World models of environment, agent and joint agent-environment systems

Manuel Baltieri[1,2] **Filippo Torresan**[1] **Yivan Zhang**[3]
Alexander Boyd[4] **Fernando E. Rosas**[2,5,6,7]

[1] **Araya Inc.**
[2] **Department of Informatics, University of Sussex**
[3] **The University of Tokyo**
[4] **Beyond Institute for Theoretical Science (BITS)**
[5] **Sussex AI and Sussex Centre for Consciousness Science, University of Sussex**
[6] **Department of Brain Sciences, Imperial College London**
[7] **Centre for Eudaimonia and Human Flourishing, University of Oxford**

manuel_baltieri@araya.org

Abstract

World models are a central component of model-based reinforcement learning. They are usually discussed in terms of what variables they predict, such as observations, rewards, states, latent or information states. We argue that there is a prior distinction: which channel they model. We consider three cases: the environment channel $O_{:} \mid A_{:}$, the agent channel $A_{:} \mid O_{:}$, and the realised joint process $(A, O)_{:}$, equivalently viewed as a channel with no inputs. Using computational mechanics, we define canonical predictive models for these three cases as ϵ-transducers or ϵ-machines. Canonical environment models recover standard predictive state representations, while the other two give analogous notions of canonical models for the agent and the joint system. We then build canonical support-restricted environment and agent models induced by closed-loop coupling, whose predictive equivalences range over continuations supported by the realised interaction. The key structural result is that canonical support-restricted environment states factor through the canonical joint causal states, and their transition structure is induced directly from the joint model; the agent-side construction is dual. Finally, we give a POMDP/controller example in which the unrestricted environment model has infinitely many states while the canonical support-restricted model induced by the coupling is finite. The framework clarifies what different world models are models of, and how coupling and support restriction can change their canonical predictive structure and complexity.

1 Introduction

World models are a central construct of modern (deep) reinforcement learning [20, 35, 15, 22]. Their formalisation usually relies on state or action-state abstractions, focusing mostly on fully-observable models [46, 1]. While often focused on modelling environments [42, 15], world models of different kinds have increasingly emerged as relevant for agents to model themselves [21, 27], other agents [64], joint agent-environment [37] or even more complex systems [65].

This proliferation suggests that world models should be distinguished not only by which variables they predict, see e.g. [42] for world models of the environment, but also by which channel they model. In the setting of agent-environment interaction, we consider three cases: the environment channel $O_{:} \mid A_{:}$, the agent channel $A_{:} \mid O_{:}$, and the realised joint process $(A, O)_{:}$, equivalently viewed as the null-input channel $(A, O)_{:} \mid \mathbf{1}_{:}$, i.e. a channel with trivial inputs ($\mathbf{1} = \{*\}$). These answer different predictive questions: how future observations depend on future actions, how future

Preprint.

actions depend on future observations, and what future action-observation traces are generated by the coupled interaction itself. They are not just different target variables for the same kind of model, as in e.g. [42]. They induce different causal equivalence relations, and therefore different notions of abstraction, model complexity, and use cases.

Distinguishing these channels and their corresponding models can be useful for different purposes. For AI interpretability, understanding what RL agents effectively model, either by construction, or as a side effect of a particular architecture, loss function, or training method, is an open challenge [7, 13]. This is particularly important for evaluating and supervising how RL agents form and use world models and beliefs of different kinds for decision making. Making the channel explicit is especially important for interpretability. Before asking whether the latent state learned by an AI model is a belief state, a predictive state, or an abstraction of some model, one must ask what it is a state *of*: a model of the environment channel, the agent channel, or the realised joint process. A latent state that is sufficient for predicting future observations under given actions need not be sufficient for predicting the agent's own future actions, and neither need coincide with a minimal state of the realised closed-loop interaction.

Various approaches, inspired by formal verification methods and logic [13], control theory [62], and neuroscience [39], among others, have been proposed to this end, taking advantage of different features and guarantees provided by each of these frameworks. More recently, computational mechanics [12, 56], an approach with the goal of understanding the mechanics of computation and information processing, has provided a formal background and semantics to understand world models and abstractions for agents and machine learning models solving various tasks [54, 50, 8, 49, 47, 44, 53].

In this work we extend the approach of [50] focused on the environment channel, and propose a systematic characterisation of canonical predictive models organised by the channel they model. In section 2 we introduce canonical environment, agent, and joint models using ϵ-machines and ϵ-transducers, with more technical background in appendix A. On the environment side, this recovers predictive state representations used to build models without relying on given latent variable representations, with predictive states as equivalence classes of histories are identified by the predictions they assign to future action-observation tests [34, 58]. We apply the same criterion to the agent channel and to the realised joint process, giving canonical states for policy behaviour and coupled interaction as well as for environment dynamics. In section 3 we then show how closed-loop coupling induces canonical support-restricted environment and agent models, whose predictive equivalences range over continuations supported by the realised interaction. This captures a precise interpretation of world models that represent only the possible interactions that an agent has with its environment. These models are in general simpler because they don't include all the possible (counterfactual) ways an agent and environment could have interacted with under different circumstances. The main structural result is that the non-sink canonical support-restricted environment states factor through the joint causal states, with transitions induced from the joint model, and the agent-side construction is dual. This is not true in general for the case without support restriction, unless more assumptions are introduced. In section 4 we review related work before a few final discussion points are brought out in section 5.

Contributions. Our contributions are three-fold. First, we extend the viewpoint behind predictive state representations of environments to the other channels over the same interface of an agent-environment interaction, an agent channel and a joint process seen as a channel with trivial inputs, see definition 2. In this way, the (latent) states of a model are defined by different equivalence relations of action-observation histories based on their predictive content, rather than assumed to be given by a particular hidden-state presentation. This gives an input-output analogue for policy behaviour and an autonomous predictive-state model for the realised joint process. For reinforcement learning, this separates the information necessary for planning from that needed for modelling a policy or an on-policy rollout, by defining different different states depending on the task. This formalises the informal idea that states sufficient for predicting future observations are in general not the same as those sufficient for predicting future actions, or future action-observation pairs. For mechanistic interpretability, it gives instead reference targets for asking what kind of predictive state a learned representation implements. This means, for instance, that we can extend existing work on beliefs about the environment of AI models such as transformers [54, 49, 47, 44, 53] to look for beliefs the AI model might have about itself or its interaction with the environment. Second, we

(a) Environment model. (b) Agent model. (c) Joint model.

Figure 1: Different kinds of world models, represented pictorially. The joint model represents the realised coupled action-observation process.

define canonical support-restricted environment and agent models induced by a fixed closed-loop coupling definition 14 and table 2. This identifies when a model is being queried outside the support of the realised interaction, and distinguishes counterfactual world modelling from policy-conditioned prediction. Third, we show that the support-restricted environment model is always determined by the joint model: the joint model determines its non-sink states and transition structure, with only the totalisation sink added theorems 16 to 18. This shows why compact closed-loop prediction need not imply a compact unrestricted model.

2 Canonical models

This section introduces the mathematical setup used to characterize the interaction between an agent and its environment. Specifically, we formalise the predictive semantics of these interactions by defining canonical environment, agent, and joint models in terms of computational mechanics tools.

