Active Inference in String Diagrams: A Categorical Account of Predictive Processing and Free Energy
(Extended Abstract)

Sean Tull[1], Johannes Kleiner[2], and Toby St Clere Smithe[3]

[1]Quantinuum
[2]Ludwig Maximilian University of Munich
[3]VERSES Research

Predictive Processing (PP) is a framework for modelling cognition and adaptive behaviour in both biological and artificial systems [23]. A prominent sub-field is the programme of *Active Inference*, developed by Friston and collaborators [17, 15, 8, 16], which aims to provide a unified understanding of cognition which can be applied at many levels, from a single neuron to an entire brain or organism.

Central to the framework is that an agent possesses a *generative model* which explains its observations in terms of both hidden states of the world and its own actions. After receiving an observation, it then *updates* this model to determine likely hidden states (perception) and choose its actions (planning). In active inference, both forms of updating are carried out together through a form of approximate Bayesian inference, by minimizing a quantity known as *free energy* [9, 7].

Though active inference is perhaps unique in aiming to derive all aspects of cognition from a single principle (free energy minimisation), at present there are various aspects of the theory which appear ad-hoc, not following simply from applying the definitions to a given generative model. Conceptually clear formal accounts would be desirable to simplify the theory, address these issues, and for applications within AI.

Now, crucially, the generative models in PP are *compositional* (often given as 'hierarchical models' [4]), suggesting (monoidal) category theory as a natural approach. Many tools have been developed in recent years for describing causal models and probability theory entirely in terms of string diagrams in monoidal categories, or more precisely *cd-categories* [2, 10, 13, 14, 11].

In the full-length version of this work [22] we present a formal account of predictive processing and active inference entirely in terms of string diagrams interpreted in cd-categories. This includes diagrammatic treatments of the key concepts of: generative models, (Bayesian) updating, perception and planning as updating, free energy and active inference itself. This can be seen as a part of the growing field of 'categorical cybernetics' [19, 1, 18, 20], while differing from previous works, including the related work by one author [21], in formalising active inference directly within a simple string-diagrammatic setting: a first step towards a fully abstract characterization.

Generative models In more detail, we work in a cd-category $\mathbf{C}$, a symmetric monoidal category where each object comes with a distinguished *copying* morphism and *discarding* morphism. Of special interest are the *channels*, the morphisms which preserve discarding (which form a *Markov category* [10]). Concretely one may focus on the category $\mathbf{Mat}_{\mathbb{R}^+}$ of positive matrices, whose channels describe finite probability theory.

In PP, generative models are typically described as *Bayesian networks*. Various works have established these as naturally accounted for in cd-categories, with a focus on *causal* Bayesian networks (CBNs) describing causal models in the sense of Pearl [6, 13, 11]. Following the formulation of causal models co-authored by one author in [14], we define an *open generative model* $\mathbb{M}$ in $\mathbf{C}$ as a *network diagram*, a certain class of string diagram, along with an interpretation in terms of channels. Here 'open' refers to the fact that such models many have 'inputs' as well as the usual output and hidden variables, allowing them to be composed.

We use this to describe various generative models common to PP as network diagrams. The simplest case simply consists of a channel from hidden states S to observations O, with a prior σ over S. A more generic form is the discrete n time-step model, which includes a space P modelling the system's own action *policies*. Often models are composed in 'hierarchical models' [4] which we show are naturally diagrammatic.

Given a model with some prior over hidden states S an agent then wishes to *update* this in light of observations over O, through some notion of Bayesian inversion. Bayesian inversion and more general conditioning have a simple description in cd-categories [2, 10, 3]. Less well known is that for a 'soft' observation, given by a distribution over observations O, there are at least two meaningful notions of updating, as studied by Jacobs [12], accounts of which we provide; see also [5].

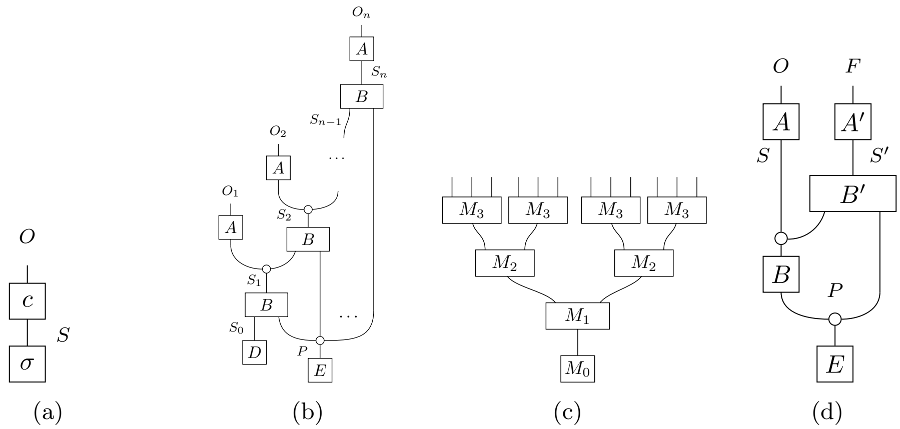
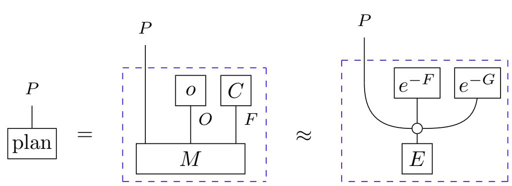


Figure 1: Generative models in PP: (a) A simple generative model from states to observations (b) Discrete time model with policies (c) Hierarchical model (d) High-level structure of model in active inference

Active inference We then apply these ingredients to account for the behaviours of a PP agent: *perception* which updates hidden states given an observation, and *planning* which updates behaviours (policies) given preferences over future outcomes, before combining both in a diagrammatic account of exact active inference. Here an agent possesses a generative model relating action policies P to present observations O and future observations F, via hidden states S, S' for each, as in Fig. 1 (d). In practice, the present and future each decompose further into discrete time steps $1, \dots, m-1$ and $m, \dots, n$, overall taking the form in Fig. 1 (b). The agent then updates their model in light of an observation and preferences for the future. Our diagrammatic account provides a conceptual overview of active inference, and formal starting point for approximate schemes.

