Cyber Kittens, or Some First Steps Towards Categorical Cybernetics

Toby St. Clere Smithe
Department of Experimental Psychology,
University of Oxford
arxiv@tsmithe.net

We define a categorical notion of cybernetic system as a dynamical realisation of a generalized open game, along with a coherence condition. We show that this notion captures a wide class of cybernetic systems in computational neuroscience and statistical machine learning, exposes their compositional structure, and gives an abstract justification for the bidirectional structure empirically observed in cortical circuits. Our construction is built on the observation that Bayesian updates compose optically, a fact which we prove along the way, via a fibred category of state-dependent stochastic channels.

1 Introduction

Those systems that we might classify as living, adaptive, or somehow intelligent all display a fundamental property: they resist or avoid perturbations that would result in their existence becoming unsustainable. This means that they must somehow be able to sense their current state of affairs (*perception*) and respond appropriately (*action*). In particular, an adaptive system should sense the relevant aspects of its current environmental state, and form expectations about the consequences of that state. In general, the interaction with the environment will be stochastic, and the statistically optimal method of 'sensing' and prediction is Bayesian inference.

Typically, however, the system has no direct access to the external state, only to sense data that indirectly have external causes. Moreover, sense data are often very high-dimensional, and predicting their consequences is underdetermined. As a result, it is common to assume that successful organisms are imbued with some kind of *generative model* of the process by which external causes generate their sense data. They can then use this model to infer those actions will bring (their beliefs about) their current state closer to those expectations: a process called *active inference*.

Systems such as these are inherently open, and often their internal models and beliefs are supposed to be structured hierarchically—that is, compositionally. The processes of prediction and action sketched here are naturally bidirectional, and indeed our first contribution in the present work is to show that Bayesian inference is abstractly structured as a *category of optics* [21, 6], the emerging canonical formalism for (open) bidirectionally structured compositional systems.

The compositional framework of *open games* [2, 13] builds on categories of optics to describe systems of motivated interacting agents, but it is substantially more general than needed for classical game theory: generalized open games naturally describe any bidirectionally structured open systems that can be associated with a measure of fitness. Consequently, such generalized open games provide a natural home for a compositional theory of interacting cybernetic systems, and using our notion of *Bayesian lens*, we characterize a number of canonical statistical models as *statistical games*.

However, mere open games themselves supply no notion of *dynamics* mediating the interactions. We therefore introduce the concept of *dynamical realisation* of an open game (Definition 4.7), as well as a

Submitted to:
ACT 2020

coherence condition that ensures such a realisation behaves as we would expect from a cybernetic system (Definition 4.8). We use these concepts to show that two prominent frameworks for active inference instantiate such categories of cybernetic systems.

Acknowledgements We thank the organizers of *Applied Category Theory 2020* for the opportunity to present this work, and the anonymous reviewers for helpful comments and questions. We also thank Bruno Gavranović, Jules Hedges, and Neil Ghani for stimulating and insightful conversations, and credit Jules Hedges for observing the correct form of the Bayesian update map in discussions at SYCO 6.

2 Bayesian Updates Compose Optically

We begin by proving that Bayesian updates compose according to the 'lens' pattern [9] that sits at the heart of categories of open games and other 'bidirectional' structures. We first show that Bayesian inversions are 'vertical' maps in a fibred category of state-dependent channels. The Grothendieck construction of this structure gives a category of lenses. Open games are commonly defined using the more general 'optics' pattern [2], and so we also show that, under the Yoneda embedding, our category of lenses is equivalently a category of optics.

Throughout the paper, we work in a general category of stochastic channels; abstractly, this corresponds to a *Markov category* [12] or *copy-delete category* [5]. Familiar examples of such categories include $\mathscr{K}\ell(\mathscr{D})$, the Kleisli category of the finitely-supported distribution monad $\mathscr{D}$, and, for 'continuous' probabiliy, $\mathscr{K}\ell(\mathscr{G})$, the Kleisli category of the Giry monad. We will write $c^\dagger_\pi := c^\dagger_{(\cdot)}(\pi)$ to indicate the **Bayesian inversion** of the channel c with respect to a state π. Then, given some $y \in Y$, $c^\dagger_\pi(y)$ is a new 'posterior' distribution on X. We will call $c^\dagger_\pi(y)$ the **Bayesian update** of π along c given y.

For a substantially expanded version of this section, including proofs and background exposition with precise definitions of Bayesian inversion, see the author's [25]. We will occasionally here refer to definitions or results in that paper.

Definition 2.1 (State-indexed categories)**.** Let $(\mathscr{C},\otimes,I)$ be a monoidal category enriched in a Cartesian closed category $\mathbf{V}$. Define the $\mathscr{C}$-state-indexed category $\mathsf{Stat} : \mathscr{C}^{\mathrm{op}} \to \mathbf{V}\text{-}\mathbf{Cat}$ as follows.

$$\mathsf{Stat} \;:\; \mathscr{C}^{\mathrm{op}} \to \mathbf{V}\text{-}\mathbf{Cat}$$

$$X \mapsto \mathsf{Stat}(X) := \begin{pmatrix} & \mathsf{Stat}(X)_0 & := & \mathscr{C}_0 \\ & \mathsf{Stat}(X)(A,B) & := & \mathbf{V}(\mathscr{C}(I,X),\mathscr{C}(A,B)) \\ \mathsf{id}_A : & \mathsf{Stat}(x)(A,A) & := & \begin{cases} \mathsf{id}_A : \mathscr{C}(I,X) \to \mathscr{C}(A,A) \\ \rho \quad \mapsto \quad \mathsf{id}_A \end{cases} \end{pmatrix} \tag{1}$$

$$f : \mathscr{C}(Y,X) \mapsto \begin{pmatrix} \mathsf{Stat}(f) : & \mathsf{Stat}(X) & \to & \mathsf{Stat}(Y) \\ & \mathsf{Stat}(X)_0 & = & \mathsf{Stat}(Y)_0 \\ & \mathbf{V}(\mathscr{C}(I,X),\mathscr{C}(A,B)) & \to & \mathbf{V}(\mathscr{C}(I,Y),\mathscr{C}(A,B)) \\ & \alpha & \mapsto & f^*\alpha : \big(\sigma : \mathscr{C}(I,Y)\big) \mapsto \big(\alpha(f \bullet \sigma) : \mathscr{C}(A,B)\big) \end{pmatrix}$$

Composition in each fibre $\mathsf{Stat}(X)$ is given by composition in $\mathscr{C}$; that is, by the left and right actions of the profunctor $\mathsf{Stat}(X)(-,=) : \mathscr{C}^{\mathrm{op}} \times \mathscr{C} \to \mathbf{V}$. Explicitly, given $\alpha : \mathbf{V}(\mathscr{C}(I,X),\mathscr{C}(A,B))$ and $\beta : \mathbf{V}(\mathscr{C}(I,X),\mathscr{C}(B,C))$, their composite is $\beta \circ \alpha : \mathbf{V}(\mathscr{C}(I,X),\mathscr{C}(A,C)) := \rho \mapsto \beta(\rho) \bullet \alpha(\rho)$. Since $\mathbf{V}$

is Cartesian, there is a canonical copier $\curlyvee : x \mapsto (x,x)$ on each object, so we can alternatively write $(\beta \circ \alpha)(\rho) = \big(\beta(-) \bullet \alpha(-)\big) \circ \curlyvee \circ \rho$. Note that we indicate composition in $\mathscr{C}$ by $\bullet$ and composition in the fibres $\mathsf{Stat}(X)$ by $\circ$.

Example 2.2. Let $\mathbf{V} = \mathbf{Meas}$ be a 'convenient' (*i.e.*, Cartesian closed) category of measurable spaces, such as the category of quasi-Borel spaces [14], let $\mathscr{P} : \mathbf{Meas} \to \mathbf{Meas}$ be a probability monad defined on this category, and let $\mathscr{C} = \mathscr{K}\ell(\mathscr{P})$ be the Kleisli category of this monad. Its objects are the objects of $\mathbf{Meas}$, and its hom-spaces $\mathscr{K}\ell(\mathscr{P})(A,B)$ are the spaces $\mathbf{Meas}(A, \mathscr{P}B)$ [12]. This $\mathscr{C}$ is a monoidal category of stochastic channels, whose monoidal unit I is the space with a single point. Consequently, states of X are just measures (distributions) in $\mathscr{P}X$. That is, $\mathscr{K}\ell(\mathscr{P})(I,X) \cong \mathbf{Meas}(1, \mathscr{P}X)$. Instantiating Stat in this setting, we obtain:

$$\mathsf{Stat} \;:\; \mathscr{K}\ell(\mathscr{P})^{\mathrm{op}} \to \mathbf{V}\text{-}\mathbf{Cat}$$

$$X \mapsto \mathsf{Stat}(X) := \left(\begin{array}{lcl} & \mathsf{Stat}(X)_0 & := \quad \mathbf{Meas}_0 \\ & \mathsf{Stat}(X)(A,B) & := \quad \mathbf{Meas}(\mathscr{P}X, \mathbf{Meas}(A, \mathscr{P}B)) \\ \mathsf{id}_A : & \mathsf{Stat}(X)(A,A) & := \quad \left\{\begin{array}{ccc} \mathsf{id}_A : \mathscr{P}X \to \mathbf{Meas}(A, \mathscr{P}A) \\ \rho \quad \mapsto \quad \eta_A \end{array}\right. \end{array}\right) \tag{2}$$

$$c : \mathscr{K}\ell(\mathscr{P})(Y,X) \mapsto \mathsf{Stat}(c) :=$$

$$\left(\begin{array}{ccccc} \mathsf{Stat}(c) : & \mathsf{Stat}(X) & \to & \mathsf{Stat}(Y) \\ & \mathsf{Stat}(X)_0 & = & \mathsf{Stat}(Y)_0 \\ & \left(\begin{array}{ccc} d^\dagger : & \mathscr{P}X & \to \mathscr{K}\ell(\mathscr{P})(A,B) \\ & \pi & \mapsto \quad d^\dagger_\pi \end{array}\right) & \mapsto & \left(\begin{array}{ccc} c^* d^\dagger : \mathscr{P}Y \to \mathscr{K}\ell(\mathscr{P})(A,B) \\ \rho \quad \mapsto \quad d^\dagger_{c \bullet \rho} \end{array}\right) \end{array}\right)$$

Each $\mathsf{Stat}(X)$ is a category of stochastic channels with respect to measures on the space X. We can write morphisms $d^\dagger : \mathscr{P}X \to \mathscr{K}\ell(\mathscr{P})(A,B)$ in $\mathsf{Stat}(X)$ as $d^\dagger_{(\cdot)} : A \overset{(\cdot)}{\bullet\!\!\to} B$, and think of them as generalized Bayesian inversions: given a measure π on X, we obtain a channel $d^\dagger_\pi : A \overset{\pi}{\bullet\!\!\to} B$ with respect to π. Given a channel $c : Y \bullet\!\!\to X$ in the base category of priors, we can pull $d^\dagger$ back along c, to obtain a Y-dependent channel in $\mathsf{Stat}(Y)$, $c^* d^\dagger : \mathscr{P}Y \to \mathscr{K}\ell(\mathscr{P})(A,B)$, which takes $\rho : \mathscr{P}Y$ to the channel $d^\dagger_{c \bullet \rho} : A \overset{c \bullet \rho}{\bullet\!\!\to} B$ defined by pushing ρ through c and then applying $d^\dagger$.

Remark 2.3. Note that by taking $\mathbf{Meas}$ to be Cartesian closed, we have $\mathbf{Meas}(\mathscr{P}X, \mathbf{Meas}(A, \mathscr{P}B)) \cong \mathbf{Meas}(\mathscr{P}X \times A, \mathscr{P}B)$ for each X, A and B, and so a morphism $c^\dagger : \mathscr{P}Y \to \mathscr{K}\ell(\mathscr{P})(X,Y)$ equivalently has the type $\mathscr{P}Y \times X \to \mathscr{P}Y$. Paired with a channel $c : Y \to \mathscr{P}X$, we have something like a Cartesian lens; and to compose such pairs, we can use the Grothendieck construction [20, 26].

