Toward Universal and Interpretable World Models for Open-ended Learning Agents

Lancelot Da Costa
VERSES AI Research Lab
lance.dacosta@verses.ai

Abstract

We introduce a generic, compositional and interpretable class of generative world models that supports open-ended learning agents. This is a sparse class of Bayesian networks capable of approximating a broad range of stochastic processes, which provide agents with the ability to learn world models in a manner that may be both interpretable and computationally scalable. This approach integrating Bayesian structure learning and intrinsically motivated (model-based) planning enables agents to actively develop and refine their world models, which may lead to developmental learning and more robust, adaptive behavior.

1 Introduction

A feature of open-ended learning agents is the ability to gradually understand wide-ranging, complex worlds. Computational accounts of cognitive development in humans and animals frame this process as approximate Bayesian inference over spaces of Bayesian networks, where inferences are continuously refined thanks to sensory data arriving from active (intrinsically motivated) interactions with the environment [1]. Current modelling approaches to development, while impressive, face scalability challenges, due to the explosive nature of inferring over *all* possible Bayesnets [2, 3]. See Appendix A for more context. Thus, a crucial challenge to enable developmental agents is finding Bayesnet spaces that can express a wide range of worlds, while being coarse enough to efficiently be searched.

2 A generic, interpretable and agentic class of generative models

Desiderata for this Bayesnet class include being: *(i)* **Sufficiently expressive** to be able to approximately express relevant naturalistic, dynamic interactions between agent and environment. *(ii)* **Sufficiently coarse** so that inference on this space can be made computationally tractable. Each Bayesnet in this class should be: *(iii)* **Interpretable** so that the agents' understanding and ensuing behaviour can be easily understood from the Bayesnets it entertains. *(iv)* **Support fast action, perception and learning**. We summarise a candidate class of Bayesnets satisfying requirements *(i)-(iv)* in Appendix B, extending earlier proposals [4].

3 Discussion

The expressivity of this class lies in its ability to represent complex stochastic processes, including those with both discrete and continuous states. By utilizing hierarchical structures, these models can capture both high-level abstractions and fine-grained details of the environment, allowing agents to form robust and generalizable representations. The interpretability of these models stems from their sparsity and encoding of causal relationships—this provides insights into the agent's decision-making, facilitating more transparent and trustworthy AI systems. This class of Bayesnets has been used to model video from raw pixels and sound files, and agents that plan from pixels [5]. Future work should seek the limits of this approach, including its scalability and ability to express relevant agent-environment interactions.

Accepted at the Intrinsically Motivated Open-ended Learning workshop at NeurIPS 2024.

References

[1] Tomer D. Ullman and Joshua B. Tenenbaum. Bayesian Models of Conceptual Development: Learning as Building Models of the World. *Annual Review of Developmental Psychology*, 2(1):533–558, 2020. _eprint: https://doi.org/10.1146/annurev-devpsych-121318-084833.

[2] Pedro A. Tsividis, Joao Loula, Jake Burga, Nathan Foss, Andres Campero, Thomas Pouncy, Samuel J. Gershman, and Joshua B. Tenenbaum. Human-Level Reinforcement Learning through Theory-Based Modeling, Exploration, and Planning. *arXiv:2107.12544 [cs]*, July 2021. arXiv: 2107.12544.

[3] Thomas Pouncy and Samuel J. Gershman. Inductive biases in theory-based reinforcement learning. *Cognitive Psychology*, 138:101509, November 2022.

[4] Karl J. Friston, Lancelot Da Costa, Alexander Tschantz, Alex Kiefer, Tommaso Salvatori, Victorita Neacsu, Magnus Koudahl, Conor Heins, Noor Sajid, Dimitrije Markovic, Thomas Parr, Tim Verbelen, and Christopher L. Buckley. Supervised structure learning, November 2023. arXiv:2311.10300 [cs].

[5] Karl Friston, Conor Heins, Tim Verbelen, Lancelot Da Costa, Tommaso Salvatori, Dimitrije Markovic, Alexander Tschantz, Magnus Koudahl, Christopher Buckley, and Thomas Parr. From pixels to planning: scale-free active inference, July 2024. arXiv:2407.20292 [cs, q-bio].

[6] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward Causal Representation Learning. *Proceedings of the IEEE*, 109(5):612–634, May 2021. Conference Name: Proceedings of the IEEE.

[7] Alison Gopnik, Clark Glymour, David M. Sobel, Laura E. Schulz, Tamar Kushnir, and David Danks. A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, 111(1):3–32, January 2004. tex.ids= gopnikTheoryCausalLearning2004a place: US publisher: American Psychological Association.