Free energy Exact Bayesian updating is computationally intractable, and so an agent typically carries out approximate updating by minimizing a quantity called *free energy*. We give a precise account of two key notions of free energy from the literature: *variational* and *expected* free energy (VFE and EFE). While often stated that an agent may perform Bayesian inversion by VFE minimization, this only holds for sharp observations. For soft observations we characterise it as a new, third form of updating called *VFE updating*.

We apply these notions to the exact active inference scheme to obtain a purely diagrammatic derivation of the well-known formula for active inference via free energy minimization. This formula is central to active inference, but previous justifications for it in the literature can be unclear (typically relying on the view that the EFE forms a 'prior'). We instead derive the formula purely graphically from the structure of the generative model, providing what we argue is the most transparent account known so far. Concretely, this is given by exactly updating the model in Fig 1. (d) in light of future preferences C and observation o, and then approximately rewriting in terms of VFE F and EFE G to yield the formula over policies π below.



$$:: \quad \pi \mapsto \sigma(\log E(\pi) - F(\pi) - G(\pi))$$

Compositional free energy Finally, we consider novel aspects of active inference proposed by the categorical formalism. In particular we give a notion of VFE applicable to an open generative model with inputs, and use this to establish *compositionality* of VFE: a system with an overall generative model composed from sub-models may mininmize global VFE by minimizing VFE locally within each component. This is crucial to apply free energy minimisation at all levels, such as from a whole brain to its individual neurons.

Outlook Overall, we hope to have provided a conceptually clear diagrammatic approach to both PP and active inference, demonstrating that as well as for reasoning with causal (and generative) models [14, 13, 11], string diagrams are natural for describing the structure of PP itself, including free energy. This work should provide useful tools for PP researchers, and an introduction for those familiar with string diagrams, and help situate PP within the context of compositional (category-theoretic) approaches to intelligence.

References

[1] Matteo Capucci, Bruno Gavranović, Jules Hedges, and Eigil Fjeldgren Rischel. Towards foundations of categorical cybernetics. *arXiv preprint arXiv:2105.06332*, 2021.

[2] Kenta Cho and Bart Jacobs. Disintegration and bayesian inversion via string diagrams. *Mathematical Structures in Computer Science*, 29(7):938–971, 2019.

[3] Bob Coecke and Robert W Spekkens. Picturing classical and quantum bayesian inference. *Synthese*, 186:651–696, 2012.

[4] Bert De Vries and Karl J Friston. A factor graph description of deep temporal active inference. *Frontiers in computational neuroscience*, 11:95, 2017.

[5] Elena Di Lavore and Mario Román. Evidential decision theory via partial markov categories. *arXiv preprint arXiv:2301.12989*, 2023.

[6] Brendan Fong. Causal theories: A categorical perspective on bayesian networks. *arXiv preprint arXiv:1301.6201*, 2013.

[7] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138, 2010.

[8] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo. Active inference: a process theory. *Neural computation*, 29(1):1–49, 2017.

[9] Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *Journal of physiology-Paris*, 100(1-3):70–87, 2006.

[10] Tobias Fritz. A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:107239, 2020.

[11] Tobias Fritz and Andreas Klingler. The d-separation criterion in categorical probability. *Journal of Machine Learning Research*, 24(46):1–49, 2023.

[12] Bart Jacobs. The mathematics of changing one's mind, via jeffrey's or via pearl's update rule. *Journal of Artificial Intelligence Research*, 65:783–806, 2019.

[13] Bart Jacobs, Aleks Kissinger, and Fabio Zanasi. Causal inference by string diagram surgery. In *International conference on foundations of software science and computation structures*, pages 313–329. Springer, 2019.

[14] Robin Lorenz and Sean Tull. Causal models in string diagrams. *arXiv preprint arXiv:2304.07638*, 2023.

[15] Thomas Parr, Giovanni Pezzulo, and Karl J Friston. *Active inference: the free energy principle in mind, brain, and behavior.* MIT Press, 2022.

[16] Noor Sajid, Philip J Ball, Thomas Parr, and Karl J Friston. Active inference: demystified and compared. *Neural computation*, 33(3):674–712, 2021.

[17] Ryan Smith, Karl J Friston, and Christopher J Whyte. A step-by-step tutorial on active inference and its application to empirical data. *Journal of mathematical psychology*, 107:102632, 2022.

[18] Toby St Clere Smithe. Compositional active inference i: Bayesian lenses. statistical games. *arXiv preprint arXiv:2109.04461*, 2021.

[19] Toby St Clere Smithe. Cyber kittens, or some first steps towards categorical cybernetics. *arXiv preprint arXiv:2101.10483*, 2021.

[20] Toby St Clere Smithe. Compositional active inference ii: Polynomial dynamics. approximate inference doctrines. *arXiv preprint arXiv:2208.12173*, 2022.

[21] Toby St Clere Smithe. Approximate inference via fibrations of statistical games. *Submission to ACT 2023*, 2023.

[22] Sean Tull, Johannes Kleiner, and Toby St Clere Smithe. *Active Inference in String Diagrams: A Categorical Account of Predictive Processing and Free Energy.* 2023.

[23] Wanja Wiese and Thomas Metzinger. Vanilla pp for philosophers: A primer on predictive processing. 2017.