Definition 2.4 ($\mathbf{GrLens}_{\mathsf{Stat}}$). Instantiating the category of Grothendieck F-lenses $\mathbf{GrLens}_F$ (see [26]) with $F = \mathsf{Stat} : \mathscr{C}^{\mathrm{op}} \to \mathbf{V}\text{-}\mathbf{Cat}$, we obtain the category $\mathbf{GrLens}_{\mathsf{Stat}}$ whose objects are pairs (X,A) of objects of $\mathscr{C}$ and whose morphisms $(X,A) \nrightarrow (Y,B)$ are elements of the set

$$\mathbf{GrLens}_{\mathsf{Stat}}\big((X,A),(Y,B)\big) \cong \mathscr{C}(X,Y) \times \mathbf{V}\big(\mathscr{C}(I,X), \mathscr{C}(B,A)\big)\,. \tag{3}$$

The identity Stat-lens on (Y,A) is $(\mathsf{id}_Y, \mathsf{id}_A)$, where by abuse of notation $\mathsf{id}_A : \mathscr{C}(I,Y) \to \mathscr{C}(A,A)$ is the constant map id_A defined in (1) that takes any state on Y to the identity on A. The sequential composite of $(c, c^\dagger) : (X,A) \nrightarrow (Y,B)$ and $(d, d^\dagger) : (Y,B) \nrightarrow (Z,C)$ is the Stat-lens $\big((d \bullet c), (c^\dagger \circ c^* d^\dagger)\big) : (X,A) \nrightarrow (Z,C)$ with $(d \bullet c) : \mathscr{C}(X,Z)$ and where $(c^\dagger \circ c^* d^\dagger) : \mathbf{V}\big(\mathscr{C}(I,X), \mathscr{C}(C,A)\big)$ takes a state $\pi : \mathscr{C}(I,X)$ on X to the

channel $c^\dagger_\pi \bullet \mathrm{d}^\dagger_{c\bullet\pi}$. If we think of the notation $(\cdot)^\dagger$ as denoting the operation of forming the Bayesian inverse of a channel (in the case where $A = X$, $B = Y$ and $C = Z$), then the main result of this section is to show that $(d \bullet c)^\dagger_\pi \overset{d\bullet c\bullet\pi}{\sim} c^\dagger_\pi \bullet \mathrm{d}^\dagger_{c\bullet\pi}$, where $\overset{d\bullet c\bullet\pi}{\sim}$ denotes $(d \bullet c \bullet \pi)$-almost-equality [25, Definition 2.5].

In order to give an optical form for $\mathbf{GrLens}_{\mathsf{Stat}}$, we need to find two $\mathscr{M}$-actegories with a common category of actions $\mathscr{M}$. Let $\hat{\mathscr{C}}$ and $\check{\mathscr{C}}$ denote the categories $\hat{\mathscr{C}} := \mathbf{V}\text{-}\mathbf{Cat}(\mathscr{C}^{\mathrm{op}}, \mathbf{V})$ and $\check{\mathscr{C}} := \mathbf{V}\text{-}\mathbf{Cat}(\mathscr{C}, \mathbf{V})$ of presheaves and copresheaves on $\mathscr{C}$, and consider the following natural isomorphisms.

$$\begin{aligned}\mathbf{GrLens}_{\mathsf{Stat}}\big((X,A),(Y,B)\big) &\cong \mathscr{C}(X,Y) \times \mathbf{V}\big(\mathscr{C}(I,X), \mathscr{C}(B,A)\big) \\ &\cong \int^{M:\mathscr{C}} \mathscr{C}(X,Y) \times \mathscr{C}(X,M) \times \mathbf{V}\big(\mathscr{C}(I,M), \mathscr{C}(B,A)\big) \\ &\cong \int^{\hat{M}:\hat{\mathscr{C}}} \mathscr{C}(X,Y) \times \hat{M}(X) \times \mathbf{V}\big(\hat{M}(I), \mathscr{C}(B,A)\big) \end{aligned} \tag{4}$$

The second isomorphism follows by Yoneda reduction [17, 23], and the third follows by the Yoneda lemma. We take $\mathscr{M}$ to be $\mathscr{M} := \hat{\mathscr{C}}$, and define an action $\odot$ of $\hat{\mathscr{C}}$ on $\check{\mathscr{C}}$ as follows.

Definition 2.5 ($\odot$)**.** We give only the action on objects; the action on morphisms is analogous.

$$\begin{aligned} \odot : \hat{\mathscr{C}} &\to \mathbf{V}\text{-}\mathbf{Cat}(\check{\mathscr{C}}, \check{\mathscr{C}}) \\ \hat{M} &\mapsto \begin{pmatrix} \hat{M} \odot - & : & \check{\mathscr{C}} & \to & \check{\mathscr{C}} \\ & & P & \mapsto & \mathbf{V}\big(\hat{M}(I), P\big) \end{pmatrix} \end{aligned} \tag{5}$$

Functoriality of $\odot$ follows from the functoriality of copresheaves. □

Proposition 2.6. $\odot$ equips $\check{\mathscr{C}}$ with a $\hat{\mathscr{C}}$-actegory structure: unitor isomorphisms $\lambda^{\odot}_F : 1 \odot F \xrightarrow{\sim} F$ and associator isomorphisms $a^{\odot}_{\hat{M},\hat{N},F} : (\hat{M} \times \hat{N}) \odot F \xrightarrow{\sim} \hat{M} \odot (\hat{N} \odot F)$ for each $\hat{M}, \hat{N}$ in $\hat{\mathscr{C}}$, both natural in $F : \mathbf{V}\text{-}\mathbf{Cat}(\mathscr{C}, \mathbf{V})$.

We are now in a position to define the category of abstract Bayesian lenses, and show that this category coincides with the category of Stat-lenses.

Definition 2.7 (Bayesian lenses)**.** Denote by **BayesLens** the category of optics $\mathbf{Optic}_{\times,\odot}$ for the action of the Cartesian product on presheaf categories $\times : \hat{\mathscr{C}} \to \mathbf{V}\text{-}\mathbf{Cat}(\hat{\mathscr{C}}, \hat{\mathscr{C}})$ and the action $\odot : \hat{\mathscr{C}} \to \mathbf{V}\text{-}\mathbf{Cat}(\check{\mathscr{C}}, \check{\mathscr{C}})$ defined in (5). Its objects $(\hat{X}, \check{Y})$ are pairs of a presheaf and a copresheaf on $\mathscr{C}$, and its morphisms $(\hat{X}, \check{A}) \nrightarrow (\hat{Y}, \check{B})$ are abstract *Bayesian lenses*—elements of the type

$$\mathbf{Optic}_{\times,\odot}\Big((\hat{X}, \check{A}), (\hat{Y}, \check{B})\Big) = \int^{\hat{M}:\hat{\mathscr{C}}} \hat{\mathscr{C}}(\hat{X}, \hat{M} \times \hat{Y}) \times \check{\mathscr{C}}(\hat{M} \odot \check{B}, \check{A}) \tag{6}$$

Given $v : \mathscr{C}(X,Y)$ and $u : \mathbf{V}(\mathscr{C}(I,X), \mathscr{C}(B,A))$, we denote the corresponding element of this type by $\langle v \,|\, u \rangle$. A Bayesian lens $(\hat{X}, \check{X}) \nrightarrow (\hat{Y}, \check{Y})$ is called a **simple** Bayesian lens.

Proposition 2.8. BayesLens is a category of lenses; a definition is given in [25, §2.2.1].

Proposition 2.9 (Stat-lenses are Bayesian lenses)**.** Let $\hat{(\cdot)} : \mathscr{C} \hookrightarrow \mathbf{V}\text{-}\mathbf{Cat}(\mathscr{C}^{\mathrm{op}}, \mathbf{V})$ denote the Yoneda embedding and $\check{(\cdot)} : \mathscr{C} \hookrightarrow \mathbf{V}\text{-}\mathbf{Cat}(\mathscr{C}, \mathbf{V})$ the coYoneda embedding. Then

$$\mathbf{Optic}_{\times,\odot}\Big((\hat{X}, \check{A}), (\hat{Y}, \check{B})\Big) \cong \mathbf{GrLens}_{\mathsf{Stat}}\Big((X,A),(Y,B)\Big) \tag{7}$$

so that $\mathbf{GrLens}_{\mathsf{Stat}}$ is equivalent to the full subcategory of $\mathbf{Optic}_{\times,\odot}$ on representable (co)presheaves.

Remark 2.10. We will often abuse notation by indicating representable objects in **BayesLens** by their representations in $\mathscr{C}$. That is, we will write (X,A) instead of $(\hat{X},\check{A})$ where this would be unambiguous.

Proposition 2.11. BayesLens is a symmetric monoidal category. The monoidal product $\otimes$ is inherited from $\mathscr{C}$; the unit object is the pair (I,I) where I is the unit object in $\mathscr{C}$. For more details on the structure, see [21] or [19].

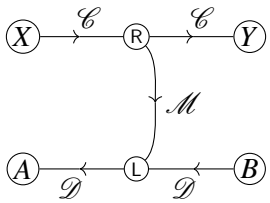
Definition 2.12 (Exact and approximate Bayesian lens)**.** Let $\left\langle c \,\middle|\, c^\dagger \right\rangle : (X,X) \nrightarrow (Y,Y)$ be a simple Bayesian lens. We say that $\left\langle c \,\middle|\, c^\dagger \right\rangle$ is **exact** if c admits Bayesian inversion and, for each $\pi : I \to\!\!\!\!\bullet\, X$ such that $c \bullet \pi$ has non-empty support, $c^\dagger_\pi$ is the Bayesian inversion of c with respect to π. Simple Bayesian lenses that are not exact are said to be **approximate**.

Lemma 2.13. Let $\left\langle c \,\middle|\, c^\dagger \right\rangle$ and $\left\langle d \,\middle|\, d^\dagger \right\rangle$ be sequentially composable exact Bayesian lenses. Then the contravariant component of the composite lens $\left\langle d \,\middle|\, d^\dagger \right\rangle \fatsemi \left\langle c \,\middle|\, c^\dagger \right\rangle \cong \left\langle d \bullet c \,\middle|\, c^\dagger \circ c^* d^\dagger \right\rangle$ is, up to $d \bullet c \bullet \pi$-almost-equality, the Bayesian inversion of $d \bullet c$ with respect to any state π on the domain of c such that $c \bullet \pi$ has non-empty support. That is to say, *Bayesian updates compose optically*: $(d \bullet c)^\dagger_\pi \overset{d \bullet c \bullet \pi}{\sim} c^\dagger_\pi \bullet d^\dagger_{c \bullet \pi}$.

3 Open Games for General Optics

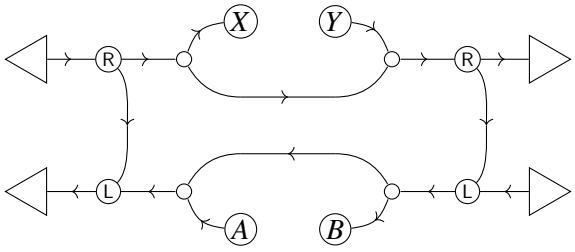
In this section, we supply mild generalizations of the structures underlying open games, building on those in [2]; at first, then, we consider games over arbitrary categories of optics $\mathbf{Optic}_{\circledR,\circledL}$. Subsequently, we use games over Bayesian lenses (in the category of optics **BayesLens** introduced above) to exemplify a number of canonical statistical concepts, such as maximum likelihood estimation and the variational autoencoder, and clarify their compositional structure using the notion of *optimization game* (Definition 3.22). Owing to space constraints, we omit most proofs in this section; they will appear in a full paper expanding the present abstract, and can be supplied at the request of the reader.

Observation 3.1. In the graphical calculus for the compact closed bicategory of profunctors **Prof** [22], the hom object $\mathbf{Optic}_{\circledR,\circledL}((X,A),(Y,B))$ has the depiction



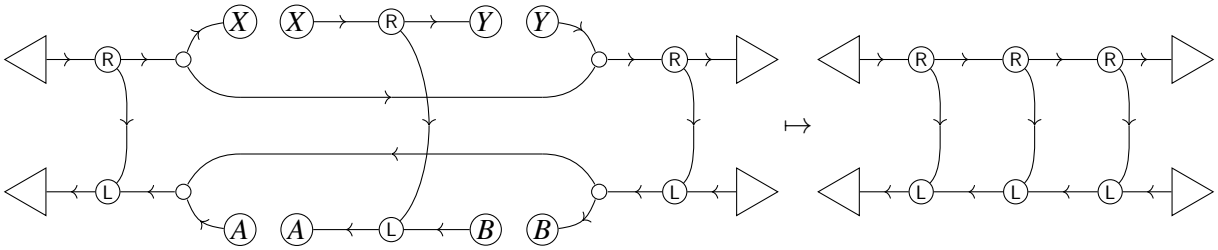
where the types on the wires are the 0-cells of **Prof**, the monoidal actions Ⓡ and Ⓛ are depicted as (co)monoids, and the states and effects are (co)representable functors on the objects X,A,Y,B, treated as profunctors.

