{R}) \to \mathbb{R}$.

As explained in [17], the monad E induced by the composite adjunction (5.1) can be described in various ways; the most direct is as follows. We write $\mathcal{E}X$ for the set of *charges* $\omega \colon \mathcal{P}X \to [0, 1]$, i.e., functions such that $\omega(X) = 1$ and $\omega(A \cup B) = \omega(A) + \omega(B)$ whenever $A, B \subseteq X$ are disjoint. The action of $\mathcal{E}$ on morphisms is given by pushforward, $(\mathcal{E}f)(\omega)(B) = \omega(f^{-1}(B))$; the monad unit $\eta_X \colon X \to \mathcal{E}X$ takes $x \in X$ to the Dirac distribution with $\eta_X(x)(A) = 1$ if $x \in A$ and $\eta_X(x)(A) = 0$ otherwise; while the monad multiplication is given by a suitable notion of integration against a valuation:

$$\mu_X(\omega)(A) = \int_{\tau \in \mathcal{E}X} \tau(A) \, \mathrm{d}\omega \ .$$

The details may be found in [17]; however, we will not need them to describe hypernormalisation for E. Rather, we need only make E into a linear exponential monad, which we can do from the perspective of the category of algebras using:

Proposition 31. *[35, Theorem 4] The composite adjunction in* (5.1) *is monadic.*

Thus E-algebras can be identified with compact convex subsets of locally convex vector spaces; and so if we can understand finite coproducts of these, we can obtain the desired linear exponential structure on E.

Proposition 32. *If $A \subseteq V$ and $B \subseteq W$ are objects of $\mathcal{KC}\mathrm{onv}$, then their coproduct may be given as*

$$\{(ra, r^*b, r) : r \in [0, 1], a \in A, b \in B\} \subseteq V \oplus W \oplus \mathbb{R} \ . \tag{5.2}$$

Proof. This is [31, Proposition 7]. □

At the level of underlying sets, the coproduct of A and B in (5.2) is given by
$$\{(a,0,1) : a \in A\} + \{(ra, r^*b, r) : r \in (0,1), a \in A, b \in B\} + \{(0,b,0) : b \in B\}$$
which is clearly isomorphic to $A + (0,1) \times A \times B + B = A \star B$. In a similar way, the coherence constraints for the coproduct in $\mathcal{KC}\mathrm{onv}$ lift the coherence constraints for the convex monoidal structure on $\mathcal{S}\mathrm{et}$, and so we have:

Proposition 33. *The expectation monad on* $\mathcal{S}\mathrm{et}$ *is a linear exponential monad with respect to the convex monoidal structure.*

It follows that the expectation monad admits a notion of hypernormalisation. To calculate this, we first define, like before, the *normalisation* of a valuation $\omega\colon \mathcal{P}A \to [0,1]$ with $\omega(A) > 0$ to be the normalised valuation $\overline{\omega}$ with $\overline{\omega}(U) = \omega(U)/\omega(A)$. Noting that each set $\mathcal{E}A$ admits a structure of abstract convex space where $r(\omega_1,\omega_2)(U) = r\omega_1(U) + r^*\omega_2(U)$, we may now describe the hypernormalisation map $\mathcal{N}\colon \mathcal{E}(\Sigma_i A_i) \to \mathcal{E}(\Sigma_i \mathcal{E}A_i)$ as in (2.2) by
$$\mathcal{N}(\omega) = \sum_{\substack{1\leqslant i\leqslant n\\ \omega_i(A_i)>0}} \omega_i(A_i)\cdot \iota_i(\overline{\omega_i})\ .$$

5.2. **The Radon monad.** The expectation monad E on $\mathcal{S}\mathrm{et}$ arose from the composite adjunction in (5.1); on the other hand, the left-hand adjunction in (5.1) induces the *Radon monad* R on $\mathcal{KH}$. Again, this was introduced implicitly in [35], with the details now being provided by [25].

Definition 34. Let X be a compact Hausdorff space. A *Radon probability measure* on X is a probability measure $\omega\colon \Sigma_X \to [0,1]$ on the Borel σ-algebra of X such that $\omega(M) = \sup\{\omega(K) : K \subseteq M, K \text{ compact}\}$ for all $M \in \Sigma_X$. We write $\mathcal{R}(X)$ for the space of Radon probability measures on X with the weak topology: the coarsest topology such that, for each continuous $f\colon X \to \mathbb{R}$, the integration map $\omega \mapsto \int f \mathrm{d}\omega$ is continuous as a function $\mathcal{R}(X) \to \mathbb{R}$.

The remaining aspects of the Radon monad R on $\mathcal{KH}$ are much as before: the action on morphisms is by pushforward, the monad unit selects the Dirac valuations, and the multiplication is given by integration against a measure. To obtain our notion of hypernormalisation, we will again exploit monadicity, using:

Proposition 35. *([35, Theorem 4]) The left-hand adjunction in* (5.1) *is monadic.*

So, identifying the category of R-algebras with the category $\mathcal{KC}\mathrm{onv}$, it only remains to relate the coproduct (5.2) of compact convex spaces with a suitable monoidal structure on $\mathcal{KH}$. This will be the well-known *topological join*:

Definition 36. The *join* of two topological spaces X and Y is the quotient space of the product space $[0,1] \times X \times Y$ under the smallest equivalence relation $\sim$ for which $(0,x,y) \sim (0,x',y)$ and $(1,x,y) \sim (1,x,y')$ for all $x,x' \in X$ and $y,y' \in Y$.

We can realise the topological join $X \star Y$ as the set $X + (0,1) \times X \times Y + Y$, with a basis for the topology generated by sets of three forms
$$U + (0,a) \times U \times Y + \emptyset \qquad \emptyset + (a,b) \times U \times V + \emptyset \qquad \emptyset + (b,1) \times X \times V + V$$

for all rationals $0 < a < b < 1$ and all $U \subseteq X$, $V \subseteq Y$ open. Presented in this way, it is easy to see that topological join is part of a monoidal structure $(\star, 0)$ on $\mathcal{KH}$ which lifts the convex monoidal structure on $\mathcal{S}\mathrm{et}$. On the other hand, comparing with the formula (5.2), we conclude that the forgetful functor $\mathcal{KC}\mathrm{onv} \to \mathcal{KH}$ sends coproduct to topological join, and so we obtain:

Proposition 37. *The Radon monad on $\mathcal{KH}$ is a linear exponential monad with respect to the join monoidal structure $(\star, 0)$ on $\mathcal{KH}$.*

We thus obtain hypernormalisation maps $\mathcal{N}\colon \mathcal{R}(\Sigma_i A_i) \to \mathcal{R}(\Sigma_i \mathcal{R} A_i)$ whose action on a distribution ω is obtained in much the same way as previously. The key point is that we get continuity of $\mathcal{N}$ for free: something which otherwise would have required some fairly messy calculation.

5.3. **The Kantorovich monad.** For our final example of probabilistic hypernormalisation, we consider the *Kantorovich monad* [36] on the category $\mathcal{C}\mathrm{Met}_1$ whose objects are complete metric spaces which are 1-bounded (i.e., $d(x, y) \leqslant 1$ for all x, y) and whose morphisms are 1-Lipschitz mappings (i.e., $d(fx, fy) \leqslant d(x, y)$ for all x, y).