In the following, uppercase letters (e.g. X, Y) are used to denote random variables and lowercase (e.g. x, y) their realisations, calligraphic letters (e.g. $\mathcal{X}, \mathcal{Y}$) denote the sets over which they take values, and the symbol Δ (as in $\Delta(\mathcal{X}), \Delta(\mathcal{Y})$) is used to denote the collection of all distributions over those sets. We use the shorthand notation $p(x \mid y) = \Pr(X = x \mid Y = y)$ to express probabilities when there is no risk of ambiguity, and assume that equalities of the form $p(x \mid y, z) = p(x \mid y)$ hold for all realisations that can take place with non-zero probability. $\mathbb{N} = \{0, 1, 2, \ldots\}$ corresponds to zero-based numbering, and we use the following abbreviations: $\boldsymbol{x}_{t:t+L} = (x_t \ldots, x_{t:t+L-1})$, $\boldsymbol{x}_{:t} = \boldsymbol{x}_{-\infty:t}$, $\boldsymbol{x}_{t:} = \boldsymbol{x}_{t:\infty}$, and $\boldsymbol{x}_{:} = \boldsymbol{x}_{-\infty:\infty}$. Note that the left index is always inclusive, the right always exclusive. We will often treat an autonomous stochastic process $X_{:}$ as a channel with a trivial input alphabet $\mathbf{1} = \{*\}$, identifying $X_{:}$ with the null-input channel $X_{:} \mid \mathbf{1}_{:}$ when convenient.

We initially consider how agents and environments interact over a given interface with observation symbols $O_t \in \mathcal{O}$ and action symbols $A_t \in \mathcal{A}$.

Definition 1 (Agent-environment interface [41, 2])**.** The interface is a pair of finite sets, $(\mathcal{A}, \mathcal{O})$ such that $|\mathcal{A}| \geq 2, |\mathcal{O}| \geq 2$.

The predictive objects considered in this work can be described using channels: two channels as in [18], together with the realised joint process viewed as a null-input channel.

Definition 2 (Predictive channels over an agent-environment interface)**.** For an interaction over the interface $(\mathcal{A}, \mathcal{O})$, we consider three associated channels, the *environment channel* $O_{:} \mid A_{:}$ which predicts observations from actions, the *agent channel* $A_{:} \mid O_{:}$ which predicts actions from observations, and the realised joint process $(A, O)_{:}$, written as the null-input channel $(A, O)_{:} \mid \mathbf{1}_{:}$.

Next, we define the corresponding canonical models in terms of causal states for these three channels. These should be seen as optimal, in the computational mechanics sense of unique, unifilar, minimal and predictive, world models for different channels, extending intuitions and classifications applied to environment models (fig. 1a), see e.g. [42], to agent (fig. 1b) and joint (fig. 1c) models.

Channel	Description	Input process	Output process	Goal
$O_{:} \mid A_{:}$	environment	$A_{:}$	$O_{:}$	predict observations from actions
$A_{:} \mid O_{:}$	agent	$O_{:}$	$A_{:}$	predict actions from observations
$(A,O)_{:} \mid \mathbf{1}_{:}$	joint	$\mathbf{1}_{:}$	$(A,O)_{:}$	predict realised coupled traces

Table 1: Channels studied in this paper. The channel, rather than only the predicted variables, determines what kind of world model is being defined.

2.1 Canonical environment model

We now define the environment's channel causal states, $\mathcal{S}^E$, see appendix A, with an equivalence relation $\sim_E$ of histories of actions and observations. For simplicity, we write $H_{:t} = (A,O)_{:t}$ for action-observation histories, with realisations $h_{:t}$ and history space $\mathcal{H}_{:t}$. The equivalence relation is:

$$h_{:t} \sim_E h'_{:t} \iff \Pr(O_{t:} \mid A_{t:}, H_{:t} = h_{:t}) = \Pr(O_{t:} \mid A_{t:}, H_{:t} = h'_{:t}) \tag{1}$$

and it induces a surjective map $\epsilon_E : H_{:t} \to \mathcal{S}^E$ given by $\epsilon_E(h_{:t}) = s^E = \{h'_{:t} \mid h_{:t} \sim_E h'_{:t}\}$. These causal states are minimal sufficient statistics for prediction of future observations given future actions and past actions and observations, see section A.1.2. Notice how conditioning on $A_{t:}$ in the environment channel is part of the predictive equivalence quantifier: it enforces equality of predictive behaviour for all possible future input continuations. This is particularly relevant when the environment is later coupled to an agent: the coupling may support only a subset of possible future input continuations, leading to a different partition of histories, see section 3.

The ϵ_E-map induces dynamics over causal states captured by stochastic matrices $\mathcal{T}^E = \{T^{(o|a)}\}_{o\in\mathcal{O},a\in\mathcal{A}}$ given by:

$$T^{(o|a)}_{s^E \to \bar{s}^E} = \Pr(S^E_{t+1} = \bar{s}^E, O_t = o \mid S^E_t = s^E, A_t = a). \tag{2}$$

Definition 3 (Canonical environment model)**.** A canonical environment model is the ϵ-transducer $(\mathcal{A}, \mathcal{O}, \mathcal{S}^E, \mathcal{T}^E)$ of the environment channel $O_{:} \mid A_{:}$.

Remark 4 (Relation to predictive state representations)**.** The canonical environment model can be read as the all-tests, presentation-independent semantic for idealised predictive state representations. This relation has been noted in, among others, [55, 63, 50]. In the predictive state representation literature, a test is a finite future action-observation experiment: future actions are supplied as inputs, and the model predicts the probability of the corresponding future observations [58]. A predictive state representation represents a history by the predictions it assigns to such tests, here the causal state is the equivalence class of histories that agree on all future tests. Finite-dimensional predictive state representations typically choose a finite set of core tests or a particular presentation, whereas the ϵ-transducer gives the optimal minimal unifilar presentation of the environment channel.

2.2 Canonical agent model

Consider an agent's causal states, $\mathcal{S}^M$, defined by the equivalence relation $\sim_M$:

$$h_{:t} \sim_M h'_{:t} \iff \Pr(A_{t+1:} \mid O_{t:}, H_{:t} = h_{:t}) = \Pr(A_{t+1:} \mid O_{t:}, H_{:t} = h'_{:t}) \tag{3}$$

inducing a surjective map $\epsilon_M : H_{:t} \to \mathcal{S}^M$ given by $\epsilon_M(h_{:t}) = s^M = \{h'_{:t} \mid h_{:t} \sim_M h'_{:t}\}$. These causal states are minimal sufficient statistics for prediction of future actions given future observations and past actions and observations, see section A.1.2. The shift from O_t to A_{t+1} reflects the convention that the current observation is treated as the input for the agent's next action. The ϵ_M-map induces dynamics over causal states captured by stochastic matrices $\mathcal{T}^M = \{T^{(a|o)}\}_{a\in\mathcal{A},o\in\mathcal{O}}$ given by:

$$T^{(a|o)}_{s^M \to \bar{s}^M} = \Pr(S^M_{t+1} = \bar{s}^M, A_{t+1} = a \mid S^M_t = s^M, O_t = o). \tag{4}$$

Definition 5 (Canonical agent model)**.** We define a canonical agent model to be the ϵ-transducer $(\mathcal{O}, \mathcal{A}, \mathcal{S}^M, \mathcal{T}^M)$ of an agent channel $A_{:} \mid O_{:}$.

The canonical agent model is obtained by applying the same predictive-state construction to the agent channel $A_{:} \mid O_{:}$ that predictive state representations apply to the environment channel $O_{:} \mid A_{:}$.

