Toward a reusable architecture for intelligent agents

Alexander D. Shaw[1]

[1]Department of Psychology, University of Exeter, Exeter, UK

Abstract

Active inference provides a powerful framework for understanding how agents perceive, learn, and act by maintaining an internal model of the world [1, 2]. However, most existing implementations rely on generative models that are designed for specific tasks or domains [3, 4]. These task specific models limit the ability of agents to generalise, adapt to new environments, or scale beyond narrow applications.

Taking inspiration from the human brain, which appears to solve the problem of generalisation through a combination of modularity, hierarchy, and embodiment [5, 6], we propose a list of core ingredients. We outline seven core components that we believe are essential for a scalable and reusable generative model, including compositional structure, temporal abstraction, attention and precision, meta-level learning, separation of self and world, and mechanisms for learning model structure itself.

The discretisation of these core components hints at the necessity for a series of interconnected generative models organised on a graph-like structure with (some) shared variables, states and parameters, similar to the functional specialisation, integration, and connectivity of the brain [7]. We propose that these components offer a blueprint for constructing intelligent systems that can generalise across tasks and domains.

1 Introduction

Active inference offers a unifying framework for perception, action, and learning, grounded in the idea that intelligent agents maintain an internal model of the world [1, 8]. This internal, *generative* model allows agents to predict the sensory consequences of their actions, infer the hidden causes of observations, and select behaviours that minimise uncertainty or surprise [9]. Moreover, it provides a principled way to think about how agents come to understand and interact with their environment, by continuously updating beliefs and acting to fulfil prior expectations [10].

In recent years, active inference has shown promise in a variety of domains, from sensorimotor control to decision making and simulated exploration [4, 11, 12]. However, the generative models used in these applications are typically hand crafted and tightly coupled to specific tasks. For example, controlling a robotic limb, navigating a 2D grid, or playing a video game, each implementation tends to build a bespoke model tailored to that particular setting.

While effective in narrow contexts, this approach poses a major obstacle to scaling active inference toward more *general* forms of intelligence. A truly adaptive agent needs to be able to reuse and repurpose its internal model across a variety of domains; for example, environments or goals, without needing a complete redesign for each new task. This leads us to a central, yet still largely open question: *What does a general purpose generative model actually look like?*

This question is especially relevant in the context of *artificial general intelligence* (AGI). If active inference is to serve as a foundation for AGI, it must support agents that can generalise, not just in policy or behaviour, but in the structure of their internal models [3, 13]. Moving beyond task specific models requires a shift toward a more principled, reusable (or adaptive) architecture.

In this paper, we explore this challenge through a neuro-inspired lens. Taking an anthropomorphic approach, we draw on principles from systems neuroscience and theoretical models of cortical function [6, 7] to outline seven core components that we believe are necessary for a general-purpose generative model. These include modularity, temporal hierarchy, compositional structure, precision weighted attention, meta -structuring, separation of world and self models, and structure-learning mechanisms.

Together, these ingredients form the backbone of a flexible and scalable generative model capable of supporting intelligent behaviour across tasks and domains.

2 What does *general-purpose* mean in an active inference context?

Within the active inference framework, agents rely on internal generative models to predict sensory input, infer hidden causes (*...it's wet out, did it rain?*), and guide action (*...I should get the washing in*) [1, 8]. In practice, however, these models are usually constructed for a single, well-defined task, often with fixed states, modalities, and parameters [3, 4].

In typical implementations, the generative model is constructed as a non-linear state-space system with fixed states and parameters:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \boldsymbol{\omega}(t) \tag{1}$$

$$\mathbf{y}(t) = g(\mathbf{x}(t), \boldsymbol{\theta}) + \boldsymbol{\nu}(t) \tag{2}$$

Here, $\mathbf{x}(t)$ denotes the latent state vector, $\mathbf{u}(t)$ the control input, and $\boldsymbol{\theta}$ the parameters of the model. The function f describes the system dynamics and g maps those latent states to observations $\mathbf{y}(t)$. Process noise $\boldsymbol{\omega}(t)$ and observation noise $\boldsymbol{\nu}(t)$ represent uncertainty.

