Open dynamical systems as coalgebras for polynomial functors, with application to predictive processing

Toby St. Clere Smithe

11 May 2022

We present categories of open dynamical systems with general time evolution as categories of coalgebras opindexed by polynomial interfaces, and show how this extends the coalgebraic framework to capture common scientific applications such as ordinary differential equations, open Markov processes, and random dynamical systems. We then extend Spivak's operad **Org** to this setting, and construct associated monoidal categories whose morphisms represent hierarchical open systems; when their interfaces are simple, these categories supply canonical comonoid structures. We exemplify these constructions using the 'Laplace doctrine', which provides dynamical semantics for active inference, and indicate some connections to Bayesian inversion and coalgebraic logic.

1. Background

1.1. Closed dynamical systems and Markov processes

In this brief section, we recall a 'behavioural' approach to dynamical systems originally due (we believe) to Lawvere; for a pedagogical account, see [1]. These systems are 'closed' in the sense that they do not require environmental interaction for their evolution, but they nonetheless form the starting point for our categories of more open systems.

Definition 1.1. Let $(\mathbb{T}, +, 0)$ be a monoid, representing time. Let $X : \mathcal{E}$ be some space, called the *state space*. Then a *closed dynamical system* ϑ *with state space* X *and time* $\mathbb{T}$ is an action of $\mathbb{T}$ on X. When $\mathbb{T}$ is also an object of $\mathcal{E}$, then this amounts to a morphism $\vartheta : \mathbb{T} \times X \to X$ (or equivalently, a time-indexed family of X-endomorphisms, $\vartheta(t) : X \to X$), such that $\vartheta(0) = \mathsf{id}_X$ and $\vartheta(s + t) = \vartheta(s) \circ \vartheta(t)$.

Proposition 1.2. When time is discrete, as in the case $\mathbb{T} = \mathbb{N}$, any dynamical system ϑ is entirely determined by its action at $1 : \mathbb{T}$. That is, letting the state space be X, we have $\vartheta(t) = \vartheta(1)^{\circ t}$ where $\vartheta(1)^{\circ t}$ means "compose $\vartheta(1) : X \to X$ with itself t times".

Example 1.3. Suppose $X : U \to TU$ is a vector field on U, with a corresponding solution (integral curve) $\chi_x : \mathbb{R} \to U$ for all $x : U$; that is, $\chi'(t) = X(\chi_x(t))$ and $\chi_x(0) = x$. Then letting the point x vary, we obtain a map $\chi : \mathbb{R} \times U \to U$. This χ is a closed dynamical system with state space U and time $\mathbb{R}$.

Proposition 1.4. Closed dynamical systems with state spaces in $\mathcal{E}$ and time $\mathbb{T}$ are the objects of the functor category $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})$, where $\mathbf{B}\mathbb{T}$ denotes the delooping of the monoid $\mathbb{T}$. Morphisms of dynamical systems are therefore natural transformations.

We will also often be interested in dynamical systems whose evolution has 'side-effects', such as the generation (or 'mixing') of uncertainty or randomness. We will largely model such systems as Kleisli maps or coalgebras of monads modelling these side-effects. In the case of uncertainty, the monads will be so-called *probability monads*, which we will often denote by $\mathcal{P}$. Such a monad $\mathcal{P} : \mathcal{E} \to \mathcal{E}$ can often be thought of as taking each set or space $X : \mathcal{E}$ to the set (or space) $\mathcal{P}X$ of probability distributions over X, and each morphism to the corresponding 'pushforwards' map; the monad multiplication is given by "averaging out" uncertainty, and the unit takes a point to the 'Dirac' distribution over it. With these ideas in mind, we can extend the concepts above to cover Markov chains and Markov processes.

Example 1.5 (Closed Markov chains and Markov processes)**.** A closed *Markov chain* is given by a map $X \to \mathcal{P}X$, where $\mathcal{P} : \mathcal{E} \to \mathcal{E}$ is a probability monad on $\mathcal{E}$; this is equivalently a $\mathcal{P}$-coalgebra with time $\mathbb{N}$, and an object in $\mathbf{Cat}\big(\mathbf{B}\mathbb{N}, \mathcal{K}\ell(\mathcal{P})\big)$. With more general time $\mathbb{T}$, one obtains closed *Markov processes*: objects in $\mathbf{Cat}\big(\mathbf{B}\mathbb{T}, \mathcal{K}\ell(\mathcal{P})\big)$. More explicitly, a closed Markov process is a time-indexed family of Markov kernels; that is, a morphism $\vartheta : \mathbb{T} \times X \to \mathcal{P}X$ such that, for all times $s, t : \mathbb{T}$, $\vartheta_{s+t} = \vartheta_s \bullet \vartheta_t$ as a morphism in $\mathcal{K}\ell(\mathcal{P})$. Note that composition $\bullet$ in $\mathcal{K}\ell(\mathcal{P})$ is given by the Chapman-Kolmogorov equation, so this means that

$$\vartheta_{s+t}(y|x) = \int_{x':X} \vartheta_s(y|x')\, \vartheta_t(\mathrm{d}x'|x)\,.$$

1.2. Polynomial functors

We will use *polynomial functors* to model the interfaces of our open systems, following Spivak and Niu [2]. We will assume these to be functors $\mathcal{E} \to \mathcal{E}$ for a locally Cartesian closed category $\mathcal{E}$, but we will typically assume that $\mathcal{E}$ is furthermore concrete, and often that it is in fact $\mathbf{Set}$.

Definition 1.6. Let $\mathcal{E}$ be a locally Cartesian closed category, and denote by y^A the representable copresheaf $y^A := \mathcal{E}(A, -) : \mathcal{E} \to \mathcal{E}$. A *polynomial functor* p is a coproduct of representable functors, written $p := \sum_{i:p(1)} y^{p_i}$, where $p(1) : \mathcal{E}$ is the indexing object. The category of polynomial functors in $\mathcal{E}$ is the full subcategory $\mathbf{Poly}_{\mathcal{E}} \hookrightarrow [\mathcal{E}, \mathcal{E}]$ of the $\mathcal{E}$-copresheaf category spanned by coproducts of representables. A morphism of polynomials is therefore a natural transformation.

Remark 1.7. Every polynomial functor $P : \mathcal{E} \to \mathcal{E}$ corresponds to a bundle $p : E \to B$ in $\mathcal{E}$, for which $B = P(1)$ and for each $i : P(1)$, the fibre p_i is $P(i)$. We will henceforth elide the distinction between a copresheaf P and its corresponding bundle p, writing $p(1) := B$ and $p[i] := p_i$, where $E = \sum_i p[i]$. A natural transformation $f : p \to q$ between copresheaves therefore corresponds to a map of bundles. In the case of polynomials, by the Yoneda lemma, this map is given by a 'forwards' map $f_1 : p(1) \to q(1)$ and a family of 'backwards' maps $f^\# : q[f_1(\text{-})] \to p[\text{-}]$ indexed by $p(1)$, as in the left diagram below. Given $f : p \to q$ and $g : q \to r$, their composite $g \circ f : p \to r$ is as in the right diagram below.

$$\begin{array}{ccccc} E & \xleftarrow{f^\#} & f^*F & \longrightarrow & F \\ {\scriptstyle p}\downarrow & & \downarrow & \lrcorner & \downarrow{\scriptstyle q} \\ B & = & B & \xrightarrow{f_1} & C \end{array} \qquad\qquad \begin{array}{ccccc} E & \xleftarrow{(gf)^\#} & f^*g^*G & \longrightarrow & G \\ {\scriptstyle p}\downarrow & & \downarrow & \lrcorner & \downarrow{\scriptstyle r} \\ B & = & B & \xrightarrow{g_1 \circ f_1} & D \end{array}$$

where $(gf)^\#$ is given by the $p(1)$-indexed family of composite maps $r[g_1(f_1(\text{-}))] \xrightarrow{f^*g^\#} q[f_1(\text{-})] \xrightarrow{f^\#} p[\text{-}]$.

We can interpret the type $p(1)$ to be a set or space of 'configurations' or 'positions' of a p-shaped system, and each $p[i]$ to be the available 'inputs' or 'directions' available to the system when it is in configuration/position i.

We now recall a handful of useful facts about polynomials and their morphisms, each of which is explained in Spivak and Niu [2] and summarized in Spivak [3].

Proposition 1.8. Polynomial morphisms $p \to y$ correspond to sections $p(1) \to \sum_i p[i]$ of the corresponding bundle p.

Proposition 1.9. There is an embedding of $\mathcal{E}$ into $\mathbf{Poly}_{\mathcal{E}}$ given by taking objects $X : \mathcal{E}$ to the linear polynomials $Xy : \mathbf{Poly}_{\mathcal{E}}$ and morphisms $f : X \to Y$ to morphisms $(f, \mathsf{id}_X) : Xy \to Yy$.

Proposition 1.10. There is a symmetric monoidal structure $(\otimes, y)$ on $\mathbf{Poly}_{\mathcal{E}}$ that we call tensor, and which is given on objects by $p \otimes q := \sum_{i:p(1)} \sum_{j:q(1)} y^{p[i] \times q[j]}$ and on morphisms $f := (f_1, f^\#) : p \to p'$ and $g := (g_1, g^\#) : q \to q'$ by $f \otimes g := (f_1 \times g_1, f^\# \times g^\#)$.

Proposition 1.11. $(\mathbf{Poly}_{\mathcal{E}}, \otimes, y)$ is symmetric monoidal closed, with internal hom denoted $[-, =]$. Explicitly, we have $[p, q] = \sum_{f:p \to q} y^{\sum_{i:p(1)} q[f_1(i)]}$. Given an object $A : \mathcal{E}$, we have $[Ay, y] \cong y^A$.

Proposition 1.12. The composition of polynomial functors $q \circ p : \mathcal{E} \to \mathcal{E} \to \mathcal{E}$ induces a monoidal structure on $\mathbf{Poly}_{\mathcal{E}}$, which we denote $\triangleleft$, and call 'composition' or 'substitution'. Its unit is again y. Famously, $\triangleleft$-comonoids correspond to categories and their comonoid homomorphisms are cofunctors [4]. If $\mathbb{T}$ is a monoid, then the comonoid structure on $y^{\mathbb{T}}$ corresponds witnesses it as the category $\mathbf{B}\mathbb{T}$. Monomials of the form Sy^S can be equipped with a canonical comonoid structure witnessing the codiscrete groupoid on S.

2. Open dynamical systems as polynomial coalgebras

2.1. Deterministic systems

Definition 2.1. A deterministic open dynamical system with interface p, state space S and time $\mathbb{T}$ is a polynomial morphism $\beta : Sy^S \to [\mathbb{T}y, p]$ such that, for any section $\sigma : p \to y$, the induced morphism

$$Sy^S \xrightarrow{\beta} [\mathbb{T}y, p] \xrightarrow{[\mathbb{T}y, \sigma]} [\mathbb{T}y, y] \xrightarrow{\sim} y^{\mathbb{T}}$$

is a comonoid homomorphism.

To see how such a morphism β is like an 'open' version of the closed dynamical systems introduced above, note that by the tensor-hom adjunction, β can equivalently be written with the type $\mathbb{T}y \otimes Sy^S \to p$. In turn, such a morphism corresponds to a pair (β^o, β^u), where β^o is the component 'on positions' with the type $\mathbb{T} \times S \to p(1)$, and β^u is the component 'on directions' with the type $\sum_{t:\mathbb{T}} \sum_{s:S} p[\beta^o(t, s)] \to S$. We will call the map β^o the *output map*, as it chooses an output position for each state and moment in time; and we will call the map β^u the *update map*, as it takes a state $s : S$, a quantity of time $t : \mathbb{T}$, and an 'input' in $p[\beta^o(t, s)]$, and returns a new state. We might imagine the new state as being given by evolving the system from s for time t, and the input as being given at the position corresponding to (s, t).

But it is not sufficient to consider merely such pairs $\beta = (\beta^o, \beta^u)$ to be our open dynamical systems, for we need them to be like 'open' monoid actions: evolving for time t then for time s must be equivalent to evolving for time $t + s$, given the same inputs. It is fairly easy to prove the following proposition, whose proof we defer until after establishing the categories $\mathbf{Coalg}^{\mathbb{T}}(p)$.

Proposition 2.2. Comonoid homomorphisms $Sy^S \to y^{\mathbb{T}}$ correspond bijectively with closed dynamical systems with state space $S : \mathcal{E}$, in the sense given by functors $\mathbf{B}\mathbb{T} \to \mathcal{E}$.

This establishes that seeking such a comonoid homomorphism will give us the monoid action property that we seek, and so it remains to show that a composite comonoid homomorphism of the form $[\mathbb{T}y, \sigma] \circ \beta$ is a closed dynamical system with the "right inputs". Unwinding this composite, we find that the condition that it be a comonoid homomorphism corresponds to the requirement that, for any $t : \mathbb{T}$, the *closure* $\beta^\sigma : \mathbb{T} \times S \to S$ of β by σ given by

$$\beta^\sigma(t) := S \xrightarrow{\beta^o(t)^*\sigma} \sum_{s:S} p[\beta^o(t,s)] \xrightarrow{\beta^u} S$$

constitutes a closed dynamical system on S. The idea here is that σ gives the 'context' in which we can make an open system closed, thereby formalizing the "given the same inputs" requirement above.

