The World Is Bigger: Interaction Within a World

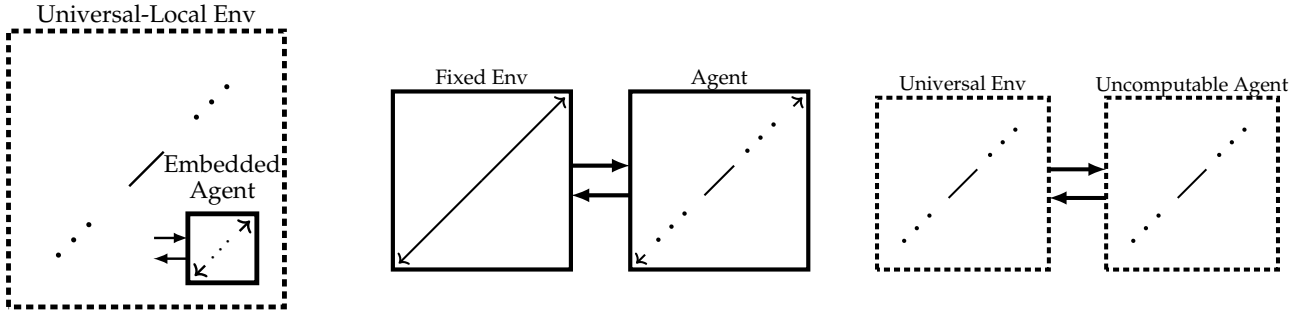
Alex Lewandowski[1,*], **Aditya A. Ramesh**[2], **Edan Meyer**[1], **Saurabh Kumar**[3],
Dale Schuurmans[1,4,5], **Marlos C. Machado**[1,5]
[1]University of Alberta, [2]IDSIA, [3]Stanford University
[4]Google Deepmind, [5]CIFAR AI Chair

Abstract

Continual learning is often motivated by the idea, known as the big world hypothesis, that the "world is bigger" than the agent. Recent problem formulations capture this idea by explicitly constraining an agent relative to the environment. These constraints lead to solutions in which the agent continually adapts to best use its limited capacity, rather than converging to a fixed solution. However, explicit constraints can be ad hoc, difficult to incorporate, and limiting to the effectiveness of scaling up the agent's capacity. In this paper, we characterize a general problem setting in which an agent of any capacity is implicitly constrained. In particular, we consider the implicit constraint faced by an agent embedded in an environment. We introduce a *universal-local environment* to embed such an agent using computational universality and transition dynamics that depend on a local neighbourhood of the state-space. The embedded agent is implicitly constrained relative to its environment, represented as a partially observable Markov decision process. We then propose *interactivity* as a measure of an embedded agent's ability to adapt its future behaviour, conditioned on its past behaviour, using Kolmogorov complexity. Using the fact that an agent's interactivity is bounded by its capacity, we conjecture that maximizing interactivity is a continual learning problem from the perspective of any agent.

Keywords: Continual learning, big world hypothesis, computational constraints.

Acknowledgements



(a) This work (b) Traditional RL (c) AIXI

Figure 1: **Comparing the agent's relationship to the environment in our work, traditional RL, and AIXI.** This work introduces a universal-local environment, in which agents of varying sizes are embedded and implicitly constrained. Traditional RL involves a fixed environment and agents of varying size, where the agent is often unconstrained by being "bigger" than the considered environment. AIXI involves a computationally universal environment and an uncomputable agent, both of which are unconstrained.

1 Introduction: Big World Hypothesis

The big world hypothesis states that, in many learning problems, the environment is much larger—more "complex"—than the agent, meaning that the agent cannot represent the optimal solution [9]. An implication of this hypothesis is that agents faced with a big world should track an ever-changing approximation rather than trying to learn the optimal fixed solution [18]. A formalization of the big world property could provide a problem formulation for continual learning, similar to the role that Markov decision processes play in reinforcement learning.