In most applications, the structure of f and g, including the dimensionality of $\mathbf{x}$, the number and type of parameters $\boldsymbol{\theta}$, and the form of the nonlinearities, is handcrafted for the task at hand [14, 15]. This constrains generalisation, as the model cannot adapt to new tasks or environments without structural redesign.

By contrast, a *general-purpose* generative model should allow for flexible inference and control across multiple domains and task structures, without requiring a complete redesign. Rather than being tied to a specific environment or objective, such a model should provide a reusable substrate that supports:

- **Cross-domain reuse:** The ability to operate across different sensory modalities, state spaces, or environments without needing (manual) structural redesign [12].

- **Compositional generalisation:** The ability to recombine learned components (e.g., objects, causes, policies) to solve new tasks [16, 13].
- **Context sensitivity:** The ability to flexibly adapt inference and action depending on goals or environmental structure [17].
- **Structural plasticity:** The ability to extend or revise the model over time, learning not just new parameters, but adapting structure [18].

In other words, generality in this context does not imply a universal or fully agnostic model, but one that supports adaptive reuse, incremental learning, and principled composition of beliefs across changing settings. This requires generative architectures that are not monolithic but modular, dynamically assembled, and context-aware [19].

3 Principles from the brain: modularity, hierarchy, and embodiment

What kinds of architectural principles would be necessary for a general-purpose generative model? In this section, we take inspiration from the brain, which provides a biologically grounded example of a system capable of adaptive, context-sensitive inference across tasks, timescales, and sensory modalities [6, 20].

The human brain is not a monolithic processor, but modular, with distinct and interacting regions specialised for sensory, motor, interoceptive, and cognitive functions [21, 22]. It is also hierarchical, with fast dynamics and low-level prediction errors flowing upward, and slower, more abstract beliefs and goals flowing downward [23, 24]. Moreover, the brain is embodied; that is, tightly coupled to the body and environment, and constantly learning and updating its internal model through action [25, 26].

We argue that these properties are not incidental but essential for general intelligence. They provide *scaffolding* for the kinds of structural priors, abstraction, and reuse that general-purpose generative models require. Drawing on this view, we propose seven interconnected components (Figure 1) that we believe form the basis of a reusable architecture for active inference agents.

3.1 Modularity and factorisation

A general-purpose generative model should not be a single monolithic structure, but rather a collection of interacting modules; each responsible for distinct but complementary functions, such as perception, planning, interoception, or memory. These modules mirror the functional specialisation observed in the brain [7], where areas like visual cortex, motor cortex, and prefrontal cortex each maintain their own partial models of the world, yet remain deeply interconnected. Crucially, these modules do not operate in isolation but share some latent states and parameters, allowing for joint inference and mutual constraint, while still retaining local autonomy. This factorised structure promotes scalability, reuse, and robustness, enabling submodels to be adapted or replaced without re-engineering the whole system.