With this conceptual framework in mind, we are in a position to render open dynamical systems on p with time $\mathbb{T}$ into a category, which we will denote by $\mathbf{Coalg}^{\mathbb{T}}(p)$. Its objects will be pairs (S, β) with $S : \mathcal{E}$ and β an open dynamical on p with state space S; we will often write these pairs equivalently as triples (S, β^o, β^u), making explicit the output and update maps. Morphisms will be maps of state spaces that commute with the dynamics:

Proposition 2.3. Open dynamical systems over p with time $\mathbb{T}$ form a category, denoted $\mathbf{Coalg}^{\mathbb{T}}(p)$. Its morphisms are defined as follows. Let $\vartheta := (X, \vartheta^o, \vartheta^u)$ and $\psi := (Y, \psi^o, \psi^u)$ be two dynamical systems over p. A morphism $f : \vartheta \to \psi$ consists in a morphism $f : X \to Y$ such that, for any time $t : \mathbb{T}$ and global section $\sigma : p(1) \to \sum_{i:p(1)} p[i]$ of p, the following naturality squares commute:

$$\begin{array}{ccccc} X & \xrightarrow{\vartheta^o(t)^*\sigma} & \sum_{x:X} p[\vartheta^o(t,x)] & \xrightarrow{\vartheta^u(t)} & X \\ {\scriptstyle f}\downarrow & & & & \downarrow{\scriptstyle f} \\ Y & \xrightarrow[\psi^o(t)^*\sigma]{} & \sum_{y:Y} p[\psi^o(t,y)] & \xrightarrow[\psi^u(t)]{} & Y \end{array}$$

The identity morphism id_ϑ on the dynamical system ϑ is given by the identity morphism id_X on its state space X. Composition of morphisms of dynamical systems is given by composition of the morphisms of the state spaces.

Proof. We need to check unitality and associativity of composition. This amounts to checking that the composite naturality squares commute. But this follows immediately, since the composite of two commutative diagrams along a common edge is again a commutative diagram. □

We can alternatively state Proposition 2.2 as follows, noting that the polynomial y corresponds to a trivial interface, exposing no configuration to any environment nor receiving any signals from it:

Proposition 2.4. $\mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(y)$ is equivalent to the classical category $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})$ of closed dynamical systems in $\mathcal{E}$ with time $\mathbb{T}$.

Proof. The trivial interface y corresponds to the trivial bundle $\mathsf{id}_1 : 1 \to 1$. Therefore, a dynamical system over y consists of a choice of state space S along with a trivial output map $\vartheta^o = \bar{\top} : \mathbb{T} \times S \to 1$ and a time-indexed update map $\vartheta^u : \mathbb{T} \times S \to S$. This therefore has the form of a classical closed dynamical system, so it remains to check the monoid action. There is only one section of id_1, which is again id_1. Pulling this back along the unique map $\vartheta^o(t) : S \to 1$ gives $\vartheta^o(t)^* \mathsf{id}_1 = \mathsf{id}_S$. Therefore the requirement that, given any section σ of y, the maps $\vartheta^u \circ \vartheta^o(t)^*\sigma$ form an action means in turn that so does $\vartheta^u : \mathbb{T} \times S \to S$. Since the pullback of the unique section id_1 along the trivial output map $\vartheta^o(t) = \bar{\top} : S \to 1$ of any dynamical system in $\mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(y)$ is the identity of the corresponding state space id_S, a morphism $f : (\vartheta(*), \vartheta^u, \bar{\top}) \to (\psi(*), \psi^u, \bar{\top})$ in $\mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(y)$ amounts precisely to a map $f : \vartheta(*) \to \psi(*)$ on the state spaces in $\mathcal{E}$ such that the naturality condition $f \circ \vartheta^u(t) = \psi^u(t) \circ f$ of Proposition 1.4 is satisfied, and every morphism in $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})$ corresponds to a morphism in $\mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(y)$ in this way. □

Now that we know that our concept of open dynamical system subsumes closed systems, let us consider some more examples.

Example 2.5. Consider a dynamical system $(S, \vartheta^o, \vartheta^u)$ with outputs but no inputs. Such a system has a 'linear' interface $p := Iy$ for some $I : \mathcal{E}$; alternatively, we can write its interface p as the 'bundle' $\mathsf{id}_I : I \to I$. A section of this bundle must again be id_I, and so $\vartheta^o(t)^* \, \mathsf{id}_I = \mathsf{id}_S$. Once again, the update maps collect into to a closed dynamical system in $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})$; just now we have outputs $\vartheta^o : \mathbb{T} \times S \to p(1) = I$ exposed to the environment.

Proposition 2.6. When time is discrete, as with $\mathbb{T} = \mathbb{N}$, any open dynamical system $(X, \vartheta^o, \vartheta^u)$ over p is entirely determined by its components at $1 : \mathbb{T}$. That is, we have $\vartheta^o(t) = \vartheta^o(1) : X \to p(1)$ and $\vartheta^u(t) = \vartheta^u(1) : \sum_{x:X} p[\vartheta^o(x)] \to X$. A discrete-time open dynamical system is therefore a triple $(X, \vartheta^o, \vartheta^u)$, where the two maps have types $\vartheta^o : X \to p(1)$ and $\vartheta^u : \sum_{x:X} p[\vartheta^o(x)] \to X$.

Proof. Suppose σ is a section of p. We require each closure ϑ^σ to satisfy the flow conditions, that $\vartheta^\sigma(0) = \mathsf{id}_X$ and $\vartheta^\sigma(t + s) = \vartheta^\sigma(t) \circ \vartheta^\sigma(s)$. In particular, we must have $\vartheta^\sigma(t + 1) = \vartheta^\sigma(t) \circ \vartheta^\sigma(1)$. By induction, this means that we must have $\vartheta^\sigma(t) = \vartheta^\sigma(1)^{\circ t}$ (compare Proposition 1.2). Therefore we must in general have $\vartheta^o(t) = \vartheta^o(1)$ and $\vartheta^u(t) = \vartheta^u(1)$. □

Example 2.7. Suppose $\dot{x} = f(x, a)$ and $b = g(x)$, with $f : X \times A \to TX$ and $g : X \to B$. Then, as for the 'closed' vector fields of Example 1.3, this induces an open dynamical system $(X, \smallint f, g) : \mathbf{Coalg}^{\mathbb{R}}(By^A)$, where $\smallint f : \mathbb{R} \times X \times A \to X$ returns the (X, A)-indexed solutions of f.

Example 2.8. The preceding example is easily extended to the case of a general polynomial interface. Suppose similarly that $\dot{x} = f(x, a_x)$ and $b = g(x)$, now with $f : \sum_{x:X} p[g(x)] \to TX$ and $g : X \to p(1)$. Then we obtain an open dynamical system $(X, \smallint f, g) : \mathbf{Coalg}^{\mathbb{R}}_{\mathsf{id}}(p)$, where now $\smallint f : \mathbb{R} \times \sum_{x:X} p[g(x)] \to X$ is the 'update' and $g : X \to p(1)$ the 'output' map.

It is quite straightforward to extend the construction of $\mathbf{Coalg}^{\mathbb{T}}(p)$ to an opindexed category $\mathbf{Coalg}^{\mathbb{T}}$; we unravel this opindexing explicitly in the appendix (Proposition A.1).

Proposition 2.9. $\mathbf{Coalg}^{\mathbb{T}}$ extends to an opindexed category $\mathbf{Coalg}^{\mathbb{T}} : \mathbf{Poly}_{\mathcal{E}} \to \mathbf{Cat}$. On objects (polynomials), it returns the categories above. On morphisms of polynomials, we simply post-compose: given $\varphi : p \to q$ and $\beta : Sy^S \to [\mathbb{T}y, p]$, obtain $Sy^S \to [\mathbb{T}y, p] \to [\mathbb{T}y, q]$ in the obvious way.

At this point, the reader may be wondering in what sense these open dynamical systems are coalgebras. To see this, observe that a polynomial morphism $Sy^S \to q$ is equivalently a map $S \to q(S)$: that is to say, a q-coalgebra. By setting $q = [\mathbb{T}y, p]$, we see the connection immediately; to make it clear, in Proposition A.2, we spell it out for the case $\mathbb{T} = \mathbb{N}$.

2.2. Open Markov processes via stochastic polynomials

Just as coalgebras $S \to pS$ correspond to discrete-time deterministic open dynamical systems, coalgebras $S \to p\mathcal{P}S$ correspond to discrete-time *stochastic* dynamical systems when $\mathcal{P}$ is a probability monad as introduced above. We have already seen that 'closed' Markov chains correspond to maps $S \to \mathcal{P}S$, and that Markov processes in general time correspond to functors $\mathbf{B}\mathbb{T} \to \mathcal{K}\ell(\mathcal{P})$. Our task in this section is therefore to connect these two perspectives, extending the categories of deterministic coalgebras $\mathbf{Coalg}^{\mathbb{T}}(p)$.

Working concretely, it is not hard to spot the relevant adjustment. We therefore make the following definition.

Definition 2.10. Let $M : \mathcal{E} \to \mathcal{E}$ be a monad on the category $\mathcal{E}$, and let $p : \mathbf{Poly}_{\mathcal{E}}$ be a polynomial in $\mathcal{E}$. Let $(\mathbb{T}, +, 0)$ be a monoid in $\mathcal{E}$, representing time. Then a *pM-coalgebra with time* $\mathbb{T}$ consists in a triple $\vartheta := (S, \vartheta^o, \vartheta^u)$ of a state space $S : \mathcal{E}$ and two morphisms $\vartheta^o : \mathbb{T} \times S \to p(1)$ and $\vartheta^u : \sum_{t:\mathbb{T}} \sum_{s:\mathbb{S}} p[\vartheta^o(t,s)] \to MS$, such that, for any section $\sigma : p(1) \to \sum_{i:p(1)} p[i]$ of p, the maps $\vartheta^\sigma : \mathbb{T} \times S \to MS$ given by

$$\sum_{t:\mathbb{T}} S \xrightarrow{\vartheta^o(-)^*\sigma} \sum_{t:\mathbb{T}} \sum_{s:S} p[\vartheta^o(-,s)] \xrightarrow{\vartheta^u} MS$$

constitute an object in the functor category $\mathbf{Cat}\big(\mathbf{B}\mathbb{T}, \mathcal{K}\ell(T)\big)$, where $\mathbf{B}\mathbb{T}$ is the delooping of $\mathbb{T}$ and $\mathcal{K}\ell(T)$ is the Kleisli category of T. Once more, we call the closed system ϑ^σ, induced by a section σ of p, the closure of ϑ by σ.

As before, such pM-coalgebras form a category; and these categories in turn are opindexed by polynomials.

Proposition 2.11. *pM-coalgebras with time $\mathbb{T}$ form a category, denoted $\mathbf{Coalg}^{\mathbb{T}}_M(p)$. Its morphisms are defined as follows. Let $\vartheta := (X, \vartheta^o, \vartheta^u)$ and $\psi := (Y, \psi^o, \psi^u)$ be two pM-coalgebras. A morphism $f : \vartheta \to \psi$ consists in a morphism $f : X \to Y$ such that, for any time $t : \mathbb{T}$ and global section $\sigma : p(1) \to \sum_{i:p(1)} p[i]$ of p, the following naturality squares commute:*

$$\begin{array}{ccccc} X & \xrightarrow{\vartheta^o(t)^*\sigma} & \sum_{x:X} p[\vartheta^o(t,x)] & \xrightarrow{\vartheta^u(t)} & MX \\ {\scriptstyle f}\downarrow & & & & \downarrow{\scriptstyle Mf} \\ Y & \xrightarrow[\psi^o(t)^*\sigma]{} & \sum_{y:Y} p[\psi^o(t,y)] & \xrightarrow[\psi^u(t)]{} & MY \end{array}$$

The identity morphism id_ϑ on the pM-coalgebra ϑ is given by the identity morphism id_X on its state space X. Composition of morphisms of pM-coalgebras is given by composition of the morphisms of the state spaces.

Proposition 2.12. *$\mathbf{Coalg}^{\mathbb{T}}_M(p)$ extends to an opindexed category, $\mathbf{Coalg}^{\mathbb{T}}_M(-) : \mathbf{Poly}_{\mathcal{E}} \to \mathbf{Cat}$. Suppose $\varphi : p \to q$ is a morphism of polynomials. We define a corresponding functor $\mathbf{Coalg}^{\mathbb{T}}_M(\varphi) : \mathbf{Coalg}^{\mathbb{T}}_M(p) \to \mathbf{Coalg}^{\mathbb{T}}_M(q)$ as follows. Suppose $(X, \vartheta^o, \vartheta^u) : \mathbf{Coalg}^{\mathbb{T}}_M(p)$ is an object (pM-coalgebra) in $\mathbf{Coalg}^{\mathbb{T}}_M(p)$. Then $\mathbf{Coalg}^{\mathbb{T}}_M(\varphi)(X, \vartheta^o, \vartheta^u)$ is defined as the triple $(X, \varphi_1 \circ \vartheta^o, \vartheta^u \circ \vartheta^{o*}\varphi^\#)$: $\mathbf{Coalg}^{\mathbb{T}}_M(q)$, where the two maps are explicitly the following composites:*

$$\mathbb{T} \times X \xrightarrow{\vartheta^o} p(1) \xrightarrow{\varphi_1} q(1)\,, \qquad \sum_{t:\mathbb{T}} \sum_{x:X} q[\varphi_1 \circ \vartheta^o(t,x)] \xrightarrow{\vartheta^{o*}\varphi^\#} \sum_{t:\mathbb{T}} \sum_{x:X} p[\vartheta^o(t,x)] \xrightarrow{\vartheta^u} MX\,.$$

On morphisms, $\mathbf{Coalg}^{\mathbb{T}}_M(\varphi)(f) : \mathbf{Coalg}^{\mathbb{T}}_M(\varphi)(X, \vartheta^o, \vartheta^u) \to \mathbf{Coalg}^{\mathbb{T}}_M(\varphi)(Y, \psi^o, \psi^u)$ is given by the same underlying map $f : X \to Y$ of state spaces.