Remark 6 (Action-predictive state representations)**.** For the environment channel, a finite test fixes a future action word and asks for the probability of a future observation word. For the agent channel, the corresponding agent-side test fixes a future observation word and asks for the probability of a future action word. Concretely, for finite words $o_{t:t+L}$ and $a_{t+1:t+L+1}$,

$$\Pr(A_{t+1:t+L+1} = a_{t+1:t+L+1} \mid O_{t:t+L} = o_{t:t+L}, H_{:t} = h_{:t}). \tag{5}$$

Thus two histories are agent-causally equivalent when they give the same predictions for all such future observation-action tests. In this sense, the canonical agent model is a predictive-state model of policy or controller behaviour: it represents histories by how the agent would act under all possible future observation continuations.

As in the case of the environment channel, conditioning on $O_{t:}$ is part of the predictive equivalence quantifier: it compares predictive behaviour across all possible future input (observation) continuations. This does not mean that the agent observes the future. In section 3, we instead consider support-restricted variants, where only observation continuations supported by the coupled environment are considered. If one further weights or selects observation continuations by desirability, related constructions connect to control-as-inference [28], planning-as-inference [31], and active-inference perspectives [4, 38].

2.3 Canonical joint model

Finally, we consider joint agent-environment models. This is inspired by the idea of embedded agency [14], where agents model themselves as part of a world that includes them and their environment. Following the convention introduced above, the joint process can also be viewed as a null-input channel, $(A, O)_{:} \mid \mathbf{1}_{:}$. Unlike the environment and agent channels, however, it has no non-trivial exogenous input: it describes the realised coupled action-observation autonomous process. We start defining an equivalence relation $\sim_J$ of the joint pasts given by:

$$h_{:t} \sim_J h'_{:t} \iff \Pr((A, O)_{t:} \mid H_{:t} = h_{:t}) = \Pr((A, O)_{t:} \mid H_{:t} = h'_{:t}) \tag{6}$$

which gives a surjective map $\epsilon_J : (\mathcal{A} \times \mathcal{O})_{:t} \to \mathcal{S}^J$ defined by $\epsilon_J(h_{:t}) = s^J = \{h'_{:t} \mid h_{:t} \sim_J h'_{:t}\}$. The resulting states $\mathcal{S}^J$ are the causal states of the joint ϵ-machine [6]. Transitions between these causal states are given by stochastic matrices $\mathcal{T}^J = \{T^{(a,o)}\}_{(a,o)\in\mathcal{A}\times\mathcal{O}}$:

$$T^{(a,o)}_{s^J \to \bar{s}^J} = \Pr(S^J_{t+1} = \bar{s}^J, (A, O)_t = (a, o) \mid S^J_t = s^J). \tag{7}$$

Definition 7 (Canonical joint model)**.** A canonical joint model of the joint process $(A, O)_{:}$ is the unique minimal unifilar machine presentation given by the tuple $(\mathcal{A} \times \mathcal{O}, \mathcal{S}^J, \mathcal{T}^J)$.

Remark 8 (Joint predictive states and policy-induced processes)**.** Under the null-input-channel convention, the joint model is the autonomous counterpart of the environment and agent channel models. Its causal states represent histories by their predictions of future action-observation traces, rather than by predictions of future observations under action inputs or future actions under observation inputs. In a fully observable MDP with a fixed policy, this reduces to the familiar policy-induced Markov chain, or Markov reward process if rewards are included [59, 45]. More generally, coupling a partially observable environment to a policy or finite-state controller induces an autonomous stochastic process over action-observation traces [36, 30]. The joint model is the canonical minimal predictive presentation of this realised process.

2.4 Relations among the three canonical models

Since the environment, agent, and joint models are all built from the same action-observation histories, one might expect their causal states to coincide, or at least to be related by a simple construction. This is not true in general. The reason is that the three models are minimal for different prediction problems: the environment model predicts future observations under future action inputs, the agent model predicts future actions under future observation inputs, while the joint model predicts future action-observation traces of the realised coupled process. These are different equivalence relations on histories, cf. eqs. (1), (3) and (6), and therefore need not induce the same partition. We introduce a running example to illustrate this point.

The example is deliberately minimal, but it isolates a common pattern in reinforcement learning. In partially observable environments, arbitrary (counterfactual) action continuations can require tracking a rich belief state, while a particular policy, controller, action mask, option structure, or dataset support may realise only a much smaller language of continuations. Thus a model of the environment under arbitrary action inputs can have a different predictive structure from a model of the action-observation traces generated by a fixed coupling. This means that a compact representation learned from rollouts need not be a representation of the full counterfactual environment dynamics; it may instead encode the realised policy-environment process. The binary construction below is a small illustration of this phenomenon.

Example 9. Consider a binary environment with latent state $E_t \in \{0, 1\}$, binary actions, and binary observations. Action 0 resets the next latent state to a fair coin, while action 1 holds the latent state fixed. The observation is a noisy readout of the new latent state, with noise parameter $\eta \in (0, 1/2)$. For the unrestricted environment channel $O_{:} \mid A_{:}$, arbitrary future action continuations must be considered. In particular, arbitrarily long runs of action 1 generate infinitely many distinct posterior beliefs over E_t. These posterior beliefs give different predictions for future observations, and hence correspond to infinitely many distinct environment causal states. Now couple this environment to a deterministic finite-state controller with memory states $\{\alpha, \beta, \gamma\}$. The controller emits action 0 in state α, and action 1 in states β and γ. Its memory update is such that, after a reset, it performs one hold action and sometimes one further hold action before returning to reset. As an agent channel $A_{:} \mid O_{:}$, this controller has only finitely many causal states: its finitely many memory modes are predictively sufficient for future action prediction. The realised joint process is also finite-state. Once the controller is coupled to the environment, arbitrarily long hold continuations are no longer realised. Only finitely many posterior configurations appear in the coupled action-observation process, and the joint ϵ-machine has finitely many causal states. Thus, in this example, the unrestricted environment model is infinite, the agent model is finite, and the joint model is finite.

In this example, we can see that the joint model cannot easily obtained simply by identifying the causal states of the environment and agent models, because these are sufficient statistics of histories for different future predictions, nor by taking an obvious product, sum, or monotone combination of their state spaces. A product bound can, for instance, be recovered after adding extra structure, see appendix D. In general however, the separately defined environment and agent causal states need not retain the correlation, timing, or support information required for the pair (S_t^E, S_t^M) to be sufficient for predicting the realised joint process. Moreover, in examples such as the one above, the bound is not informative because the unrestricted environment model is already infinite.

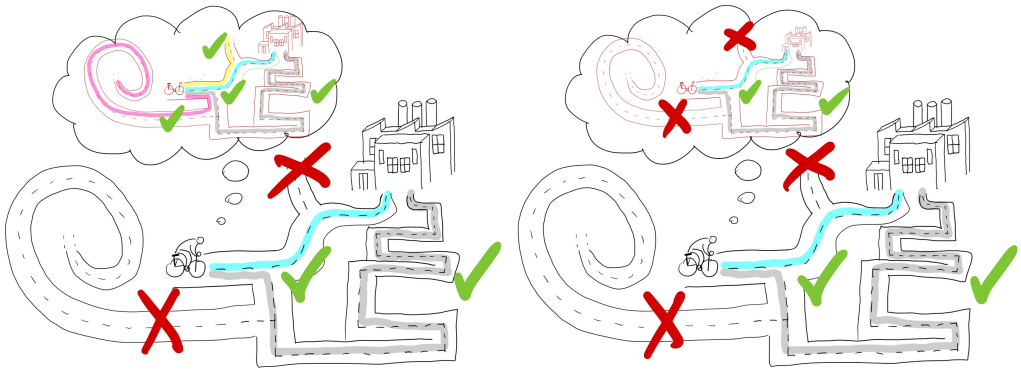
The next section shows that a more robust relation appears after support restriction. While the unrestricted models are not related in any obvious way, the canonical support-restricted environment and agent models induced by the realised coupling do factor through the joint model.