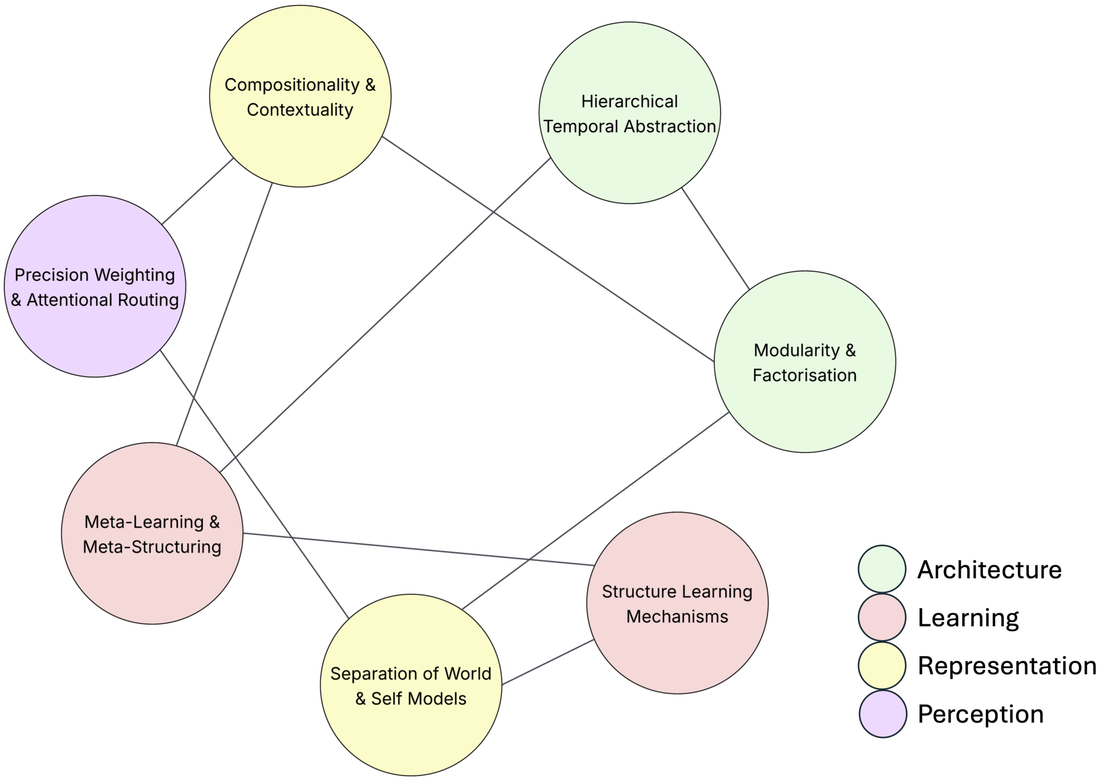

Figure 1: The seven interrelated ingredients of a general-purpose generative model for Active Inference. Each node is colour-coded by functional domain: Architecture (green), Learning (red), Representation (yellow), and Perception (purple). Edges indicate conceptual and functional dependencies between components, many of which may involve shared parameters (e.g. precision estimates, modular priors) or shared hidden states (e.g. beliefs about temporal structure, agency, or context). These shared elements reflect the deep interdependence of mechanisms required for flexible, scalable, and adaptive inference.

3.2 Hierarchical temporal abstraction

A second key principle is hierarchy; specifically, hierarchical organisation over time. The brain maintains generative models operating across nested temporal scales, from fast sensory fluctuations to slowly evolving beliefs about context, goals, and self [27, 28]. In a general-purpose architecture, this translates to stacked generative models, each capturing dynamics at a different timescale. Lower levels handle rapid sensorimotor loops, while higher levels encode temporally extended policies, environmental contingencies, or abstract narratives. This organisation supports flexible planning and allows agents to generalise behaviours across contexts or environments.

3.3 Compositionality and contextuality

A third ingredient is compositionality; that is, the ability to construct complex generative predictions by combining simpler, reusable components. Rather than learning entirely new models for each environment or task, a general-purpose architecture should represent entities, events, and relationships as modular elements that can be recombined across contexts [16]. This *compositional structure* supports powerful inductive generalisation,

Component	Description
Modularity and factorisation	Independent modules for sensory, motor, planning, etc., that share latent variables or pass messages
Hierarchical temporal abstraction	Layered generative models operating across fast-to-slow temporal scales
Compositionality and contextuality	Reusable representations of entities and relations, modulated or reconfigured based on context
Precision-weighting and attentional routing	Dynamic allocation of computational resources via precision estimates that prioritise relevant inference
Meta-learning or meta-structuring	Adaptive learning of the model structure and priors themselves, not just parameters
Separation of world model and self model	Distinct representations for internal (e.g., proprioceptive) and external (e.g., exteroceptive) causes, enabling simulation and planning
Structure learning mechanisms	Mechanisms for expanding, pruning, or reorganising model components based on uncertainty or Bayesian evidence

Table 1: Seven key components of a general-purpose generative model for active inference.

such as inferring unseen combinations of familiar parts [13]. However, compositionality must be complemented by contextuality; the capacity to modulate, gate, or reconfigure submodels based on situational demands [29]. Together, these features enable flexible reuse without rigid or fixed coupling, and support the emergence of structured, context-sensitive behaviour.

4 Designing modules

Designing a generative model that supports general intelligence requires more than just modularity and hierarchy. It also involves embedding mechanisms that allow the model to dynamically allocate resources, adapt its internal structure, and distinguish between self-generated and external signals. In this section, we highlight three additional components that make such a system flexible, efficient, and capable of continual learning.