Definition 3.2 (Generalized context)**.** The context functor $C : \mathbf{Optic}_{\circledR,\circledL}{}^{\mathrm{op}} \times \mathbf{Optic}_{\circledR,\circledL} \to \mathbf{Set}$ takes the pair of optical objects $((X,A),(Y,B))$ to the type with depiction



The triangles depict the (co)presheaves on the monoidal unit I in the underlying actegories. The action on morphisms (*i.e.*, optics) is by precomposition on the left and postcomposition on the right. Functoriality follows accordingly. □

We can compose a context with an optic to obtain a 'closed' system, as follows:



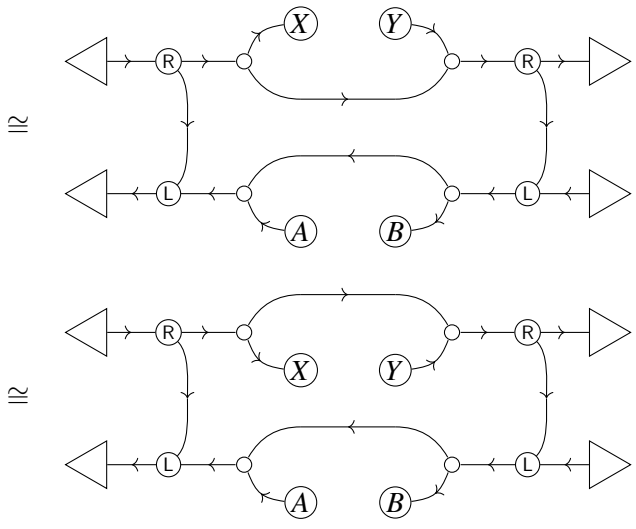
Conjecture 3.3. It is easy to show that a context on $((X,A),(Y,B))$ is equivalently a state $(I,I) \nrightarrow ((X,A),(Y,B))$ in the monoidal category of 'double lenses', $\mathbf{Lens}_{\mathbf{Optic}_{\circledR,\copyright}}$ [2]. Rendering this graphically leads us to the following conjecture: categories of double optics are instances of the *doubling* or *CP* construction from categorical quantum mechanics (*cf.* [8, 7]).

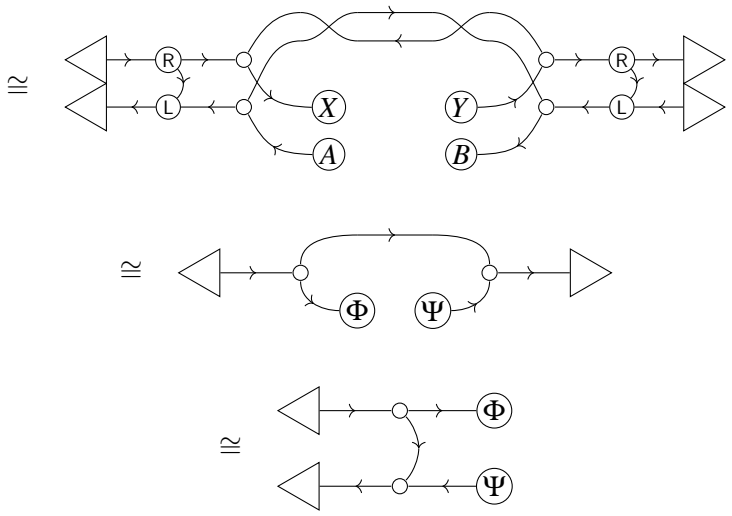
Observation 3.4 (Justification for Conjecture 3.3)**.** Assuming $\mathbf{Optic}_{\circledR,\copyright}$ is monoidal, a context on $((X,A),(Y,B))$ is equivalently a monoidal lens $((I,I),(I,I)) \nrightarrow ((X,A),(Y,B))$ over $\mathbf{Optic}_{\circledR,\copyright}$:

$$\int^{(M,N):\mathbf{Optic}_{\circledR,\copyright}} \mathbf{Optic}_{\circledR,\copyright}\big((I,I),(M,N)\otimes(X,A)\big) \times \mathbf{Optic}_{\circledR,\copyright}\big((M,N)\otimes(Y,B),(I,I)\big)$$
$$\cong \mathbf{Optic}_{\otimes,\otimes}\Big(\big((I,I),(I,I)\big),\big((X,A),(Y,B)\big)\Big)$$

where $\otimes$ is here the induced monoidal product on $\mathbf{Optic}_{\circledR,\copyright}$. Graphically, we can represent this isomorphism as follows:

$$\int^{(M,N):\mathbf{Optic}_{\circledR,\copyright}} \mathbf{Optic}_{\circledR,\copyright}\big((I,I),(M,N)\otimes(X,A)\big) \times \mathbf{Optic}_{\circledR,\copyright}\big((M,N)\otimes(Y,B),(I,I)\big)$$

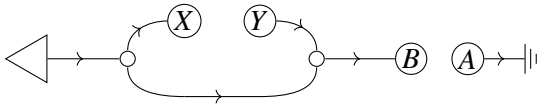




$$\cong \mathbf{Optic}_{\otimes,\otimes}\Big(\big((I,I),(I,I)\big),\big((X,A),(Y,B)\big)\Big)$$

where for the second isomorphism we used the symmetry of the monoidal product assumed on the ® actegory. The fourth isomorphism suggests that the symmetric monoidal product $\otimes$ on $\mathbf{Optic}_{\text{®},\text{©}}$ is a 'quantum spider' in the terminology of **(author?)** [7].

Proposition 3.5. Let $\mathscr{C}$ and $\mathscr{D}$ be the (monoidal) actegories underlying $\mathbf{Optic}_{\text{®},\text{©}}$, and denote their respective monoidal units by $I_{\mathscr{C}}$ and $I_{\mathscr{D}}$. If these unit objects are terminal in their respective categories, then the contexts $C((X,A),(Y,B))$ simplify to



where we have depicted the representable presheaf on $I_{\mathscr{D}}$ as $\overline{\top}$ to indicate that A is just discarded. Consequently, in this case, a context is just an optic $(I,B) \nrightarrow (X,Y)$.

Proof of Proposition 3.5. The proof that contexts simplify when the underlying monoidal units are terminal is a simple exercise in the graphical calculus: just throw away any hom-objects with terminal codomains, and simplify the resulting diagrams. □

Definition 3.6 (Generalized open game)**.** Let (X,A) and (Y,B) be objects in any symmetric monoidal category of optics $\mathbf{Optic}_{\text{®},\text{©}}$. Let Σ be a $\mathscr{U}$-category, for any base of enrichment $\mathscr{U}$ such that $\mathscr{U}$-**Prof** is compact closed. An **open game** from (X,A) to (Y,B) with strategies in Σ, denoted $G : (X,A) \xrightarrow{\Sigma} (Y,B)$, is given by:

(a) a play function $P : \Sigma_0 \to \mathbf{Optic}_{\text{®},\text{©}}((X,A),(Y,B))$; and

(b) a best response function $B : C((X,A),(Y,B)) \to \mathscr{U}\text{-}\mathbf{Prof}(\Sigma,\Sigma)$.

Given a strategy $\sigma : \Sigma$, we will often write $\langle v \,|\, u \rangle_\sigma$ or similar to denote its image under P. A strategy is an **equilibrium** in a context $\langle \pi \,|\, k \rangle$ if it is a fixed point of $B(\langle \pi \,|\, k \rangle)$.

Roughly speaking, the 'best responses' to a strategy σ in a context is are those strategies τ such that choosing τ would result in performance at the game at least as good as choosing σ; equilibrium strategies are those for which such deviation would not improve performance.

Remark 3.7. Note: whereas classic open games use a best-response relation, we categorify that here to a best-response *relator* (in the terminology of [17]; *i.e.*, a 'proof-relevant' relation), so that we can describe the trajectories witnessing the computation of equilibria, rather than their mere existence.

Proposition 3.8. Generalized open games over the symmetric monoidal category of optics $\mathbf{Optic}_{®,©}$ with strategies enriched in $\mathscr{U}$ form a symmetric monoidal category denoted $\mathbf{Game}(\mathscr{U},®,Ⓛ)$.

Proof of Proposition 3.8. Since we assume $\mathscr{U}$-**Prof** to be compact closed, the proof is a straightforward generalization of that given by **(author?)** [2], replacing best-response relations by best-response *relators* [17] (*i.e.*, 'proof-relevant' relations). The play function of a sequentially composite game is given by the sequential composition of the corresponding optics, and the associated best response function is given by relators on the product of the strategies using a natural notion of 'local context'. The symmetric monoidal structure is given similarly. We refer the reader to **(author?)** [2] for the details. □

Since our games are only a mild generalization of those of [2], we refer the reader to §3.10 of that paper for an idea of the proof of the foregoing proposition, which goes through analogously. The sequential composition of games is given by the sequential composition of optics, with the best response to the composite being the product of the best responses to the factors. Similarly, parallel composition is given by the monoidal product of optics, and the best response to the composite is again the product of the best responses to the factors.

We now consider some games over **BayesLens** that supply the building blocks of the archetypal cybernetic systems to be considered in §4. For now, we will take the strategies simply to be discrete categories (*i.e.*, sets), as in the standard formulation of open games. Consequently, we will take the codomain of the best response function to be $\mathbf{Set}(\Sigma,\mathbf{Set}(\Sigma,2))$, for each strategy type Σ. We assume the ambient category of stochastic channels is semicartesian, so that the monoidal unit is the terminal object.

Remark 3.9. All the games we will consider henceforth will have play functions whose codomains restrict to the representable subcategory $\mathbf{GrLens}_{\mathsf{Stat}}$ of **BayesLens**; in this work, we do not use the extra generality afforded by **BayesLens**, except insofar as it grants us the use of string diagrams in **Prof**, which we find helpful for reasoning intuitively about these systems. The generality of optics *is* however used in the 'game-theoretic' games of [2], and in future work we hope to relate the cybernetic systems of this paper to the game-theoretic setting of that earlier work.

Remark 3.10. All the statistical games considered in this paper will be 'atomic' in the sense of [2]: in particular, the best response functions we consider will be constant, meaning that, in any context, the set of best strategies does not depend on the 'current' choice of strategy. Permitting such dependence will be important in future work, however, when we consider how cybernetic systems interact, and hence respond to each other.

Example 3.11. A Bayesian lens of the form $(I,I) \nrightarrow (X,X)$ is fully specified by a state $\pi : I \to\!\!\!\!\bullet\, X$. A context for such a lens is given by a lens $\langle ! \,|\, k \rangle : (I,X) \nrightarrow (X,X)$ where $! : I \to\!\!\!\!\bullet\, I$ is the unique map and $k : X \to\!\!\!\!\bullet\, X$ is any endochannel on X. A **maximum likelihood game** is any game whose play function has codomain in Bayesian lenses of this form $(I,I) \nrightarrow (X,X)$ for any $X : \mathscr{C}$, and whose best response function is isomorphic to

$$B(\langle ! \,|\, k \rangle) = \langle \rho \,|\, ! \rangle_\sigma \mapsto \left\{ \langle \pi \,|\, ! \rangle_\tau \,\middle|\, \pi \in \operatorname*{arg\,max}_{\pi : I \to\!\!\!\!\bullet\, X} \mathop{\mathbb{E}}_{k \bullet \pi} [\pi] \right\}$$

where $\mathbb{E}$ is the canonical expectation operator (*i.e.* algebra evaluation) associated to states in $\mathscr{C}$, and where we have written $\langle \rho \,|\, ! \rangle_\sigma$ and $\langle \pi \,|\, ! \rangle_\tau$ to denote the images of the strategies σ and τ under the play function. Intuitively, then, the best response is given by the strategy that maximises the likelihood of the state obtained from the context k.

Remark 3.12. In what follows, we assume that the underlying category $\mathscr{C}$ of stochastic channels *admits density functions*. Informally, a density function for a stochastic channel $c : X \to\!\!\!\!\bullet\, Y$ is a measurable function $p_c : Y \times X \to [0,1]$ whose values are the probabilities (or probability densities) $p_c(y|x)$ at each pair $(y,x) : Y \times X$. We say that the value $p_c(y|x)$ is the probability (or probability density) of y *given* x. In a category such as $\mathscr{K}\ell(\mathscr{D}_{\leq 1})$, whose objects are sets and whose morphisms $X \to\!\!\!\!\bullet\, Y$ are functions $X \to \mathscr{D}(Y+1)$, a density function for $c : X \to\!\!\!\!\bullet\, Y$ is a morphism $Y \otimes X \to\!\!\!\!\bullet\, I$; note that in $\mathscr{K}\ell(\mathscr{D}_{\leq 1})$, I is not terminal. In the finitely-supported case, density functions are effectively equivalent to channels, but this is not the case in the continuous setting, where they are of most use. For more on this, see [25, §2.1.4].