3 Canonical support-restricted models from closed-loop coupling

The canonical environment and agent models of section 2 are *unrestricted* channel models: their causal equivalences compare predictive behaviour under arbitrary future input continuations. In a realised closed-loop agent-environment interaction, however, not all such continuations are supported [23, 9]. An agent may be limited by its policy class, controller, embodiment, action mask, or choices, see fig. 2. Dually, an environment restricts which future observation continuations are available to the agent. We call the resulting models *canonical support-restricted models*. We first give the environment-side construction. The agent-side construction is obtained dually by exchanging actions and observations.

The natural support-restricted environment object is initially partial. For an ordinary history, the realised coupling determines which action queries are supported. On supported action queries, the environment-side prediction is the ordinary conditional observation law induced by the joint process. On unsupported action queries, no ordinary environment prediction is defined by the realised coupling. We totalise [57, 25] this partial channel by adjoining a symbol $\perp$, which is emitted by the totalised channel when an action query is unsupported. Intuitively, an action can be unsupported because the coupled agent cannot execute it, its controller or policy assigns it zero probability, or it is excluded by an action mask or other constraint.

Let $\widetilde{\mathcal{O}} \coloneqq \mathcal{O} \cup \{\perp\}$ be the augmented observation alphabet, and let the set of extended histories be $\widetilde{\mathcal{H}}_{:t} \coloneqq (\mathcal{A} \times \widetilde{\mathcal{O}})_{:t}$. Call $\widetilde{h}_{:t} \in \widetilde{\mathcal{H}}_{:t}$ *ordinary* if it contains no occurrence of $\perp$. In that case we identify



(a) Canonical environment model. (b) Canonical support-restricted environment model.

Figure 2: Unrestricted and support-restricted prediction. The unrestricted model considers arbitrary future continuations through the interface, including counterfactual paths not generated by the coupled agent, i.e. for instance paths on a map that an agent doesn't consider or never takes. The canonical support-restricted model keeps only continuations supported by the realised coupling: unsupported queries are marked as failures and sent to the totalisation sink $\perp$.

it with the unique ordinary history $h_{:t} \in \mathcal{H}_{:t}$ having the same symbols. To make the domain of this partial channel explicit, we first define which finite future action continuations are supported by the realised joint process at a given ordinary history. This finite-horizon definition describes the support language of the coupling. The one-step case is then used to define the transition kernel below, because the support-restricted channel is specified one queried action at a time.

Definition 10 (Realisable action continuations)**.** For an ordinary history $h_{:t} \in \mathcal{H}_{:t}$ and $k \in \mathbb{N}^+$, the set of k-step realisable action continuations is

$$\mathcal{A}_k(h_{:t}) \coloneqq \left\{ a_{t:t+k} \in \mathcal{A}^k \;\middle|\; \exists o_{t:t+k} \in \mathcal{O}^k : \Pr\big((A,O)_{t:t+k} = (a,o)_{t:t+k} \mid H_{:t} = h_{:t}\big) > 0 \right\}. \quad (8)$$

An infinite future action continuation is realisable, in the cylinder-support sense, when all of its finite prefixes are realisable: $a_{t:} \in \mathcal{A}_\infty(h_{:t}) \iff a_{t:t+k} \in \mathcal{A}_k(h_{:t})$ for all $k \in \mathbb{N}^+$. The one-step realisable action set, used in the transition kernel below, is the case $k = 1$:

$$\mathcal{A}_1(h_{:t}) = \left\{ a \in \mathcal{A} \;\middle|\; \exists o \in \mathcal{O} : \Pr\big((A,O)_t = (a,o) \mid H_{:t} = h_{:t}\big) > 0 \right\}. \quad (9)$$

Equivalently, with the marginal $q(a \mid h_{:t}) \coloneqq \sum_{o \in \mathcal{O}} \Pr\big((A,O)_t = (a,o) \mid H_{:t} = h_{:t}\big)$, we have $a \in \mathcal{A}_1(h_{:t}) \iff q(a \mid h_{:t}) > 0$.

The finite-horizon sets $\mathcal{A}_k(h_{:t})$ describe the future action support induced by the realised coupling. The one-step set $\mathcal{A}_1(h_{:t})$ is enough for the recursive transducer definition: after a supported action and an ordinary observation, support is recomputed at the next step. If a queried future action sequence first leaves this support, the totalised channel emits $\perp$ at that step and then remains in the sink. For $k = 1$, this recovers the usual on-policy action support when the realised joint process is generated by a fixed policy coupled to an environment.

Example 11 (On-policy one-step actions)**.** Suppose the realised joint process is generated by coupling an environment with a policy $\pi(a \mid h_{:t}) = \Pr(A_t = a \mid H_{:t} = h_{:t})$. If the one-step joint law factors as $\Pr\big((A,O)_t = (a,o) \mid H_{:t} = h_{:t}\big) = \pi(a \mid h_{:t}) \Pr\big(O_t = o \mid H_{:t} = h_{:t}, A_t = a\big)$, then

$$\mathcal{A}_1(h_{:t}) = \{ a \in \mathcal{A} \mid \pi(a \mid h_{:t}) > 0 \text{ and } \exists o \in \mathcal{O} : \Pr(O_t = o \mid H_{:t} = h_{:t}, A_t = a) > 0 \}. \quad (10)$$

In the usual case where the environment produces some observation with nonzero probability after every action in the support of π, this reduces to

$$\mathcal{A}_1(h_{:t}) = \operatorname{supp} \pi(\cdot \mid h_{:t}). \quad (11)$$

Thus, for a deterministic policy, the one-step realisable action set is the singleton containing the on-policy action: $\mathcal{A}_1(h_{:t}) = \{\pi(h_{:t})\}$.

Definition 12 (Support-restricted environment channel)**.** The support-restricted environment channel $\widetilde{O}_{:} \mid A_{:}$ is defined on extended histories $\widetilde{h}_{:t} \in \widetilde{\mathcal{H}}_{:t}$ as follows:

1. if $\widetilde{h}_{:t}$ contains an occurrence of $\perp$, then for every $a \in \mathcal{A}, \Pr\big(\widetilde{O}_t = \perp \mid A_t = a, \widetilde{H}_{:t} = \widetilde{h}_{:t}\big) = 1$.

2. if $\widetilde{h}_{:t}$ is ordinary, identified with $h_{:t} \in \mathcal{H}_{:t}$, then for every $a \in \mathcal{A}$ and $\widetilde{o} \in \widetilde{\mathcal{O}}$,

$$\Pr\big(\widetilde{O}_t = \widetilde{o} \mid A_t = a, \widetilde{H}_{:t} = h_{:t}\big) := \begin{cases} \dfrac{\Pr\big((A,O)_t = (a,o) \mid H_{:t} = h_{:t}\big)}{\sum_{\bar{o} \in \mathcal{O}} \Pr\big((A,O)_t = (a,\bar{o}) \mid H_{:t} = h_{:t}\big)}, & \text{if } \widetilde{o} = o \in \mathcal{O} \text{ and } a \in \mathcal{A}_1(h_{:t}), \\ 1, & \text{if } \widetilde{o} = \perp \text{ and } a \notin \mathcal{A}_1(h_{:t}), \\ 0, & \text{otherwise.} \end{cases}$$

Proposition 13 (Totalisation of the partial channel)**.** *The channel of definition 12 is the totalisation of the partial support-restricted environment channel induced by the realised joint process. On realisable actions it agrees with the ordinary conditional observation law induced by the joint process. On unrealisable actions it emits $\perp$, and after $\perp$ has occurred all future outputs are $\perp$.*