4.1 Precision-weighting and attention

To operate efficiently in complex environments, a general-purpose generative model must be able to selectively attend to relevant information. This is achieved through precision-weighting, which dynamically modulates the influence of prediction errors based on their estimated reliability [30]. In the brain, this is thought to correspond to attentional mechanisms that amplify or suppress signals across cortical hierarchies [28, 31]. Within a modular architecture, precision estimates help steer inference toward the most informative states or modalities, allowing the system to flexibly prioritise different submodels, depending on context.

4.2 Meta-learning or meta-structuring

Intelligent agents must not only learn the *parameters* of their generative models, but also learn how to learn; that is, how to structure their own models over time. Meta-learning provides a way for agents to discover priors over model architecture, update hyperparameters governing learning rates or structure, and adaptively modify their inference routines based on experience [32, 33]. In neuro-inspired terms, this might correspond to hypermodels or slow changing cortical systems that shape how local circuits adapt [14]. Within a reusable architecture, meta-learning allows the agent to refine its own inductive biases and accumulate structured knowledge over time [34].

4.3 Separation of world model and self model

Another key requirement is the ability to distinguish between the causes of sensory input that originate in the external world (exogenous) and those generated by the agent itself (endogenous). This separation of world and self models enables internal simulation; the capacity to imagine actions before executing them, predict their consequences, and differentiate self-generated signals from those arising externally [25]. The brain achieves this through specialised pathways for proprioception and interoception [35]. In a general purpose generative model, similar distinctions are necessary for modelling embodied agents that learn from experience while maintaining an integrated sense of self. This division supports planning, credit assignment, and the development of counterfactual reasoning [36].

5 Learning structure vs learning parameters

Most machine learning methods focus on adjusting parameters within a fixed model architecture. This includes updating weights, biases, or transition probabilities; typically under the assumption that the structure of the model is already known and correct. However, both biological and artificial general intelligence demand a more flexible approach. A truly adaptive agent must not only update beliefs about *hidden states and parameters*, but also infer the *structure* of its internal model [37, 13].

Structure-learning involves discovering which variables should be included in the model, how they are connected, and how new components can be integrated or redundant ones removed. This is far more challenging than parameter learning, both conceptually and mathematically. It requires evaluating different model topologies, often with combinatorially large search spaces and sparse evidence [38]. In active inference, this means inferring not just the values of existing variables, but what variables should exist in the first place.

This distinction is visualised in Figure 1, where we propose seven high-level generative modules as key ingredients for general intelligence. The black lines in the diagram represent an initial assumed pattern of connectivity. Learning parameters would correspond to adjusting the strength or precision of these connections. In contrast, structure learning asks a deeper question: *which connections should exist at all*? Which modules should be linked? Which should remain conditionally independent? How can new modules be added without breaking the coherence of the system? Structure learning is thus a kind of meta-inference; not just learning within a model, but learning the model itself.

Techniques such as Bayesian model comparison, model averaging, or graph-based priors can support this process [39, 40]. In a modular architecture, structure learning provides the glue that links new components into existing systems while maintaining coherence, scalability, and adaptability over time.

6 Integration with existing neuro-AI work

The ideas proposed here resonate with several strands of ongoing work in neuro-AI, particularly those that aim to move beyond narrow, domain-specific tasks and hand-engineered architectures. In the active inference literature, recent models have begun to incorporate deep temporal hierarchies, modular structures, and even basic forms of structure learning [5, 13]. These efforts reflect a growing recognition that flexibility and reuse require generative models that are compositional, adaptive, and embedded.

The emphasis on modularity, temporal abstraction, and meta-structuring aligns with proposals for agents that learn causal structure through experience [13]. In parallel, the Free Energy Principle has been increasingly used to frame the functional segregation and integration of brain regions [28], offering a theoretical basis for the kind of architecture envisioned here.

Outside of active inference, related ideas are emerging across domains: object-orientated reinforcement learning [41, 42], graph-based cognitive architectures [43], and systems neuroscience approaches to cortical computation [44, 45] all converge on similar principles. What unites these perspectives is the recognition that intelligent behaviour depends not just on data or training, but on structured, interpretable models of the world that can be reused across time and contexts.

By formalising these insights in terms of interconnected generative models, the aim is to contribute a blueprint that unifies these efforts under the active inference framework.