Explicit constraints on the agent have been previously considered in continual learning as a means of capturing the big world hypothesis. For example, in continual learning experiments, it is common practice to constrain what the agent can store [14], or the capacity of its function approximator [13]. Information theory provides a framework to formalize explicit agent constraints [11, 10]. However, outside of simple and well-specified pairs of agent and environment, these constraints can be difficult to characterize without knowledge of the true information-theoretic quantities involved between the state maintained by the agent and its future sensory stream from the environment. In addition, explicit constraints hinder the effectiveness of scaling up the agent's capacity, which has been a source of progress in machine learning more broadly [6].

In contrast to explicit constraints, our approach considers the implicit constraint that arises from an agent embedded in an environment (see Figure 1a). The embedded aspect of all intelligent systems, by existing in the physical world, is not often considered to be part of the problem formulation [5]. However, the physical world is a clear example of a world bigger than any agent, suggesting that embedded agency may be useful in formulating the big world hypothesis.

To provide a general environment in which an agent can be embedded, we define a *universal-local environment*. This environment is a Markov process that is computationally universal—capable of simulating any computation—where the transition dynamics can be localized to a neighbourhood of the state-space. Our approach is similar to universal artificial intelligence [8], which considers a computationally universal environment to explore the limits of the theoretically optimal, but uncomputable, AIXI agent [7].

To define an embedded agent, we consider an embedded automaton simulated within the state-space of our universal-local environment. This automaton interacts with a partially observable Markov decision process, defined on the boundary between the automaton and the rest of the universal-local environment. We then propose *interactivity* that measures an embedded automaton's ability to adapt its future behaviour, conditioned on its past behaviour, using Kolmogorov complexity. Interactivity is similar to previously considered intrinsic motivation objectives [3, 16], and specifically predictive information [2, 17]. However, interactivity differs because of its formulation in terms of behaviours using Kolmogorov complexity. This makes interactivity better suited to sequential decision making in the constrained and partially observable setting that we consider.

2 A Bigger World: Computational Universality and Locality

We begin by defining a general notion of an environment that is unviversal in which an agent can be locally embedded. Specifically, *environment* is used to refer to a general history-based process that is defined over a finite set of symbols, and without an explicit notion of agent.

Definition 1. *An environment, $\mathcal{E} = (\Sigma, \mathbb{C})$, is a discrete process defined over a finite symbol-set, Σ, that maps a string of symbols, $\sigma_{0:t-1} = \sigma_0\sigma_1 \cdots \sigma_{t-1}$, to the next-symbol that extends the string, $\sigma_t \in \Sigma$, using the construction function, $\sigma_t = \mathbb{C}(\sigma_{0:t-1})$.*

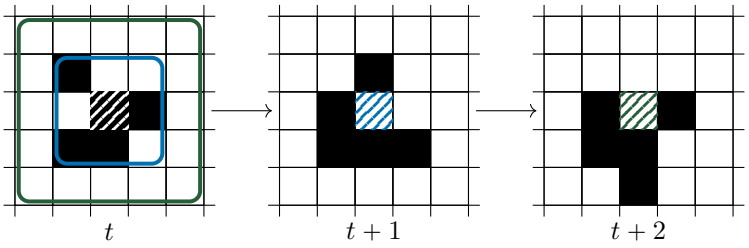


Figure 2: **Conway's Game of Life is a cellular automaton and an example of a universal-local environment.** The state-space is an infinite 2D grid, in which cells live (black) with 2 or 3 neighbours, but die (white) otherwise, and dead cells with 3 neighbours become alive. The blue and green borders (*left*) correspond to neighbourhoods that determine the middle cell at time-steps $t+1$ (*middle*) and $t+2$ (*right*). Longer-term transition dynamics depend on larger neighbourhoods.

An environment is computationally universal if it is equivalent to a universal Turing machine, meaning that it is capable of simulating any computation given a suitable initial string of symbols. Such an environment can also be represented as a Markov process, $\mathcal{M}(\mathcal{E}) = (\Omega, \mathbb{U})$, defined over the countably infinite state-space, Ω, in which the state, $\omega_t \in \Omega$, is updated using the transition function, $\omega_{t+1} = \mathbb{U}(\omega_t)$.

2.1 Defining Locality with Boundaried Markov Processes

Intuitively, locality ensures that the environment's transition dynamics on a restricted portion of the state-space. Specifically, we use the term substate-space to refer to the portion of the state-space restricted to a finite index-set.