The opindexed category $\mathbf{Coalg}^{\mathbb{T}}_M(p)$ clearly generalizes $\mathbf{Coalg}^{\mathbb{T}}$, since we can always take $M = \mathsf{id}_{\mathcal{E}}$. Yet these concrete definitions obscure the more elegant representation of the objects of $\mathbf{Coalg}^{\mathbb{T}}$ as morphisms $Sy^S \to [\mathbb{T}y, p]$. Our task is therefore to find a setting in which a similar representation is possible; to do so, we generalize $\mathbf{Poly}_{\mathcal{E}}$ so that the backwards components of its morphisms may incorporate 'side-effects' modelled by M. We will call the corresponding category $\mathbf{Poly}_M$, and will find that instantiating $\mathbf{Coalg}^{\mathbb{T}}$ in $\mathbf{Poly}_M$ recovers $\mathbf{Coalg}^{\mathbb{T}}_M(p)$.

We begin by recalling that $\mathbf{Poly}_{\mathcal{E}}$ is equivalent to the category of Grothendieck lenses for the self-indexing [5, 2]: $\mathbf{Poly}_{\mathcal{E}} \cong \int \mathcal{E}/-^{\,\mathrm{op}}$, where the opposite is taken pointwise on each $\mathcal{E}/B$. We will define $\mathbf{Poly}_M$ by analogy, using the following indexed category. Suppose M is a commutative monad on $\mathcal{E}$ and let ι denote the identity-on-objects inclusion $\mathcal{E} \hookrightarrow \mathcal{K}\ell(M)$ given on morphisms by post-composing with the unit η of the monad structure. For ease of exposition in this short paper, we will assume here that $\mathcal{E} = \mathbf{Set}$.

Definition 2.13. Define the indexed category $\mathcal{E}_M/- : \mathcal{E}^{\,\mathrm{op}} \to \mathbf{Cat}$ as follows. On objects $B : \mathcal{E}$, we define $\mathcal{E}_M/B$ to be the full subcategory of $\mathcal{K}\ell(M)/B$ on those objects $\iota p : E \nrightarrow B$ which correspond to maps $E \xrightarrow{p} B \xrightarrow{\eta_B} MB$ in the image of ι. Now suppose $f : C \to B$ is a map in $\mathcal{E}$. We define $\mathcal{E}_M/f : \mathcal{E}_M/B \to \mathcal{E}_M/C$ as follows. The functor $\mathcal{E}_M/f$ takes objects $\iota p : E \nrightarrow B$ to $\iota(f^*p) : f^*E \nrightarrow C$ where f^*p is the pullback of p along f in $\mathcal{E}$, included into $\mathcal{K}\ell(M)$ by ι.

To define the action of $\mathcal{E}_M/f$ on morphisms $\alpha : (E, \iota p : E \nrightarrow B) \to (F, \iota q : F \nrightarrow B)$, note that since we must have $\iota q \bullet \alpha = \iota p$, α must correspond to a family of maps $\alpha_x : p[x] \to Mq[x]$ for $x : B$. Then we can define $(\mathcal{E}_M/f)(\alpha)$ pointwise as $(\mathcal{E}_M/f)(\alpha)_y := \alpha_{f(y)} : p[f(y)] \to Mq[f(y)]$ for $y : C$.

Definition 2.14. We define $\mathbf{Poly}_M$ to be the category of Grothendieck lenses for $\mathcal{E}_M/-$. That is, $\mathbf{Poly}_M := \int \mathcal{E}_M/-^{\,\mathrm{op}}$, where the opposite is again taken pointwise.

Unwinding this definition, we find that the objects of $\mathbf{Poly}_M$ are the same polynomial functors as constitute the objects of $\mathbf{Poly}_{\mathcal{E}}$. The morphisms $f : p \to q$ are pairs $(f_1, f^\#)$, where $f_1 : B \to C$ is a map in $\mathcal{E}$ and $f^\#$ is a family of morphisms $q[f_1(x)] \nrightarrow p[x]$ in $\mathcal{K}\ell(M)$, making the following diagram commute:

$$\begin{array}{ccccc}
\sum_{x:B} Mp[x] & \xleftarrow{f^\#} & \sum_{b:B} q[f_1(x)] & \longrightarrow & \sum_{y:C} q[y] \\
\downarrow{\scriptstyle \eta_B{}^*p} & & \downarrow & \lrcorner & \downarrow{\scriptstyle q} \\
B & = & B & \xrightarrow{f_1} & C
\end{array}$$

Remark 2.15. Note that the tensor $\otimes$ extends to $\mathbf{Poly}_M$: on objects, it is defined identically to the tensor on $\mathbf{Poly}_{\mathcal{E}}$. On morphisms $f := (f_1, f^\#) : p \to q$ and $g := (g_1, g^\#) : p' \to q'$, we define the tensor $f \otimes g$ to have forwards component $f_1 \times g_1$ as before, and the backwards components are defined by $(f \otimes g)^\#_{(x,x')} := q[f_1(x)] \times q'[g_1(x')] \to Mp[x] \times Mp'[x'] \to M\big(p[x] \times p'[x']\big)$, where the second arrow is given by the commutativity of the monad M. On the other hand, we only get an internal hom satisfying the adjunction $\mathbf{Poly}_M(p \otimes q, r) \cong \mathbf{Poly}_M(p, [q, r])$ when the backwards components of morphisms $p \otimes q \to r$ are 'uncorrelated' between p and q.

Remark 2.16. For $\mathbf{Poly}_M$ to behave faithfully like the category $\mathbf{Poly}_{\mathcal{E}}$ of polynomial functors and their morphisms, we should want the substitution functors $\mathcal{E}_M/f : \mathcal{E}_M/C \to \mathcal{E}_M/B$ to have left and right adjoints. Although we do not spell it out here, it is quite straightforward to exhibit the left adjoints. On the other hand, writing f^* as shorthand for $\mathcal{E}_M/f$, we can see that a right adjoint only obtains in restricted circumstances. Denote the putative right adjoint by $\Pi_f : \mathcal{E}_M/B \to \mathcal{E}_M/C$, and for $\iota p : E \nrightarrow B$ suppose that $(\Pi_f E)[y]$ is given by the set of 'partial sections' $\sigma : f^{-1}\{y\} \to TE$ of p over $f^{-1}\{y\}$ as in the commutative diagram:

$$\begin{array}{ccccc}
 & \overset{\sigma}{\swarrow} & f^{-1}\{y\} & \longrightarrow & \{y\} \\
 & & \downarrow & \lrcorner & \downarrow \\
TE & \xrightarrow{\eta_B{}^*p} & B & \xrightarrow{f} & C
\end{array}$$

Then we would need to exhibit a natural isomorphism $\mathcal{E}_M/B(f^*D, E) \cong \mathcal{E}_M/C(D, \Pi_f E)$. But this will only obtain when the 'backwards' components $h^\#_y : D[y] \to M(\Pi_f E)[y]$ are in the image of ι—otherwise, it is not generally possible to pull $f^{-1}\{y\}$ out of M.

Despite these restrictions, we do have enough structure at hand to instantiate $\mathbf{Coalg}^{\mathbb{T}}$ in $\mathbf{Poly}_M$. The only piece remaining is the composition product $\lhd$, but for our purposes it suffices to define its action on objects, which is identical to its action on objects in $\mathbf{Poly}_{\mathcal{E}}$[1], and then consider $\lhd$-comonoids in $\mathbf{Poly}_M$. The comonoid laws force the structure maps to be deterministic (*i.e.*, in the image of ι), and so $\lhd$-comonoids in $\mathbf{Poly}_M$ are just $\lhd$-comonoids in $\mathbf{Poly}_{\mathcal{E}}$.

[1]We leave the full exposition of $\lhd$ in $\mathbf{Poly}_M$ to the forthcoming extended version of this paper.

Finally, we note that we can define morphisms $\beta : Sy^S \to [\mathbb{T}y, p]$: these again just correspond to morphisms $\mathbb{T}y \otimes Sy^S \to p$, and the condition that the backwards maps be uncorrelated between $\mathbb{T}y$ and p is satisfied because $\mathbb{T}y$ has a trivial exponent. Unwinding such a β according to the definition of $\mathbf{Poly}_M$ indeed gives precisely a pair (β^o, β^u) of the requisite types; and a comonoid homomorphism $Sy^S \to y^{\mathbb{T}}$ in $\mathbf{Poly}_M$ is precisely a functor $\mathbf{B}\mathbb{T} \to \mathcal{K}\ell(M)$, thereby establishing equivalence between the objects of $\mathbf{Coalg}^{\mathbb{T}}(p)$ established in $\mathbf{Poly}_M$ and the objects of $\mathbf{Coalg}^{\mathbb{T}}_M(p)$. The equivalence between the hom-sets is established by a similar unwinding. All told, in this section, we have sketched the proof of the following theorem:

Theorem 2.17. Constructing $\mathbf{Coalg}^{\mathbb{T}}(p)$ in $\mathbf{Poly}_M$ yields a category equivalent to $\mathbf{Coalg}^{\mathbb{T}}_M(p)$.

2.3. Random dynamical systems and bundle systems

In the analysis of stochastic systems, it is often fruitful to consider two perspectives: on one side, one considers explicitly the evolution of the distribution of the states of the system, by following (for instance) a Markov process, or Fokker-Planck equation. On the other side, one considers the system as if it were a deterministic system, perturbed by noisy inputs, giving rise to the frameworks of stochastic differential equations and associated *random dynamical systems*.

Whereas a (closed) Markov process is typically given by the action of a 'time' monoid on an object in a Kleisli category of a probability monad, a (closed) random dynamical system is given by a *bundle* of closed dynamical systems, where the base system is equipped with a probability measure which it preserves: the idea being that a random dynamical system can be thought of as a 'random' choice of dynamical system on the total space at each moment in time, with the base measure-preserving system being the source of the randomness [6].

This idea corresponds in non-dynamical settings to the notion of *randomness pushback* [7, Def. 11.19], by which a stochastic map $f : A \to \mathcal{P}\, B$ can be presented as a deterministic map $f^\flat : \Omega \times A \to B$ where (Ω, ω) is a probability space such that, for any $a : A$, pushing ω forward through $f^\flat(\text{-}, a)$ gives the state $f(a)$; that is, ω induces a random choice of map $f^\flat(\omega, \text{-}) : A \to B$. Similarly, under 'nice' conditions, random dynamical systems and Markov processes do coincide, although they have different suitability in applications.

In this section, we sketch how the generalized-coalgebraic structures developed above extend also to random dynamical systems, though with most details deferred to the Appendix. By observing that we can also 'open up' the base system of a random dynamical system, we obtain furthermore a notion of *open bundle system*: a bundle of dynamical systems that is coherently 'open' over polynomials both in the total space and the base space.

Definition 2.18. Suppose $\mathcal{E}$ is a category equipped with a probability monad $\mathcal{P} : \mathcal{E} \to \mathcal{E}$ and a terminal object $1 : \mathcal{E}$. A probability space in $\mathcal{E}$ is an object of the slice $1/\mathcal{K}\ell(\mathcal{P})$ of the Kleisli category of the probability monad under 1.

Remark 2.19. In order to consider polynomials in $\mathcal{E}$, we will later assume again that it is locally Cartesian closed. A simple example of a locally Cartesian closed category equipped with a probability monad is the category $\mathbf{Set}$ equipped with the monad $\mathcal{D}$ taking each set to the set of finitely-supported probability distributions upon it.

Proposition 2.20. There is a forgetful functor $1/\mathcal{K}\ell(\mathcal{P}) \to \mathcal{E}$ taking probability spaces (B, β) to the underlying spaces B and their morphisms $f : (A, \alpha) \to (B, \beta)$ to the underlying maps $f : A \to \mathcal{P}\, B$. We will write B to refer to the space in $\mathcal{E}$ underlying a probability space (B, β), in the image of this forgetful functor.

Definition 2.21. Let (B, β) be a probability space in $\mathcal{E}$. A closed metric or measure-preserving dynamical system (ϑ, β) on (B, β) with time $\mathbb{T}$ is a closed dynamical system ϑ with state space $B : \mathcal{E}$ such that, for all $t : \mathbb{T}$, $\mathcal{P}\,\vartheta(t) \circ \beta = \beta$; that is, each $\vartheta(t)$ is a (B, β)-endomorphism in $1/\mathcal{K}\ell(\mathcal{P})$.

Proposition 2.22. Closed measure-preserving dynamical systems in $\mathcal{E}$ with time $\mathbb{T}$ form the objects of a category $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})_{\mathcal{P}}$ whose morphisms $f : (\vartheta, \alpha) \to (\psi, \beta)$ are maps $f : \vartheta(*) \to \psi(*)$ in $\mathcal{E}$ between the state spaces that preserve both flow and measure, as in the following commutative diagram, which also indicates their composition:

$$
\begin{array}{ccccccc}
 & & \mathcal{P}\,\vartheta(*) & \xrightarrow{\mathcal{P}\,\vartheta(t)} & \mathcal{P}\,\vartheta(*) & & \\
 & \overset{\alpha}{\nearrow} & \big\downarrow{\scriptstyle \mathcal{P}\,f} & & {\scriptstyle \mathcal{P}\,f}\big\downarrow & \overset{\alpha}{\nwarrow} & \\
1 & \xrightarrow[\beta]{} & \mathcal{P}\,\psi(*) & \xrightarrow[\mathcal{P}\,\psi(t)]{} & \mathcal{P}\,\psi(*) & \xleftarrow[\beta]{} & 1 \\
 & \underset{\gamma}{\searrow} & \big\downarrow{\scriptstyle \mathcal{P}\,g} & & {\scriptstyle \mathcal{P}\,g}\big\downarrow & \underset{\gamma}{\swarrow} & \\
 & & \mathcal{P}\,\lambda(*) & \xrightarrow[\mathcal{P}\,\lambda(t)]{} & \mathcal{P}\,\lambda(*) & &
\end{array}
$$

Definition 2.23. Let (ϑ, β) be a closed measure-preserving dynamical system. A closed random dynamical system over (ϑ, β) is an object of the slice category $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})/\vartheta$; it is therefore a bundle of the corresponding functors.