7 Research directions and open challenges

This paper outlines a conceptual architecture for general-purpose generative models inspired by the structure and function of the brain. While we believe the seven components described here are necessary, steps remain in turning this blueprint into a working system.

First, the space of possible structures is vast. Learning which modules to include, how to connect them, and when to activate them presents significant computational and theoretical challenges. Although there has been progress in structure learning and neural architecture research [37, 46], most approaches remain brittle or task-specific.

Second, there are very few benchmarks for evaluating general-purpose generative models. Unlike supervised learning tasks, "*generalisability*" is hard to quantify. Progress may require developing environments where agents must flexibly reuse generative components, adapt to changing goals, and transfer knowledge between domains [16].

Third, implementing meta-structuring; where the agent not only learns parameters but learns how to modify its own architecture, remains largely unexplored in active inference. This is both a technical and philosophical challenge, raising questions about self-modelling, plasticity, and control [47].

Finally, scaling these ideas will require careful trade-offs between tractability and flexibility. Structure learning must not lead to combinatorial explosion, and attention must be guided without overwhelming the agent with uncertainty. Despite these obstacles, the principles outlined here offer a roadmap by drawing on the brain's solutions: modularity, abstraction, embodiment, and reuse.

References

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

[2] Karl Friston, Thomas Parr, and Beata de Vries. Graphical brain: Belief propagation and active inference. *Network Neuroscience*, 1(4):381–414, 2017.

[3] R. Conor Heins, M. Berk Mirza, Thomas Parr, Karl J. Friston, Igor Kagan, and Arezoo Pooresmaeili. Deep active inference and scene construction. *Frontiers in Artificial Intelligence*, 3:509354, 2020. doi: 10.3389/frai.2020.509354. Received November 1, 2019; Accepted September 10, 2020; e-Collection 2020.

[4] Alexander Tschantz, Beren Millidge, Anil K Seth, and Christopher L Buckley. Learning action-oriented models through active inference. In *Proceedings of the 2020 Conference on Cognitive Computational Neuroscience*, 2020.

[5] Karl Friston, Richard Rosch, Thomas Parr, Cathy Price, and Howard Bowman. Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*, 90: 486–501, 2018.

[6] Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013.

[7] Olaf Sporns and Richard F. Betzel. Modular brain networks. *Annual Review of Psychology*, 67:613–640, 2016. doi: 10.1146/annurev-psych-122414-033634.

[8] Christopher L Buckley, Chang Sub Kim, Simon McGregor, and Anil K Seth. The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81:55–79, 2017.

[9] Karl Friston. Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B*, 364(1521):1211–1221, 2009.

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

[11] Kai Ueltzhoffer. Deep active inference. *Biological Cybernetics*, 112(6):547–573, 2018.

[12] 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 preprint arXiv:2112.01871, 2021. URL `https://arxiv.org/abs/2112.01871`. Under review in an IEEE journal.

[13] Alexander Tschantz, Anil K. Seth, and Christopher L. Buckley. Learning action-oriented models through active inference. *PLOS Computational Biology*, 16 (4):e1007805, 2020. doi: 10.1371/journal.pcbi.1007805.

[14] Karl Friston et al. Variational free energy and the laplace approximation. *NeuroImage*, 34(1):220–234, 2008.

[15] Thomas Parr and Karl J. Friston. Generalised free energy and active inference. *Biological Cybernetics*, 113(5–6):495–513, 2019. doi: 10.1007/s00422-019-00805-w. URL `https://doi.org/10.1007/s00422-019-00805-w`.

[16] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, 2017.

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

[18] Karl Friston, Luca Da Costa, Danijar Hafner, Casper Hesp, and Thomas Parr. Sophisticated inference. *Neural Computation*, 33(5):713–763, 2021. doi: 10.1162/neco_a_01351.

[19] Beren Millidge, Alexander Tschantz, and Christopher L. Buckley. Whence the expected free energy? arXiv preprint arXiv:2004.08128, Apr–Sep 2020. URL `https://arxiv.org/abs/2004.08128`. Version v5 (final), submitted Apr 17, 2020; updated Jul 5, 2020; accepted Sep 28, 2020.

[20] Karl Friston, Rick A Adams, Laurent Perrinet, and Michael Breakspear. Perceptions as hypotheses: saccades as experiments. *Frontiers in Psychology*, 3:151, 2012.