Definition 2. *A substate-space, Ω_Λ, is defined as a restriction of the state-space, Ω, to a finite index-set, $\texttt{Idx}(\Omega_\Lambda) := \Lambda$ where $|\Lambda| < \infty$, such that $\Omega_\Lambda = \{\omega_\Lambda : \omega \in \Omega\}$ where $\omega_\Lambda = \{\omega_i\}_{i \in \Lambda}$. We use square set notation to denote operations on the index set, such as $\Omega_\Lambda \sqsubseteq \Omega$ to denote the inclusion of the index-set, $\Lambda \subseteq \texttt{Idx}(\Omega)$, and the union of index-sets, $\Omega_{\Lambda_1} \sqcup \Omega_{\Lambda_2} = \Omega_{\Lambda_1 \cup \Lambda_2}$.*

We now consider the environment's transition restricted to a generic substate-space, $X \sqsubseteq \Omega$, without reference to the specific index-set, $\texttt{Idx}(X)$. In particular, we define a boundaried Markov process in which the one-step transition dynamics, $\mathbb{U}_X$, depend on another substate-space, $B_X \sqsubseteq \Omega$, referred to as the boundary-space for a given substate-space, X.

Definition 3. *A boundaried Markov process, $\mathcal{M}_X = (X, B_X, \mathbb{U}_X)$, is a discrete process in which the substate-space, X, and boundary-space, B_X, define the one-step transition of the substate-space, $x_{t+1} = \mathbb{U}_X(x_t, b_t)$, for $x_{t+1}, x_t \in X$ and $b_t \in B_X$.*

The boundary-space is defined for one-step dynamics; A larger boundary-space is generally needed for multi-step transition dynamics. This is because the current substate, $x_t \in X$, and the current boundary, $b_t \in B_X$, only define the next-substate, $x_{t+1} \in X$, and not the next-boundary, $b_{t+1} \in B_X$. We use this fact to define a local environment that consists of nested boundaried Markov processes.

Definition 4 (Locality)**.** *A universal Markov environment is local if, for any two proper substate-spaces, $W \sqsubsetneq X \sqsubseteq \Omega$, there exists boundaried Markov processes with corresponding index-sets that are properly contained, $W \sqcup B_W \sqsubsetneq X \sqcup B_X$.*

Thus, a *universal-local environment* is a universal Markov environment that is also local. This environment is capable of simulating arbitrary computations, and any bounded computation is localized to a portion of the environment's state-space. It can be understood as a computationally universal Markov process in which longer-term dynamics are a function of a larger portion of the state-space.

2.2 Example of a Universal-Local Environment: Conway's Game of Life

Conway's Game of Life is an example of a universal-local environment [4]. This environment is computationally universal because, within Conway's Game of Life, a universal Turing machine can be simulated [1, 15]. A substate-space in Conway's Game of Life is a finite subset of locations on the grid, specifying the possible values taken by the cells at those locations. The one-step transition dynamics on any substate-space depend on the adjacent neighbourhood of that substate-space, which defines the boundary-space (see Figure 2). Conway's game of life is local because if one substate-space contains another, then the boundary-spaces (the adjacent neighbourhood of the substate-spaces) are also also contained.

While Conway's Game of Life has the potential to simulate any computation using its local dynamics, we are not suggesting to program an agent within it. We only point out Conway's Game of Life as a proof-of-existence for universal-local environments. Instead, we will consider and formalize the implicit constraints faced by an agent if it were embedded in such an environment.

3 Embedded Agents as Localized Computations

A universal-local environment can simulate arbitrary computations, which we use to define an embedded automaton, $\mathcal{A}$, on the environment's state-space, Ω. Moreover, due to locality, the embedded automaton can be localized to a substate-space, $A \sqsubseteq \Omega$.