Example 2.24. The solutions $X(t, \omega; x_0) : \mathbb{R}_+ \times \Omega \times M \to M$ to a stochastic differential equation $\mathrm{d}X_t = f(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t$, where $W : \mathbb{R}_+ \times \Omega \to M$ is a Wiener process in M, define a random dynamical system $\mathbb{R}_+ \times \Omega \times M \to M : (t, \omega, x) \mapsto X(t, \omega; x_0)$ over the Wiener base flow $\theta : \mathbb{R}_+ \times \Omega \to \Omega : (t, \omega) \mapsto W(s + t, \omega) - W(t, \omega)$ for any $s : \mathbb{R}_+$.

Definition 2.25. Let (θ, β) be a closed measure-preserving dynamical system in $\mathcal{E}$ with time $\mathbb{T}$, and let $p : \mathbf{Poly}_{\mathcal{E}}$ be a polynomial in $\mathcal{E}$. Write $\Omega := \theta(*)$ for the state space of θ, and let $\pi : S \to \Omega$ be an object (bundle) in $\mathcal{E}/\Omega$. An open random dynamical system over (θ, β) on the interface p with state space $\pi : S \to \Omega$ and time $\mathbb{T}$ consists in a pair of morphisms $\vartheta^o : \mathbb{T} \times S \to p(1)$ and $\vartheta^u : \sum_{t:\mathbb{T}} \sum_{s:S} p[\vartheta^o(t, s)] \to S$, such that, for any global section $\sigma : p(1) \to \sum_{i:p(1)} p[i]$ of p, the maps $\vartheta^\sigma : \mathbb{T} \times S \to S$ defined as

$$\sum_{t:\mathbb{T}} S \xrightarrow{\vartheta^o(-)^*\sigma} \sum_{t:\mathbb{T}} \sum_{s:S} p[\vartheta^o(-, s)] \xrightarrow{\vartheta^u} S$$

form a closed random dynamical system in $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})/\theta$, in the sense that, for all $t : \mathbb{T}$ and sections σ, the following diagram commutes:

$$
\begin{array}{ccccc}
S & \xrightarrow{\vartheta^o(t)^*\sigma} & \sum_{s:S} p[\vartheta^o(t, s)] & \xrightarrow{\vartheta^u(t)} & S \\
{\scriptstyle \pi}\big\downarrow & & & & \big\downarrow{\scriptstyle \pi} \\
\Omega & & \xrightarrow[\theta(t)]{} & & \Omega
\end{array}
$$

Proposition 2.26. Let (θ, β) be a closed measure-preserving dynamical system in $\mathcal{E}$ with time $\mathbb{T}$, and let $p : \mathbf{Poly}_{\mathcal{E}}$ be a polynomial in $\mathcal{E}$. Open random dynamical systems over (θ, β) on the interface p form the objects of a category $\mathbf{RDyn}^{\mathbb{T}}(p, \theta)$. See Definition A.3 in the Appendix for details.

Proposition 2.27. The categories $\mathbf{RDyn}^{\mathbb{T}}(p, \theta)$ collect into a doubly-indexed category of the form $\mathbf{RDyn}^{\mathbb{T}} : \mathbf{Poly}_{\mathcal{E}} \times \mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})_{\mathcal{P}} \to \mathbf{Cat}$. See Proposition A.4 in the Appendix for details.

By allowing the base systems of open random dynamical systems instead to be arbitrary dynamical systems, and then by opening them up similarly, one obtains notions of *open bundle dynamical system*, and correspondingly doubly-opindexed categories over pairs of polynomials. Representing these categories concisely, as we did for the categories $\mathbf{Coalg}^{\mathbb{T}}_{M}(p)$, is the subject of on-going work, and so we defer the details to the Appendix, in Definition A.5, and Propositions A.6, A.7, and A.8.

3. Hierarchical systems via generalized Org

In order to exhibit the main example of this paper, we will need to construct, from the opindexed categories of $p\mathcal{P}$-coalgebras introduced above, monoidal categories whose objects represent the interfaces of hierarchical systems and whose morphisms represent the hierarchical systems themselves. Informally put, we will think of a morphism $p \to q$ in such a category as "a q-shaped system with a p-shaped hole". In order to achieve this, we will in turn adopt and generalize the operad $\mathbf{Org}$ introduced by Spivak [8].

Definition 3.1 (Following Spivak [8, Def. 2.19])**.** We define a (category-enriched, symmetric, coloured) operad, $\mathbf{Org}_M^{\mathbb{T}}$. Its objects are polynomials, and for any tuple of polynomials $(p_1, \ldots, p_k; p')$ of at least length 2, the hom category $\mathbf{Org}_M^{\mathbb{T}}(p_1, \ldots, p_k; p')$ is given by $\mathbf{Coalg}_M^{\mathbb{T}}([p_1 \otimes \cdots \otimes p_k, p'])$. Note that $\{y \to [\mathbb{T}y, [p,p]]\} \cong \{\mathbb{T}y \to [p,p]\}$. On any given interface p, the identity coalgebra is therefore given by the morphism $\mathbb{T}y \to [p,p]$ that constantly outputs id_p and has trivial backwards component. To define composition, we use the canonical maps $[p,q] \otimes [q,r] \to [p,r]$ and $[p,q] \otimes [p',q'] \to [p \otimes p', q \otimes q']$, the pseudofunctoriality of $\mathbf{Coalg}_M^{\mathbb{T}}(-)$, and the laxators $\mathbf{Coalg}_M^{\mathbb{T}}(p) \times \mathbf{Coalg}_M^{\mathbb{T}}(q) \to \mathbf{Coalg}_M^{\mathbb{T}}(p \otimes q)$; since each of these components is associative and unital, the composition is well-defined.

Remark 3.2. Spivak's original definition of $\mathbf{Org}$ corresponds to the case where $M = \mathsf{id}_{\mathcal{E}}$ and $\mathbb{T} = \mathbb{N}$.

For our present purposes, all that is required is to obtain from $\mathbf{Org}$ a (monoidal) (bi)category[2]. We therefore restrict $\mathbf{Org}_M^{\mathbb{T}}$ to a bicategory $\mathbf{Hier}$ whose objects are again polynomials and whose hom-categories from p to q are given by $\mathbf{Org}_M^{\mathbb{T}}(p,q)$; it inherits a monoidal structure from the monoidal category associated to the symmetric operad $\mathbf{Org}_M^{\mathbb{T}}$. We will write $\mathbf{Hier}|_{\mathcal{E}}$ to denote the restriction of $\mathbf{Hier}$ to the linear polynomials Ay.

To bring things a little down to earth, first consider a general system $\beta : p \to q$ in $\mathbf{Hier}$. Recall that $[p,q] = \sum_{f:p\to q} y^{\sum_{i:p(1)} q[f_1(i)]}$. β is therefore given by a choice of state space X along with a pair of maps $\beta^o : \mathbb{T} \times X \to \mathbf{Poly}_M(p,q)$ and $\beta^u : \sum_{t:\mathbb{T}} \sum_{x:X} \sum_{i:p(1)} q[\beta^o(t,s)_1(i)] \to MX$.

To make this a little more comprehensible again, suppose $p = Ay^S$ and $q = By^T$. Then $\mathbf{Poly}_M(p,q) = \mathcal{E}(A,B) \times \mathcal{E}(A \times T, S)$, and so by the universal property of the product, β^o is equivalently given by a pair of maps: a 'forwards' output map $\beta_1^o : \mathbb{T} \times X \times A \to B$ and a 'backwards' output map $\beta_2^o : \mathbb{T} \times X \times A \times T \to S$; if this reminds you of a category of lenses, then this is no surprise: the subcategory of $\mathbf{Poly}_{\mathcal{E}}$ on the monomials Ay^S is indeed the category of bimorphic lenses in $\mathcal{E}$. Finally, the update map simplifies to $\beta^u : \mathbb{T} \times X \times A \times T \to MX$, which updates the state given 'forwards' inputs in A and 'backwards' inputs in T. We might denote the subcategory of $\mathbf{Hier}$ on such linear polynomials as $\mathbf{HiBi}$, to indicate 'hierarchical bidirectional' systems.

Taking one further step down the ladder of complexity, we briefly consider systems $\beta : Ay \to By$ in $\mathbf{Hier}|_{\mathcal{E}}$: these are just hierarchical bidirectional systems where $S = T = 1$. Therefore, in this case, the backwards output map becomes trivial, leaving only a forwards output map $\beta^o : \mathbb{T} \times X \times A \to B$ and an update map taking inputs in A, $\beta^u : \mathbb{T} \times X \times A \to MX$. By filling in the A-inputs, we get a system with B-outputs, corresponding to the informal intuition with which we opened this section: we have a B-shaped system with an A-shaped hole. Composition of these systems corresponds to placing systems in parallel using $\otimes$ and plugging interfaces into holes of the matching shape.

We end this section by briefly sketching the canonical $\otimes$-comonoid structure on $\mathbf{Hier}|_{\mathcal{E}}$, making $\mathbf{Hier}|_{\mathcal{E}}$ into a 'semi-Markov' [7] or 'copy-discard' [9] category. Note that, if a system has the trivial state space 1, then (i) tensoring with it is a no-op, and (ii) it has a trivial update map (assuming that $M1 \cong 1$[3]). Thus, for each object Ay, we obtain a discarding system $\bar{\top}_A : Ay \to 1y$ by taking the trivial state space, trivial update map, and trivial output map. The copying system $\curlyvee_A : Ay \to (A \times A)y$

[2]and in fact, we won't even really need to make use of the monoidal or bicategorical structures here!

[3]This condition is satisfied when M is a probability monad like the finite-support distribution monad, for instance.

again has trivial state space and update map, but now the output map $\Upsilon^o_A : \mathbb{T} \times A \to A \times A$ is given by the constant copying map $(t, a) \mapsto (a, a)$. It is then straightforward to check the comonoid laws.

4. Dynamical Bayesian inversion

One consequence of $\mathbf{Hier}|_{\mathcal{E}}$ being a copy-discard category is that we can instantiate an abstract form of Bayes' rule there, giving rise to a notion of when one $p\,\mathcal{P}$-coalgebraic system can be seen to be 'predicting' or 'inverting' another. In general, Bayes' rule is expressed as an equality between morphisms, but this is too strong for dynamical systems, which 'black-box' their state spaces: that is to say, we should consider two morphisms (systems) 'equal' when they are observationally equivalent—or, more precisely, when they are related by a (quasi-)bisimulation.

Definition 4.1. We define a family of relations $\sim$ that we collectively call *quasi-bisimilarity.* Given systems $\vartheta := (X, \vartheta^o, \vartheta^u)$ and $\psi := (Y, \psi^o, \psi^u)$ in $\mathbf{Coalg}^{\mathbb{T}}_{\mathcal{P}}(p)$ and a section σ of p, we first define the *trace*[4] or *trajectory* of ϑ given σ as the morphism

$$\mathsf{tr}(\theta, \sigma) := \mathbb{T} \times X \xrightarrow{\vartheta^o(\text{-})^*\sigma} \sum_{t:\mathbb{T}} \sum_{x:X} p[\vartheta^o(t,x)] \xrightarrow{(\vartheta^u)^{\rhd \mathbb{T}}} \mathbb{T} \times \mathcal{P}\,X \xrightarrow{\mathcal{P}\,\vartheta^o} \mathcal{P}\,p(1)\,.$$

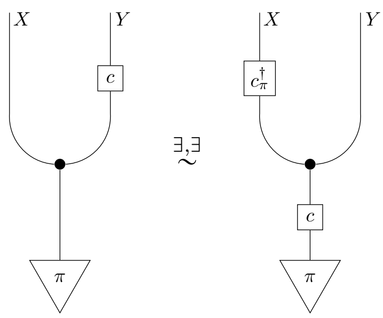
Supposing $\alpha : 1 \to \mathcal{P}\,X$ and $\beta : 1 \to \mathcal{P}\,Y$ to be corresponding initial states, we define $\vartheta \overset{\alpha,\beta}{\sim} \psi$ as the relation

$$\vartheta \overset{\alpha,\beta}{\sim} \psi \iff \forall \sigma : \Gamma(p).\, \forall t : \mathbb{T}.\, \mathsf{tr}(\vartheta, \sigma)(t) \bullet \alpha = \mathsf{tr}(\psi, \sigma)(t) \bullet \beta\,,$$

where we write $g \bullet f$ to indicate Kleisli composition $g \bullet f = \mu \circ \mathcal{P}\,g \circ f$ (where μ is the multiplication of the monad $\mathcal{P}$). We write $\vartheta \overset{\exists,\exists}{\sim} \psi$ when there exists some α, β such that $\vartheta \overset{\alpha,\beta}{\sim} \psi$, and likewise for $\vartheta \overset{\forall,\forall}{\sim} \psi$, $\vartheta \overset{\forall,\exists}{\sim} \psi$, and $\vartheta \overset{\exists,\forall}{\sim} \psi$.

In light of this definition, we can define an appropriate notion of Bayesian inversion for $\mathbf{Hier}|_{\mathcal{E}}$:

Definition 4.2. We say that a system $c : Xy \to Yy$ in $\mathbf{Hier}|_{\mathcal{E}}$ *admits Bayesian inversion* with respect to $\pi : y \to Xy$, if there exists a system $c^\dagger_\pi : Yy \to Xy$ satisfying the equation [9, eq. 5]:



We call $c^\dagger_\pi$ the *Bayesian inversion* of c with respect to π, and call the defining relation the *dynamical Bayes' rule.*

5. The Laplace doctrine of predictive processing

In real-world systems, however, even such quasi-bisimulation is too strong. In the setting of computational neuroscience, it is proposed [10, 11] that certain neural circuits implement *approximate* Bayesian inference by optimizing certain statistical games [12]. A statistical game consists of a Bayesian lens—a

[4]Note that this is in analogy with the *coalgebraic trace*, not the trace of *traced monoidal categories*.

pair of a 'forwards' stochastic channel $A \to \mathcal{P}\, B$ and a 'backwards' inversion $\mathcal{P}\, A \times B \to \mathcal{P}\, A$— equipped with a loss function to evaluate the systems predictive performance. Embodied predictive systems such as brains then realize these games as dynamical systems. Here we sketch this functorial semantics, using a category of 'hierarchical bidirectional Stat-systems', following [13, 12].