Definition 38. Let X be a complete 1-bounded metric space. $\mathcal{K}(X)$ is the complete metric space whose elements are Radon probability measures on X, under the Kantorovich (or "earth-mover's") metric:

$$d(\omega, \pi) = \inf\{ \textstyle\int d_X(x, y)\, \mathrm{d}\mu(x, y) : \mu \in \mathcal{K}(X \times X), (\pi_1)_*(\mu) = \omega, (\pi_2)_*(\mu) = \pi \}$$

where the infimum is over joint distributions on $X \times X$ with marginals ω and π.

This operation underlies a monad on $\mathcal{C}\mathrm{Met}_1$ following the established pattern; and to exhibit a notion of hypernormalisation, we also follow the established pattern, by investigating coproducts of K-algebras. We begin with the characterisation of these algebras.

Definition 39. A *convex metric space* is a metric space X which is also a convex space in the sense of Definition 5, subject to the compatibility condition

$$d(r(x, z), r(y, z)) \leqslant r d(x, y) \quad \text{(or equally, } d(r(x, y), r(x, z)) \leqslant r^* d(y, z)\ ) . \tag{5.3}$$

We write $\mathcal{CC}\mathrm{onv}\mathcal{M}\mathrm{et}_1$ for the category of 1-bounded complete convex metric spaces, and convex 1-Lipschitz maps.

Proposition 40. *The category of K-algebras is isomorphic to $\mathcal{CC}\mathrm{onv}\mathcal{M}\mathrm{et}_1$ over $\mathcal{C}\mathrm{Met}_1$.*

Proof. This is [10, Theorem 5.2.1] (though see also [20, Theorem 10.9]). □

And now we characterise finite coproducts in this category.

Proposition 41. *The coproduct of $X, Y \in \mathcal{CC}\mathrm{onv}\mathcal{M}\mathrm{et}_1$ is the coproduct $X \star Y$ of underlying convex spaces, endowed with the metric:*

$$d_{X\star Y}(r \cdot x + r^* \cdot y, s \cdot w + s^* \cdot z) = r d(x, w) + (s - r) + (1 - s) d(y, z) \tag{5.4}$$

for any $0 \leqslant r \leqslant s \leqslant 1$; here, by convention, we allow $1 \cdot x + 0 \cdot y$ to denote $x \in X \subseteq X \star Y$, and correspondingly for $0 \cdot x + 1 \cdot y$.

Proof. It is a straightfoward calculation to show that this is indeed a complete, 1-bounded metric satisfying (5.3); indeed, it is easy to see that $d_{X\star Y}$ is really just the Kantorovich metric restricted to distributions concentrated at two points.

To exhibit the universal property of coproduct, let $f\colon X \to Z$ and $g\colon Y \to Z$ be maps in $\mathcal{CConvMet}_1$, and let $\langle f, g\rangle\colon X \star Y \to Z$ be the induced unique map of convex spaces as in Section 2.2. It suffices to show that $\langle f, g\rangle$ is 1-Lipschitz. For the most involved case, consider elements $r \cdot x + r^* \cdot y$ and $s \cdot w + s^* \cdot z$ in $X \star Y$ with $0 < r \leqslant s < 1$. We must show that

$$d\big(r(fx, gy), s(fw, gz)\big) \leqslant rd(x, w) + (s - r) + (1 - s)d(y, z) \ .$$

Now, the left-hand side is by the triangle inequality smaller than

$$d\big(r(fx, gy), r(fw, gy)\big) + d\big(r(fw, gy), s(fw, gy)\big) + d\big(s(fw, gy), s(fw, gz)\big)$$

and we calculate that $d\big(r(fx, gy), r(fw, gy)\big) \leqslant rd(fx, fw) \leqslant rd(x, w)$ and $d\big(s(fw, gy), s(fw, gz)\big) \leqslant s^*d(gy, gz) \leqslant s^*d(y, z)$ using (5.3) and contractivity of f and g. So it suffices to show that $d\big(r(fw, gy), s(fw, gy)\big) \leqslant s-r$ in Z. Writing $u = fw$ and $v = gy$, this follows by the calculation

$$\begin{aligned} d\big(r(u, v), s(u, v)\big) &= d\big(s(\tfrac{r}{s}(u, v), v), s(u, v)\big) \leqslant sd\big(\tfrac{r}{s}(u, v), u\big) \\ &= sd\big(\tfrac{s-r}{s}(v, u), \tfrac{s-r}{s}(u, u)\big) \leqslant (s - r)d(v, u) \leqslant s - r \ . \quad \square \end{aligned}$$

We are thus in the familiar situation that the $\mathcal{CMet}_1$-object underlying the coproduct of $X, Y \in \mathcal{CConvMet}_1$ depends only on the underlying $\mathcal{CMet}_1$-objects of X and Y. Thus we have:

Definition 42. The *join* of two 1-bounded complete metric spaces X and Y is the set $X \star Y = X + (0, 1) \times X \times Y + Y$ endowed with the metric (5.4). This provides the binary tensor of the *convex monoidal structure* $(\star, 0)$ on $\mathcal{CMet}_1$, whose remaining data is all lifted from the convex monoidal structure on $\mathcal{Set}$.

Proposition 43. *The Kantorovich monad on* $\mathcal{CMet}_1$ *is a linear exponential monad with respect to the join monoidal structure* $(\star, 0)$ *on* $\mathcal{CMet}_1$.

As such, we have a good notion of hypernormalisation for the Kantorovich monad. Once again, the maps $\mathcal{N}\colon \mathcal{K}(\Sigma_i A_i) \to \mathcal{K}(\Sigma_i \mathcal{K} A_i)$ are defined in the expected way—but we obtain with no extra work the fact that this is a contractive mapping, as required.

6. Combinatorial examples

In this section, we describe a range of further instances of our framework of a more combinatorial flavour, including examples such as the non-empty finite power-set monad (encoding the side-effect of finite total non-determinism) and the non-empty list monad (encoding the side-effect of ranked choice). One important situation that we fail to capture in the obvious way is that where finite probability distributions on a set are replaced by “logical distributions”—convex combinations whose coefficients are drawn not from $[0, 1]$ but from a given Boolean algebra B; however, we will see that we *can* capture this example by instantiating our framework in a category other than the category of sets. Finally in this section, we consider an example in which hypernormalisation implements an extensional collapse for computable functions on a type of streams.

6.1. **Other tricocycloids.** As noted above, our next examples of abstract hypernormalisation will arise from other tricocycloids in the category of sets. To motivate the manner in which this will happen, we first explain how we can derive the finite discrete distribution monad D from the convex tricocycloid.

First of all, the convex tricocycloid induces the convex monoidal structure $(\star, 0)$, and so, as with any symmetric monoidal structure, we can consider the commutative $\star$-monoids. It is easy to see that these are *almost* abstract convex spaces: they are sets A endowed with an operation $(0,1) \times A \to A \to A$ satisfying axioms (ii) and (iii), but not necessarily (i) from Definition 5. The missing axiom (i) is the idempotency condition $r(a,a) = a$, which we can re-find as follows.

Definition 44. A *monoidal diagonal* for a symmetric monoidal category $(\mathcal{C}, \otimes, I)$ is a monoidal natural transformation $\delta \colon \mathrm{id} \Rightarrow \otimes \circ \Delta \colon \mathcal{C} \to \mathcal{C}$; this comprises a natural family of maps $\delta_A \colon A \to A \otimes A$ rendering commutative each diagram:

$$\begin{array}{ccc} & A \otimes B & \\ {\scriptstyle \delta_A \otimes \delta_B} \swarrow & & \searrow {\scriptstyle \delta_{A \otimes B}} \\ A \otimes A \otimes B \otimes B & \xrightarrow{1 \otimes \sigma \otimes 1} & A \otimes B \otimes A \otimes B \end{array} \tag{6.1}$$

A commutative $\otimes$-monoid (A, m, e) in $\mathcal{C}$ is *idempotent* if, whenever δ is a monoidal diagonal for $(\mathcal{C}, \otimes, I)$, we have $m \circ \delta_A = \mathrm{id}_A \colon A \to A$.