[8] Noah D. Goodman, Tomer D. Ullman, and Joshua B. Tenenbaum. Learning a theory of causality. *Psychological Review*, 118(1):110–119, 2011. Place: US Publisher: American Psychological Association.

[9] Daria de Tinguy, Tim Verbelen, and Bart Dhoedt. Exploring and Learning Structure: Active Inference Approach in Navigational Agents, August 2024. arXiv:2408.05982 [cs].

[10] Eric W. Weisstein. Acyclic Digraph. Publisher: Wolfram Research, Inc.

[11] Jürgen Schmidhuber. Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, September 2010.

[12] Edward Deci and Richard M. Ryan. *Intrinsic Motivation and Self-Determination in Human Behavior*. Perspectives in Social Psychology. Springer US, New York, 1985. tex.ids= deciIntrinsicMotivationSelfDetermination1985a.

[13] K. J Åström. Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10(1):174–205, February 1965. tex.ids= astromOptimalControlMarkov1965.

[14] Andrew Barto and Richard Sutton. *Reinforcement Learning: An Introduction*. 1992.

[15] Lancelot Da Costa, Thomas Parr, Noor Sajid, Sebastijan Veselic, Victorita Neacsu, and Karl Friston. Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, 99:102447, December 2020.

[16] Thomas Parr, Giovanni Pezzulo, and Karl J. Friston. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press, Cambridge, MA, USA, March 2022.

[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, April 2022.

[18] Pablo Lanillos, Cristian Meo, Corrado Pezzato, Ajith Anil Meera, Mohamed Baioumy, Wataru Ohata, Alexander Tschantz, Beren Millidge, Martijn Wisse, Christopher L. Buckley, and

Jun Tani. Active Inference in Robotics and Artificial Agents: Survey and Challenges. *arXiv:2112.01871 [cs]*, December 2021. arXiv: 2112.01871.

[19] Thomas Parr, Dimitrije Markovic, Stefan J. Kiebel, and Karl J. Friston. Neuronal message passing using Mean-field, Bethe, and Marginal approximations. *Scientific Reports*, 9(1):1889, December 2019.

[20] Conor Heins, Beren Millidge, Lancelot Da Costa, Richard P. Mann, Karl J. Friston, and Iain D. Couzin. Collective behavior from surprise minimization. *Proceedings of the National Academy of Sciences*, 121(17):e2320239121, April 2024. Publisher: Proceedings of the National Academy of Sciences.

[21] Karl J. Friston, Jean Daunizeau, James Kilner, and Stefan J. Kiebel. Action and behavior: a free-energy formulation. *Biological Cybernetics*, 102(3):227–260, March 2010.

[22] Karl Friston. Hierarchical Models in the Brain. *PLoS Computational Biology*, 4(11):e1000211, November 2008.

[23] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, and Giovanni Pezzulo. Active Inference: A Process Theory. *Neural Computation*, 29(1):1–49, January 2017.

[24] Takuya Isomura, Hideaki Shimazaki, and Karl J. Friston. Canonical neural networks perform active inference. *Communications Biology*, 5(1):1–15, January 2022. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Dynamical systems;Learning algorithms;Machine learning Subject_term_id: dynamical-systems;learning-algorithms;machine-learning.

[25] Takuya Isomura, Kiyoshi Kotani, Yasuhiko Jimbo, and Karl J. Friston. Experimental validation of the free-energy principle with in vitro neural networks. *Nature Communications*, 14(1):4547, August 2023. Number: 1 Publisher: Nature Publishing Group.

[26] Pietro Mazzaglia, Tim Verbelen, Ozan Çatal, and Bart Dhoedt. The Free Energy Principle for Perception and Action: A Deep Learning Perspective. *Entropy*, 24(2):301, February 2022. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[27] Martin Hairer and Xue-Mei Li. Markov Processes. January 2020.

[28] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, August 2014. Google-Books-ID: VvBjBAAAQBAJ.

[29] Karl J. Friston, Tommaso Salvatori, Takuya Isomura, Alexander Tschantz, Alex Kiefer, Tim Verbelen, Magnus Koudahl, Aswin Paul, Thomas Parr, Adeel Razi, Brett Kagan, Christopher L. Buckley, and Maxwell J. D. Ramstead. Active Inference and Intentional Behaviour, December 2023. tex.ids= fristonActiveInferenceIntentional2023 arXiv: 2312.07547 [cs, q-bio].

[30] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Grundlehren der mathematischen Wissenschaften. Springer International Publishing, 2014.