We noted above that the category **HiBi** resembles a category of lenses, but it does not sufficiently resemble the category of *Bayesian* lenses: notice that the backwards maps of the latter have codomains of the form $\mathcal{P}\, A \times T \to \mathcal{P}\, S$ rather than $A \times T \to S$. For this reason, **HiBi** makes for an inadequate semantic category for predictive processing. However, all is not lost, for we can define a modification of **HiBi** by analogy to the definition of Bayesian lenses as Grothendieck lenses for the indexed category Stat of state-dependent maps [13].

Definition 5.1. Denote by $\mathbf{HiBi}_{\mathcal{P}}$ the following (semi-)(bi)category. Its objects are pairs of objects in $\mathcal{E}$, and its hom-categories $\mathbf{HiBi}_{\mathcal{P}}((A,S),(B,T))$ are given by $\mathbf{Org}^{\mathbb{T}}_{\mathcal{P}}(\mathcal{P}\, Ay^S, By^T)$. Composition is given by the following family of composite maps:

$$
\begin{aligned}
&\mathbf{HiBi}_{\mathcal{P}}((A,S),(B,T)) \times \mathbf{HiBi}_{\mathcal{P}}((B,T),(C,U)) \\
&= \mathbf{Org}^{\mathbb{T}}_{\mathcal{P}}(\mathcal{P}\, Ay^S, By^T) \times \mathbf{Org}^{\mathbb{T}}_{\mathcal{P}}(\mathcal{P}\, By^T, Cy^U) \\
&= \mathbf{Coalg}^{\mathbb{T}}_{\mathcal{P}}([\mathcal{P}\, Ay^S, By^T]) \times \mathbf{Coalg}^{\mathbb{T}}_{\mathcal{P}}([\mathcal{P}\, By^T, Cy^U]) \\
&\to \mathbf{Coalg}^{\mathbb{T}}_{\mathcal{P}}([\mathcal{P}\, Ay^S, By^T] \otimes [\mathcal{P}\, By^T, Cy^U]) \\
&\to \mathbf{Coalg}^{\mathbb{T}}_{\mathcal{P}}([\mathcal{P}\, Ay^S, \mathcal{P}\, By^T] \otimes [\mathcal{P}\, By^T, Cy^U]) \\
&\to \mathbf{Coalg}^{\mathbb{T}}_{\mathcal{P}}([\mathcal{P}\, Ay^S, Cy^U]) \\
&= \mathbf{HiBi}_{\mathcal{P}}((A,S),(C,U))
\end{aligned}
$$

where the fourth line is generated from the monadic unit $\eta_B : B \to \mathcal{P}\, B$ by $\mathbf{Coalg}([\mathcal{P}\, y^S, (\eta_B)y^T])$.

Remark 5.2. Note that we say 'semi-'(bi)category: this is because $\mathbf{HiBi}_{\mathcal{P}}$ does not have identities. This is not problematic for our work here; and of course $\mathbf{Org}^{\mathbb{T}}_{\mathcal{P}}$ itself does have identities.

We are now in a position to sketch the 'Laplace doctrine' of dynamical semantics for approximate inference. We first recall the notion of D-Bayesian inference game [12]:

Definition 5.3 (Bayesian inference)**.** Let $D : \mathcal{K}\ell(\mathcal{P})(I,X) \times \mathcal{K}\ell(\mathcal{P})(1,X) \to \mathbb{R}$ be a measure of divergence between states on X. Then a (simple) *D-Bayesian inference* game is a statistical game $(X,X) \to (Y,Y)$ with fitness function $\phi : \mathcal{K}\ell(\mathcal{P})(1,X) \times \mathcal{K}\ell(\mathcal{P})(Y,X) \to \mathbb{R}$ given by $\phi(\pi,k) = \mathbb{E}_{y\sim k\bullet c\bullet\pi}\left[D\left(c'_\pi(y), c^\dagger_\pi(y)\right)\right]$, where (c,c') constitutes the lens part of the game and $c^\dagger_\pi$ is the exact inversion of c with respect to π.

Write D_{KL} for the Kullback-Leibler divergence. Given a D_{KL}-Bayesian inference game $(\gamma,\rho,\phi) : (X,X) \to (Y,Y)$ where X and Y are Euclidean spaces and whose forward and backward channels are constrained to output Gaussian distributions, the Laplace doctrine returns a hierarchical bidirectional Stat-system minimizing an upper bound on the divergence between each approximate posterior ρ_π and the 'true' posterior $\gamma^\dagger_\pi$, for any Gaussian state $\pi : \mathcal{P}\, X$.

Remark 5.4. Note that the statistical properties of the system are not the focus of this paper: this doctrine is merely being used to illustrate the coalgebraic framework.

The Laplace doctrine hinges on the following approximation, whose proof we defer to A.9.

Lemma 5.5 (Laplace approximation)**.** Given a D_{KL}-Bayesian inference game $(\gamma,\rho,\phi) : (X,X) \to (Y,Y)$ with forwards channel $\gamma : X \to \mathcal{P}\, Y$ constrained to emit Gaussian distributions, write $\mu_\gamma(x)$:

$\mathbb{R}^{|Y|}$ for the mean of $\gamma(x)$ and $\Sigma_\gamma(x) : \mathbb{R}^{|Y|\times|Y|}$ for its covariance matrix, and assume that for all $y : Y$, the eigenvalues of $\Sigma_{\rho_\pi}(y)$ are small.

Then the loss $\phi : \mathcal{K}\ell(\mathcal{P})(1, X) \times \mathcal{K}\ell(\mathcal{P})(Y, X) \to \mathbb{R}$ is approximately bounded from above by

$$\begin{aligned}\phi(\pi, k) &= \underset{y\sim k\bullet\gamma\bullet\pi}{\mathbb{E}}\left[D\left(\rho_\pi(y), \gamma_\pi^\dagger(y)\right)\right] \\ &\leqslant \underset{y\sim k\bullet\gamma\bullet\pi}{\mathbb{E}}\left[D\left(\rho_\pi(y), \gamma_\pi^\dagger(y)\right) - \log p_{\gamma\bullet\pi}(y)\right] \\ &= \underset{y\sim k\bullet\gamma\bullet\pi}{\mathbb{E}}\left[\mathcal{F}(y)\right] \approx \underset{y\sim k\bullet\gamma\bullet\pi}{\mathbb{E}}\left[\mathcal{F}^L(y)\right]\end{aligned}$$

where $\mathcal{F}$ is called the *free energy* and where $\mathcal{F}^L$ is its *Laplace approximation*,

$$\begin{aligned}\mathcal{F}^L(y) &= E_{(\pi,\gamma)}\left(\mu_{\rho_\pi}(y), y\right) - S_X\left[\rho_\pi(y)\right] \\ &= -\log p_\gamma(y|\mu_{\rho_\pi}(y)) - \log p_\pi(\mu_{\rho_\pi}(y)) - S_X\left[\rho_\pi(y)\right]\end{aligned} \tag{1}$$

where $S_x[\rho_\pi(y)] = \mathbb{E}_{x\sim\rho_\pi(y)}[-\log p_{\rho_\pi}(x|y)]$ is the Shannon entropy of $\rho_\pi(y)$, and $p_\gamma : Y \times X \to [0, 1]$, $p_\pi : X \to [0, 1]$, and $p_{\rho_\pi} : X \times Y \to [0, 1]$ are density functions for γ, π, and ρ_π respectively. The approximation is valid when Σ_{ρ_π} satisfies

$$\Sigma_{\rho_\pi}(y) = \left(\partial_x^2 E_{(\pi,\gamma)}\right)\left(\mu_{\rho_\pi}(y), y\right)^{-1}\,. \tag{2}$$

With this approximation in hand, and given such a statistical game (γ, ρ, ϕ), we will construct a hierarchical bidirectional Stat-system $\mathsf{Laplace}(\gamma, \rho, \phi)$ performing approximate stochastic gradient descent on the loss function, with respect to the statistical parameters of the inversions ρ_π. We will work in discrete time, $\mathbb{T} = \mathbb{N}$, although all of what follows can be done in continuous time, $\mathbb{T} = \mathbb{R}_+$, by replacing the discrete update steps by stochastic differential equations.

Since the entropy $S_X\left[\rho_\pi(y)\right]$ depends only on the variance $\Sigma_{\rho_\pi}(y)$, to optimize the mean $\mu_{\rho_\pi}(y)$ it suffices to consider only the energy $E_{(\pi,\gamma)}(\mu_{\rho_\pi}(y), y)$. We have

$$\begin{aligned}E_{(\pi,\gamma)}(x, y) &= -\log p_\gamma(y|x) - \log p_\pi(x) \\ &= -\frac{1}{2}\left\langle \epsilon_\gamma(y, x), \Sigma_\gamma(x)^{-1}\epsilon_\gamma(y, x)\right\rangle - \frac{1}{2}\left\langle \epsilon_\pi(x), {\Sigma_\pi}^{-1}\epsilon_\pi(x)\right\rangle \\ &\quad + \log\sqrt{(2\pi)^{|Y|}\det\Sigma_\gamma(x)} + \log\sqrt{(2\pi)^{|X|}\det\Sigma_\pi}\end{aligned}$$

and a straightforward computation shows that

$$\partial_x E_{(\pi,\gamma)}(x, y) = -\partial_x\mu_\gamma(x)^T\Sigma_\gamma(x)^{-1}\epsilon_\gamma(y, x) + {\Sigma_\pi}^{-1}\epsilon_\pi(x)\,.$$

Let $\eta_\gamma(y, x) := \Sigma_\gamma(x)^{-1}\epsilon_\gamma(y, x)$ and $\eta_\pi(x) := {\Sigma_\pi}^{-1}\epsilon_\pi(x)$, so that

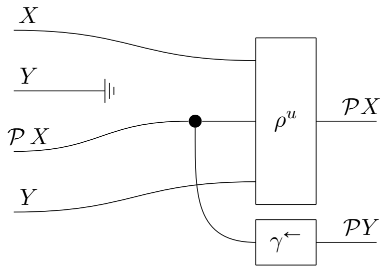
$$\partial_x E_{(\pi,\gamma)}(x, y) = -\partial_x\mu_\gamma(x)^T\eta_\gamma(y, x) + \eta_\pi(x)\,. \tag{3}$$

Note that $E_{(\pi,\gamma)}$ defines a function $X \times Y \to \mathbb{R}$. We will use the domain $X \times Y$ of this function as the state space of our system. To avoid ambiguity, we will write $\overrightarrow{X}$ to indicate the space X when it is used as an input in the 'forwards' direction, and $\overleftarrow{Y}$ to indicate the space Y when it is used as an input in the 'backwards' direction.

Our system $\mathsf{Laplace}(\gamma, \rho, \phi)$ will therefore have the type

$$\begin{aligned}&(X \times Y, \quad \beta_1^o : X \times Y \times \mathcal{P}\overrightarrow{X} \to \overrightarrow{Y}, \\ &\beta_2^o : X \times Y \times \mathcal{P}\overrightarrow{X} \times \overleftarrow{Y} \to \overleftarrow{X}, \\ &\beta^u : X \times Y \times \mathcal{P}\overrightarrow{X} \times \overleftarrow{Y} \to \mathcal{P}(X \times Y)).\end{aligned}$$

We define β_1^o to be the projection of the second factor Y of the state space onto Y, and β_2^o to be the projection of the first factor X onto X. The update map $\beta^u : X \times Y \times \mathcal{P}\overrightarrow{X} \times \overleftarrow{Y} \to \mathcal{P}(X \times Y)$ is then given by composing the commutativity (or 'double strength') of the monad $\mathcal{P}$, $\mathsf{dst} : \mathcal{P}X \times \mathcal{P}Y \to \mathcal{P}(X \times Y)$, after the following map (represented as a string diagram in $\mathcal{E}$):



where $(-)^{\leftarrow} := \mu^{\mathcal{P}} \circ \mathcal{P}(-)$ denotes Kleisli extension (for $\mu^{\mathcal{P}}$ the multiplication of the monad $\mathcal{P}$), so that $\gamma^{\leftarrow} := \mu_Y^{\mathcal{P}} \circ \mathcal{P}(\gamma) : \mathcal{P}X \to \mathcal{P}Y$.

In turn, the map $\rho^u : X \times \mathcal{P}X \times Y \to \mathcal{P}X$ is defined by

$$\rho^u : X \times \mathcal{P}X \times Y \to \mathbb{R}^{|X|} \times \mathbb{R}^{|X| \times |X|} \hookrightarrow \mathcal{P}X$$
$$(x, \pi, y) \mapsto \left(x - \lambda \partial_x E_{(\pi,\gamma)}(x, y), \Sigma^*_{\rho_\pi}(y)\right)$$

where the inclusion into $\mathcal{P}X$ picks the Gaussian state with the given statistical parameters, where $\lambda : \mathbb{R}_+$ is some choice of "learning rate", where $\Sigma^*_{\rho_\pi}(y)$ is as above and in Equation (2), and where $\partial_x E_{(\pi,\gamma)}(x, y)$ is as in Equation (3).