When $(\mathcal{C}, \otimes, I)$ is $(\mathcal{S}\mathrm{et}, \times, 1)$, the only monoidal diagonal is the usual diagonal $(\mathrm{id}, \mathrm{id}) \colon A \to A \times A$, and so idempotency in the above sense coincides with idempotency in the usual sense. On the other hand:

Lemma 45. *A commutative monoid in* $(\mathcal{S}\mathrm{et}, \star, 0)$ *is idempotent just when it satisfies axiom (i) as well as axioms (ii) and (iii) in Definition 5—in other words, just when it is an abstract convex space.*

Proof. By considering naturality with respect to maps $1 \to A$, we see that the possible monoidal diagonals $\delta_r \colon \mathrm{id} \Rightarrow \star \circ \Delta$ are indexed by $r \in [0,1]$; we have that δ_0, δ_1 are the left and right coproduct coprojections $A \to A + \big((0,1) \times A \times A\big) + A$, and that $\delta_r(a) = (r, a, a)$ for $0 < r < 1$. It follows that a commutative $\star$-monoid is idempotent just when it satisfies the additional axiom $r(a,a) = a$ for all $0 < r < 1$—that is, just when it is an abstract convex space. □

So from the convex tricocycloid, we can obtain abstract convex spaces as the the idempotent commutative monoids for the associated monoidal structure $(\star, 0)$, and obtain the monad D as the free idempotent commutative $\star$-monoid monad. More generally, for any symmetric tricocycloid H with structure maps

$$\begin{aligned} v \colon H^2 &\mapsto H^2 \\ (r,s) &\mapsto (rs, r \diamond s) \end{aligned} \qquad\qquad \begin{aligned} \gamma \colon H &\mapsto H \\ r &\mapsto r^* \ , \end{aligned}$$

we can consider idempotent commutative monoids for the associated monoidal structure $\star_H$; these will be sets A equipped with an operation $H \times A \times A \to A$, written like before as $r, a, b \mapsto r(a,b)$, that satisfies

$$r(a,a) = a \qquad r(a,b) = r^*(b,a) \qquad r(s(a,b),c) = (rs)\big(a, (r \diamond s)(b,c)\big) \ .$$

Since this is clearly algebraic structure, we obtain a monad D_H on $\mathcal{S}\mathrm{et}$ whose algebras are these "H-convex sets". In fact, this monad is *automatically* linear exponential for the $\star_H$-monoidal structure, and so admits a notion of hypernormalisation. To justify this last claim, we first use the well known result of [9] that the tensor product on any symmetric monoidal category $(\mathcal{C}, \otimes, I)$ lifts to the category of commutative monoids, and there becomes coproduct; so free commutative monoid monads are always linear exponential. Because we consider *monoidal* diagonals, the idempotent commutative monoids are always closed under this lifted tensor product, so that free *idempotent* commutative monoid monads are also linear exponential; in particular, any monad D_H obtained as above is linear exponential, as claimed.

We now illustrate this with some other examples of symmetric tricocycloids in $\mathcal{S}\mathrm{et}$. In giving these examples, we can exploit the fact that D_H is linear exponential to calculate the action of the monads D_H. Indeed, it is clear that $\mathcal{D}_H(1) \cong 1$, whence for any *finite* set $n \cong 1 + \cdots + 1$ we have $\mathcal{D}_H(n) \cong 1 \star_H \cdots \star_H 1$; and to extend to *infinite* sets, we note that the theory of H-convex sets is finitary, so that $\mathcal{D}_H(A)$ is the directed colimit of the sets $\mathcal{D}_H(n)$ for all finite $n \subseteq A$.

Example 46. Consider the one-element symmetric tricocycloid 1. In this case, the induced monoidal structure is given by $A \star_1 B = A + A \times B + B$, which it is also fruitful to think of as

$$A \star_1 B = \Big((A + \{\bot\}) \times (B + \{\bot\}) \Big) \setminus (\bot, \bot) \ . \tag{6.2}$$

In this case, the idempotent commutative $\star_1$-monoids are *join-semilattices* (possibly without bottom element), the associated monad is the non-empty finite powerset monad P^+_f, and hypernormalisation $\mathcal{N} \colon \mathcal{P}^+_f(A + B) \to \mathcal{P}^+_f(\mathcal{P}^+_f A + \mathcal{P}^+_f B)$ is given by

$$\begin{aligned} \{a_1, \dots, a_n\} &\mapsto \{\{a_1, \dots, a_n\}\} \\ \{b_1, \dots, b_m\} &\mapsto \{\{b_1, \dots, b_m\}\} \\ \{a_1, \dots, a_n, b_1, \dots, b_m\} &\mapsto \{\{a_1, \dots, a_n\}, \{b_1, \dots, b_m\}\} \ . \end{aligned} \tag{6.3}$$

Example 47. Let X be a topological space, and let H be the set of continuous functions $X \to (0,1)$ with the tricocycloid structure given pointwise as in $(0,1)$. In this case, a typical example of a H-convex set is the set of global sections of a sheaf of vector spaces on X; while the action of the monad D_H is given by

$$\mathcal{D}_H(A) = \{S \subseteq A \text{ finite}, \omega \colon X \to \Delta_S \text{ continuous}\}$$

where Δ_S is the interior of the standard topological $|S|$-simplex, given by a singleton when $|S| = 1$, and by $\{r \in (0,1)^S : \sum_{x \in S} r(s) = 1\}$ otherwise. In other words, the elements of $\mathcal{D}_H(A)$ are those of A, together with all non-trivial "finite convex combinations"

$$\sum_{1 \leqslant i \leqslant n} f_i \cdot a_i$$

where each $a_i \in A$, and where the f_i's are continuous maps $X \to (0,1)$ satisfying $\sum_i f_i = 1$ (i.e., constituting a partition of unity). In this case, the notion of hypernormalisation carries over *mutatis mutandis* from the motivating case, where arithmetic on the coefficients f_i is done pointwise in $(0,1)$.

Before continuing, we take a slight detour in order to deepen our understanding of tricocycloids. As we have just seen, we can obtain the discrete distribution monad from the tricocycloid $(0,1)$. However, as shown in [15], we may also obtain it in an apparently different way: from the *effect monoid* structure on $[0,1]$.

Definition 48. (cf. [8]) A *partial commutative monoid* is a set M with a constant 0 and partial binary operation $\ovee\colon M \times M \rightharpoonup M$ satisfying the axioms:

$$r \ovee 0 \simeq r \simeq 0 \ovee r \qquad (r \ovee s) \ovee t \simeq r \ovee (s \ovee t) \qquad r \ovee s \simeq s \ovee r$$

where $\simeq$ denotes *Kleene equality* of partially defined functions, i.e., one side is defined just when the other is, and they are then equal. M is an *effect algebra* if it is equipped with a constant 1 and (total) unary operation $(-)^\perp$ such that:

(i) For all $r \in M$, the element $r^\perp$ is unique such that $r \ovee r^\perp \simeq 1$;
(ii) If $r \ovee 1$ is defined, then $r = 0$.