Proof. This follows directly from the two clauses of definition 12. □

We now define the canonical predictive states of this channel in the standard way. For extended histories $\widetilde{h}_{:t}, \widetilde{h}'_{:t} \in \widetilde{\mathcal{H}}_{:t}$, write

$$\widetilde{h}_{:t} \sim_{\epsilon_{\mathrm{sr}-E}} \widetilde{h}'_{:t} \iff \Pr\big(\widetilde{O}_{t:} \mid A_{t:}, \widetilde{H}_{:t} = \widetilde{h}_{:t}\big) = \Pr\big(\widetilde{O}_{t:} \mid A_{t:}, \widetilde{H}_{:t} = \widetilde{h}'_{:t}\big). \tag{12}$$

Let $\epsilon_{\mathrm{sr}-E}(\widetilde{h}_{:t}) := [\widetilde{h}_{:t}]_{\sim_{\epsilon_{\mathrm{sr}-E}}}$ and $\mathcal{S}^{\mathrm{sr}-E} := \widetilde{\mathcal{H}}_{:t} / \sim_{\epsilon_{\mathrm{sr}-E}}$. The induced transition matrices are

$$T^{\mathrm{sr}-E,(\widetilde{o}|a)}_{s \to s'} := \Pr\big(S^{\mathrm{sr}-E}_{t+1} = s', \widetilde{O}_t = \widetilde{o} \mid S^{\mathrm{sr}-E}_t = s, A_t = a\big). \tag{13}$$

Definition 14 (Canonical support-restricted environment model)**.** The canonical support-restricted environment model is the ϵ-transducer $(\mathcal{A}, \widetilde{\mathcal{O}}, \mathcal{S}^{\mathrm{sr}-E}, \mathcal{T}^{\mathrm{sr}-E})$ of the support-restricted environment channel $\widetilde{O}_{:} \mid A_{:}$.

This ϵ-transducer always contains a *sink state*, which is the equivalence class of all extended histories containing at least one occurrence of $\perp$, generated by an unsupported action query.

Proposition 15 (The totalisation sink)**.** *All extended histories containing at least one occurrence of $\perp$ belong to the same causal state, denoted by $s^E_\perp$. It is absorbing, and for every $a \in \mathcal{A}$, $T^{\mathrm{sr}-E,(\perp|a)}_{s^E_\perp \to s^E_\perp} = 1$.*

Proof. See appendix B. □

The canonical support-restricted agent model is defined dually, by restricting future observation continuations to those supported by the realised coupling and totalising unsupported observation queries with an agent-side sink. Interestingly, the restrictions of future actions for the environment and of future observations for the agent model, which use information from the joint model, lead to some interesting relations among these models, as proven next.

3.1 Relations among canonical support-restricted models and the canonical joint model

Here we look at the relations among these canonical support-restricted models and the canonical joint model, proving a result that cannot be obtained for their unrestricted counterparts without extra structure (e.g. a factorisation assumptions) as seen in appendix D. First we show that the non-sink states of the canonical support-restricted environment are induced by the causal states of the joint process.

Theorem 16 (Canonical support-restricted environment states factor through the joint ϵ-machine causal states)**.** *Let $\epsilon_J : \mathcal{H}_{:t} \to \mathcal{S}^J$ be the causal-state map of the canonical joint model. If ordinary histories $h_{:t}, h'_{:t} \in \mathcal{H}_{:t}$ satisfy $\epsilon_J(h_{:t}) = \epsilon_J(h'_{:t})$, this implies that $\epsilon_{\mathrm{sr}-E}(h_{:t}) = \epsilon_{\mathrm{sr}-E}(h'_{:t})$. In other*

words, there exists a unique surjective map $\psi : \mathcal{S}^J \to \mathcal{S}^{\mathrm{sr}-E} \setminus \{s^E_\perp\}$ *such that for every ordinary history* $h_{:t} \in \mathcal{H}_{:t}$, $\epsilon_{\mathrm{sr}-E}(h_{:t}) = \psi\big(\epsilon_J(h_{:t})\big)$, *i.e.*

$$\begin{array}{ccc} \mathcal{H}_{:t} & \xrightarrow{\ \epsilon_J\ } & \mathcal{S}^J \\ & \searrow^{\epsilon_{\mathrm{sr}-E}} & \downarrow \psi \\ & & \mathcal{S}^{\mathrm{sr}-E} \setminus \{s^E_\perp\} \end{array} \tag{14}$$

Proof. See appendix B. □

This gives $|\mathcal{S}^{\mathrm{sr}-E} \setminus \{s^E_\perp\}| \le |\mathcal{S}^J|$, i.e. $|\mathcal{S}^{\mathrm{sr}-E}| \le |\mathcal{S}^J| + 1$ considering the totalisation sink. Thus the ordinary canonical support-restricted environment states are quotients of joint causal states. The transition structure is similarly also induced from the canonical joint model.

Theorem 17 (Transitions induced from the canonical joint model)**.** *Let* $s \in \mathcal{S}^{\mathrm{sr}-E} \setminus \{s^E_\perp\}$ *and choose any* $s^J \in \psi^{-1}(s)$. *Write* $p^J(a, o \mid s^J) \coloneqq \Pr((A, O)_t = (a, o) \mid S^J_t = s^J)$, *and define the marginal* $q(a \mid s^J) \coloneqq \sum_{o \in \mathcal{O}} p^J(a, o \mid s^J)$. *For* $p^J(a, o \mid s^J) > 0$, *let* $\delta^J(s^J, a, o)$ *denote the unique successor joint state given by unifilarity. The transitions of the canonical support-restricted environment model are given, independent of the choice of representative* $s^J \in \psi^{-1}(s)$, *by:*

1. *If* $q(a \mid s^J) = 0$, $T^{\mathrm{sr}-E,(\perp|a)}_{s \to s^E_\perp} = 1$ *and* $T^{\mathrm{sr}-E,(o|a)}_{s \to s'} = 0$ *for all* $o \in \mathcal{O}$ *and* $s' \in \mathcal{S}^{\mathrm{sr}-E}$.

2. *If* $q(a \mid s^J) > 0$, $T^{\mathrm{sr}-E,(\perp|a)}_{s \to s'} = 0$ *for all* $s' \in \mathcal{S}^{\mathrm{sr}-E}$ *and for all* $o \in \mathcal{O}$,

$$T^{\mathrm{sr}-E,(o|a)}_{s \to s'} = \begin{cases} \dfrac{p^J(a, o \mid s^J)}{q(a \mid s^J)} \mathbf{1}\big\{s' = \psi\big(\delta^J(s^J, a, o)\big)\big\}, & \text{if } p^J(a, o \mid s^J) > 0, \\ 0, & \text{if } p^J(a, o \mid s^J) = 0. \end{cases}$$

3. *For the sink state, for all* $a \in \mathcal{A}$, $T^{\mathrm{sr}-E,(\perp|a)}_{s^E_\perp \to s^E_\perp} = 1$. *and* $T^{\mathrm{sr}-E,(o|a)}_{s^E_\perp \to s'} = 0$ *for all* $o \in \mathcal{O}$ *and* $s' \in \mathcal{S}^{\mathrm{sr}-E}$.

Proof. See appendix B. □

At this point the reconstruction has been proved component by component, but not yet stated as a model-level consequence. Theorem 16 gives the non-sink state space from the joint causal states, theorem 17 shows that the non-sink transition probabilities are functions of joint-state data, and proposition 15 supplies the remaining sink state and its transitions. The next theorem packages these pieces into the claim that the canonical joint model determines the entire canonical support-restricted environment model.