Definition 5. *An embedded automaton is defined by $\mathcal{A} = (A, I_A, O_A, \mathbb{U}_A, \pi_A)$, where $A \sqsubseteq \Omega$ is the internal substate-space of the automaton, $I_A, O_A \sqsubseteq B_A$ are input and output spaces defined on the boundary-space, B_A, and $\mathbb{U}_A, \pi_A$ are the automaton's transition and output function respectively.*

An embedded automaton is equivalent to an agent interacting with a (potentially reward-free) partially observable Markov decision process, if its boundary-space consists of only the input and output spaces, $I_A \sqcup O_A = B_A$. Relating this to an agent in reinforcement learning, we may think of the input-space as the observation-space,[1] the internal substate as the parameters of a function approximator, the output-space as an action-space, the transition function as a learning rule, and the output function as a policy.

By construction this agent is implicitly constrained relative to the environment by being a restricted model of computation. While every embedded agent is implicitly constrained, some may generate simple output sequences that do not require more than agent's capacity. For example, a periodic output sequence would not require more capacity than the period of the sequence. We will show, however, that agents are constrained by their finite capacity when adapting to their past input/output experience.

3.1 Interactivity as a Computational Measure of Adaptivity

An agent's capability for learning can be characterized by its ability to adapt its future behaviour using its past experience. We propose *interactivity* to measure an embedded agent's intrinsic ability to adapt its future behaviour, towards higher complexity, conditioned on its past behaviour. Specifically, we use Kolmogorov complexity to formalize this otherwise intuitive notion of adaptation and complexity.

We represent an embedded agent as an embedded automaton $\mathcal{A}$ where its input and output spaces determine its boundary-space, $I_A \cup O_A = B_A$. Thus, the behaviour of the agent is determined by the values taken on the boundary-space, $b_t = (i_t, \pi_A(i_t)) \in B_A$ where $i_t \in I_A$ and $\pi_A(i_t) \in O_A$. At any time t, the behaviour can be separated into past, $b_{0:t} = b_0 b_1 \cdots b_t$ and the T-horizon future, $b_{t+1:T} = b_{t+1} b_{t+2} \cdots b_{t+T}$.

Definition 6. *An agent's interactivity at time t is the average difference in the unconditional Kolmogorov complexity of its future behaviour and the conditional Kolmogorov complexity of its future behaviour, conditioned on its past behaviour, $\mathbb{I}^*_t(\mathcal{A}) = \lim_{T\to\infty} \frac{1}{T}\big(\mathbb{K}(b_{t+1:T}) - \mathbb{K}(b_{t+1:T}|b_{0:t})\big)$.*

That is, interactivity measures the predictable complexity of an agent's future behaviour, given its past behaviour. Interactivity is high if (i) the future behaviour, $b_{t+1:T}$, has high unconditional Kolmogorov complexity and (ii) the past behaviour, $b_{0:t}$, is predictive of this future behaviour, thereby yielding a low conditional Kolmogorov complexity. However, interactivity is low if the future behaviour has low Kolmogorov complexity, or if the past behaviour is not sufficiently predictive.

3.2 An Interactivity-Maximizing Agent Faces a Big World

The interactivity of any embedded agent is always constrained by its capacity. That is, with a given capacity, an embedded agent can only sustain a given level of interactivity. However, if the embedded agent is given more capacity, then it could use the additional capacity to increase its interactivity. Thus, the environment appears to be a big world from the perspective of an interactivity-maximizing agent.

An interactivity-maximizing agent has an ability to continually adapt its future behaviour by using its past experience. This suggests the following interactivity thesis:

Interactivity measures a general capability for continual adaptation.

We refer to this as the interactivity thesis, rather than a hypothesis, to reflect its speculative and philosophical nature. An agent's capability for continual adaptation with low interactivity is limited because its future behaviour is either: i) simple, or ii) complex, but not predictable from its past experience. In either case, the thesis stresses the relative notion of capabilities. A simple agent could be capable of some adaptation, but its capabilities would be greater if its past experience was used to produce more complex behaviour. Moreover, an agent that produces complex behaviour could only be recognized as an adaptation if this complexity can be attributed, via prediction, to its past experience. Embracing the interactivity thesis naturally leads to a relative spectrum of possible adaptive agents.

[1] The input-space may also provide an external reward to the automaton, but this need not be the case.