An effect algebra is an *effect monoid* if it comes equipped with a (total) binary operation $r, s \mapsto r \cdot s$ which is associative and has unit 1, and which distributes over $\ovee$; i.e., we have equalities

$$r \cdot 0 = 0 = 0 \cdot r \qquad r \cdot (s \ovee t) \simeq' (r \cdot s) \ovee (r \cdot t) \qquad (r \ovee s) \cdot t \simeq' (r \cdot s) \ovee (r \cdot t)$$

where $\simeq'$ means "if the left-hand side is defined, then so is the right-hand side, and they are equal".

Examples 49. (i) $[0,1]$ is an effect monoid where $0,1$ have their usual meanings; $\cdot$ is ordinary multiplication; $r \ovee s$ is given by $r+s$ if this sum lies in $[0,1]$, and is undefined otherwise; and where $r^\perp = 1 - r$.

(ii) Any Boolean algebra is an effect monoid, where $0,1$ are $\bot, \top$, where $\cdot$ is intersection $\wedge$, where $r \ovee s$ is given by $r \vee s$ if $r \wedge s = \bot$, and is undefined otherwise; and where $r^\perp$ is the complement of r.

As shown in [15], any effect monoid M induces a monad D_M on $\mathcal{S}\mathrm{et}$ whose action on objects is given by

$$\mathcal{D}_M(A) = \{\, \omega\colon A \to M \mid \mathrm{supp}(\omega) \text{ finite and } \textstyle\bigovee_{a \in A} \omega(a) \simeq 1 \,\} \tag{6.4}$$

and whose remaining structure is defined by analogy with the discrete distribution monad; in particular, if $M = [0,1]$, we get the discrete distribution monad itself.

Remark 50. It is natural to ask how these two constructions of D—from the tricocycloid $(0,1)$ and the effect monoid $[0,1]$—relate to each other. More generally, we may ask whether the monad D_M associated to an effect monoid M also arises as the monad D_H associated to some tricocycloid, or vice versa. This question was investigated in detail by Kaddar [19]. One of his main results is that the assignments

$$M \mapsto M \setminus \{0,1\} \qquad \text{and} \qquad H \mapsto H \amalg \{0,1\}$$

give a bijection between tricocycloids satisfying certain cancellativity properties, and effect monoids with *normalisation*, i.e., effect monoids such that for all a, b with $a \otimes b$ defined and $a \neq 1$, there is a unique c with $b = a^{\perp}c$. In particular, this construction relates the tricocycloid $(0, 1)$ and the effect monoid $[0, 1]$.

An intuitive way of understanding this correspondence is by way of the notion of pseudo-operad from Definition 24. Any effect monoid M gives a symmetric pseudo-operad with underlying sets

$$H_n = \{(m_1, \dots, m_n) \in (M \setminus \{0, 1\})^n : m_1 \otimes \cdots \otimes m_n \simeq 1\} \ , \tag{6.5}$$

and structure maps $\circ_i$ and σ defined as in (4.12) and (4.13). By Lemma 25, this pseudo-operad yields a tricocycloid just when each map $\circ_i \colon H_n \times H_m \to H_{n+m-1}$ is invertible: which is *exactly* the condition that M admit normalisation.

Although this is not proven in detail in [19], it flows from the above understanding that the monad D_M of an effect monoid with normalisation coincides with the monad D_H of the corresponding tricocycloid; the point is that the n-ary "M-convex combination operations" for D_M can via normalisation be decomposed into composites of *binary* convex combinations.

6.2. **Ranked choice.** In Example 46, we discussed hypernormalisation for the non-empty finite powerset monad P_f^+, which encodes total, finite non-deterministic choice. An obvious variant on this side-effect is what one might call *ranked choice*, encoded not by P_f^+, but the monad of non-empty lists.

Definition 51. Given a set X, we write X^+ for the set of non-empty lists of elements of X. The assignment $X \mapsto X^+$ underlies a monad on $\mathcal{S}\mathrm{et}$, whose unit $X \to X^+$ sends x to the singleton list (x), and whose multiplication is given by concatenation of lists.

Our goal in this short section is to show that this monad admits hypernormalisation, and to describe it explicitly. Let us first observe that the algebras for the monad $(–)^+$ are *semigroups*: sets endowed with an associative binary operation. The following lemma characterises finite coproducts of semigroups.

Lemma 52. *The initial semigroup is the empty set* 0. *Given semigroups M and N, their coproduct in the category of semigroups is*

$$M \star N = \{(x_1, \dots, x_n) \in (M + N)^+ : x_i \in M \iff x_{i+1} \in N\} \ , \tag{6.6}$$

i.e., the set of non-empty lists which alternate between elements of M and N, with operation

$$(x_1, \dots, x_n)\cdot(y_1, \dots, y_m) = \begin{cases} (x_1, \dots, x_n y_1, \dots, y_m) & \text{if } x_n, y_1 \in M \text{ or } x_n, y_1 \in N \\ (x_1, \dots, x_n, y_1, \dots, y_m) & \text{otherwise.} \end{cases}$$

The coprojections $N \to M * N \leftarrow M$ send an element $n \in N$ (resp. $m \in M$) to the singleton list (n) (resp., (m)).

The definition (6.6) does not rely on the semigroup structure of M and N, only on their underlying set, and in much the same way, the associativity, unitality and symmetry constraints of the coproduct monoidal structure on semigroups do not rely on the semigroup structures involved. We thus have:

Proposition 53. *The operation* $\star$ *of* (6.6) *underlies a monoidal structure* $(\star, \mathbf{0})$ *on* $\mathcal{S}\mathrm{et}$*, with respect to which the non-empty list monad* $(\text{–})^+$ *is linear exponential.*

The associated hypernormalisation maps $(A+B)^+ \to (A^+ + B^+)^+$ implement the basic equality $(a+b)^* = a^*(ba^*)^*$ of Kleene algebra: they take a non-empty list of $A+B$'s, and subdivide it into non-empty lists drawn exclusively from either A or B. More explicitly, if for $(x_1, \dots, x_n) \in (A+B)^+$ we let $1 \leqslant i_1 < \cdots < i_k < n$ enumerate those indices i for which x_i and x_{i+1} lie in different summands of $A+B$, then $\mathcal{N}((x_1, \dots, x_n))$ is given by

$$((x_1, \dots, x_{i_1}), (x_{i_1+1}, \dots, x_{i_2}), \dots, (x_{i_{k-1}+1}, \dots, x_{i_k}), (x_{i_k+1}, \dots, x_n)) \ .$$

6.3. **Logical hypernormalisation.** So far we have said nothing about the second example of an effect monoid from Examples 49, that of a Boolean algebra. One might hope this example to be associated to some kind of "logical hypernormalisation". As we will now see, this is not directly possible; however, it does become possible if we change our perspective appropriately.