Theorem 18 (Joint reconstruction of the canonical support-restricted environment model)**.** *The canonical joint model determines the canonical support-restricted environment model* $(\mathcal{A}, \widetilde{\mathcal{O}}, \mathcal{S}^{\mathrm{sr}-E}, \mathcal{T}^{\mathrm{sr}-E})$ *of definition 14. More explicitly, the interface fixes* $\mathcal{A}$, *the totalisation construction fixes* $\widetilde{\mathcal{O}} = \mathcal{O} \cup \{\perp\}$, *theorem 16 and proposition 15 determine the state space, and theorem 17 and proposition 15 determine the transition family.*

Proof. See appendix B. □

The agent-side reconstruction is dual: exchanging actions and observations gives the corresponding map from joint causal states to non-sink support-restricted agent states and the induced transition structure.

The abstract construction above changes the future action continuations against which environment histories are compared. To see this in practical terms, we return to the running example makes this concrete: the ordinary futures retained by the support-restricted environment model are exactly those compatible with the controller's finite pattern of resets and holds.

	Environment side	Agent side
All future continuations	**Canonical environment model** Channel: $O_{:} \mid A_{:}$ Predicts future observations under arbitrary future action continuations.	**Canonical agent model** Channel: $A_{:} \mid O_{:}$ Predicts future actions under arbitrary future observation continuations.
Realisable continuations only	**Canonical support-restricted environment model** Channel: $\widetilde{O}_{:} \mid A_{:}$ Predicts future observations under action continuations supported by the realised coupling, unsupported action queries emit an environment-side $\perp$.	**Canonical support-restricted agent model** Channel: $\widetilde{A}_{:} \mid O_{:}$ *(dually)* Predicts future actions under observation continuations supported by the realised coupling, unsupported observation queries emit an agent-side $\perp$.

Table 2: Environment and agent models along two axes. The horizontal axis distinguishes which channel is modelled. The vertical axis distinguishes whether predictive equivalence ranges over all future input continuations or only continuations supported by the realised coupling.

Support restriction in the running example. Recall (example 9) that the unrestricted environment model is infinite because arbitrary hold continuations allow unbounded accumulation of evidence about the latent environment state. The support-restricted construction considers a different object: the environment model induced by the actual coupling to the finite-state controller. In this controller, action 0 is realisable in mode α, while action 1 is realisable in modes β and γ. Thus, querying action 1 from histories in mode α, or querying action 0 from histories in modes β or γ, is unsupported and is sent to the sink state. The important mechanism is that the controller bounds hold continuations: after at most two hold actions, it returns to reset. Therefore the canonical support-restricted environment model never has to represent arbitrarily long hold continuations as ordinary futures. The realised joint process has five causal states, see appendix C. The non-sink canonical support-restricted environment states are their images under the factorisation map ψ and are also five, see appendix C. Together with the totalisation sink $s^E_\perp$, this gives six canonical support-restricted environment states in the example.

This illustrates theorem 18. The example-specific mechanism is that the controller bounds the supported hold continuations. The general point is that the ordinary canonical support-restricted environment states are induced from the joint causal states, while unsupported action queries are recorded by the sink. Thus the same setup can have an infinite unrestricted environment model but a finite canonical support-restricted environment model induced by the realised coupling.

4 Related work

Most computational-mechanics treatments of world models in ML focus on the environment channel, giving machine or transducer presentations of environment-side predictive structure [54, 50, 8, 49, 47, 44, 53]. This connects to observable operator models [26, 60], predictive state representations [34, 58], belief MDPs and bisimulation-style abstractions [50], while ϵ-transducers give the canonical minimal unifilar presentation of the environment channel [6]. AIXI also provides a Bayesian environment-model perspective for exhaustive planning [24], and information-theoretic quantities such as empowerment [29] and directed-information/plasticity constructions [2] similarly depend on which agent-environment channel is being considered. Models of the environment are often simply called world models [20, 15]. We however separate these from agent and joint models. Agent-side predictive models are related to self-predictive universal AI [11], policy distillation [52], plasticity as the dual of empowerment [2], and habit-formation models predicting actions from sensorimotor histories [17, 16]. In partially observable RL, policies can be viewed as maps from histories to actions [40], while finite-state controllers give explicit stateful presentations of such policies and are often used to build or interpret recurrent policies [36, 30]; our agent model is the corresponding canonical predictive-state construction for the channel $A_{:} \mid O_{:}$. Joint models relate to embedded agency [14, 37], coarse-grainings between autonomous systems and open components, Bayesian inversions based on the internal model principle [3, 5], and computationally embedded settings such as universal-local environments [33]. Finally, support restriction is related to the distinction between arbitrary counterfactual prediction and prediction under policy-, dataset-, controller-, or coupling-induced support. On the environment side this is analogous to on-/off-policy and offline model learning, where data may come from older policies, experts, teachers, or other sources that the

current agent cannot reproduce [61, 51, 43, 32], and to adaptive offline RL outside the original data support [19]. It also resonates with π-bisimulations or on-policy bisimulations [10], which contrast with standard bisimulation abstractions that quantify over all actions rather than only actions realised by a policy.

5 Conclusions

In this work we introduced a formal characterisation of world models according to the channel they model: the environment channel, the agent channel, or the realised joint process viewed as a null-input channel. Using tools from computational mechanics, we defined canonical models that are unique, minimal, and unifilar. On the environment side, this connects to predictive state representations and belief-MDP-style abstractions, on the agent side, to action-predictive models, policy distillation, and finite-state controllers, and on the joint side, to policy-induced processes and embedded-agency perspectives.

By considering the realised joint process, we then defined canonical support-restricted environment and agent models induced by closed-loop coupling. These models differ from ordinary canonical environment or agent models because their predictive equivalences are defined relative to continuations supported by the coupling, with unsupported queries handled by totalisation. On the environment side, we proved that the non-sink canonical support-restricted states factor through the joint causal states, and that the transition structure is induced from the joint model. The example illustrates that the same finite system can have an infinite unrestricted environment model but a finite canonical support-restricted environment model.

This approach can be further extended to formalise situations when, for instance, an agent considers agents other than itself, such as in behavioural cloning, or joint models of a different agent and a different environment. An interesting avenue for future work is the application of our definitions to the analysis of trained AI models, particularly when we move past the idealised scenarios of canonical models treated here and consider what *beliefs* can be attributed to an AI system. This can be achieved, for instance, using mixed-state presentations of both machines [48] and transducers [50], which would yield agent- and joint-side belief-based counterparts of environment models such as belief MDPs that could be used to identify more structure in, e.g. transformers [54, 49, 47, 44, 53].

Acknowledgments

Redacted for double blind peer review.

The authors thank Martin Biehl and Scott Garrabrant for inspiring discussions and useful feedback. M.B. was supported by Advanced Research + Invention Agency (ARIA) through project code MSAI-SE01-P011.

References

[1] D. Abel. A Theory of Abstraction in Reinforcement Learning, Mar. 2022.

[2] D. Abel, M. Bowling, A. Barreto, W. Dabney, S. Dong, S. Hansen, A. Harutyunyan, K. Khetarpal, C. Lyle, R. Pascanu, G. Piliouras, D. Precup, J. Richens, M. Rowland, T. Schaul, and S. Singh. Plasticity as the Mirror of Empowerment, Oct. 2025.