[31] Thomas Parr, Jakub Limanowski, Vishal Rawji, and Karl Friston. The computational neurology of movement under active inference. *Brain*, 144(6):1799–1818, June 2021. tex.ids= parrComputationalNeurologyMovement2021.

[32] Scott Linderman, Matthew Johnson, Andrew Miller, Ryan Adams, David Blei, and Liam Paninski. Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pages 914–922. PMLR, April 2017. ISSN: 2640-3498.

[33] Patrick Kidger. On Neural Differential Equations, February 2022. arXiv:2202.02435 [cs, math, stat].

[34] Thomas Parr and Karl J. Friston. The Discrete and Continuous Brain: From Decisions to Movement—And Back Again. *Neural Computation*, 30(9):2319–2347, September 2018.

[35] Karl J. Friston, Thomas Parr, and Bert de Vries. The graphical brain: Belief propagation and active inference. *Network Neuroscience*, 1(4):381–414, December 2017.

A Current challenges in developmental agents

Engineering agents that scalably learn models of the world remains a relatively open challenge [6], despite the fact that cognitive development of humans and animals is well-studied within computational cognitive science [1]. Several strands of work in computational cognitive science have converged to the idea that the developmental process is a process of approximate Bayesian inference about explanations for the world (i.e. Bayesian networks) and how the agent interacts with it, where inferences are gradually refined by actively sampling new, informative data (e.g. through intrinsic motivation) [1, 5, 7, 8, 9]. Engineering this simple picture is difficult as the search space of possible explanations for the world explodes combinatorially in the number of variables being modelled [10]. To illustrate the problem further, we summarise the current theory-based reinforcement learning work [2, 3], which simulates developmental agents by combining Bayesian structure learning and intrinsically motivated (model-based) planning.

In foundational papers [2, 3], an agent maintains Bayesian beliefs about probabilistic programs (which encode Bayesnets) that might explain the world. The agent gradually refines refines its inferences by actively seeking new data through a mixture of exploration and exploitation (i.e. intrinsic motivation [11, 12]). The authors deployed this agent in a suite of simplified Atari games, and found not only that their agent averaged human learning efficiency across the games (after comparing with data from human participants), but the agents' learning trajectories were relatively similar to that of humans. This work serves as a proof of concept that combining inference about the structure of the world with intrinsically motivated model-based planning can achieve relatively human-like learning and behaviour.

The fundamental limitation of existing theory-based RL work is that their agents [2, 3] consider a search space of explanatory hypotheses about the world that is the whole set of programs (up to a certain length) that can be generated from the code grammar generating the data generating process. This is an extremely large search space even for the simplified Atari environments their agents face and highlights the current limitations of this work: 1) In complex environments, the space of programs that can be generated from the code grammar generating the data-generating process may be far too large to be searchable, 2) in real environments the modeler does not know the generative process and cannot easily form a space of candidate explanations that contains the data generating process.

Thus a fundamental question is what might be a 'universal' set of primitives and compositional rules to produce a space of candidate explanations for the world (i.e. Bayesnets) satisfying requirements *(i)-(iv)* in Section 2. Indeed, there is already a tension between requirements *(i)-(2)* and a significant difficulty lies in striking the right balance. Asking what a universal space of Bayesnets might look like, we first consider the existing literature: Spaces of **probabilistic programs** are easily made extremely expressive, but it is not clear how to do so while keeping them coarse enough for inference to remain tractable. Probabilistic programs are not always easily interpretable, and, barring specific assumptions, do not support efficient perception and learning, as Bayesian inference over states and parameters may require sampling. One example of probabilistic programs that might satisfy these requirements—to a first approximation—are **hierarchical discrete and continuous state partially observed Markov decision processes** (POMDPs) [13, 14].

This space of Bayesnets is very expressive in being able to reproduce a wide variety of behavioural simulations and empirical data. For example, nearly all modeling work in the active inference literature alone, which spans nearly two decades, employed Bayesnets that are built by hierarchically stacking these two types of layers [15, 16, 17, 18]. The resulting Bayesnets support fast action, perception and learning where inference about states and parameters is done through fast variational inference procedures [15, 16, 19, 20, 21], which have a degree of biological plausibility in being able to reproduce a wide range of features from real neural dynamics, e.g. [22, 23, 24, 25]. Barring the use of neural networks for expressing non-linearities in these layers [26], each of them furnishes an interpretable model of dynamics.