Example 54. Let B be a non-trivial Boolean algebra; we may view B as an effect monoid as in Examples 49, and so may form the associated monad D_B as in (6.4), with action on objects given by:

$$\mathcal{D}_B(X) = \{\omega \colon X \to B \mid \omega \text{ has finite support and } \mathrm{im}(\omega) \text{ is a partition of } B\} \ .$$

The Eilenberg–Moore algebras for D_B admit a characterisation via binary operations due to Bergman [2, Theorem 14]: they are sets A endowed with an operation $B \times A \times A \to A$, written $r, a, b \mapsto r(a, b)$ as usual, satisfying the axioms:

$$\text{(i) } r(a,a) = a \qquad \text{(ii) } r(a,b) = r^\perp(b,a) \qquad \text{(iii) } 0(a,b) = b$$
$$\text{(iv) } r(r(a,b),c) = r(a,c) \qquad \text{(iv) } r(s(a,b),b) = (rs)(a,b) \ .$$

But in fact, in the presence of (i)–(iii), axioms (iv)–(v) may be replaced by:

$$\text{(iv)}' \ r(s(a,b),c) = (sr)(a, r(b,c)) \ .$$

The easier direction is that taking $b = c$ in (iv)′ yields (v), while taking $s = r^\perp$ and using (iii) yields (iv). Conversely, if we assume (iv) and (v), then we obtain (iv)′ by the calculation

$$\begin{aligned}(sr)(a, r(b,c)) &= s(r(a, r(b,c)), r(b,c)) = s(r(a,c), r(b,c)) \\ &= r(s(a,b), s(c,c)) = r(s(a,b), c)\end{aligned}$$

using, in turn: (v); the equality $r(a, r(b,c)) = r(a,c)$ obtained from (iv) and (ii); the non-trivial equality $r(s(w,x), s(y,z)) = s(r(w,y), r(x,z))$ of [2, Proposition 11]; and (i). Thus, if we write $r(a,b)$ as $a +_r b$, then this alternate axiomisation becomes:

$$a +_r a = a \quad a +_r b = b +_{r^\perp} a \quad (a +_s b) +_r c = a +_{sr} (b +_r c) \quad a +_0 b = b$$

which are the axioms U1–U3, U7 for the operation of *guarded union* in the theory of guarded Kleene algebra with tests [32].

However, the monad D_B does not admit a notion of "logical hypernormalisation". Indeed, consider an element $\omega \in \mathsf{D}_B(X+Y)$. We have the (complementary) elements $\omega(X) = \bigvee_{x\in X} \omega(x)$ and $\omega(Y) = \bigvee_{y \in Y} \omega(y)$ of B, and from this may attempt to produce an element $\mathcal{N}(\omega) \in \mathsf{D}_B(\mathsf{D}_B X + \mathsf{D}_B Y)$. We focus on the interesting case where neither $\omega(X)$ nor $\omega(Y)$ are $\bot$; here, writing ω_X and ω_Y for the restriction of ω to X and Y, we would like to take

$$\mathcal{N}(\omega) = \omega(X) \cdot \overline{\omega_X} + \omega(Y) \cdot \overline{\omega_Y} \ .$$

The issue is that there is no obvious meaning to be assigned to $\overline{\omega_X}$ or $\overline{\omega_Y}$. Indeed, ω_X represents a sum $\sum_i r_i \cdot x_i$ of elements of X weighted by disjoint elements $r_i \in B$ of total weight $\omega(X)$, and there is no sensible way of distributing the missing weight $\omega(Y)$ among the r_i's to obtain a normalised distribution $\overline{\omega_X}$.

While this is disappointing, it turns out that if we change our perspective slightly, we *can* get a notion of logical hypernormalisation; we now describe this, though since the answer we get is perhaps less satisfying than one would like, we omit some of the finer details. To motivate our new perspective, note that elements of $\mathsf{D}_B(X)$ can be seen as total computations that switch on an element of the Boolean algebra B to return an element in X. However, it also make sense to consider *partial* computations, like ω_X and ω_Y above, where one branch of the switch is left undefined. We can encode this situation by considering not sets, but rather *B-labelled sets*.

Definition 55. Let B be a non-trivial Boolean algebra. A *B-labelled set* is a pair $\mathbf{X} = (X, |{-}|_{\mathbf{X}})$ where X is a set and $|{-}|_{\mathbf{X}} \colon X \to B \setminus \{\bot\}$; we call $|x|_{\mathbf{X}}$ the *(logical) mass* of x. We write $\mathcal{S}\mathrm{et}_B$ for the category of B-labelled sets and mass-preserving functions.

Definition 56. A *finitely supported logical subdistribution* on a B-labelled set $\mathbf{X}$ is a function $\omega \colon X \to B$ such that $\mathsf{supp}(\omega)$ is non-empty and finite, the image of ω is pairwise-disjoint, and $\omega(x) \leqslant |x|_{\mathbf{X}}$ for all x. The set of logical subdistributions on $\mathbf{X}$ becomes a B-labelled set $\mathrm{s}\mathcal{D}_B(\mathbf{X})$ on defining the logical mass of ω to be $\omega(X)$, where like before we write $\omega(A) = \bigvee_{a \in A} \omega(a)$ for any $A \subseteq X$.

The assignment $\mathbf{X} \mapsto \mathrm{s}\mathcal{D}_B(\mathbf{X})$ now underlies a monad $\mathrm{s}\mathsf{D}_B$ on $\mathcal{S}\mathrm{et}_B$, whose action on objects is given by pushforward as in (2.5), and whose unit and multiplication are given by:

$$\begin{aligned} \eta_{\mathbf{X}} \colon \mathbf{X} &\to \mathrm{s}\mathcal{D}_B(\mathbf{X}) & \mu_{\mathbf{X}} \colon \mathrm{s}\mathcal{D}_B(\mathrm{s}\mathcal{D}_B(\mathbf{X})) &\to \mathrm{s}\mathcal{D}_B(\mathbf{X}) \\ x &\mapsto |x|_{\mathbf{X}} \cdot x & \textstyle\sum_i r_i \cdot \omega_i &\mapsto \big(x \mapsto \textstyle\bigvee_i r_i \wedge \omega_i(x) \big) \ . \end{aligned}$$

Here, we write logical subdistributions as formal convex sums as in (2.1); so, for example, $\eta_{\mathbf{X}}(x)$ is the Dirac distribution with $x \mapsto |x|_{\mathbf{X}}$ and $y \mapsto \bot$ for $y \neq x$.

As with the finite distribution monad on $\mathcal{S}\mathrm{et}$, the algebras for the logical distribution monad on $\mathcal{S}\mathrm{et}_B$ are well known: they are *sheaves* on the Boolean algebra B as characterised, for example, in [2, Proposition 1]. The relevance of this for us is that we can exploit the well-known characterisation of coproducts in categories of sheaves to give a concrete description of finite coproducts in the category of $\mathrm{s}\mathsf{D}_B$-algebras.

Lemma 57. *The initial* sD_B*-algebra is the empty* B*-labelled set* $\mathbf{0}$*; while if* $\mathbf{X}$ *and* $\mathbf{Y}$ *are* sD_B*-algebras, then their coproduct* $\mathbf{X} \star \mathbf{Y}$ *is given by the* B*-labelled set*

$$\mathbf{X} \star \mathbf{Y} = X + \{(x, y) \in X \times Y : |x|_{\mathbf{X}} \wedge |y|_{\mathbf{Y}} = \bot\} + Y \ , \tag{6.7}$$

where the logical mass of an element in the first, second and third summand is given respectively by $|x|_{\mathbf{X}}$*;* $|x|_{\mathbf{X}} \vee |y|_{\mathbf{Y}}$*; and* $|y|_{\mathbf{Y}}$*, and where we endow this* B*-labelled set with* sD_B*-algebra structure in an analogous way to Lemma 7.*