Observe that the factor ρ^u performs approximate stochastic gradient descent on the free energy: for a given input $y : Y$, the mean trajectory of the system follows the update law $\mu_\rho \mapsto \mu_\rho - \lambda \partial_{\mu_\rho} E_{(\pi,\gamma)}(\mu_\rho, y)$, and, when $\Sigma_\rho(y) = \Sigma^*_\rho(y)$, we have $\partial_{\mu_\rho} E_{(\pi,\gamma)}(\mu_\rho, y) \approx \partial_{\mu_\rho} \mathcal{F}(y)$. Note also that the update map ρ^u depends on a prior, just as the inversion map ρ of the lens (γ, ρ) does.

A full treatment of the Laplace doctrine will appear in a forthcoming sequeal to the author's [12].

6. Conclusions; current and future work

In this work we have sketched a framework for treating open dynamical systems of a general nature as coalgebras for certain polynomial functors or—in the case of systems with side-effects such as randomness—certain generalizations thereof. Although we have attempted to give a wide overview of the applicability of these structures, with a particular focus on the adaptive systems of primary interest to the author, we are aware that we have barely scratched the surface of their use and relationships. Here, we briefly list some avenues of current and future work.

Our current principal focus is on exploring the connections between these structures and other compositional treatments of dynamical systems. In particular, relating our categories to the respective frameworks of Myers [14], Libkind [15] and Baez and colleagues (*e.g.*, [16]). Evidently, the structures presented here are most closely in line with the approaches explored by Spivak [8, 17], and are particularly interested in generalizing his topos-theoretic perspective: given that the category of discrete-time deterministic systems over a polynomial p forms a topos, we suspect that so too does $\mathbf{Coalg}^{\mathbb{T}}(p)$. We are also seeking the connections between these putative topoi and the topoi of behaviour types [17] as well as with coalgebraic logic [18], particularly in its modal forms. We hope that we can further develop the theory of $\mathbf{Poly}_M$ to support some of these methods, too.

Finally, there are a number of ways in which this framework should be made more elegant. In particular, we hope to cast a number of properties instead as structures, including the comonoid-homomorphism property of our main definition, and the explicit definitions of random and bundle dynamical systems. With particular respect to the latter, we expect there to be an inductive story of nested parameterization, which appears to the author to have an opetopic shape closely connected to the $\mathbf{Para}$ construction [19].

7. References

1. Lawvere, F.W., and Schnauel, S.H.: Conceptual mathematics : a first introduction to categories. Cambridge University Press, Cambridge, UK New York (2009). DOI: 10.1017/CBO9780511804199

2. Spivak, D.I., and Niu, N.: Polynomial Functors: A General Theory of Interaction (2021)

3. Spivak, D.I.: A reference for categorical structures on **Poly**. (2022)

4. Ahman, D., and Uustalu, T.: Directed Containers as Categories. EPTCS 207, 2016, pp. 89-98 (2016). DOI: 10.4204/EPTCS.207.5

5. Spivak, D.I.: Poly: An abundant categorical setting for mode-dependent dynamics. (2020)

6. Arnold, L.: Random Dynamical Systems. Springer Berlin Heidelberg (1998)

7. Fritz, T.: A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. (2019)

8. Spivak, D.I.: Learners' languages. (2021)

9. Cho, K., and Jacobs, B.: Disintegration and Bayesian Inversion via String Diagrams. Math. Struct. Comp. Sci. 29 (2019) 938-971 (2017). DOI: 10.1017/S0960129518000488

10. Friston, K.: A theory of cortical responses. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 360(1456), 815–836 (2005). DOI: 10.1098/rstb.2005.1622

11. Buckley, C.L., Kim, C.S., McGregor, S., and Seth, A.K.: The free energy principle for action and perception: A mathematical review. Journal of Mathematical Psychology 81, 55–79 (2017)

12. St. Clere Smithe, T.: Compositional Active Inference I: Bayesian Lenses. Statistical Games. (2021)

13. St. Clere Smithe, T.: Bayesian Updates Compose Optically. (2020)

14. Myers, D.J.: Double Categories of Open Dynamical Systems. In: Applied Category Theory 2020 (2020)

15. Libkind, S.: An Algebra of Resource Sharing Machines. (2020)

16. Baez, J.C., Courser, K., and Vasilakopoulou, C.: Structured versus Decorated Cospans. (2021)

17. Schultz, P., Spivak, D.I., and Vasilakopoulou, C.: Dynamical Systems and Sheaves. Applied Categorical Structures (2016)

18. Jacobs, B.: Introduction to Coalgebra. Cambridge University Press (2017)

19. Capucci, M., Gavranović, B., and St. Clere Smithe, T.: Parameterized Categories and Categories by Proxy. In: Category Theory 2021 (2021)

A. Extra proofs and structures

Proposition A.1. $\mathbf{Coalg}^{\mathbb{T}}(p)$ extends to a polynomially-indexed category, $\mathbf{Coalg}^{\mathbb{T}} : \mathbf{Poly}_{\mathcal{E}} \to \mathbf{Cat}$. Suppose $\varphi : p \to q$ is a morphism of polynomials. We define a corresponding functor $\mathbf{Coalg}^{\mathbb{T}}(\varphi) : \mathbf{Coalg}^{\mathbb{T}}(p) \to \mathbf{Coalg}^{\mathbb{T}}(q)$ as follows. Suppose $(X, \vartheta^o, \vartheta^u) : \mathbf{Coalg}^{\mathbb{T}}(p)$ is an object (dynamical system) in $\mathbf{Coalg}^{\mathbb{T}}(p)$. Then $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(X, \vartheta^o, \vartheta^u)$ is defined as the triple $(X, \varphi_1 \circ \vartheta^o, \vartheta^u \circ \vartheta^{o*}\varphi^{\#}) : \mathbf{Coalg}^{\mathbb{T}}(q)$, where the two maps are explicitly the following composites:

$$\mathbb{T} \times X \xrightarrow{\vartheta^o} p(1) \xrightarrow{\varphi_1} q(1)\,, \qquad \sum_{t:\mathbb{T}} \sum_{x:X} q[\varphi_1 \circ \vartheta^o(t,x)] \xrightarrow{\vartheta^{o*}\varphi^{\#}} \sum_{t:\mathbb{T}} \sum_{x:X} p[\vartheta^o(t,x)] \xrightarrow{\vartheta^u} X\,.$$

On morphisms, $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(f) : \mathbf{Coalg}^{\mathbb{T}}(\varphi)(X, \vartheta^o, \vartheta^u) \to \mathbf{Coalg}^{\mathbb{T}}(\varphi)(Y, \psi^o, \psi^u)$ is given by the same underlying map $f : X \to Y$ of state spaces.

Proof. We need to check that $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(X, \vartheta^o, \vartheta^u)$ satisfies the flow conditions of Definition 2.1, that $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(f)$ satisfies the naturality condition of Proposition 2.3, and that $\mathbf{Coalg}^{\mathbb{T}}$ is functorial with respect to polynomials. We begin with the flow condition. Given a section $\tau : q(1) \to \sum_{j:q(1)} q[j]$ of q, we require the closures $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(\vartheta)^\tau : \mathbb{T} \times X \to X$ given by

$$\sum_{t:\mathbb{T}} X \xrightarrow{\vartheta^o(-)^*\tau} \sum_{t:\mathbb{T}} \sum_{x:X} q[\varphi_1 \circ \vartheta^o(t,x)] \xrightarrow{\vartheta^{o*}\varphi^\#} \sum_{t:\mathbb{T}} \sum_{x:X} p[\vartheta^o(t,x)] \xrightarrow{\vartheta^u} X$$

to satisfy $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(\vartheta)^\tau(0) = \mathsf{id}_X$ and $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(\vartheta)^\tau(s{+}t) = \mathbf{Coalg}^{\mathbb{T}}(\varphi)(\vartheta)^\tau(s) \circ \mathbf{Coalg}^{\mathbb{T}}(\varphi)(\vartheta)^\tau(t)$. Note that the following diagram commutes, by the definition of $\varphi^\#$,

$$\begin{array}{ccccc}
\sum_{i:p(1)} p[i] & \xleftarrow{\varphi^\#} & \sum_{i:p(1)} q[\varphi_1(i)] & \xleftarrow{\varphi_1^*\tau} & p(1) \\
\downarrow{\scriptstyle p} & & \downarrow{\scriptstyle \varphi_1^* q} & \swarrow\!\!\!\!= & \\
p(1) & = & p(1) & &
\end{array}$$

so that $\varphi^\# \circ \varphi_1^*\tau$ is a section of p. Therefore, letting $\sigma := \varphi^\# \circ \varphi_1^*\tau$, for $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(\vartheta)^\tau$ to satisfy the flow condition for τ reduces to ϑ^σ satisfying the flow condition for σ. But this is given *ex hypothesi* by Definition 2.1, for any such section σ, so $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(\vartheta)^\tau$ satisfies the flow condition for τ. And since τ was any section, we see that $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(\vartheta)$ satisfies the flow condition generally.

The proof that $\mathbf{Coalg}^{\mathbb{T}}(\varphi)(f)$ satisfies the naturality condition of Proposition 2.3 proceeds similarly. Supposing again that τ is any section of q, we require the following diagram to commute for any time $t : \mathbb{T}$:

$$\begin{array}{ccccccc}
X & \xrightarrow{\vartheta^o(t)^*\varphi_1^*\tau} & \sum_{x:X} q[\varphi_1 \circ \vartheta^o(t,x)] & \xrightarrow{\vartheta^o(t)^*\varphi^\#} & \sum_{x:X} p[\vartheta^o(t,x)] & \xrightarrow{\vartheta^u(t)} & X \\
\downarrow{\scriptstyle f} & & & & & & \downarrow{\scriptstyle f} \\
Y & \xrightarrow{\psi^o(t)^*\varphi_1^*\tau} & \sum_{y:Y} q[\varphi_1 \circ \psi^o(t,x)] & \xrightarrow{\vartheta^o(t)^*\varphi^\#} & \sum_{y:Y} p[\psi^o(t,x)] & \xrightarrow{\psi^u(t)} & Y
\end{array}$$

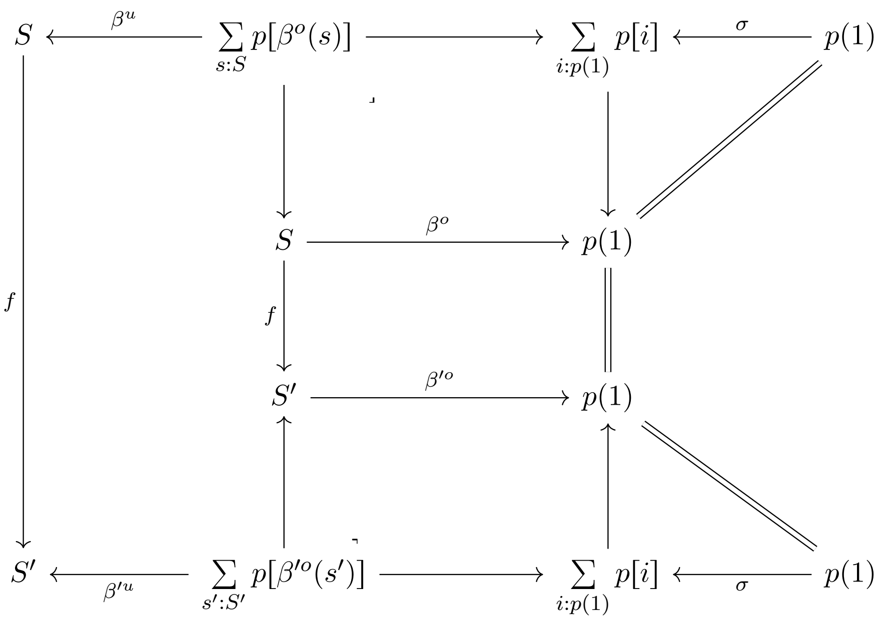
Again letting $\sigma := \varphi^\# \circ \varphi_1^*\tau$, we see that this diagram reduces exactly to the diagram in Proposition 2.3 by the functoriality of pullback, and since f makes that diagram commute, it must also make this diagram commute.

Finally, to show that $\mathbf{Coalg}^{\mathbb{T}}$ is functorial with respect to polynomials amounts to checking that composition and pullback are functorial; but this is a basic result of category theory. □

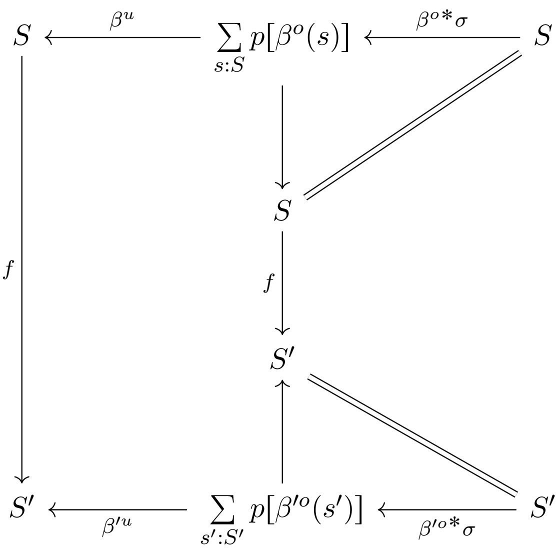
Proposition A.2. When $\mathbb{T} = \mathbb{N}$, the category $\mathbf{Coalg}^{\mathbb{N}}(p)$ of open dynamical systems over p with time $\mathbb{N}$ is equivalent to the topos p-$\mathbf{Coalg}$ of p-coalgebras [8].