B Details on generic, interpretable and agentic class of Bayesnets

The Bayesnet class we consider here as potential explanations for the world is built from a simple set of primitives and compositional rules. We claim that it satisfies the requirements *(i)-(iv)* from Section 2. *(i)* Its expressivity lies in its ability to represent complex stochastic processes, including those with both discrete and continuous states. *(ii)* Its sparsity stems from significant inductive biases on the structrures of the Bayesnets being considered. *(iii)* The interpretability of each Bayesnet in this class

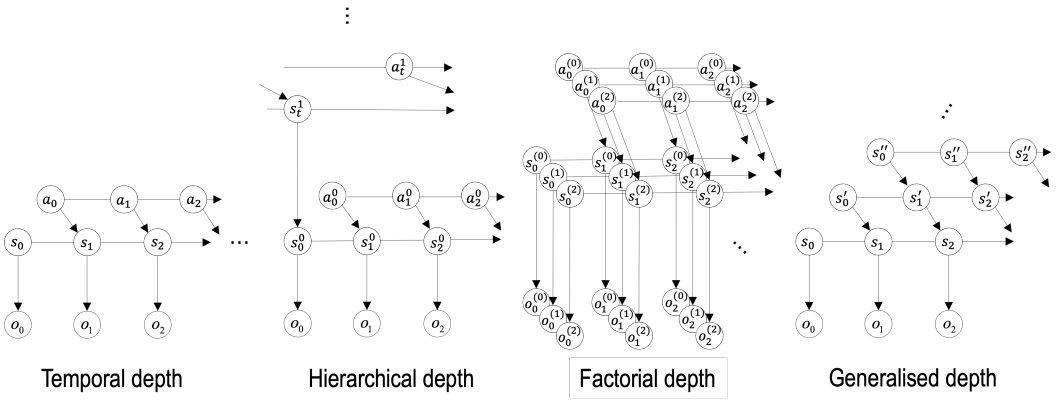

Figure 1: **Space of discrete-state Bayesian networks.** The basic module for expressing discrete-state agent-environment interactions is a partially observed Markov decision process (POMDP). This module can have an arbitrary finite temporary horizon (i.e. temporal depth). Several such modules can be stacked atop each other finitely many times (i.e. hierarchical depth), thereby expressing multi-scale semi-Markovian latent dynamics. We can specify multiple co-evolving factors (i.e. factorial depth; e.g. position and colour of an object). In any given layer, a number of auxiliary (i.e. generalised) states can be added, accounting for velocity, acceleration, and higher orders of motion in the latent states (i.e. generalised depth) to express semi-Markovian processes [4]. In each of these layers, the highest generalised states may or may not be actions denoted by a (cf. controllable states [4]) while all other states are uncontrollable (i.e. part of the environment) denoted by s. The controllable and uncontrollable states cause observations denoted by o. The parameters of these Bayesian networks (not shown) correspond to the parameters of the causal maps as well as the initial distributions over states [15]. These parameters as well as the graphical structure of these Bayesnets to be inferred from data (e.g. past actions and observations). The resulting approximate posterior belief then informs action, which follows a mixture of goal-seeking and information gathering (i.e. intrinsic motivation) imperatives. Please see [4, 29] for more details on temporal, hierarchical and factorial depth—and examples of controllable versus uncontrollable states.

stems from their sparsity as well as encoding of causal relationships. Finally, and as we will see, these Bayesnets support fast action, perception and decision-making *(iv)*.

The proposed Bayesnet class is built by hierarchically assembling a set of basic structural modules satisfying requirements *(iii)-(iv)*, which together, express a wide range of dynamic agent-environment interactions. In what follows, we detail each of these building blocks developed specifically to express a large class of stochastic processes on either discrete or continuous states.

B.1 Discrete dynamics

Markov processes are a fairly ubiquitous class of stochastic processes [27]. All Markov processes on discrete states have simple transition dynamics that are given by linear algebra. When these transitions also depend on actions, we obtain Markov decision processes [28]. When states are partially observed and observations depend only on the current latent state, we obtain POMDPs. We can add auxiliary latent states to those POMDPs [4] (i.e. the equivalent of momentum, acceleration etc) to account for the effect of memory in the system, producing semi-Markovian POMDPs. Lastly, we can stack these layers hierarchically to express multi-scale semi-Markovian processes in latent space. In summary, hierarchical discrete POMDPs extended in this way provide a very generic class of models for agent-environment interactions on discrete states. See Figure 1 for a graphical representation of discrete extended POMDPs and their various degrees of freedom.