Since the underlying B-labelled set of $\mathbf{X} \star \mathbf{Y}$ relies only on the underlying B-labelled sets of $\mathbf{X}$ and $\mathbf{Y}$, and not on their sD_B-algebra structure, we may expect that $(\star, \mathbf{0})$ extends to a monoidal structure on $\mathcal{S}\mathrm{et}_B$ with respect to which sD_B is linear exponential. This is in fact the case; the only non-trivial point is showing that unitality, associativity and symmetry constraints for $\star$ also descend to $\mathcal{S}\mathrm{et}_B$, which may be done by direct calculation. We thus obtain:

Proposition 58. *The operation* $\star$ *of* (6.7) *underlies a monoidal structure* $(\star, \mathbf{0})$ *on* $\mathcal{S}\mathrm{et}_B$*, with respect to which the logical subdistribution monad* sD_B *is linear exponential.*

The induced hypernormalisation maps $\mathcal{N} \colon \mathrm{s}\mathcal{D}_B(\Sigma_{i \in I} X_i) \to \mathrm{s}\mathcal{D}_B(\Sigma_{i \in I} \mathrm{s}\mathcal{D}_B(X_i))$ can be described as follows. An element of $\mathrm{s}\mathcal{D}_B(\Sigma_i X_i)$ of domain b is a function $\omega \colon \Sigma_i X_i \to B$ of finite support whose image is a partition of $b \in B$, such that $\omega(x) \leqslant |x|_{\mathbf{X}_i}$ for all $x \in X_i$. For each i, we have the elements $\omega(X_i) \in B$, which themselves constitute an I-fold partition of b; and we also have for each i the restricted function $\omega_i = \omega|_{X_i} : X_i \to B$, which, so long as $\omega(X_i) \neq \bot$, is an element of $\mathcal{D}X_i$ of domain $\omega(X_i)$. As such, we can define $\mathcal{N}(\omega)$ to be the element

$$\sum_{\substack{i \in I \\ \omega(X_i) \neq \bot}} \omega(X_i) \cdot \iota_i(\omega_i) \in \mathrm{s}\mathcal{D}_B(\Sigma_{i \in I} \mathrm{s}\mathcal{D}_B(X_i)) \ .$$

In this way, we have sidestepped the problem of normalising the subdistributions ω_i to total ones, as we would have needed to do for the monad D_B on $\mathcal{S}\mathrm{et}$; instead, we can allow them to remain as subdistributions. One way to see this is that we have decoupled the two aspects of hypernormalisation of probability distributions from each other: on the one hand, collecting like terms to obtain an outer distribution; and on the other, normalising the inner sub-distributions to genuine distributions. Our logical hypernormalisation does the former, but not the latter; it is for this reason that we say that our solution is perhaps not as satisfactory as one may like—but, we hope, still of some value, in particular for providing an example of hypernormalisation beyond the category of sets.

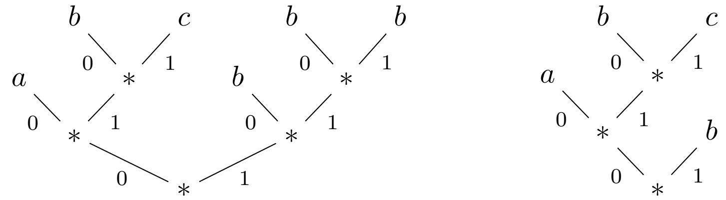
6.4. **Normalisation of continuous functions on streams.** For our final example, we describe an instance of hypernormalisation of a different flavour; in it, the hypernormalisation map will allow us to normalise programs encoding continuous functions defined on a type of streams.

By a *stream* over a finite alphabet B, we mean simply an $\mathbb{N}$-indexed family $\vec{b} = (b_0, b_1, \dots) \in B^{\mathbb{N}}$. The set $\mathsf{Stream}(B)$ of streams can be topologised, as usual, as a product of discrete spaces, and then a function $f \colon \mathsf{Stream}(B) \to A$ to a finite discrete space A is continuous precisely when it is *locally constant*:

$$\forall \vec{b} \in \mathsf{Stream}(B).\, \exists n \in \mathbb{N}.\, \forall \vec{c} \in \mathsf{Stream}(B).\, b_{\leqslant n} = c_{\leqslant n} \implies f(\vec{b}) = f(\vec{c}) \ ;$$

where we define $b_{\leqslant n} = (b_0, \dots, b_n)$. In other words, a continuous f uses only finitely many tokens of its input to compute its output.

It is an idea which goes back to Brouwer (though see [13] for a modern treatment) that such functions can be represented, non-uniquely, by well-founded B-ary branching trees with leaves labelled in A, i.e., elements of the initial algebra $T_B(A) = \mu X.\, X^B + A$. For example, the binary tree to the left in:

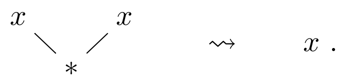


is a decision tree encoding the function $f\colon \mathsf{Stream}(\{0,1\}) \to \{a,b,c\}$ with

$$f(00\dots) = a\ , \qquad f(010\dots) = f(1\dots) = b\ , \quad \text{and} \quad f(011\dots) = c\ .$$

However, this tree is an inefficient encoding, since, on receiving a first input token of 1, we request needless additional input tokens before returning the output value b. There is an obvious algorithm which normalises such decision trees to maximally efficient ones, in this case the tree right above: starting from the leaves, we recursively contract all subtrees as left below to leaves as to the right:

(6.8)



While this normalisation algorithm is clearly motivated from a computational perspective, it is less clear how to understand it structurally. We will show that, in fact, it arises as an instance of hypernormalisation.

To see it in this way, we consider the monad T_B whose value at A is the set $T_B(A)$ of B-ary branching trees with leaves in A. This is the *free monad* on the endofunctor $X \mapsto X^B$ of $\mathcal{S}\mathrm{et}$, so that T_B-algebras are exactly *B-ary magmas*: sets X endowed with a B-ary operation $X^B \to X$. We will show that T_B is a linear exponential monad on $\mathcal{S}\mathrm{et}$ for a suitable monoidal structure, and so admits hypernormalisation. From this, we obtain for any finite set A a function

$$(6.9) \quad T_B(A) \cong T_B(\Sigma_{a\in A} 1) \xrightarrow{\mathcal{N}} T_B(\Sigma_{a\in A} T_B 1) \xrightarrow{T_B(\Sigma_{a\in A} !)} T_B(\Sigma_{a\in A} 1) \cong T_B(A)$$

which we will see implements the normalisation algorithm described above.

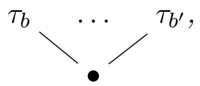
To find the monoidal structure on $\mathcal{S}\mathrm{et}$ for which T_B is linear exponential, the first step, as usual, is to characterise finite coproducts of T_B-algebras. In the following result, and henceforth, if $\tau \in T_B(A)$ and $A' \subseteq A$, then we say that τ is *A'-labelled* if it lies in $T_B(A') \subseteq T_B(A)$.