Proof. p-$\mathbf{Coalg}$ has as objects pairs (S, β) where $S : \mathcal{E}$ is an object in $\mathcal{E}$, $\beta : S \to p \lhd S$ is a morphism of polynomials (interpreting S as the constant copresheaf on the set S), and $\lhd$ denotes the composition monoidal product in $\mathbf{Poly}_{\mathcal{E}}$ (*i.e.*, composing the corresponding copresheaves $\mathcal{E} \to \mathcal{E}$). A straightforward computation shows that, interpreted as an object in $\mathcal{E}$, $p \lhd S$ corresponds to $\sum_{i:p(1)} S^{p[i]}$. By the universal property of the dependent sum, a morphism $\beta : S \to \sum_{i:p(1)} S^{p[i]}$ therefore corresponds bijectively to a pair of maps $\beta^o : S \to p(1)$ and $\beta^u : \sum_{s:S} p[\beta^o(s)] \to X$. By Proposition 2.6, such a pair is equivalently a discrete-time open dynamical system over p with state space S: that is, the objects of p-$\mathbf{Coalg}$ are in bijection with those of $\mathbf{Coalg}^{\mathbb{N}}(p)$.

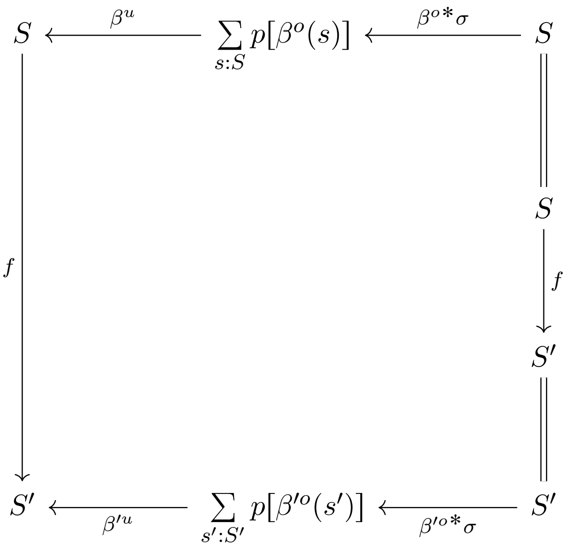
Next, we show that the hom-sets p-$\mathbf{Coalg}\big((S,\beta),(S',\beta')\big)$ and $\mathbf{Coalg}^{\mathbb{N}}(p)\big((S,\beta^o,\beta^u),(S',\beta'^o,\beta'^u)\big)$ are in bijection. A morphism $f : (S,\beta) \to (S',\beta')$ of p-coalgebras is a morphism $f : S \to S'$ between the state spaces such that $\beta' \circ f = (p \lhd f) \circ \beta$. Unpacking this, we find that this means the following diagram in $\mathcal{E}$ must commute for any section σ of p:



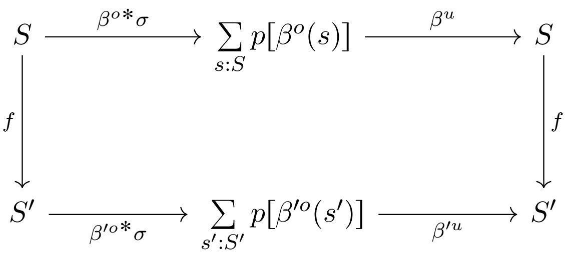
Pulling the arbitrary section σ back along the 'output' maps β^o and β'^o means that the following commutes:



Forgetting the vertical projections out of the pullbacks gives:

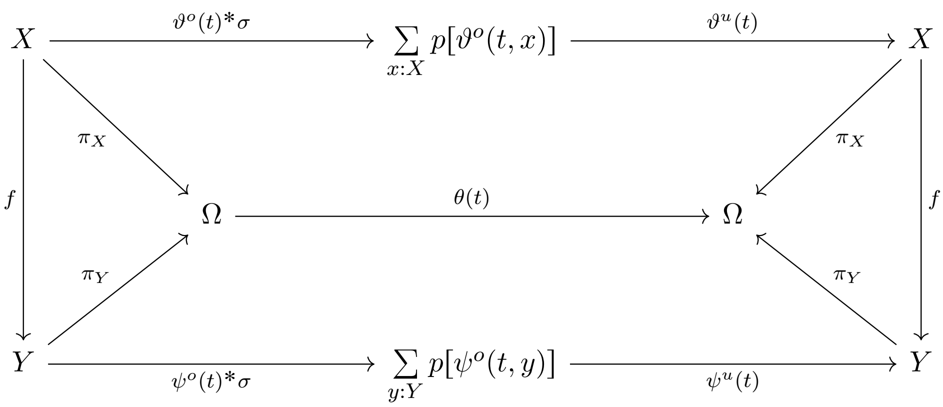


Finally, by collapsing the identity maps and reflecting the diagram horizontally, we obtain



which we recognize from Proposition 2.3 as the defining characteristic of a morphism in $\mathbf{Coalg}^{\mathbb{N}}(p)$. Finally, we note that each of these steps is bijective, and so we have the desired bijection of hom-sets. $\square$

Definition A.3 (Category of open random dynamical systems over p)**.** Writing $\vartheta := (\pi_X, \vartheta^o, \vartheta^u)$ and $\psi := (\pi_Y, \psi^o, \psi^u)$, a morphism $f : \vartheta \to \psi$ is a map $f : X \to Y$ in $\mathcal{E}$ making the following diagram commute for all times $t : \mathbb{T}$ and sections σ of p:



Identities are given by the identity maps on state-spaces. Composition is given by pasting of diagrams.

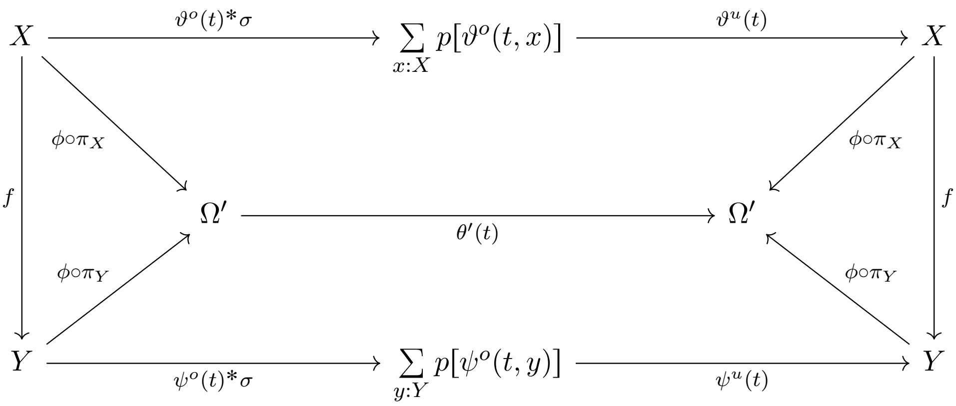
Proposition A.4 (Opindexed category of open random dynamical systems over polynomials)**.** By the universal property of the product $\times$ in $\mathbf{Cat}$, it suffices to define the actions of $\mathbf{RDyn}^{\mathbb{T}}$ separately on morphisms of polynomials and on morphisms of closed measure-preserving systems.

Suppose therefore that $\varphi : p \to q$ is a morphism of polynomials. Then, for each measure-preserving system $(\theta, \beta) : \mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})_{\mathcal{P}}$, we define the functor $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta) : \mathbf{RDyn}^{\mathbb{T}}(p, \theta) \to \mathbf{RDyn}^{\mathbb{T}}(q, \theta)$ as follows. Let $\vartheta := (\pi_X : X \to \Omega, \vartheta^o, \vartheta^u) : \mathbf{RDyn}^{\mathbb{T}}(p, \theta)$ be an object (open random dynamical system) in $\mathbf{RDyn}^{\mathbb{T}}(p, \theta)$. Then $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(\vartheta)$ is defined as the triple $(\pi_X, \varphi_1 \circ \vartheta^o, \vartheta^u \circ \varphi^{o*}\varphi^{\#}) : \mathbf{RDyn}^{\mathbb{T}}(q, \theta)$, where the two maps are explicitly the following composites:

$$\mathbb{T} \times X \xrightarrow{\vartheta^o} p(1) \xrightarrow{\varphi_1} q(1)\,, \qquad \sum_{t:\mathbb{T}} \sum_{x:X} q[\varphi_1 \circ \vartheta^o(t,x)] \xrightarrow{\vartheta^{o*}\varphi^{\#}} \sum_{t:\mathbb{T}} \sum_{x:X} p[\vartheta^o(t,x)] \xrightarrow{\vartheta^u} X\,.$$

On morphisms $f : (\pi_X : X \to \Omega, \vartheta^o, \vartheta^u) \to (\pi_Y : Y \to \Omega, \psi^o, \psi^u)$, the image $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(f) : \mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(\pi_X, \vartheta^o, \vartheta^u) \to \mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(\pi_Y, \psi^o, \psi^u)$ is given by the same underlying map $f : X \to Y$ of state spaces.

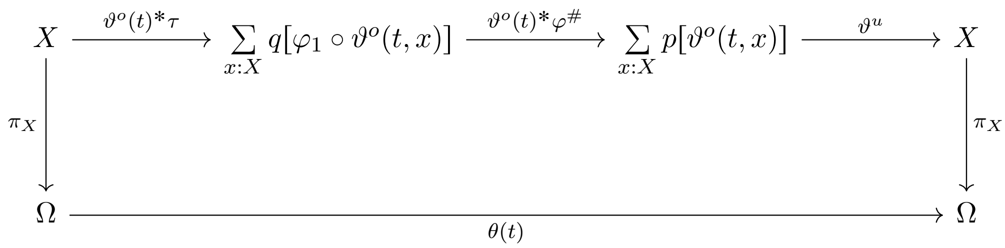
Next, suppose that $\phi : (\theta, \beta) \to (\theta', \beta')$ is a morphism of closed measure-preserving dynamical systems, and let $\Omega' := \theta'(*)$ be the state space of the system θ'. By Proposition 2.22, the morphism ϕ corresponds to a map $\phi : \Omega \to \Omega'$ on the state spaces that preserves both flow and measure. Therefore, for each polynomial $p : \mathbf{Poly}_{\mathcal{E}}$, we define the functor $\mathbf{RDyn}^{\mathbb{T}}(p, \phi) : \mathbf{RDyn}^{\mathbb{T}}(p, \theta) \to \mathbf{RDyn}^{\mathbb{T}}(p, \theta')$ by post-composition. That is, suppose given open random dynamical systems and morphisms over (p, θ) as in the diagram of Proposition 2.26. Then $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)$ returns the following diagram:



That is, $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)(\vartheta) := (\phi \circ \pi_X, \vartheta^o, \vartheta^u)$ and $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)(f)$ is given by the same underlying map $f : X \to Y$ on state spaces.

Proof. We need to check: the naturality condition of Definition 2.25 for both $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(\vartheta)$ and $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)(\vartheta)$; functoriality of $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)$ and $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)$; and (pseudo)functoriality of $\mathbf{RDyn}^{\mathbb{T}}$ with respect to both morphisms of polynomials and of closed measure-preserving systems.

We begin by checking that the conditions of Definition 2.25 are satisfied by the objects $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(\pi_X, \vartheta^o, \vartheta^u) : \mathbf{RDyn}^{\mathbb{T}}(q, \theta)$ and morphisms $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(f) : \mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(\pi_X, \vartheta^o, \vartheta^u) \to \mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(\pi_Y, \psi^o, \psi^u)$ in the image of $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)$. Given a section $\tau : q(1) \to \sum_{j:q(1)} q[j]$ of q, we need to check that the closure $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(\vartheta)^{\tau}$ forms a closed random dynamical system in $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})/\theta$. That is to say, for all $t : \mathbb{T}$ and sections τ, we need to check that the following naturality square commutes:



As before, we find that $\varphi^{\#} \circ \varphi_1^* \tau$ is a section of p, so that commutativity of the diagram above reduces to commutativity of the diagram in Definition 2.25. Similarly, given a morphism $f : (\pi_X, \vartheta^o, \vartheta^u) \to$

(π_Y, ψ^o, ψ^u), we need to check that the diagram in Proposition 2.26 induced for $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(f)$ commutes for all times $t : \mathbb{T}$ and sections τ of q. But given such a section τ, the diagram for $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)(f)$ reduces to that for f and the section $\varphi^{\#} \circ \varphi_1^* \tau$ of p, which commutes *ex hypothesi*; and functoriality of $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)$ follows immediately.

Next, we check that the conditions of Definition 2.25 are satisfied in the image of $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)$. It is clear by the definition of the action of $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)$ that the condition that the diagram in Proposition A.1 commutes is satisfied, from which it follows by pasting that $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)$ is functorial. We therefore just have to check the induced diagram in Definition 2.25 commutes. Consider the following diagram:

$$\begin{array}{ccccc}
X & \xrightarrow{\vartheta^o(t)^*\sigma} & \sum_{x:X} p[\vartheta^o(t,x)] & \xrightarrow{\vartheta^u(t)} & X \\
\Big\downarrow{\scriptstyle \pi_X} & & & & \Big\downarrow{\scriptstyle \pi_X} \\
\Omega & & \xrightarrow{\quad\theta(t)\quad} & & \Omega \\
\Big\downarrow{\scriptstyle \phi} & & & & \Big\downarrow{\scriptstyle \phi} \\
\Omega' & & \xrightarrow{\quad\theta'(t)\quad} & & \Omega'
\end{array}$$

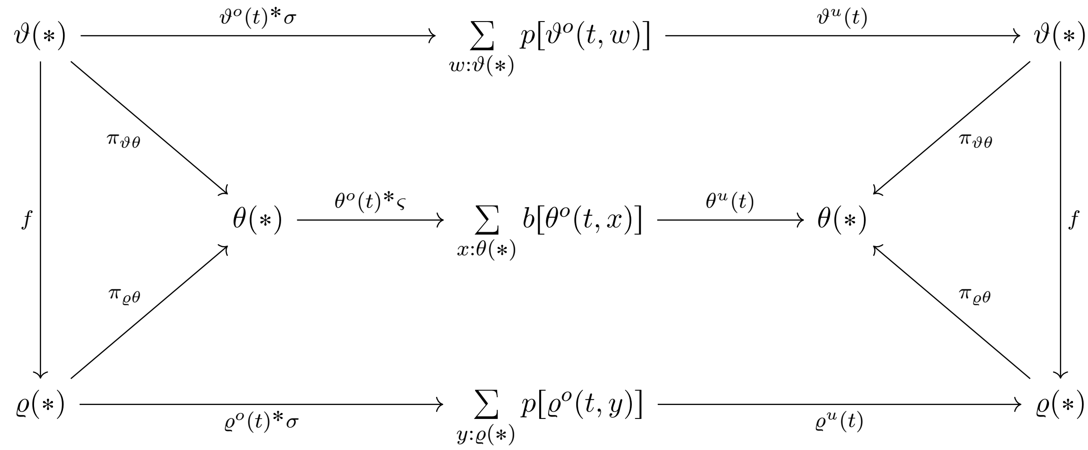
The top square commutes *ex hypothesi*, the bottom square commutes by the definition of morphism of closed measure-preserving dynamical systems (Proposition 2.22), and the outer square is the induced diagram we need to check, which therefore commutes by the pasting of commuting squares.