B.2 Continuous dynamics

For expressing continuous dynamics, the situation is somewhat more difficult. Repeating the construction of the discrete state-space model seems hardly possible because continuous-space Markov processes are given by linear operators in infinite (as opposed to finite) dimensional spaces [30]. A working alternative is to restrict ourselves to a more manageable but still very expressive class of processes. We can consider continuous POMDPs with latent dynamics given by stochastic differential equations (SDEs), which is another very expressive class of stochastic processes. However, for expressing a wide range of agent environment interactions, these SDEs must break detailed balance and possibly be driven by non-Markovian noise (fluctuations in the wind or in the ocean)[31]. There

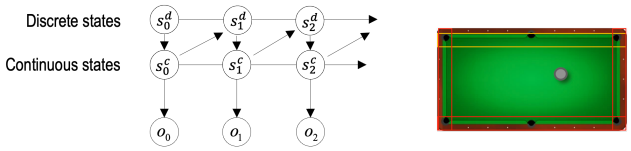

Figure 2: **Recurrent switching linear dynamical systems (rsLDS).** This figure illustrates rsLDS as a way of interpretably parameterising fairly arbitrary stochastic differential equations (SDEs) as a switching linear SDE. Consider for example a generative model used to play pool. This must be able to express the non-linear trajectories in a game of pool. The behaviour of a ball on a pool table depends on whether the ball is by one of the four walls (i.e. bouncing), or not. In each of these sectors the dynamic of the ball is captured by a simple dynamical equation. The rsLDS [32] is a simple generative model that is able to express these kinds of trajectories, where a (switching) discrete state expresses in which of these sectors the ball is. This discrete state furnishes a linear drift and volatility to the linear SDE describing the continuous motion of the ball (cf. continuous states), which yields continuous observations. Note the so-called recurrent connection (i.e. causal map) from continuous to discrete states; this connection enables the continuous dynamic to influence the discrete switching: if the continuous trajectory of the ball collisions with a wall the discrete state switches so that the continuous dynamic of the ball changes course. The rsLDS layer may be supplemented with active states, acting on the discrete latent states, thereby emulating a continuous partially observable Markov decision process. This continuous POMDP can easily be extended by varying the temporal, hierarchical, factorial and generalised depth as in Figure 1, furnishing a generic model of continuous dynamics. Please see [32] for more details on the current rsLDS architecture.

is a remarkably expressive class of SDEs supporting non-linearities, non-Markovian noise and broken detailed balance—that is, many times differentiable stochastic differential equations—for which POMDPs with these latent dynamics support fast and biologically plausible update rules for action, perception and learning [16, 20, 22]. These continuous POMDPs form a very expressive space of continuous-state Bayesian networks by varying the temporal, hierarchical, factorial and generalised depth as in Figure 1.

One important challenge remains: to parameterise the non-linearities in these POMDPs (e.g. flows of SDEs) without sacrificing interpretability, and learn these parameterisations from data. A promising approach is to express non-linear SDEs with recurrent switching linear dynamical systems (rsLDS; see Figure 2) [32]; that is, switching mixtures of linear SDEs, because one could use a very fine grained piece-wise linear approximation to recover arbitrary non-linearities, as necessary. The advantage of using switching linear SDEs is that they are interpretable and afford relatively scalable exact Bayesian inference [32].[1] However, current rsLDS architectures are restricted to approximating the dynamics of non-linear so-called 'diffusion' SDEs discretised with an Euler scheme [32], which do not feature non-Markovian noise by definition. Looking forward, it seems desirable to extend the rsLDS architecture to express SDEs with more arbitray noise signals, perhaps by adding generalised coordinates (velocity, acceleration, and higher orders of motion etc) [20, 22]. This would entail introducing generalised depth into the current rsLDS layer. Doing this should furnish a very expressive yet sparse class of Bayesnets for continuous-state dynamics that satisfies the basic requirements *(i)-(iv)*.

B.3 Hierarchical mixed dynamics

Stacking hierarchies of discrete layers atop hierarchies of continuous layers yields mixed generative models that can express rich non-linearities and dynamics at several levels of abstraction. Despite the absence of traditional neural networks in these hierarchies, one can think of them as a neural network where the layers are discrete and continuous POMDPs and the computations are of efficient approximate Bayesian inference. Hierarchies of these layers may be interpretable as they represent nested processes operating at different timescales. These hierarchical structures are compatible with views of the brain as entertaining discrete-state low-dimensional abstract dynamics that condition the high-dimensional continuous representations closer to sensory input [34, 35].

[1] An alternative way of inferring non-linearities would be by parameterising them with Bayesian neural networks (i.e. neural SDEs [33]), but this would likely hinder interpretability.