A natural first generalization of maximum likelihood games takes us from states $I \to\!\!\!\!\bullet\, X$ to channels $Z \to\!\!\!\!\bullet\, X$; that is, from 'elements' to 'generalized elements' in the covariant (forwards) part of the lens. Unlike Bayesian lenses $(I,I) \nrightarrow (X,X)$, lenses $(Z,Z) \nrightarrow (X,X)$ admit nontrivial contravariant components, which we think of as generalized Bayesian inversions. Consequently, our first generalization is a notion of 'Bayesian inference game'. A context $\langle \pi \,|\, k \rangle : (I,X) \nrightarrow (Z,X)$ for a Bayesian lens $(Z,Z) \nrightarrow (X,X)$ then constitutes a 'prior' state $\pi : I \to\!\!\!\!\bullet\, Z$ and a 'continuation' channel $k : X \to\!\!\!\!\bullet\, X$ which together witness the closure of the otherwise open system.

Example 3.13. Fix a channel $c : Z \to\!\!\!\!\bullet\, X$ with associated density function $p_c : X \times Z \to \mathbb{R}_+$ and a measure of divergence between states on Z, $D : \mathscr{C}(I,Z) \times \mathscr{C}(I,Z) \to \mathbb{R}$. A corresponding (generalized) **simple Bayesian inference game** is any game whose play function has codomain $\mathbf{BayesLens}((Z,Z),(X,X))$ and whose best response function is isomorphic to

$$\begin{aligned} B(\langle \pi \,|\, k \rangle) = \left\langle d \,\middle|\, d' \right\rangle_\sigma \mapsto \left\{ \left\langle c \,\middle|\, c' \right\rangle_\tau \middle| c' \in \operatorname*{arg\,min}_{c' : \mathbf{V}(\mathscr{C}(I,Z),\mathscr{C}(X,Z))} \mathop{\mathbb{E}}_{x \sim k \bullet c \bullet \pi} \left[\mathop{\mathbb{E}}_{z \sim c'_\pi(x)} [-\log p_c(x|z)] + D(c'_\pi(x), \pi) \right] \right\} \\ = \left\langle d \,\middle|\, d' \right\rangle_\sigma \mapsto \left\{ \left\langle c \,\middle|\, c' \right\rangle_\tau \middle| c' \in \operatorname*{arg\,min}_{c' : \mathbf{V}(\mathscr{C}(I,Z),\mathscr{C}(X,Z))} \left(\mathop{\mathbb{E}}_{z \sim c'_\pi \bullet k \bullet c \bullet \pi} \left[-\int_X \log p_c(\mathrm{d}k \bullet c \bullet \pi | z) \right] \right. \right. \\ \left. \left. + D(c'_\pi \bullet k \bullet c \bullet \pi, \pi) \right) \right\} \end{aligned}$$

where $\pi : I \to\!\!\!\!\bullet\, Z$ and $k : X \to\!\!\!\!\bullet\, X$, and where the notation $z \sim \pi$ means "z distributed according to the state π". Note that the second line follows from the first by linearity of expectation.

Proposition 3.14 ([16, Thm. 1])**.** When D is chosen to be the Kullback-Leibler divergence D_{KL}, minimizing the objective function defining a simple Bayesian inference game is equivalent to computing an (exact) Bayesian inversion.

Corollary 3.15. Given two Bayesian inference games $G : (Z,Z) \nrightarrow (Y,Y)$ and $H : (Y,Y) \nrightarrow (X,X)$, we can compose them sequentially to obtain a game $H \odot G : (Z,Z) \nrightarrow (X,X)$, which we will call a **hierarchical Bayesian inference game**. It is then an immediate consequence of Lemma 2.13 that, in any given context for which the forwards channels admit Bayesian inversion, the best response to the composite game $H \odot G$ (that is, the optimal inversion of the composite channel) is given simply by (the composition of) the best responses to the factors H and G. Consequently, Bayesian inference games are closed under composition.

Similarly, given a channel $c : Z \otimes Y \nrightarrow X$, we can consider the **marginal Bayesian inference game** in which the objective is to compute the inversion of the channel onto just one of the factors Z or Y in the domain.

Example 3.16 (Variational autoencoder game)**.** Fix a family $\mathscr{F} \hookrightarrow \mathscr{C}(Z,X)$ of forward channels and a family $\mathscr{P} \hookrightarrow \mathscr{C}(X,Z)$ of backward channels such that each $c : \mathscr{F}$ admits a density function $p_c : X \otimes Z \to \mathbb{R}_+$ and each $d : \mathscr{P}$ admits a density function $q : Z \otimes X \to \mathbb{R}_+$; think of these families as determining parameterizations of the channels. We take our strategy type to be $\Sigma = \mathscr{F} \times \mathscr{P}$. A **simple variational autoencoder game** $(Z,Z) \xrightarrow{\Sigma} (X,X)$ is any game with play function $P : \Sigma \to \mathbf{BayesLens}((Z,Z),(X,X))$ and whose best response function is isomorphic to

$$B(\langle \pi \,|\, k \rangle) = \left\langle d \,\middle|\, d' \right\rangle_\sigma \mapsto \left\{ \left\langle c \,\middle|\, c' \right\rangle_\tau \middle| (c,c') \in \underset{\substack{c \in \mathscr{F}, \\ c' \in \mathbf{V}(\mathscr{C}(I,Z),\mathscr{P})}}{\arg\min} \mathop{\mathbb{E}}_{x \sim k \bullet c \bullet \pi} \mathop{\mathbb{E}}_{z \sim c'_\pi(x)} \left[\log \frac{q(z|x)}{p_c(x|z) p_\pi(z)} \right] \right\}$$

where $\pi : I \nrightarrow Z$ admits a density function $p_\pi : Z \to \mathbb{R}_+$, $q : Z \otimes X \to \mathbb{R}_+$ is a density function associated to c'_π, and k has type $X \nrightarrow X$.

Proposition 3.17. A best response to a variational autoencoder game is a stochastic channel $c : \mathscr{F}$ that maximises the likelihood of the state observed through the continuation k under the assumption that the generative process is in $\mathscr{F}$, along with an inverse channel $c'_\pi : \mathscr{P}$ that best approximates the exact Bayesian inverse $c^\dagger_\pi$ under the constraint of being in $\mathscr{P}$.

Proof of Proposition 3.17. We seek to show that giving a best response to a simple variational autoencoder game is equivalent to maximizing the likelihood of the observed data under the assumption that the data is generated by the process $k \bullet c \bullet \pi$, with c constrained to lie in $\mathscr{F}$ and with a (possibly approximate) inverse c'_π constrained to lie in $\mathscr{P}$. The intuition is that an agent responding to such a game builds a model of the process generating the data such that the agent can invert the model to infer the latent causes of the observed data. Typically, therefore, one chooses families $\mathscr{F}$ and $\mathscr{P}$ that are computationally tractable at the cost of some approximation.

As in the maximum likelihood game (Example 3.11), maximising the likelihood of data assumed to be generated by $k \bullet c \bullet \pi$ means maximising $\mathbb{E}_{k \bullet c \bullet \pi}[c \bullet \pi]$. We assume that c is represented by the density function $p_c : X \times Z \to \mathbb{R}_+$, π is represented by $p_\pi : Z \to \mathbb{R}_+$, and c'_π is represented by $q : Z \times X \to \mathbb{R}_+$. The density function for the composite state $c \bullet \pi : I \nrightarrow X$ is then given by $p_{c \bullet \pi}(x) = \int_Z p_c(x|z) p_\pi(z) \mathrm{d}z$. Consequently, we have $\mathbb{E}_{k \bullet c \bullet \pi}[c \bullet \pi] = \mathbb{E}_{k \bullet c \bullet \pi}[p_{c \bullet \pi}]$. Moreover, since log is monotonic, we can just as well minimize $\mathbb{E}_{k \bullet c \bullet \pi}[-\log p_{c \bullet \pi}]$.

Let $p_\omega : Z \times X \to \mathbb{R}_+$ be given by $p_\omega(z,x) = p_c(x|z) p_\pi(z)$, and suppose the exact Bayesian inversion of c with respect to π, denoted $c^\dagger_\pi$, has density function $p_{c^\dagger_\pi} : Z \times X \to \mathbb{R}_+$.

We then have the following equalities:

$$\begin{aligned}
-\log p_{c \bullet \pi}(x) &= \mathop{\mathbb{E}}_{z \sim c'_\pi(x)} [-\log p_{c \bullet \pi}(x)] \\
&= \mathop{\mathbb{E}}_{z \sim c'_\pi(x)} \left[-\log \frac{p_\omega(z,x)}{p_{c^\dagger_\pi}(z|x)} \right] \\
&= \mathop{\mathbb{E}}_{z \sim c'_\pi(x)} \left[-\log \frac{p_\omega(z,x)}{q(z|x)} \frac{q(z|x)}{p_{c^\dagger_\pi}(z|x)} \right] \\
&= - \mathop{\mathbb{E}}_{z \sim c'_\pi(x)} \left[\log \frac{p_\omega(z,x)}{q(z|x)} \right] - D_{KL}\left[c'_\pi(x), c^\dagger_\pi(x) \right]
\end{aligned}$$

so that

$$\mathop{\mathbb{E}}_{z\sim c'_\pi(x)}\left[\log\frac{q(z|x)}{p_c(x|z)p_\pi(z)}\right] = D_{KL}\left[c'_\pi(x), c^\dagger_\pi(x)\right] - \log p_{c\bullet\pi}(x)\,. \tag{8}$$

This quantity is called the *free energy*, from an analogy with statstical mechanics, or *evidence upper bound* [15], from machine learning. Minimizing the free energy entails both minimizing the divergence between the model posterior $c'_\pi(x)$ and the exact Bayesian posterior $c^\dagger_\pi(x)$ (given an observation x), as well as maximizing the likelihood of the observation x under the assumed generative model c. If the exact posterior $c^\dagger_\pi$ is in $\mathscr{F}$, then the divergence term has a minimum at 0 where the two posteriors are equal almost everywhere; in this case, the objective absolutely maximises the likelihood of the data under the model c.

In general, we are not just interested in a single observation $x : X$, but in a distribution (*i.e.* state) induced on X by some external process, represented here by the continuation k. Thus, taking the free energy under the expectation induced by k gives precisely the objective of the simple variational autoencoder game. □

Proposition 3.18. Variational autoencoder games generalize inference games for the Kullback-Leibler divergence. More precisely, the objective function defining autoencoder games is of the same form as that defining inference games (3.13) when $D = D_{KL}$.

Proof of Proposition 3.18. We start from the free energy (8):

$$\begin{aligned}\mathop{\mathbb{E}}_{z\sim c'_\pi(x)}\left[\log\frac{q(z|x)}{p_c(x|z)p_\pi(z)}\right] &= \mathop{\mathbb{E}}_{z\sim c'_\pi(x)}\left[\log q(z|x) - \log p_c(x|z) - \log p_\pi(z)\right]\\ &= \mathop{\mathbb{E}}_{z\sim c'_\pi(x)}\left[-\log p_c(x|z)\right] + D_{KL}\left[c'_\pi(x), \pi\right]\,.\end{aligned}$$

Taking the expectation induced by the continuation k gives a generalization of the objective of a simple Bayesian inference game, with D taken to be the Kullback-Leibler divergence D_{KL}, and with c' constrained to lie in $\mathscr{P}$ and c allowed to vary in $\mathscr{F}$. □

This prompts the following generalization:

Example 3.19 (Generalized autoencoder game)**.** Fix two families of channels $\mathscr{F}, \mathscr{P}$ and a strategy type Σ as in Example 3.16. Then a (generalized) **simple autoencoder game** $(Z,Z) \xrightarrow{\Sigma} (X,X)$ is any game with play function $P : \Sigma \to \mathbf{BayesLens}((Z,Z),(X,X))$ and whose best response function is isomorphic to

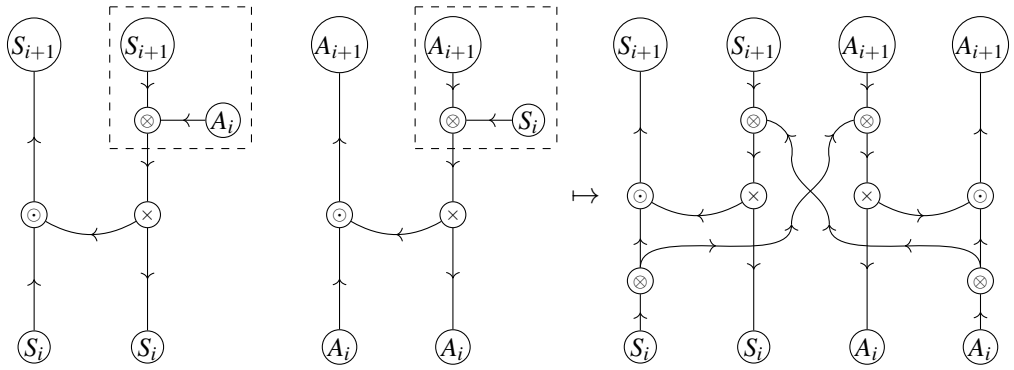
$$B(\langle \pi \,|\, k\rangle) = \left\langle d \,\middle|\, d'\right\rangle_\sigma \mapsto \left\{ \left\langle c \,\middle|\, c'\right\rangle_\tau \,\middle|\, (c,c') \in \mathop{\arg\min}_{\substack{c\in\mathscr{F},\\ c'\in\mathbf{V}(\mathscr{C}(I,Z),\mathscr{P})}} \left(\mathop{\mathbb{E}}_{z\sim c'_\pi\bullet k\bullet c\bullet\pi}\left[-\int_X \log p_c(\mathrm{d}k\bullet c\bullet\pi|z)\right] + D(c'_\pi\bullet k\bullet c\bullet\pi, \pi)\right)\right\}$$

where π and k have respective types $I \nrightarrow Z$ and $X \nrightarrow X$, and D is any measure of divergence between states.