[3] M. Baltieri, M. Biehl, M. Capucci, and N. Virgo. A Bayesian Interpretation of the Internal Model Principle, Mar. 2025.

[4] M. Baltieri and C. L. Buckley. On Kalman-Bucy filters, linear quadratic control and active inference, May 2020.

[5] M. Baltieri and K. Suzuki. Mathematical approaches to the study of agents, 2025.

[6] N. Barnett and J. P. Crutchfield. Computational Mechanics of Input–Output Processes: Structured Transformations and the \textit{ε}-Transducer. *Journal of Statistical Physics*, 161(2):404–451, Oct. 2015.

[7] L. Bereska and E. Gavves. Mechanistic Interpretability for AI Safety – A Review, Aug. 2024.

[8] A. Boyd, F. Nowak, D. Hyland, M. Baltieri, and F. E. Rosas. From monoliths to modules: Decomposing transducers for efficient world modelling, Dec. 2025.

[9] C. L. Buckley and T. Toyoizumi. A theory of how active behavior stabilises neural activity: Neural gain modulation by closed-loop environmental feedback. *PLoS Computational Biology*, 14(1):e1005926, 2018.

[10] P. S. Castro. Scalable Methods for Computing State Similarity in Deterministic Markov Decision Processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10069–10076, Apr. 2020.

[11] E. Catt, J. Grau-Moya, M. Hutter, M. Aitchison, T. Genewein, G. Delétang, K. Li, and J. Veness. Self-Predictive Universal AI. *Advances in Neural Information Processing Systems*, 36:27181–27198, Dec. 2023.

[12] J. P. Crutchfield. The Calculi of Emergence: Computation Dynamics, and Induction The Calculi of Emergence: Computation, Dynamics, and Induction. *Physica D*, 1994.

[13] D. d. Dalrymple, J. Skalse, Y. Bengio, S. Russell, M. Tegmark, S. Seshia, S. Omohundro, C. Szegedy, B. Goldhaber, N. Ammann, A. Abate, J. Halpern, C. Barrett, D. Zhao, T. Zhi-Xuan, J. Wing, and J. Tenenbaum. Towards Guaranteed Safe AI: A Framework for Ensuring Robust and Reliable AI Systems, May 2024.

[14] A. Demski and S. Garrabrant. Embedded Agency, Oct. 2020.

[15] J. Ding, Y. Zhang, Y. Shang, Y. Zhang, Z. Zong, J. Feng, Y. Yuan, H. Su, N. Li, N. Sukiennik, F. Xu, and Y. Li. Understanding World or Predicting Future? A Comprehensive Survey of World Models. *ACM Computing Surveys*, 58(3):57:1–57:38, Sept. 2025.

[16] M. Egbert and L. Canamero. Habit-based regulation of essential variables. In *Artificial Life Conference Proceedings*, pages 168–175. MIT Press, 2014.

[17] M. D. Egbert and X. E. Barandiaran. Modeling habits as self-sustaining patterns of sensorimotor behavior. *Frontiers in Human Neuroscience*, 8, Aug. 2014.

[18] L. J. Fiderer, P. C. Barth, I. D. Smith, and H. J. Briegel. The Work Capacity of Channels with Memory: Maximum Extractable Work in Percept-Action Loops, Apr. 2025.

[19] D. Ghosh, A. Ajay, P. Agrawal, and S. Levine. Offline RL Policies Should Be Trained to be Adaptive. In *Proceedings of the 39th International Conference on Machine Learning*, pages 7513–7530. PMLR, June 2022.

[20] D. Ha and J. Schmidhuber. World Models, Mar. 2018.

[21] N. Haber, D. Mrowca, S. Wang, L. F. Fei-Fei, and D. Yamins. Learning to Play With Intrinsically-Motivated, Self-Aware Agents. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[22] D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640(8059):647–653, Apr. 2025.

[23] R. Held and A. Hein. Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 56(5):872–876, 1963.

[24] M. Hutter. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Texts in Theoretical Computer Science. Springer, Berlin ; New York, 2005.

[25] B. Jacobs. *Introduction to Coalgebra: Towards Mathematics of States and Observation*. Cambridge University Press, 1 edition, 2017.

[26] H. Jaeger. Discrete-time, discrete-valued observable operator models: A tutorial. Technical report, International University Bremen, 1998.

[27] A. Kanervisto, D. Bignell, L. Y. Wen, M. Grayson, R. Georgescu, S. Valcarcel Macua, S. Z. Tan, T. Rashid, T. Pearce, Y. Cao, A. Lemkhenter, C. Jiang, G. Costello, G. Gupta, M. Tot, S. Ishida, T. Gupta, U. Arora, R. W. White, S. Devlin, C. Morrison, and K. Hofmann. World and Human Action Models towards gameplay ideation. *Nature*, 638(8051):656–663, Feb. 2025.

[28] H. J. Kappen, V. Gómez, and M. Opper. Optimal control as a graphical model inference problem. *Machine Learning*, 87(2):159–182, May 2012.

[29] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Empowerment: A universal agent-centric measure of control. In *Evolutionary Computation, 2005. The 2005 IEEE Congress On*, volume 1, pages 128–135. IEEE, 2005.

[30] A. Koul, A. Fern, and S. Greydanus. Learning Finite State Representations of Recurrent Policy Networks. In *International Conference on Learning Representations*, Sept. 2018.

[31] S. Levine. Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review, May 2018.

[32] S. Levine, A. Kumar, G. Tucker, and J. Fu. Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems, Nov. 2020.

[33] A. Lewandowski, A. A. Ramesh, E. Meyer, D. Schuurmans, and M. C. Machado. The World Is Bigger! A Computationally-Embedded Perspective on the Big World Hypothesis, Dec. 2025.

[34] M. Littman and R. S. Sutton. Predictive Representations of State. In *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2001.

[35] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, and J. Morimoto. Deep learning, reinforcement learning, and world models. *Neural Networks*, 152:267–275, Aug. 2022.

[36] N. Meuleau, L. Peshkin, K.-E. Kim, and L. P. Kaelbling. Learning finite-state controllers for partially observable environments. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, UAI'99, pages 427–436, San Francisco, CA, USA, July 1999. Morgan Kaufmann Publishers Inc.

[37] A. Meulemans, R. Nasser, M. Wołczyk, M. A. Weis, S. Kobayashi, B. Richards, G. Lajoie, A. Steger, M. Hutter, J. Manyika, R. A. Saurous, J. Sacramento, and B. A. y Arcas. Embedded Universal Predictive Intelligence: A coherent framework for multi-agent learning, Nov. 2025.

[38] B. Millidge, A. Tschantz, A. K. Seth, and C. L. Buckley. On the relationship between active inference and control as inference. In *International Workshop on Active Inference*, pages 3–11. Springer, 2020.

[39] P. Mineault, N. Zanichelli, J. Z. Peng, A. Arkhipov, E. Bingham, J. Jara-Ettinger, E. Mackevicius, A. Marblestone, M. Mattar, A. Payne, S. Sanborn, K. Schroeder, Z. Tavares, A. Tolias, and A. Zador. NeuroAI for AI Safety, Apr. 2025.

[40] K. Murphy. Reinforcement Learning: An Overview, Sept. 2025.

[41] D. J. Myers. *Categorical Systems Theory*. 2021.

[42] T. Ni, B. Eysenbach, E. Seyedsalehi, M. Ma, C. Gehring, A. Mahajan, and P.-L. Bacon. Bridging State and History Representations: Understanding Self-Predictive RL, Apr. 2024.