Finally, we check that $\mathbf{RDyn}^{\mathbb{T}}$ is functorial with respect to morphisms of polynomials and morphisms of closed measure-preserving dynamical systems. These reduce to checking that pullback and composition are functorial, which we again leave to the dedicated reader. □

Definition A.5 (Open bundle dynamical system)**.** Let $p, b : \mathbf{Poly}_{\mathcal{E}}$ be polynomials in $\mathcal{E}$, and let $\theta := (\theta(*), \theta^o, \theta^u) : \mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(b)$ be an open dynamical system over b. An open bundle dynamical system over (p, b, θ) is a pair $(\pi_{\vartheta\theta}, \vartheta)$ where $\vartheta := (\vartheta(*), \vartheta^o, \vartheta^u) : \mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(p)$ is an open dynamical system over p and $\pi_{\vartheta\theta} : \vartheta(*) \to \theta(*)$ is a bundle in $\mathcal{E}$, such that, for all time $t : \mathbb{T}$ and sections σ of p and ς of b, the following diagrams commute, thereby inducing a bundle of closed dynamical systems $\pi^{\sigma\varsigma}_{\vartheta\theta} : \vartheta^{\sigma} \to \theta^{\varsigma}$ in $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})$:

$$\begin{array}{ccccc}
\vartheta(*) & \xrightarrow{\vartheta^o(t)^*\sigma} & \sum_{w:\vartheta(*)} p[\vartheta^o(t,w)] & \xrightarrow{\vartheta^u(t)} & \vartheta(*) \\
\Big\downarrow{\scriptstyle \pi_{\vartheta\theta}} & & & & \Big\downarrow{\scriptstyle \pi_{\vartheta\theta}} \\
\theta(*) & \xrightarrow{\theta^o(t)^*\varsigma} & \sum_{x:\theta(*)} b[\theta^o(t,x)] & \xrightarrow{\theta^u(t)} & \theta(*)
\end{array}$$

Proposition A.6 (Category of open bundle dynamical systems over (p, b))**.** Let $p, b : \mathbf{Poly}_{\mathcal{E}}$ be polynomials in $\mathcal{E}$, and let $\theta := (\theta(*), \theta^o, \theta^u) : \mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(b)$ be an open dynamical system over b. Open bundle dynamical systems over (p, b, θ) form the objects of a category $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta)$. Morphisms $f : (\pi_{\vartheta\theta}, \vartheta) \to (\pi_{\varrho\theta}, \varrho)$ are maps $f : \vartheta(*) \to \varrho(*)$ in $\mathcal{E}$ making the following diagram commute for all times $t : \mathbb{T}$ and sections σ of p and ς of b:



That is, f is a map on the state spaces that induces a morphism $(\pi_{\vartheta\theta}, \vartheta^\sigma) \to (\pi_{\varrho\theta}, \varrho^\sigma)$ in $\mathbf{Cat}(\mathbf{B}\mathbb{T}, \mathcal{E})/\theta^\varsigma$ of bundles of the closures. Identity morphisms are the corresponding identity maps, and composition is by pasting.

Proposition A.7 (Opindexed category of open bundle dynamical systems)**.** Varying the polynomials p in $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta)$ induces an opindexed category $\mathbf{BunDyn}^{\mathbb{T}}(-, b, \theta) : \mathbf{Poly}_{\mathcal{E}} \to \mathbf{Cat}$. On polynomials p, it returns the categories $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta)$ of Proposition A.6. On morphisms $\varphi : p \to q$ of polynomials, define the functors $\mathbf{BunDyn}^{\mathbb{T}}(\varphi, b, \theta) : \mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta) \to \mathbf{BunDyn}^{\mathbb{T}}(q, b, \theta)$ as in Proposition 2.27. That is, suppose $(\pi_{\vartheta\theta}, \vartheta) : \mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta)$ is object (open bundle dynamical system) in $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta)$, where $\vartheta := (\vartheta(*), \vartheta^o, \vartheta^u)$. Then its image $\mathbf{BunDyn}^{\mathbb{T}}(\varphi, b, \theta)(\pi_{\vartheta\theta}, \vartheta)$ is defined as the pair $(\pi_{\vartheta\theta}, \varphi\vartheta)$, where $\varphi\vartheta := (\vartheta(*), \phi_1 \circ \vartheta^o, \vartheta^u \circ \vartheta^{o*}\varphi^\#)$. On morphisms $f : (\pi_{\vartheta\theta}, \vartheta) \to (\pi_{\varrho\theta}, \varrho)$, $\mathbf{BunDyn}^{\mathbb{T}}(\varphi, b, \theta)(f)$ is again given by the same underlying map $f : \vartheta(*) \to \varrho(*)$ of state spaces.

Proof. The proof amounts to the proof for Proposition 2.27 that $\mathbf{RDyn}^{\mathbb{T}}(\varphi, \theta)$ constitutes an indexed category, except that the closed base dynamical system θ of that Proposition is here replaced, for any section ς of b, by the closure θ^ς by ς of the open dynamical system $\theta : \mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(b)$ of the present Proposition. The proof goes through accordingly, since the relevant diagrams are guaranteed to commute for any such ς by the conditions in Definition A.5 and Proposition A.6. □

Proposition A.8 (Doubly-opindexed category of open bundle dynamical systems)**.** Letting the base system θ also vary induces a doubly-opindexed category $\mathbf{BunDyn}^{\mathbb{T}}(-, b, =) : \mathbf{Poly}_{\mathcal{E}} \times \mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(b) \to \mathbf{Cat}$. Given a polynomial $p : \mathbf{Poly}_{\mathcal{E}}$ and morphism $\phi : \theta \to \rho$ in $\mathbf{Coalg}^{\mathbb{T}}_{\mathsf{id}}(b)$, the functor $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \phi) : \mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta) \to \mathbf{BunDyn}^{\mathbb{T}}(p, b, \rho)$ is defined by post-composition, as in Proposition 2.27 for the action of $\mathbf{RDyn}^{\mathbb{T}}$ on morphisms of the base systems there. More explicitly, such a morphism ϕ corresponds to a map $\phi : \theta(*) \to \rho(*)$ of state spaces in $\mathcal{E}$. Given an object $(\pi_{\vartheta\theta}, \vartheta)$ of $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta)$, we define $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \phi)(\pi_{\vartheta\theta}, \vartheta) := (\phi \circ \pi_{\vartheta\theta}, \vartheta)$. Given a morphism $f : (\pi_{\vartheta\theta}, \vartheta) \to (\pi_{\varrho\theta}, \varrho)$ in $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \theta)$, its image $\mathbf{BunDyn}^{\mathbb{T}}(p, b, \phi)(f) : (\phi \circ \pi_{\vartheta\theta}, \vartheta) \to (\phi \circ \pi_{\varrho\theta}, \varrho)$ is given by the same underlying map $f : \vartheta(*) \to \varrho(*)$ of state spaces.

Proof. As for Proposition A.7, the proof here amounts to the proof for Proposition 2.27 that $\mathbf{RDyn}^{\mathbb{T}}(p, \phi)$ constitutes an indexed category, except again the closed systems are replaced by (the appropriate closures of) open ones, and the measure-preserving structure is forgotten. □

Proof A.9 (Proof of the Laplace approximation). First note that the KL divergence is bounded from above by the free energy since $\log p_{\gamma \bullet \pi}(y)$ is always negative.

Next, we can write the density functions as:

$$\begin{aligned}\log p_\gamma(y|x) &= \frac{1}{2}\left\langle \epsilon_\gamma, {\Sigma_\gamma}^{-1}\epsilon_\gamma \right\rangle - \log\sqrt{(2\pi)^{|Y|}\det\Sigma_\gamma} \\ \log p_{\rho_\pi}(x|y) &= \frac{1}{2}\left\langle \epsilon_{\rho_\pi}, {\Sigma_{\rho_\pi}}^{-1}\epsilon_{\rho_\pi} \right\rangle - \log\sqrt{(2\pi)^{|X|}\det\Sigma_{\rho_\pi}} \\ \log p_\pi(x) &= \frac{1}{2}\left\langle \epsilon_\pi, {\Sigma_\pi}^{-1}\epsilon_\pi \right\rangle - \log\sqrt{(2\pi)^{|X|}\det\Sigma_\pi}\end{aligned}$$

where for clarity we have omitted the dependence of Σ_γ on x and Σ_{ρ_π} on y, and where

$$\begin{aligned}\epsilon_\gamma &: Y \times X \to Y : (y,x) \mapsto y - \mu_\gamma(x)\,, \\ \epsilon_{\rho_\pi} &: X \times Y \to X : (x,y) \mapsto x - \mu_{\rho_\pi}(y)\,, \\ \epsilon_\pi &: X \times 1 \to X : (x,*) \mapsto x - \mu_\pi\,.\end{aligned}$$

Then, note that we can write the free energy $\mathcal{F}(y)$ as the difference between expected energy and entropy:

$$\begin{aligned}\mathcal{F}(y) &= \mathop{\mathbb{E}}_{x\sim\rho_\pi(y)}\left[\log\frac{p_{\rho_\pi}(x|y)}{p_\gamma(y|x)\cdot p_\pi(x)}\right] \\ &= \mathop{\mathbb{E}}_{x\sim\rho_\pi(y)}\left[-\log p_\gamma(y|x) - \log p_\pi(x)\right] - S_X\left[\rho_\pi(y)\right] \\ &= \mathop{\mathbb{E}}_{x\sim\rho_\pi(y)}\left[E_{(\pi,\gamma)}(x,y)\right] - S_X\left[\rho_\pi(y)\right]\end{aligned}$$

Next, since the eigenvalues of $\Sigma_{\rho_\pi}(y)$ are small for all $y : Y$, we can approximate the expected energy by its second-order Taylor expansion around the mean $\mu_{\rho_\pi}(y)$:

$$\begin{aligned}\mathcal{F}(y) \approx E_{(\pi,\gamma)}(\mu_{\rho_\pi}(y), y) &+ \frac{1}{2}\left\langle \epsilon_{\rho_\pi}\left(\mu_{\rho_\pi}(y), y\right),\ \left(\partial_x^2 E_{(\pi,\gamma)}\right)\left(\mu_{\rho_\pi}(y), y\right)\cdot\epsilon_{\rho_\pi}\left(\mu_{\rho_\pi}(y), y\right)\right\rangle \\ &- S_X\big[\rho_\pi(y)\big]\,.\end{aligned}$$

where $\left(\partial_x^2 E_{(\pi,\gamma)}\right)(\mu_{\rho_\pi}(y), y)$ is the Hessian of $E_{(\pi,\gamma)}$ with respect to x evaluated at $(\mu_{\rho_\pi}(y), y)$.

Note that

$$\left\langle \epsilon_{\rho_\pi}\left(\mu_{\rho_\pi}(y), y\right),\ \left(\partial_x^2 E_{(\pi,\gamma)}\right)\left(\mu_{\rho_\pi}(y), y\right)\cdot\epsilon_{\rho_\pi}\left(\mu_{\rho_\pi}(y), y\right)\right\rangle = \operatorname{tr}\left[\left(\partial_x^2 E_{(\pi,\gamma)}\right)\left(\mu_{\rho_\pi}(y), y\right)\ \Sigma_{\rho_\pi}(y)\right]\,, \tag{4}$$

that the entropy of a Gaussian measure depends only on its covariance,

$$S_X\big[\rho_\pi(y)\big] = \frac{1}{2}\log\det\left(2\pi\, e\,\Sigma_{\rho_\pi}(y)\right)\,,$$

and that the energy $E_{(\pi,\gamma)}(\mu_{\rho_\pi}(y), y)$ does not depend on $\Sigma_{\rho_\pi}(y)$. We can therefore write down directly the covariance $\Sigma^*_{\rho_\pi}(y)$ minimizing $\mathcal{F}(y)$ as a function of y. We have

$$\partial_{\Sigma_{\rho_\pi}}\mathcal{F}(y) \approx \frac{1}{2}\left(\partial_x^2 E_{(\pi,\gamma)}\right)\left(\mu_{\rho_\pi}(y), y\right) + \frac{1}{2}{\Sigma_{\rho_\pi}}^{-1}\,.$$

Setting $\partial_{\Sigma_{\rho_\pi}}\mathcal{F}(y) = 0$, we find the optimum

$$\Sigma^*_{\rho_\pi}(y) = \left(\partial_x^2 E_{(\pi,\gamma)}\right)\left(\mu_{\rho_\pi}(y), y\right)^{-1}\,.$$

Finally, on substituting $\Sigma^*_{\rho_\pi}(y)$ in equation (4), we obtain the desired expression

$$\mathcal{F}(y) \approx E_{(\pi,\gamma)}\left(\mu_{\rho_\pi}(y), y\right) - S_X\left[\rho_\pi(y)\right] =: \mathcal{F}^L(y)\,.$$

□