[21] Jerry A Fodor. *The Modularity of Mind*. MIT Press, 1983.

[22] Karl Friston. A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, 360(1456):815–836, 2005.

[23] Andre M Bastos, W Martin Usrey, Rick A Adams, George R Mangun, Pascal Fries, and Karl J Friston. Canonical microcircuits for predictive coding. *Neuron*, 76(4): 695–711, 2012.

[24] Rajesh PN Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

[25] Karl Friston, Jean Daunizeau, James Kilner, and Stefan J Kiebel. Action and behavior: a free-energy formulation. *Biological Cybernetics*, 102(3):227–260, 2011.

[26] Andy Clark. *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press, 2015.

[27] Stefan J Kiebel, Jean Daunizeau, and Karl J Friston. A hierarchy of time-scales and the brain. *PLoS Computational Biology*, 4(11):e1000209, 2008.

[28] Karl Friston. Hierarchical models in the brain. *PLOS Computational Biology*, 4(11): e1000211, 2008. doi: 10.1371/journal.pcbi.1000211. Epub 2008 Nov 7.

[29] M. Berk Mirza, Maell Cullen, Thomas Parr, Sukhi Shergill, and Rosalyn J. Moran. Contextual perception under active inference. *Scientific Reports*, 11(1):16223, 2021. doi: 10.1038/s41598-021-95510-9. Received 04 July 2020; Accepted 17 June 2021; Published online 10 August 2021.

[30] Heidi Feldman and Karl Friston. Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4:215, 2010.

[31] Emma Holmes, Thomas Parr, Timothy D. Griffiths, and Karl J. Friston. Active inference, selective attention, and the cocktail party problem. *Neuroscience & Biobehavioral Reviews*, 131:1288–1304, 2021. doi: 10.1016/j.neubiorev.2021.09.038. Epub 2021 Oct 21; Published online Dec 2021.

[32] Jane X Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Rémi Munos, Charles Blundell, Dharshan Kumaran, and Matthew Botvinick. Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*, 2016.

[33] Yan Duan, John Schulman, Xi Chen, Peter L Bartlett, Ilya Sutskever, and Pieter Abbeel. Rl^2: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779*, 2016.

[34] Ron Amit and Ron Meir. Meta-learning by adjusting priors based on extended pac-bayes theory. arXiv preprint arXiv:1711.01244 [stat.ML], 2017. URL `https://arxiv.org/abs/1711.01244`. Accepted to ICML 2018.

[35] Anil K Seth. Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11):565–573, 2013.

[36] Micah Allen and Karl J. Friston. From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, 195(6):2459–2482, 2018. doi: 10.1007/s11229-016-1288-5. Epub 2016 Dec 22.

[37] Samuel H. Rudy and Themistoklis P. Sapsis. Sparse methods for automatic relevance determination. *Physica D: Nonlinear Phenomena*, 418:132843, 2021. doi: 10.1016/j.physd.2021.132843.

[38] Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022): 1279–1285, 2011.

[39] Will D Penny, Klaas E Stephan, Jean Daunizeau, Maria J Rosa, Karl J Friston, Timothy M Schofield, and Alexander P Leff. Comparing families of dynamic causal models. *PLOS Computational Biology*, 6(3):e1000709, 2010.

[40] Matthew James Beal. *Variational algorithms for approximate Bayesian inference*. PhD thesis, University of Cambridge, 2003.

[41] Carlos Diuk, Andrew Cohen, and Michael L Littman. An object-oriented representation for efficient reinforcement learning. In *Proceedings of the 25th International Conference on Machine Learning*, pages 240–247, 2008.

[42] Vinicius Zambaldi, David Raposo, Adam Santoro, Victor Bapst, Yujia Li, Igor Babuschkin, Karl Tuyls, Dan Toyama, Alexander de G. Matthews, Martin Tan, et al. Relational deep reinforcement learning. *arXiv preprint arXiv:1806.01830*, 2018.

[43] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

[44] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):356–365, 2016.

[45] Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258, 2017.

[46] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search: A survey. *Journal of Machine Learning Research*, 20(55):1–21, 2019.

[47] Anil K. Seth, Tomasz Korbak, and Alexander Tschantz. A continuity of markov blanket interpretations under the free-energy principle. *Behavioral and Brain Sciences*, 45, 2022.