Lemma 59. *The initial B-ary magma is the empty set. If $(X, \xi \colon X^B \to X)$ and $(Y, \gamma \colon Y^B \to Y)$ are B-ary magmas, then their coproduct in the category of B-ary magmas is given by*

$$X \odot Y := \{\, \tau \in T_B(X+Y) : \tau \text{ has no non-trivial } X\text{- or } Y\text{-labelled subtree} \,\} \ .$$

The magma operation $\theta \colon (X \odot Y)^B \to X \odot Y$ takes a family of trees $(\tau_b : b \in B)$ to the disjoint union of the τ_b's, joined together at a fresh root vertex,

$$\theta(\tau_b : b \in B) =$$



except in the cases where this would create a non-trivial X- or Y-labelled subtree. These exceptional cases are where:

- *For each $b \in B$, the tree τ_b is a one-vertex tree $\bullet_{x_b}$ labelled by some $x_b \in X$; in this case, we take $\theta(\tau_b : b \in B) = \xi(x_b : b \in B)$.*
- *For each $b \in B$, the tree τ_a is a one-vertex tree $\bullet_{y_b}$ labelled by some $y_b \in Y$; in this case, we take $\theta(\tau_b : b \in B) = \gamma(y_b : b \in B)$.*

Proof. It is clear that the empty B-ary magma is initial. As for the binary coproduct, by its construction the magma operation on $X \odot Y$ is well-defined, and the two maps $\iota_1 \colon X \to X \odot Y \leftarrow Y : \iota_2$ sending an element of X or Y to the corresponding one-vertex labelled tree are magma homomorphisms. Moreover, given B-ary magma morphisms $f \colon X \to Z$ and $g \colon Y \to Z$, the unique magma morphism $\langle f, g \rangle \colon X \odot Y \to Z$ with $\langle f, g \rangle \iota_1 = f$ and $\langle f, g \rangle \iota_2 = g$ is given as the composite of the inclusion $X \odot Y \hookrightarrow T_B(X+Y)$ with the unique B-ary magma morphism $T_B(X+Y) \to Z$ extending the function $\langle f, g \rangle \colon X + Y \to Z$. □

Note that the set $X \odot Y$ underlying the coproduct of (X, ξ) and (Y, γ) does not rely on ξ and γ, but only on the underlying sets X and Y. So, like before, we may posit the existence of a symmetric monoidal structure $(\odot, 0)$ on $\mathcal{S}\mathrm{et}$ for which T_B is linear exponential. Once again, the only point requiring work is defining the associativity, unitality and symmetry isomorphisms, and, like before, we concentrate on the case of associativity. For this, we follow the idea of Lemma 25 by interposing a ternary tensor product $X \odot Y \odot Z \subseteq T_B(X+Y+Z)$ composed of those trees without any non-trivial X-, Y- or Z-labelled subtree.

Lemma 60. *For any sets X, Y, Z we have isomorphisms*

$$(X \odot Y) \odot Z \xrightarrow{\ \ell\ } X \odot Y \odot Z \xleftarrow{\ r\ } X \odot (Y \odot Z)$$

Proof (sketch). An element $\tau \in (X \odot Y) \odot Z$ is an $(X \odot Y) + Z$-labelled tree, and each $X \odot Y$-leaf is itself an $X + Y$-labelled tree. These data are equally encapsulated by an $X + Y + Z$-labelled tree with a collection of vertices marked: namely, the roots of the $X \odot Y$-trees at the leaves of the original τ. Forgetting this vertex marking yields the element $\ell(\tau) \in X \odot Y \odot Z$; and to show invertibility of the ℓ so defined, we need to reconstruct τ's marking uniquely from $\ell(\tau)$. But this is easy; it is uniquely characterised by the following properties:

(i) The subtree above every marked vertex is $X + Y$-labelled;
(ii) Every leaf vertex is above some marked vertex;

(iii) No non-leaf vertex has all of its B children marked,

and we can obtain it via the following algorithm: first mark each X- or Y-leaf; then recursively move markings towards the root until (ii) is satisfied. This must terminate by well-foundedness. This defines the invertible ℓ; now r is dual. □

We can thus take $r^{-1}\ell\colon (X \odot Y) \odot Z \to X \odot (Y \odot Z)$ as the desired associativity constraint. We may proceed similarly to verify the remaining details in:

Proposition 61. *There is a monoidal structure* $(\odot, 0)$ *on* Set*, with respect to which the B-ary magma monad* T_B *is linear exponential.*

The associated hypernormalisation maps $\mathcal{N}\colon T_B(\Sigma_i X_i) \to T_B(\Sigma_i T_B(X_i))$ may be described as follows. An element $\tau \in T_B(\Sigma_i X_i)$ is a $\Sigma_i X_i$-labelled B-ary tree. There is a unique way of marking vertices in τ such that:

(i) The subtree above any marked vertex is X_i-labelled for some i;
(ii) No vertex has all B of its children marked.

On constructing this marking, the subtree above each marked vertex is an element of $\Sigma_i T_B(X_i)$; so viewing each such subtree as a leaf labelled in $\Sigma_i T_B(X_i)$, we have obtained the element $\mathcal{N}(\tau) \in T_B(\Sigma_i T_B(X_i))$.

More intuitively, if we think of $\tau \in T_B(\Sigma_i X_i)$ as a decision tree computing a continuous function $f\colon \mathsf{Stream}(B) \to \Sigma_i X_i$, then $\mathcal{N}(\tau) \in T_B(\Sigma_i T_B(X_i))$ computes a function $f'\colon \mathsf{Stream}(B) \to \Sigma_i T_B(X_i)$ as follows: given $S \in \mathsf{Stream}(B)$, we run the computation of $f(S)$ using τ, and halt at the precise moment that the summand $X_i \subseteq \sum_i X_i$ in which $f(S)$ lies has been determined. We then return as $f'(S)$ the X_i-labelled subtree lying above the halting vertex, i.e., the continuation of the computation of $f(S)$ as an element of the set X_i.

We now use our understanding of hypernormalisation to describe the map $T_B(A) \to T_B(A)$ of (6.9). This first applies $\mathcal{N}\colon T_B(\Sigma_{a\in A} 1) \to T_B(\Sigma_{a\in A} T_B 1)$, whose effect on $\tau \in T_B(A)$ is to mark the roots of the largest subtrees whose leaves are all labelled with a single element $a \in A$. It then applies the function $T_B(\Sigma_a !)\colon T_B(\Sigma_{a\in A} T_B 1) \to T_B(\Sigma_{a\in A} 1)$, which has the effect of collapsing the marked vertex at the root of each $\{a\}$-labelled subtree to the bare leaf a. The endofunction of $T_B(A)$ so resulting acts on a tree τ precisely by carrying out the contractions in (6.8)—in other words, it normalises τ to its most efficient representative, as desired.

In [13], the authors use the monad T_B to describe not only continuous functions $\mathsf{Stream}(B) \to A$, but also continuous functions $\mathsf{Stream}(B) \to \mathsf{Stream}(A)$; whereas the former were encoded by elements of $T_B(A)$, the latter are encoded by elements of the final coalgebra $F = \nu X.\, T_B(A \times X)$. Rather like before, the elements of F are only *intensional* representations of continuous functions; but also like before, there is a maximally efficient such representation of any given function, to which any given element of F may be normalised via the hypernormalisation maps. Discussing this in detail must await further work, but let us at least sketch the key construction. Let $\varphi\colon F \to T_B(A \times F)$ be the coalgebra map of the final coalgebra, and consider the composite

$$\psi\colon F \xrightarrow{\varphi} T_B(A \times F) \xrightarrow{\cong} T_B(\Sigma_{a\in A} F) \xrightarrow{\mathcal{N}} T_B(\Sigma_{a\in A} T_B(F)) \xrightarrow{\cong} T_B(A \times T_B(F)) \ .$$

This extends uniquely to a map of free T_B-algebras

$$\psi^\sharp = T_B(F) \xrightarrow{T_B(\psi)} T_B T_B(A \times T_B(F)) \xrightarrow{\mu} T_B(A \times T_B(F))$$

which makes $T_B(F)$ into a $T_B(A \times -)$-coalgebra. Thus we induce a unique map of coalgebras $T_B(F) \to F$; and precomposing this with the unit of T_B yields a map $F \to F$ which implements the desired normalisation operation.