As with Bayesian inference games, we can generalize simple autoencoder games to **hierarchical** and **marginal** autoencoder games via the corresponding sequential and parallel compositions.

The foregoing games have been purely statistically formulated, without capturing the motivating feature of an open system as something in interaction with an external environment. Nonetheless, we can model a simple open system of hierarchical **active inference** that receives stochastic inputs from an environment and emits actions stochastically into the environment, as follows.

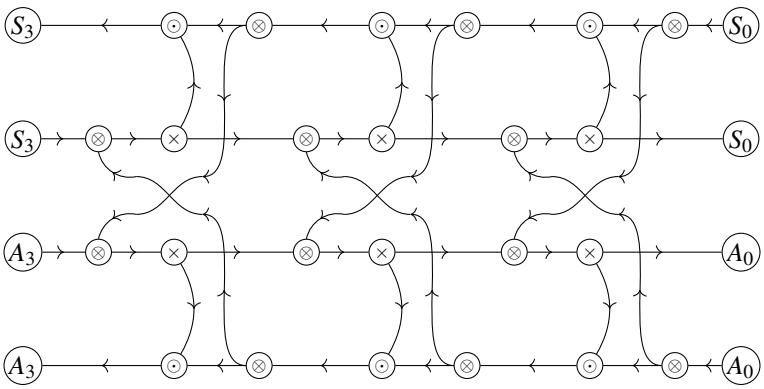
Example 3.20 (Active inference game)**.** Let $\{S_i\}_i$ be set of spaces of sensory data indexed by hierarchical levels of abstraction i (for instance, the levels of abstraction might range from representations of whole objects to fine details about their texture); similarly, let $\{A_i\}_i$ be a set of spaces of possible actions similarly hierarchically organized. Consider the marginal autoencoder games $(S_{i+1} \otimes A_i, S_{i+1}) \nrightarrow (S_{i+1}, S_{i+1})$ and $(A_{i+1} \otimes S_i, A_{i+1}) \nrightarrow (A_{i+1}, A_{i+1})$ coupled via the symmetric monoidal structure $\otimes$ of $\mathscr{C}$:



giving a composite game $(S_{i+1} \otimes A_{i+1}, S_{i+1} \times A_{i+1}) \nrightarrow (S_i \otimes A_i, S_i \times A_i)$. Recall from [2, §§3.7-3.8] that a composite game is given by the (sequential and parallel) composition of optics, with best-response given by the product of the best-responses of the factors.

Note that the Bayesian posterior inferred by such a game has independent factors on S_{i+1} and A_{i+1}. This is not merely a diagrammatic convenience, but coincides with a common 'mean field' simplification in the modelling literature [3, 15]. The dashed box is a functorial box [18] depicting the Yoneda embedding; recall that optics in **BayesLens** were defined over (co)presheaves, and so here we needed to lift the monoidal product on $\mathscr{C}$ into a diagram over its presheaf category $\mathbf{Cat}(\mathscr{C}^{\mathrm{op}}, \mathbf{Set})$.

Next, compose these games along the hierarchy indexed by i, to obtain a game $(S_N \otimes A_N, S_N \times A_N) \nrightarrow (S_0 \otimes A_0, S_0 \times A_0)$, such as an element of the following object:



Given a context with a strong prior about expected sensory states and a continuation that responds to an action of type A_0 by feeding back a state on S_0, the best response can be shown to be that which selects actions that, under the current state, maximize the likelihood of obtaining the expected 'goal' state [3, 11].

Remark 3.21. We have framed each of these statistical procedures as optimization problems not only to suggest a link to the utility-maximising agents of game theory, but also because it suggests the use of iterative methods to compute best responses; note that computational tractability is an important motivation in the proof of Proposition 3.17.

The question of providing such dynamical or, thinking of game composition as an algebra for building complex systems, 'coalgebraic' semantics for (generalized) optimization games is the topic of the next section. We first formalize this notion.

Definition 3.22. An **optimization game** is any open game whose best response function can be defined by a function of the form $\Sigma \times C \xrightarrow{\pi} M \xrightarrow{\varphi} P$, where Σ is a strategy type, C a context type, M any space, and P a poset. We call φ the **fitness function**, and think of π as projecting systems into a space whose points can be assigned a fitness. The best response function of an optimization game can then be defined by giving the subset of strategies contextually maximizing fitness, for each context $c : C$.

4 Cybernetic Systems and Dynamical Realisation

In this section, we begin to answer the question of precisely how the optimization games of the previous section may be realized in physical systems, such as brains or computers. More formally, this means we seek open dynamical systems whose input and output types correspond to the domain and codomain types of the foregoing games, such that there is a correspondence between the behaviours of the abstract games and their dynamical realisations, and such that the evolutions of the internal states of the dynamical systems correspond to strategic improvements in game-playing: by concentrating on optimization games, a natural measure of such improvement is encoded in the fitness function underlying the best-response relator.

We do not require that there is a correspondence between internal states of the realisations and strategies for the corresponding games, but we do require that the fitness functions extend to the the total state spaces of the closure of a realisation induced by the context. When there *is* a correspondence between internal states and strategies, we can take advantage of Definition 3.6 and interpret trajectories over the state space as trajectories over strategies witnessing the strategic improvement.

We begin by sketching categories of dynamical games, and then use these ideas to define preliminary notions of open cybernetic systems and categories thereof. We consider principally single systems whose underlying games are atomic (in the sense of Remark 3.10), and leave the study of the behaviour of interacting cybernetic systems to future work. Once more, we omit proofs in this section; they will appear in a paper to follow.

Definition 4.1 (Discrete-time dynamical system over $\mathscr{C}$; after [24, 6])**.** A **discrete-time dynamical system** over $\mathscr{C}$ with state space $S : \mathscr{C}$, input type $A : \mathscr{C}$ and output type $B : \mathscr{C}$ is a lens $(S,S) \nrightarrow (B,A)$ over $\mathscr{C}$, *i.e.* in the following optical hom object:

$$\int^{M:\mathscr{C}} \mathbf{Comon}(\mathscr{C})(S, M \otimes B) \times \mathscr{C}(M \otimes A, S) \cong \mathbf{Comon}(\mathscr{C})(S,B) \times \mathscr{C}(S \otimes A, S)$$

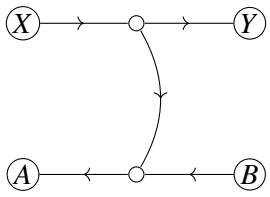
where the isomorphism follows by Yoneda reduction. Note that this requires that the 'output' map of the dynamical system is a comonoid homomorphism in $\mathscr{C}$ and hence deterministic in a category of stochastic channels.

Definition 4.2 (Category of discrete-time dynamical systems)**.** We define a category $\mathbf{Dyn}_{\mathscr{C}}$ whose objects are the objects of $\mathscr{C}$ and whose morphisms, denoted $A \xrightarrow{S} B$, are discrete-time dynamical systems; the symbol above the arrow denotes the internal state space. Hom objects are given by

$$\mathbf{Dyn}_{\mathscr{C}}(A,B) = \sum_{S:\mathscr{C}} \mathbf{Comon}(\mathscr{C})(S,B) \times \mathscr{C}(S \otimes A, S)\,.$$

Identity dynamical systems on each $A : \mathscr{C}$ are the 'no-op' dynamical systems $A \xrightarrow{A} A$ given by identity optics $\mathsf{id}_A : (A,A) \nrightarrow (A,A)$. Associativity and unitality of composition is inherited from the category of optics underlying Definition 4.1; a symmetric monoidal structure is similarly inherited. □

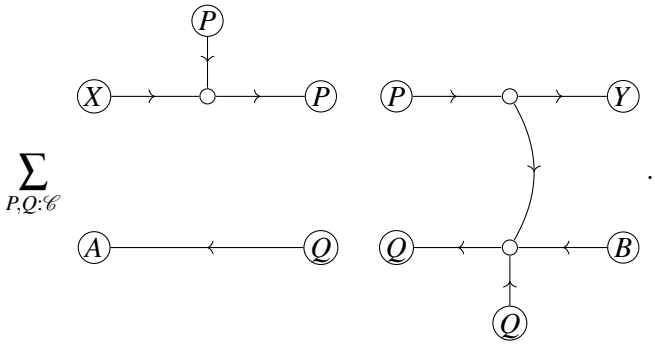
Definition 4.3 (Lenses over dynamical systems; after [21])**.** The category of (monoidal) lenses over $\mathscr{C}$-dynamical systems has as objects pairs (X,A) of objects in $\mathscr{C}$ and as morphisms, **dynamical lenses** $(X,A) \nrightarrow (Y,B)$, elements of the type



$$\int^{M:\mathscr{C}} \mathbf{Dyn}_{\mathscr{C}}(X, M \otimes Y) \times \mathbf{Dyn}_{\mathscr{C}}(M \otimes B, A)$$

$$\cong$$

$$\sum_{P,Q:\mathscr{C}} \int^{M:\mathscr{C}} \mathscr{C}(P \otimes X, P) \times \mathbf{Comon}(\mathscr{C})(P, M \otimes Y) \times \mathscr{C}(Q \otimes M \otimes B, Q) \times \mathbf{Comon}(\mathscr{C})(Q, A)$$

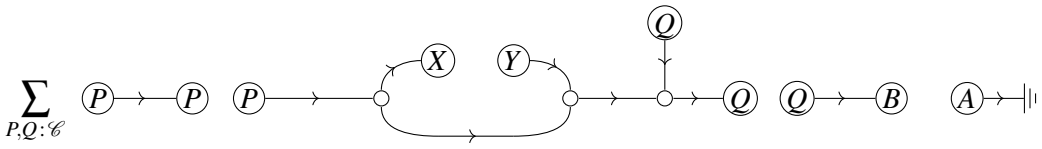


That is, a dynamical lens is a pair of dynamical systems coupled along some 'residual' type.

Remark 4.4. At this point we begin to run into sizes issues. However, for the purposes of this paper, we will simply assume that a satisfactory resolution of these matters is at hand; for instance, that there is a hierarchy of Grothendieck universes such that the coends over ('large') sums in the preceding definition constitute accessible objects.

We now expand the definition of context in the dynamical setting. We will see that a dynamical context is simply a closure of an open dynamical system: that is, a 'larger' system into which a 'smaller' open dynamical system can plug such that the composite is a closed (but still uninitialized) system.

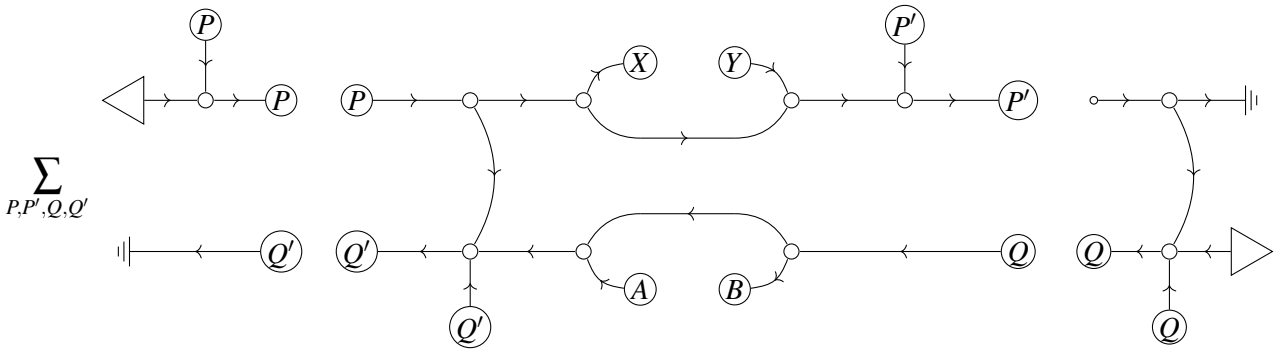
Proposition 4.5. If I is terminal in $\mathscr{C}$, a context for a dynamical lens $(X,A) \nrightarrow (Y,B)$ is an element of the following type, denoted $\tilde{C}\big((X,A),(Y,B)\big)$:



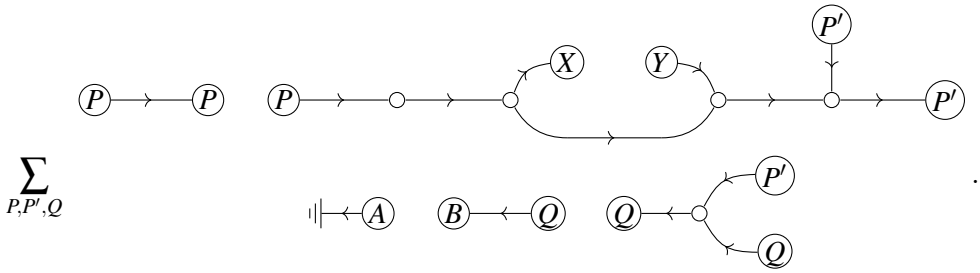
Interpreting this diagram, a context for a dynamical lens $(X,A) \nrightarrow (Y,B)$ amounts to an autonomous dynamical system with output type of the form $X \otimes M$ (for some residual type M), coupled along the

residual M to an open dynamical system with input type $Y \otimes M$ and output type B; and the A type is discarded. This is precisely what we should expect from a dynamical analogue of Proposition 3.5.