4 Discussion

In this paper, we constructed a formalism for an agent interacting with a bigger world, by considering the implicit constraint faced by an embedded agent. Our work suggests that maximizing interactivity leads to the common desideratum of the continual learning problem in which any agent that stops learning is suboptimal. The key to this formalism is the fact that interactivity does not depend on external feedback, but rather is defined in terms of the past and future behaviour of the agent. While interactivity could potentially provide a rich source of intrinsic feedback, it also introduces challenges the stability of our algorithms combining nonlinear representations, temporal difference learning, and online learning.

Maximizing interactivity provides a problem setting for studying continual learning in isolation. A promising direction is the development of an efficient algorithms for maximizing interactivity using reinforcement learning. In particular, interactivity could be predicted, similar to a value function which is conditioned on the agent's current policy and learning algorithm. Experimental evaluation in this setting also requires special consideration. Holding the agent fixed for evaluation, as is commonly done in machine learning, is not be appropriate given that interactivity is defined as an online objective. In addition, standard approaches to hyperparameter tuning may not be feasible for evaluating the long-term performance of a continual learning agent [12]. Overcoming these obstacles would require re-evaluation of several components of empirical practice in machine learning, and we thus leave an empirical investigation for future work.

We close with the following conjecture regarding interactivity and its utility as a general objective in an arbitrary environment: if an agent is capable of sustaining a particular level of interactivity, then it is also capable of behaviours that achieve other goals in that environment—such as maximizing external reward—that require equal or less interactivity.

References

[1] Berlekamp, E. R., Conway, J. H., and Guy, R. K. (1982). *Winning Ways for Your Mathematical Plays, Vol. 2*. Academic Press.

[2] Bialek, W., Nemenman, I., and Tishby, N. (2001). Predictability, complexity, and learning. *Neural computation*, 13(11):2409–2463.

[3] Chentanez, N., Barto, A. G., and Singh, S. (2004). Intrinsically motivated reinforcement learning. *Advances in Neural Information Processing Systems*.

[4] Conway, J. H. (1970). The game of life. *Scientific American*, 223(4):4.

[5] Demski, A. and Garrabrant, S. (2019). Embedded agency. *CoRR*, abs/1902.09469v3.

[6] Hoffmann, J. et al. (2022). Training compute-optimal large language models. *CoRR*, abs/2203.15556v1.

[7] Hutter, M. (2000). A theory of universal artificial intelligence based on algorithmic complexity. *arXiv preprint cs/0004001*.

[8] Hutter, M. (2005). *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin.

[9] Javed, K. and Sutton, R. S. (2024). The big world hypothesis and its ramifications for artificial intelligence. In *Finding the Frame Workshop at Reinforcement Learning Conference*.

[10] Kumar, S., Jeon, H. J., Lewandowski, A., and Van Roy, B. (2024). The need for a big world simulator: A scientific challenge for continual learning. In *Finding the Frame Workshop at Reinforcement Learning Conference*.

[11] Kumar, S., Marklund, H., Rao, A., Zhu, Y., Jeon, H. J., Liu, Y., and Van Roy, B. (2023). Continual learning as computationally constrained reinforcement learning. *CoRR*, abs/2307.04345.

[12] Mesbahi, G., Mastikhina, O., Panahi, P. M., White, M., and White, A. (2024). Tuning for the unknown: Revisiting evaluation strategies for lifelong rl. *CoRR*, abs/2404.02113v2.

[13] Meyer, E. J., White, A., and Machado, M. C. (2024). Harnessing discrete representations for continual reinforcement learning. *Reinforcement Learning Journal*, 2:606–628.

[14] Prabhu, A., Torr, P. H., and Dokania, P. K. (2020). GDumb: A simple approach that questions our progress in continual learning. In *European Conference on Computer Vision*.

[15] Rendell, P. (2011). A universal turing machine in Conway's game of life. In *International Conference on High Performance Computing & Simulation*.

[16] Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990-2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247.

[17] Still, S. and Precup, D. (2012). An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Biosciences*, 131:139–148.

[18] Sutton, R. S., Koop, A., and Silver, D. (2007). On the role of tracking in stationary environments. In *International Conference on Machine Learning*.