This normalisation function was considered in detail in [11], where it was explained not in terms of hypernormalisation, but rather via a kind of *normalisation-by-trace-evaluation*: here, the "trace" of an element of F is the continuous function it implements. In forthcoming work, we will clarify the link between these two approaches and, in particular, justify the form of the normalisation map on F described here. We will also see that, on swapping out the monad T_B for the monads P_f^+ and D for non-deterministic and probabilistic choice, we obtain a notion of normalisation for states of labelled transition systems and probabilistic generative systems, which can, on the one hand, be explained as a kind of normalisation-by-trace-evaluation; and on the other, can be derived from the hypernormalisation maps discussed here.

References

[1] Benton, N., Bierman, G., de Paiva, V., and Hyland, M. Linear λ-calculus and categorical models revisited. In *Computer science logic (San Miniato, 1992)*, vol. 702 of *Lecture Notes in Computer Science*. Springer, 1993, pp. 61–84.

[2] Bergman, G. M. Actions of Boolean rings on sets. *Algebra Universalis 28* (1991), 153–187.

[3] Blute, R., Lucyshyn-Wright, R. B. B., and O'Neill, K. Derivations in codifferential categories. *Cahiers de Topologie et Géométrie Différentielle Catégoriques 57* (2016), 243–279.

[4] Blute, R. F., Cockett, J. R. B., Lemay, J.-S. P., and Seely, R. A. G. Differential categories revisited. *Applied Categorical Structures* (2019).

[5] Blute, R. F., Cockett, J. R. B., and Seely, R. A. G. Differential categories. *Mathematical Structures in Computer Science 16* (2006), 1049–1083.

[6] Day, B. On closed categories of functors. In *Reports of the Midwest Category Seminar IV*, vol. 137 of *Lecture Notes in Mathematics*. Springer, 1970, pp. 1–38.

[7] Day, B., and Street, R. Lax monoids, pseudo-operads, and convolution. In *Diagrammatic morphisms and applications*, vol. 318 of *Contemporary Mathematics*. American Mathematical Society, 2003.

[8] Foulis, D. J., and Bennett, M. K. Effect algebras and unsharp quantum logics. *Foundations of Physics 24* (1994), 1331–1352.

[9] Fox, T. Coalgebras and Cartesian categories. *Communications in Algebra 4* (1976), 665–667.

[10] Fritz, T., and Perrone, P. A probability monad as the colimit of spaces of finite samples. Preprint, available as arXiv:1712.05363, 2019.

[11] Garner, R. Stream processors and comodels. In *9th Conference on Algebra and Coalgebra in Computer Science, CALCO 2021, Salzburg, Austria* (2021), F. Gadducci and A. Silva, Eds., vol. 211 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, pp. 15:1–15:17.

[12] Girard, J.-Y. Linear logic. *Theoretical Computer Science 50* (1987), 1–101.

[13] Hancock, P., Pattinson, D., and Ghani, N. Representations of stream processors using nested fixed points. *Logical Methods in Computer Science 5* (2009), 3:9, 17.

[14] Jacobs, B. Convexity, duality and effects. In *Theoretical computer science*, vol. 323 of *IFIP Adv. Inf. Commun. Technol.* Springer, 2010, pp. 1–19.

[15] Jacobs, B. Probabilities, distribution monads, and convex categories. *Theoretical Computer Science 412* (2011), 3323–3336.

[16] Jacobs, B. Hyper normalisation and conditioning for discrete probability distributions. *Logical Methods in Computer Science 13* (2017), Paper No. 17, 29.
[17] Jacobs, B., Mandemaker, J., and Furber, R. The expectation monad in quantum foundations. *Information and Computation 250* (2016), 87–114.
[18] Jacobs, B., Westerbaan, B., and Westerbaan, B. States of convex sets. In *Foundations of software science and computation structures*, vol. 9034 of *Lecture Notes in Computer Science*. Springer, Heidelberg, 2015, pp. 87–101.
[19] Kaddar, Y. Tricocycloids, effect monoids and effectuses. Research internship report, available as https://younesse.net/assets/M2_report.pdf., 2019.
[20] Mardare, R., Panangaden, P., and Plotkin, G. Quantitative algebraic reasoning. In *Proceedings of the 31st Annual ACM-IEEE Symposium on Logic in Computer Science (LICS 2016)* (2016), ACM, New York, p. 10.
[21] Markl, M. Models for operads. *Communications in Algebra 24* (1996), 1471–1500.
[22] May, J. P. *The geometry of iterated loop spaces*, vol. 271 of *Lecture Notes in Mathematics*. Springer, 1972.
[23] McIver, A., Morgan, C., Smith, G., Espinoza, B., and Meinicke, L. Abstract channels and their robust information-leakage ordering. In *Principles of Security and Trust* (2014), M. Abadi and S. Kremer, Eds., Springer, pp. 83–102.
[24] Melliès, P.-A. Categorical semantics of linear logic. In *Interactive models of computation and program behavior*, vol. 27 of *Panor. Synthèses*. Soc. Math. France, Paris, 2009, pp. 1–196.
[25] Mislove, M. W. Probabilistic monads, domains and classical information. In *Proceedings of DCM 2011* (2011), E. Kashefi, J. Krivine, and F. van Raamsdonk, Eds., vol. 88 of *EPTCS*, pp. 87–100.
[26] Moerdijk, I. Monads on tensor categories. *Journal of Pure and Applied Algebra 168* (2002), 189–208.
[27] Moggi, E. Notions of computation and monads. *Information and Computation 93* (1991), 55–92.
[28] Neumann, W. D. On the quasivariety of convex subsets of affine spaces. *Archiv der Mathematik 21* (1970), 11–16.
[29] Plotkin, G., and Power, J. Notions of computation determine monads. In *Foundations of software science and computation structures (Grenoble, 2002)*, vol. 2303 of *Lecture Notes in Computer Science*. Springer, Berlin, 2002, pp. 342–356.
[30] Power, A. J. A general coherence result. *Journal of Pure and Applied Algebra 57* (1989), 165–173.
[31] Roumen, F., and Roy, S. Duality for convex monoids. *Order 34* (2017), 349–362.
[32] Smolka, S., Foster, N., Hsu, J., Kappé, T., Kozen, D., and Silva, A. Guarded kleene algebra with tests: Verification of uninterpreted programs in nearly linear time. *Proceedings of the ACM on Programming Languages 4* (2019).
[33] Street, R. Fusion operators and cocycloids in monoidal categories. *Applied Categorical Structures 6* (1998), 177–191.
[34] Sweedler, M. E. *Hopf algebras*. Mathematics Lecture Note Series. W. A. Benjamin, Inc., New York, 1969.
[35] Świrszcz, T. Monadic functors and convexity. *Bulletin de l'Académie Polonaise des Sciences. Série des Sciences Mathématiques, Astronomiques et Physiques 22* (1974), 39–42.
[36] van Breugel, F. The metric monad for probabilistic nondeterminism. Unpublished notes, available at http://www.cse.yorku.ca/~franck/research/drafts/monad.pdf, 2005.
[37] Van Daele, A. Multiplier Hopf algebras. *Transactions of the American Mathematical Society 342* (1994), 917–932.

School of Mathematical and Physical Sciences, Macquarie University, NSW 2109, Australia
Email address: richard.garner@mq.edu.au