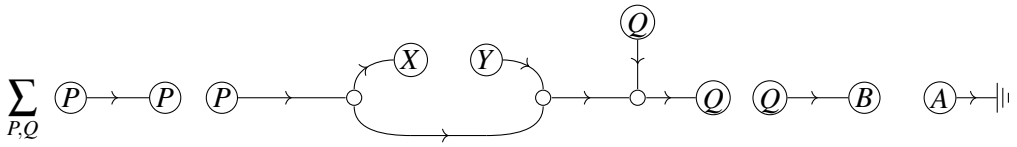
Proof of Proposition 4.5. Writing it out in full, we find that the context $\tilde{C}\big((X,A),(Y,B)\big)$ is an element of the following type:



By discarding the factors that output into the terminal object, the context reduces to



Since P' is an arbitrary choice of internal state space type, we can without loss of generality set $P' = Q$, which gives us



as required. □

Definition 4.6. A **dynamical game** is just a generalized open game (3.6) over the category of dynamical lenses. We write $(X,A) \xrightarrow{\tilde{\Sigma},S} (Y,B)$ to indicate both the strategy type $\tilde{\Sigma}$ and state space S. Dynamical games form a symmetric monoidal category in the corresponding way. For notational clarity, we will write $\tilde{G}$ for a dynamical game, $\tilde{P}$ for its play function, and $\tilde{B}$ for its best response function.

Definition 4.7 (Dynamical realisation of an open game)**.** Let $G : (X,A) \xrightarrow{\Sigma} (Y,B)$ be an open game with X,A,Y,B all objects of some symmetric monoidal category $\mathscr{C}$. A **dynamical realisation** of G is a choice of dynamical game $\tilde{G} : (X,A) \xrightarrow{\tilde{\Sigma},S} (Y,B)$ on the same objects, along with a function $[\![\cdot]\!] : C((X,A),(Y,B)) \to \tilde{C}((X,A),(Y,B))$ lifting static contexts to dynamical contexts. Given a context $\langle \pi \mid k \rangle : C((X,A),(Y,B))$, we choose a representative $\langle [\![\pi]\!] \mid [\![k]\!] \rangle \cong [\![\langle \pi \mid k \rangle]\!] : \tilde{C}((X,A),(Y,B))$ for its realisation.

A 'dynamical context' is an element of the type given in Proposition 4.5: a context for a dynamical lens. A 'static context' is simply a context for the 'static' game that is being dynamically realized. At this stage, we impose no particular requirements on the context realisation function $[\![\cdot]\!]$, except to say that in the intended semantics, $[\![\langle \pi \,|\, k \rangle]\!]$ is a (coupled, open) dynamical system that constantly emits the state π and (by some mechanism) realizes the channel k. We call such a context *stationary* as neither π nor k vary in time; future work will generalize the results of this section to *non-stationary* contexts.

Definition 4.8 (Open cybernetic systems)**.** An open **cybernetic system** is defined by the data:

- an open optimization game (Def. 3.22) $G : (X,A) \xrightarrow{\Sigma} (Y,B)$ with X,A,Y,B all objects of some symmetric monoidal category $\mathscr{C}$,
- a fitness function $\varphi_G : \Sigma \times C \to M \xrightarrow{\varphi} F$ where $C = C((X,A),(Y,B))$,
- a dynamical realisation $\big(\tilde{G} : (X,A) \xrightarrow{\tilde{\Sigma},S} (Y,B), [\![\cdot]\!] : C((X,A),(Y,B)) \to \tilde{C}((X,A),(Y,B))\big)$ of G,

satisfying the following condition for each context $\langle \pi \,|\, k \rangle : C((X,A),(Y,B))$:

- there exists a dynamical strategy $\tilde{\sigma} : \tilde{\Sigma}$, such that
- writing Z for the total state space of the autonomous dynamical system $[\![\langle \pi \,|\, k \rangle]\!] \odot \tilde{P}(\tilde{\sigma})$ induced by the context, there exists a function $v : Z \to M$ projecting Z into the 'fitness landscape' M, such that
- there exists a fitness-maximising fixed point $\zeta^* : Z$, in the sense that
- for some equilibrium strategy of the static system $\sigma^* : \mathrm{fix}\, B(\langle \pi \,|\, k \rangle)$, $\varphi(v(\zeta^*)) \leq \varphi_G(\sigma^*, \langle \pi \,|\, k \rangle)$.

A **category of open cybernetic systems** is a category of (generalized) open games such that each game is an open cybernetic system with dynamics realised in the same category $\mathscr{C}$, and such that the composite of games is a cybernetic system whose fitness-maximising fixed point projects onto fitness-maximising fixed points of each of the factors in their corresponding local contexts. (See [2, §3.7] for the definition of local context.)

The idea here is that, by using the fitness function of the underlying optimization game, the cybernetic condition forces the behaviour of the dynamical realisation to coincide with the process of iteratively improving the strategies deployed by the system in playing the game. We summarize the condition in the diagram

$$\begin{array}{ccccc} \Sigma \times C & \longrightarrow & M & \xrightarrow{\varphi} & F \\ {\scriptstyle [\![\cdot]\!]}\downarrow & & \uparrow & & \\ \tilde{\Sigma} \times \tilde{C} & \xrightarrow{\mathrm{fix}} & Z & & \end{array}$$

though this is in general ill-defined: we do not require a function $[\![\cdot]\!] : \Sigma \to \tilde{\Sigma}$, and nor do we require that the best response to $\tilde{G}$ coincides in any way with the best reponse to G. Investigating such conditions is the subject of future work; for instance, we may be interested in nested cybernetic systems, such as characterize evolution by natural selection, and how their fitness functions constrain one another. For similar reasons, we are also interested in the case where the fitness function is itself non-stationary.

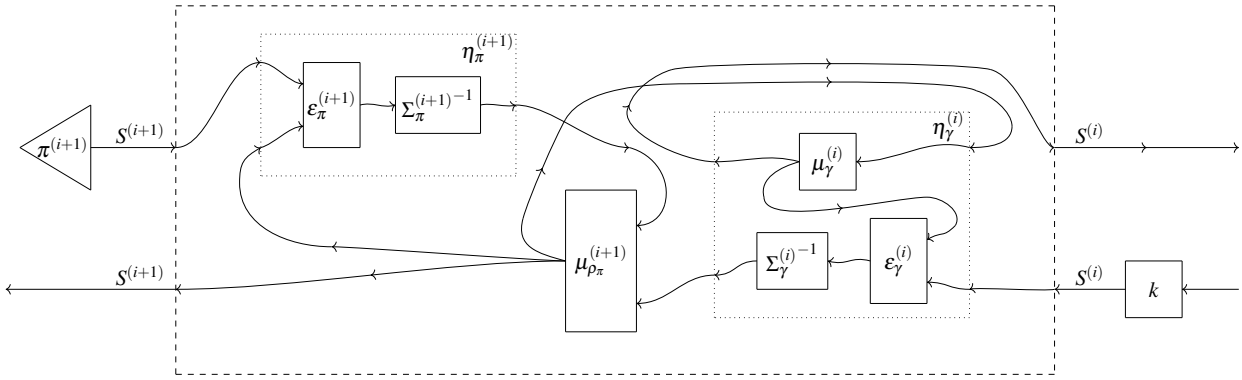
Remark 4.9. The codomain category of the cybernetic realisation functor is in general much larger than the domain category of static games, and often it makes sense to consider dynamical games in this codomain category as if they were dynamical realisations of static games, even if in fact there is no static game to which they could correspond. For instance, adaptive systems in physical environments are in general not realisations of static games because their contexts are irreducibly dynamical and thus not

the dynamical realisation of a static context; but over short time intervals, it can be productive to treat such systems as realisations of static games. In continuous time (not treated here), it is even possible to consider dynamical games that are indeed realisations of games that are static when represented in a smoothly varying coordinate system. The free-energy framework of Theorem 4.10 is an example of a category of cybernetic systems with a rich underlying category of dynamic games.

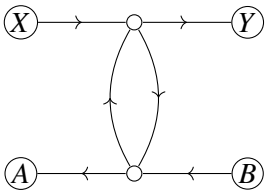
A classic category of open cybernetic systems is found in the computational neuroscience literature, as summarized in the following theorem.

Theorem 4.10. Consider the subcategory of **BayesLens** spanned by finite-dimensional Euclidean spaces, with morphisms generated (under sequential and parallel composition) by the (variational) autoencoder and inference games whose forwards and backwards channels emit Gaussian measures with high-precision. The (discrete-time) free-energy framework for action and perception [3] instantiates a category of open cybernetic systems realising games over this subcategory.

Proof of Theorem 4.10. By generalizing the exposition given by **(author?)** [3] and rendering it compositional, we show that every such game is realised by a dynamical system of the following form, depicted as a wiring diagram [24], and that the dynamics has a fixed point corresponding to a best-response to the underlying game:



While it is possible to give a dynamical realisation by a system of the form in Definition 4.3, it is more parsimonious to allow the 'forwards' view system to receive inputs from the 'backwards' update system, as well as vice versa. In a sense, we want the coupling along the residual to be bidirectional. We can depict this as the following type:



$$\cong \int^{M,N:\mathscr{C}} \mathbf{Dyn}_{\mathscr{C}}(M\otimes X, N\otimes Y)\times \mathbf{Dyn}_{\mathscr{C}}(N\otimes B, M\otimes A)$$

We can choose residuals of type $N=A$ and $M=B$, so that representatives are of the type $\mathbf{Dyn}_{\mathscr{C}}(X\otimes B, Y\otimes A)\times \mathbf{Dyn}_{\mathscr{C}}(A\otimes B, A\otimes B)$. We can then choose the second factor of the representative to be the identity system, leaving us only needing to construct a system of type $\mathbf{Dyn}_{\mathscr{C}}(X\otimes B, Y\otimes A)$, for some X, A, Y, B. We call such a realisation *bidirectionally coupled*. Reconciling this type with the type given in Definition 4.3 is the subject of ongoing work. Note that we retain the form for the dynamical context given in Proposition 4.5.

We will work in the 'convenient' setting of Example 2.2, in which $\mathscr{C} = \mathscr{K}\ell(\mathscr{P})$ for some probability monad $\mathscr{P} : \mathbf{Meas} \to \mathbf{Meas}$ on a Cartesian closed category of measurable spaces, such as the category of quasi-Borel spaces [14]. In this setting, a Bayesian lens $(X,A) \nrightarrow (Y,B)$ is given by a pair of maps of type $\mathbf{Meas}(X, \mathscr{P}Y) \times \mathbf{Meas}(\mathscr{P}X \times B, \mathscr{P}A)$. A bidirectionally coupled dynamical realisation is then of the type $\mathbf{Dyn}_{\mathscr{K}\ell(\mathscr{P})}(X \otimes B, Y \otimes A)$, or equivalently

$$\sum_{Z:\mathscr{K}\ell(\mathscr{P})} \mathscr{K}\ell(\mathscr{P})(Z \otimes X \otimes B, Z) \times \mathbf{Comon}\big(\mathscr{K}\ell(\mathscr{P})\big)(Z, Y \times A)\,.$$

A corresponding dynamical context has the type:

$$\sum_{Z_\pi, Z_k : \mathscr{K}\ell(\mathscr{P})} \int^{M:\mathscr{K}\ell(\mathscr{P})} \mathscr{K}\ell(\mathscr{P})(Z_\pi, Z_\pi) \times \mathbf{Comon}\big(\mathscr{K}\ell(\mathscr{P})\big)(Z_\pi, M \times X) \\ \times \mathscr{K}\ell(\mathscr{P})(M \otimes Z_k \otimes Y, Z_k) \times \mathbf{Comon}\big(\mathscr{K}\ell(\mathscr{P})\big)(Z_k, B)$$

We will consider first simple variational autoencoder games (as in Example 3.16), of type $\big\langle \gamma^{(i)} \,\big|\, \rho^{(i+1)} \big\rangle : (S^{(i+1)}, S^{(i+1)}) \nrightarrow (S^{(i)}, S^{(i)})$, in which the spaces $S^{(i)}$ are Euclidean, and we restrict to Gaussian measures over them. Let $\pi^{(i+1)} : \mathscr{P}S^{(i+1)}$ be a prior on $S^{(i+1)}$. We will write $p^{(i)}_\gamma : S^{(i)} \times S^{(i+1)} \to [0,1]$, $p^{(i+1)}_{\rho_\pi} : S^{(i+1)} \times S^{(i)} \to [0,1]$, and $p^{(i+1)}_\pi : S^{(i+1)} \to [0,1]$ for the corresponding density functions. The objective of such a variational autoencoder game is then

$$\underset{\substack{\gamma^{(i)} \in \Gamma^{(i)}, \\ \rho^{(i+1)} \in \mathbf{Meas}(\mathscr{P}S^{(i+1)}, \mathrm{P}^{(i+1)})}}{\arg\min} \; \mathop{\mathbb{E}}_{s^{(i)} \sim k \bullet \gamma^{(i)} \bullet \pi^{(i+1)}} \; \mathop{\mathbb{E}}_{s^{(i+1)} \sim \rho_\pi(s^{(i)})} \left[\log \frac{p_{\rho_\pi}(s^{(i+1)} | s^{(i)})}{p_\gamma(s^{(i)} | s^{(i+1)})\, p_\pi(s^{(i+1)})} \right]$$

where $\Gamma^{(i)} \subseteq \mathscr{K}\ell(\mathscr{P})(S^{(i+1)}, S^{(i)})$ and $\mathrm{P}^{(i+1)} \subseteq \mathscr{K}\ell(\mathscr{P})(S^{(i)}, S^{(i+1)})$ are the subtypes of stochastic channels emitting the corresponding Gaussian measures.

Note that these types are families of *parameterized* stochastic channels, in the sense that each $\gamma^{(i)}(s^{(i+1)})$, $\rho^{(i+1)}(s^{(i)})$ and $\pi^{(i+1)}$ can be completely specified by parameter vectors $\vartheta^{(i)}_\gamma(s^{(i+1)}) : \Theta^{(i)}_\gamma$, $\vartheta^{(i+1)}_{\rho_\pi}(s^{(i)}) : \Theta^{(i+1)}_{\rho_\pi}$, and $\vartheta^{(i+1)}_\pi : \Theta^{(i+1)}_\pi$, where we have denoted the parameter spaces by Θ. Consequently, the channels $\gamma^{(i)}$ and $\rho^{(i+1)}_\pi$ themselves are completely specified by functions $\vartheta^{(i)}_\gamma : \mathbf{Meas}(S^{(i+1)}, \Theta^{(i)}_\gamma)$ and $\vartheta^{(i+1)}_{\rho_\pi} : \mathbf{Meas}(S^{(i)}, \Theta^{(i+1)}_{\rho_\pi})$. Indeed, we can factor the channels as $\gamma^{(i)} : S^{(i+1)} \xrightarrow{\vartheta} \Theta^{(i)}_\gamma \hookrightarrow \mathscr{P}S^{(i)}$ and $\rho^{(i+1)}_\pi : S^{(i)} \xrightarrow{\vartheta} \Theta^{(i+1)}_{\rho_\pi} \hookrightarrow \mathscr{P}S^{(i+1)}$, where the inclusions map a parameter vector onto the corresponding measure.

Since each measure is Gaussian, the relevant parameters are means and covariances, which we denote by μ and Σ, so that $\vartheta^{(i)}_\gamma(s^{(i+1)}) = \big(\mu^{(i)}_\gamma, \Sigma^{(i)}_\gamma\big)(s^{(i+1)}) : \big(S^{(i)} \otimes \mathbf{FVect}_{\mathbb{R}}(S^{(i)}, S^{(i)})\big) = \Theta^{(i)}_\gamma$, and similarly for ρ_π and π. Write $|S^{(i)}|$ to denote the dimension of the space $|S^{(i)}|$, and $\langle \cdot, \cdot \rangle$ for the inner product of vectors. We can write down the (log) density functions explicitly as

$$\log p^{(i)}_\gamma(s^{(i)} | s^{(i+1)}) = \frac{1}{2} \left\langle \varepsilon^{(i)}_\gamma, {\Sigma^{(i)}_\gamma}^{-1} \varepsilon^{(i)}_\gamma \right\rangle - \log \sqrt{(2\pi)^{|S^{(i)}|} \det \Sigma^{(i)}_\gamma}$$

$$\log p^{(i+1)}_{\rho_\pi}(s^{(i+1)} | s^{(i)}) = \frac{1}{2} \left\langle \varepsilon^{(i+1)}_{\rho_\pi}, {\Sigma^{(i+1)}_{\rho_\pi}}^{-1} \varepsilon^{(i+1)}_{\rho_\pi} \right\rangle - \log \sqrt{(2\pi)^{|S^{(i+1)}|} \det \Sigma^{(i)}_{\rho_\pi}}$$

$$\log p^{(i+1)}_\pi(s^{(i+1)}) = \frac{1}{2} \left\langle \varepsilon^{(i+1)}_\pi, {\Sigma^{(i+1)}_\pi}^{-1} \varepsilon^{(i+1)}_\pi \right\rangle - \log \sqrt{(2\pi)^{|S^{(i+1)}|} \det \Sigma^{(i+1)}_\pi}$$

where for clarity we have omitted the dependence of $\Sigma_\gamma^{(i)}$ on $s^{(i+1)}$ and $\Sigma_{\rho_\pi}^{(i+1)}$ on $s^{(i)}$, and where

$$\begin{aligned}\varepsilon_\gamma^{(i)} &:= s^{(i)} - \mu_\gamma^{(i)}(s^{(i+1)})\\ \varepsilon_{\rho_\pi}^{(i+1)} &:= s^{(i+1)} - \mu_{\rho_\pi}^{(i+1)}(s^{(i)})\\ \varepsilon_\pi^{(i+1)} &:= s^{(i+1)} - \mu_\pi^{(i+1)}.\end{aligned}$$

We now construct a dynamical system performing gradient descent on the free energy

$$\begin{aligned}\varphi^{(i)} : S^{(i)} \to \mathbb{R}_+ := s^{(i)} \mapsto &\mathop{\mathbb{E}}_{s^{(i+1)}\sim\rho_\pi(s^{(i)})}\left[\log\frac{p_{\rho_\pi}(s^{(i+1)}|s^{(i)})}{p_\gamma(s^{(i)}|s^{(i+1)})\,p_\pi(s^{(i+1)})}\right]\\ &= \mathrm{H}\left[\rho_\pi^{(i+1)}(s^{(i)})\right] - \mathop{\mathbb{E}}_{s^{(i+1)}\sim\rho_\pi(s^{(i)})}\left[\log p_\gamma(s^{(i)}|s^{(i+1)}) + \log p_\pi(s^{(i+1)})\right]\\ &= \mathrm{H}\left[\rho_\pi^{(i+1)}(s^{(i)})\right] + \mathop{\mathbb{E}}_{s^{(i+1)}\sim\rho_\pi(s^{(i)})}\left[E^{(i)}(s^{(i+1)}, s^{(i)})\right].\end{aligned}$$

where $\mathrm{H}[\cdot]$ is the Shannon entropy functor and $E^{(i)}(s^{(i+1)}, s^{(i)})$ is called the *energy* of the system. Note that this defines a fitness function $\varphi^{(i)}$ for the variational autoencoder game, in the sense of Definitions 3.22 and 4.8, by

$$\varphi_G : \Gamma \times \mathrm{P} \times C \to \mathbb{R}_+ := (\gamma, \rho, \langle \pi \,|\, k\rangle) \mapsto \mathop{\mathbb{E}}_{k\bullet\gamma\bullet\pi}\left[\varphi^{(i)}\right].$$

Note also that under the assumption that the posterior $\rho_\pi^{(i+1)}(s^{(i)})$ has small variance, we can approximate the expected energy by its second-order Taylor expansion around the mean $\mu_{\rho_\pi}^{(i+1)}(s^{(i)})$, so that

$$\begin{aligned}\varphi^{(i)}(s^{(i)}) \approx E^{(i)}(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}) + \frac{1}{2}\left\langle \varepsilon_{\rho_\pi}^{(i+1)}, \left(\partial^2_{s^{(i+1)}} E^{(i)}\right)\left(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}\right)\varepsilon_{\rho_\pi}^{(i+1)}\right\rangle\\ + \mathrm{H}\left[\rho_\pi^{(i+1)}(s^{(i)})\right].\end{aligned}$$

where $\left(\partial^2_{s^{(i+1)}} E^{(i)}\right)\left(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}\right)$ is the Hessian of $E^{(i)}$ with respect to $s^{(i+1)}$ evaluated at $(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)})$. Note that

$$\left\langle \varepsilon_{\rho_\pi}^{(i+1)}, \left(\partial_{s^{(i+1)}}{}^2 E^{(i)}\right)\left(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}\right)\varepsilon_{\rho_\pi}^{(i+1)}\right\rangle = \mathrm{tr}\left[\left(\partial^2_{s^{(i+1)}} E^{(i)}\right)\left(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}\right)\Sigma_{\rho_\pi}^{(i+1)}(s^{(i)})\right],$$

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

$$\mathrm{H}\left[\rho_\pi^{(i+1)}(s^{(i)})\right] = \frac{1}{2}\log\det\left(2\pi e\,\Sigma_{\rho_\pi}^{(i+1)}(s^{(i)})\right),$$

and that the energy $E^{(i)}(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)})$ does not depend on $\Sigma_{\rho_\pi}^{(i+1)}(s^{(i)})$. We can therefore write down directly the covariance $\Sigma_{\rho_\pi}^{(i+1)^*}(s^{(i)})$ minimizing $\varphi^{(i)}(s^{(i)})$ as a function of $s^{(i)}$. We have

$$\partial_{\Sigma_{\rho_\pi}^{(i+1)}}\varphi^{(i)}(s^{(i)}) \approx \frac{1}{2}\left(\partial^2_{s^{(i+1)}} E\right)\left(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}\right) + \frac{1}{2}\Sigma_{\rho_\pi}^{(i+1)^{-1}}.$$

Setting $\partial_{\Sigma_{\rho_\pi}^{(i+1)}}\varphi^{(i)}(s^{(i)}) = 0$, we find the optimum

$$\Sigma_{\rho_\pi}^{(i+1)^*}(s^{(i)}) = \left(\partial^2_{s^{(i+1)}} E\right)\left(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}\right)^{-1}. \tag{9}$$

Next, we optimize the mean parameter $\mu_{\rho_\pi}^{(i+1)}$. It suffices to consider only the energy term $E^{(i)}(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)})$. We have

$$\begin{aligned} E^{(i)}(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}) &= -\log p_\gamma(s^{(i)}|s^{(i+1)}) - \log p_\pi(s^{(i+1)}) \\ &= -\frac{1}{2}\left\langle \varepsilon_\gamma^{(i)}, {\Sigma_\gamma^{(i)}}^{-1} \varepsilon_\gamma^{(i)} \right\rangle - \frac{1}{2}\left\langle \varepsilon_\pi^{(i+1)}, {\Sigma_\pi^{(i+1)}}^{-1} \varepsilon_\pi^{(i+1)} \right\rangle \\ &\quad + \log\sqrt{(2\pi)^{|S^{(i)}|}\det\Sigma_\gamma^{(i)}} + \log\sqrt{(2\pi)^{|S^{(i+1)}|}\det\Sigma_\pi^{(i+1)}} \end{aligned}$$

and a simple computation shows that

$$\partial_{\mu_{\rho_\pi}^{(i+1)}} E^{(i)}(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}) = -\left(\partial_{s^{(i+1)}}\mu_\gamma^{(i)}\right)(\mu_{\rho_\pi}^{(i+1)})^T {\Sigma_\gamma^{(i)}}^{-1}\varepsilon_\gamma^{(i)} + {\Sigma_\pi^{(i+1)}}^{-1}\varepsilon_\pi^{(i+1)}.$$

Define the precision-weighted error functions

$$\begin{aligned} \eta_\gamma^{(i)} : S^{(i)} \times S^{(i)} \xrightarrow{\varepsilon_\gamma^{(i)}} S^{(i)} \xrightarrow{{\Sigma_\gamma^{(i)}}^{-1}} S^{(i)} \quad \text{and} \\ \eta_\pi^{(i+1)} : S^{(i+1)} \times S^{(i+1)} \xrightarrow{\varepsilon_\pi^{(i+1)}} S^{(i+1)} \xrightarrow{{\Sigma_\pi^{(i+1)}}^{-1}} S^{(i+1)} \end{aligned}$$

so that

$$\partial_{\mu_{\rho_\pi}^{(i+1)}} E^{(i)}(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}) = -\left(\partial_{s^{(i+1)}}\mu_\gamma^{(i)}\right)(\mu_{\rho_\pi}^{(i+1)})^T \eta_\gamma^{(i)} + \eta_\pi^{(i+1)}.$$

We can therefore define a discrete-time dynamical system with update function

$$\begin{aligned} \mu_{\rho_\pi}^{(i+1)} : S^{(i+1)} \times S^{(i)} &\to S^{(i+1)} \\ \mu_{\rho_\pi}^{(i+1)}(t) \times s^{(i)} &\mapsto \mu_{\rho_\pi}^{(i+1)}(t+1) \end{aligned} \tag{10}$$

where

$$\begin{aligned} \mu_{\rho_\pi}^{(i+1)}(t+1) &= \mu_{\rho_\pi}^{(i+1)}(t) \;-\; \lambda\left(\partial_{s^{(i+1)}}\mu_\gamma^{(i)}\right)(\mu_{\rho_\pi}^{(i+1)})^T \eta_\gamma^{(i)}\left(s^{(i)}, \mu_\gamma^{(i)}(\mu_{\rho_\pi}^{(i+1)}(t))\right) \;+\; \lambda\,\eta_\pi^{(i+1)}\left(\mu_{\rho_\pi}^{(i+1)}(t), \mu_\pi^{(i+1)}\right) \\ &= \mu_{\rho_\pi}^{(i+1)}(t) \;-\; \lambda\,\partial_{\mu_{\rho_\pi}^{(i+1)}} E^{(i)}\left(\mu_{\rho_\pi}^{(i+1)}(s^{(i)}), s^{(i)}\right) \end{aligned}$$

for some 'learning rate' $\lambda : \mathbb{R}_+$, and identity output function. Note that these dynamics depend on the parameters of the prior π, thereby witnessing the exponential structure of the Bayesian inversion operation as a 'state-dependent stochastic channel' (Definition 2.4); and, less subtly, the dynamics for the inverse channel $\rho_\pi^{(i+1)}$ depend on the forwards channel $\gamma^{(i)}$.

Equations (10) and (9) together give a dynamical system with input space $S^{(i)}$ and state space $\Theta_{\rho_\pi}^{(i+1)} = S^{(i+1)} \times \mathbf{FVect}_{\mathbb{R}}(S^{(i+1)}, S^{(i+1)})$. Given an identity output function and since $\Theta_{\rho_\pi}^{(i+1)} \hookrightarrow \mathscr{P}S^{(i+1)}$, we have a stochastic dynamical system of type $\mathbf{Dyn}_{\mathscr{K}\ell(\mathscr{P})}(S^{(i)}, S^{(i+1)})$ realising the update factor (or 'recognition model') $\rho_\pi^{(i+1)}$ of the variational autoencoder game. We now give a corresponding dynamical system for the view factor (or 'generative model') $\gamma^{(i)}$. □

Remark 4.11. Typical presentations of 'active inference' under the free-energy principle are excessively complicated by the lack of attention paid to compositionality. Because the free-energy framework instantiates a *category* of open cybernetic systems, a radically simplified compositional presentation is possible. Such a presentation forms a companion to the present work.

Corollary 4.12. The free-energy framework has been used to supply a computational explanation for the pervasive bidirectionality of cortical circuits in the mammalian brain [1, 10]. A corollary of Theorem 4.10 is that this bidirectionality is furthermore justified by the abstract structure of Bayesian inference and its dynamical realisation: because Bayesian updates compose optically, a cybernetic system realising Bayesian inference compositionally must instantiate this structure. We note also that the parallel interacting bidirectional structure of the active inference game (Example 3.20) is reproduced in the cortex.

The free-energy framework realisation of autoencoder games is not unique; an alternative is found in machine learning.

Theorem 4.13. Consider the subcategory of **BayesLens** spanned by finite-dimensional Euclidean spaces, with morphisms generated (under sequential and parallel composition) by the (variational) autoencoder and inference games whose forwards and backwards channels emit exponential-family measures. The deep (variational) autoencoder framework [15] instantiates a category of open cybernetic systems realising games over this subcategory.

Increasingly, the variational autoencoder framework is used to model complete agents in machine learning, rather than merely dynamically realise static inference or learning problems. Indeed, thinking of the 'free-energy framework' as a collection of cybernetic realisations of autoencoder and active-inference games, the demonstration of the following corollary of Theorem 4.13 is unsurprising:

Corollary 4.14. The "deep active inference agent" [27] is a cybernetic system realising an active inference game in the variational autoencoder framework.

We have heretofore concentrated on 'variational Bayesian' realisations of the games introduced in §3, as they most strikingly fit the language of optimization used there. But we expect any other family of approximate inference methods to supply a corresponding category of cybernetic systems. We thus make the following conjecture.

Conjecture 4.15. Consider the subcategory of **BayesLens** spanned by finite-dimensional smooth manifolds, with morphisms generated (under sequential and parallel composition) by the generalized autoencoder and inference games. We expect sampling algorithms, such as Markov chain Monte Carlo, to supply a corresponding category of open cybernetic systems of interest.

Finally, we provide further justification for Remark 3.7.

Observation 4.16. Consider a variational autoencoder, realised as in Theorem 4.13. By choosing the parameterizations $\mathscr{F}, \mathscr{P}$ of the forwards and backwards channels to coincide with the state spaces of their dynamical realisations, and the (static) play function P to take a parameter vector to the corresponding channel, the dynamical realisation induces a trajectory over the strategy space. Such trajectories organize into sheaf whose sections are trajectories of arbitrary length [24], spans of which are again just (generalized) dynamical systems; these spans are equivalently profunctors [4]. We can thus define a best-response function valued in profunctors whose elements are trajectories witnessing deviations of strategies to 'better' strategies, and whose dynamical equilibria correspond precisely to the equilibria of the 'static' best response function.

On-going and Future Work The structures sketched in this paper are merely first steps towards a categorical theory of cybernetics. In particular, since the first draft of this work was written, we have come to believe that the preliminary notions presented here of dynamical realisation, and by extension of open cybernetic system, are substantially less elegant than they could be. On-going work is focusing on this issue. We hope that a consequence of this refinement will be that the treatment of *interacting* cybernetic

systems is simplified. In this new setting, we will also treat non-stationary systems in dynamical contexts and in continuous time, thereby supplying a general compositional treatment of (amongst other things) the 'free-energy' framework.

Finally, with respect to applications, we are interested in using these tools to realise game-theoretic games and to investigate the connections between repeated games and dynamical realisation. There are deep links with reinforcement learning to be explored, and we seek a setting for the study of nested and mutli-agent ('ecological') systems.

References

[1] A. M. Bastos, W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries & K. J. Friston (2012): *Canonical microcircuits for predictive coding*. Neuron 76(4), pp. 695–711, doi:10.1016/j.neuron.2012.10.038.

[2] Joe Bolt, Jules Hedges & Philipp Zahn (2019): *Bayesian open games*. Available at http://arxiv.org/abs/1910.03656v1.

[3] Christopher L Buckley, Chang Sub Kim, Simon McGregor & Anil K Seth (2017): *The free energy principle for action and perception: A mathematical review*. Journal of Mathematical Psychology 81, pp. 55–79, doi:10.1016/j.jmp.2017.09.004. Available at http://arxiv.org/abs/1705.09156v1.

[4] Jean Bénabou (2000): *Distributors at work*. Available at http://www.mathematik.tu-darmstadt.de/~streicher/FIBR/DiWo.pdf. Lecture notes written by Thomas Streicher.

[5] Kenta Cho & Bart Jacobs (2017): *Disintegration and Bayesian Inversion via String Diagrams*. *Math. Struct. Comp. Sci. 29 (2019) 938-971*, doi:10.1017/S0960129518000488. Available at http://arxiv.org/abs/1709.00322v3.

[6] Bryce Clarke, Derek Elkins, Jeremy Gibbons, Fosco Loregian, Bartosz Milewski, Emily Pillmore & Mario Román (2020): *Profunctor optics, a categorical update*. Available at http://arxiv.org/abs/2001.07488v1.

[7] Bob Coecke & Aleks Kissinger (2016): *Categorical Quantum Mechanics II: Classical-Quantum Interaction*. doi:10.1142/S0219749910006502. Available at http://arxiv.org/abs/1605.08617v1.

[8] Bob Coecke & Aleks Kissinger (2017): *Categorical Quantum Mechanics I: Causal Quantum Processes*. In Elaine Landry, editor: *Categories for the Working Philosopher*, chapter 12, Oxford University Press, pp. 286–328. Available at https://arxiv.org/abs/1510.05468v3.

[9] J. Nathan Foster, Michael B. Greenwald, Jonathan T. Moore, Benjamin C. Pierce & Alan Schmitt (2007): *Combinators for bidirectional tree transformations*. *ACM Transactions on Programming Languages and Systems* 29(3), p. 17, doi:10.1145/1232420.1232424.

[10] Karl Friston (2010): *The free-energy principle: a unified brain theory?* Nature Reviews Neuroscience 11(2), pp. 127–138, doi:10.1038/nrn2787.

[11] Karl Friston, Francesco Rigoli, Dimitri Ognibene, Christoph Mathys, Thomas Fitzgerald & Giovanni Pezzulo (2015): *Active inference and epistemic value*. Cognitive Neuroscience 6(4), pp. 187–214, doi:10.1080/17588928.2015.1020053.

[12] Tobias Fritz (2019): *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics*. Available at http://arxiv.org/abs/1908.07021v8.

[13] Neil Ghani, Jules Hedges, Viktor Winschel & Philipp Zahn (2016): *Compositional game theory*. *Proceedings of Logic in Computer Science (LiCS) 2018*, doi:10.1145/3209108.3209165. Available at http://arxiv.org/abs/1603.04641v3.

[14] Chris Heunen, Ohad Kammar, Sam Staton & Hongseok Yang (2017): *A Convenient Category for Higher-Order Probability Theory*. doi:10.1109/lics.2017.8005137.

[15] Diederik P. Kingma (2017): *Variational Inference & Deep Learning*. Available at https://hdl.handle.net/11245.1/8e55e07f-e4be-458f-a929-2f9bc2d169e8.

[16] Jeremias Knoblauch, Jack Jewson & Theodoros Damoulas (2019): *Generalized Variational Inference*. Available at http://arxiv.org/abs/1904.02063v4.

[17] Fosco Loregian (2015): *This is the (co)end, my only (co)friend*. Available at http://arxiv.org/abs/1501.02503v4.

[18] Paul-André Melliès (2006): *Functorial Boxes in String Diagrams*. In: *Computer Science Logic*, Springer Berlin Heidelberg, pp. 1–30, doi:10.1007/11874683_1.

[19] Joe Moeller & Christina Vasilakopoulou (2018): *Monoidal Grothendieck Construction*. Available at http://arxiv.org/abs/1809.00727v2.

[20] nLab authors (2020): *Grothendieck construction*. Available at http://ncatlab.org/nlab/show/Grothendieck+construction. Revision 62.

[21] Mitchell Riley (2018): *Categories of Optics*. Available at http://arxiv.org/abs/1809.00738v2.

[22] Mario Román (2020): *Open Diagrams via Coend Calculus*. Available at http://arxiv.org/abs/2004.04526v2.

[23] Mario Román (2020): *Profunctor optics and traversals*. Available at http://arxiv.org/abs/2001.08045v1.

[24] Patrick Schultz, David I Spivak & Christina Vasilakopoulou (2019): *Dynamical Systems and Sheaves*. *Applied Categorical Structures*, pp. 1–57, doi:10.1007/s10485-019-09565-x. Available at http://arxiv.org/abs/1609.08086v4.

[25] Toby St. Clere Smithe (2020): *Bayesian Updates Compose Optically*. Available at http://arxiv.org/pdf/2006.01631v1.

[26] David I. Spivak (2019): *Generalized Lens Categories via functors* $\mathscr{C}^{op} \to$ **Cat**. Available at http://arxiv.org/abs/1908.02202v2.

[27] Kai Ueltzhöffer (2018): *Deep Active Inference*. *Biological Cybernetics* 112(6), pp. 547–573, doi:10.1007/s00422-018-0785-7. Available at http://arxiv.org/abs/1709.02341v5.