[43] G. Ostrovski, P. S. Castro, and W. Dabney. The Difficulty of Passive Learning in Deep Reinforcement Learning. In *Advances in Neural Information Processing Systems*, volume 34, pages 23283–23295. Curran Associates, Inc., 2021.

[44] M. Piotrowski, P. M. Riechers, D. Filan, and A. S. Shai. Constrained belief updates explain geometric structures in transformer representations, Oct. 2025.

[45] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 2014.

[46] B. Ravindran. *An Algebraic Approach to Abstraction in Reinforcement Learning*. PhD thesis, University of Massachusetts Amherst, 2004.

[47] P. M. Riechers, H. R. Bigelow, E. A. Alt, and A. Shai. Next-token pretraining implies in-context learning, July 2025.

[48] P. M. Riechers and J. P. Crutchfield. Spectral simplicity of apparent complexity. I. The nondiagonalizable metadynamics of prediction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(3), Mar. 2018.

[49] P. M. Riechers, T. J. Elliott, and A. S. Shai. Neural networks leverage nominally quantum and post-quantum representations, July 2025.

[50] F. Rosas, A. Boyd, and M. Baltieri. AI in a vat: Fundamental limits of efficient world modelling for agent sandboxing and interpretability. In *Proceedings of the Second Reinforcement Learning Conference*, pages 2844–2881, Apr. 2025.

[51] S. Ross, G. Gordon, and D. Bagnell. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, June 2011.

[52] A. A. Rusu, S. G. Colmenarejo, C. Gulcehre, G. Desjardins, J. Kirkpatrick, R. Pascanu, V. Mnih, K. Kavukcuoglu, and R. Hadsell. Policy Distillation, Jan. 2016.

[53] A. Shai, L. Amdahl-Culleton, C. L. Christensen, H. R. Bigelow, F. E. Rosas, A. B. Boyd, E. A. Alt, K. J. Ray, and P. M. Riechers. Transformers learn factored representations, Feb. 2026.

[54] A. S. Shai, S. E. Marzen, L. Teixeira, A. G. Oldenziel, and P. M. Riechers. Transformers represent belief state geometry in their residual stream, May 2024.

[55] C. R. Shalizi. Methods and Techniques of Complex Systems Science: An Overview. In T. S. Deisboeck and J. Y. Kresh, editors, *Complex Systems Science in Biomedicine*, pages 33–114. Springer US, Boston, MA, 2006.

[56] C. R. Shalizi and J. P. Crutchfield. Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *Journal of Statistical Physics*, 104(3):817–879, Aug. 2001.

[57] A. Silva, F. Bonchi, M. Bonsangue, and J. J. M. M. Rutten. Generalizing determinization from automata to coalgebras. *Logical Methods in Computer Science*, Volume 9, Issue 1:1087, Mar. 2013.

[58] S. Singh, M. R. James, and M. R. Rudary. Predictive state representations: A new theory for modeling dynamical systems. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pages 512–519, Arlington, Virginia, USA, July 2004. AUAI Press.

[59] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge, Massachusetts, second edition edition, 2018.

[60] M. Thon and H. Jaeger. Links between multiplicity automata, observable operator models and predictive state representations: A unified learning framework. *The Journal of Machine Learning Research*, 16(1):103–147, 2015.

[61] F. Torresan and M. Baltieri. Disentangled representations for causal cognition. *Physics of Life Reviews*, 51:343–381, Dec. 2024.

[62] L. Ullrich, W. Zimmer, R. Greer, K. Graichen, A. C. Knoll, and M. M. Trivedi. A New Perspective on AI Safety Through Control Theory Methodologies. *IEEE Open Journal of Intelligent Transportation Systems*, 6:938–966, 2025.

[63] A. Zhang, Z. C. Lipton, L. Pineda, K. Azizzadenesheli, A. Anandkumar, L. Itti, J. Pineau, and T. Furlanello. Learning Causal State Representations of Partially Observable Environments, Feb. 2021.

[64] X. Zhou, J. Liu, A. Yerukola, H. Kim, and M. Sap. Social World Models, 2025.

[65] M. Zhuge, C. Zhao, H. Liu, Z. Zhou, S. Liu, W. Wang, E. Chang, G. L. Lan, J. Fei, W. Zhang, Y. Sun, Z. Cai, Z. Liu, Y. Xiong, Y. Yang, Y. Tian, Y. Shi, V. Chandra, and J. Schmidhuber. Neural Computers, 2026.

A Computational mechanics: a brief overview

A.1 Canonical models

A.1.1 ϵ-machines

Initially, we consider a univariate bi-infinite discrete-time stochastic process of the form $X_{:} = \{X_t\}_{t\in\mathbb{Z}}$, on a finite alphabet $\mathcal{X}$. A stochastic process is specified by a consistent family of finite-dimensional distributions

$$\Pr(X_{t:t+L} = x_{t:t+L}), \qquad x_{t:t+L} \in \mathcal{X}^L, \tag{15}$$

for all $t \in \mathbb{Z}$ and $L \in \mathbb{N}^+$, where $\mathcal{X}^L$ is the set of words of length L. Equivalently, a consistent family of finite-dimensional distributions specifies probabilities for cylinder events in the space of bi-infinite sequences $\mathcal{X}_{:}$. For a finite word $x_{t:t+L} \in \mathcal{X}^L$, the corresponding cylinder is

$$[x_{t:t+L}] \coloneqq \{x'_{:} \in \mathcal{X}_{:} \mid x'_{t:t+L} = x_{t:t+L}\}.$$

The finite-dimensional probability $\Pr(X_{t:t+L} = x_{t:t+L})$ is then the probability assigned to this cylinder. Under the usual consistency conditions, these cylinder probabilities determine a probability measure on $\mathcal{X}_{:}$.

A stochastic process $X_{:}$ is stationary if its finite-dimensional distributions are invariant under time shifts, i.e. $\Pr(X_{t:t+L} = x_{0:L}) = \Pr(X_{0:L} = x_{0:L})$ for all $t \in \mathbb{Z}$, $L \in \mathbb{N}^+$, and $x_{0:L} \in \mathcal{X}^L$. From here onwards, we only consider stationary processes. A presentation of this process can be given in terms of a *machine*.

Definition 19 (Machine)**.** A machine presentation of $X_{:}$ is given by a triple $(\mathcal{X}, \mathcal{Z}, \mathcal{T})$ where $\mathcal{Z}$ is a state space and $\mathcal{T}$ are transition dynamics by Markov kernels (or stochastic matrices in the finite case) $\mathcal{T} = \{T^{(x)}\}_{x\in\mathcal{X}}$ given by:

$$\mathcal{T} = \{T^{(x)}\}_{x\in\mathcal{X}}, \quad T^{(x)}_{z\to z'} \coloneqq \Pr(Z_{t+1} = z', X_t = x \mid Z_t = z). \tag{16}$$

When the process is time-independent, we simply write them as $T_{s\to s'} = T^{(x)}_{s\to s'}$.

Intuitively, computational mechanics gives us a way to construct particularly interesting machine presentations of a process. Informally, state variables are defined by forming a latent state space for a stationary stochastic system $X_{:}$ by quotienting histories of $X_{:}$ into equivalence classes that have identical predictive consequences. More formally, we define pasts and futures of $X_{:}$ as $X_{:t}$ and $X_{t:}$ for all $t \in \mathbb{Z}$ respectively, and an equivalence relation $\sim_\epsilon$ given by:

$$x_{:t} \sim_\epsilon x'_{:t} \iff \Pr(X_{t:} \mid X_{:t} = x_{:t}) = \Pr(X_{t:} \mid X_{:t} = x'_{:t}) \tag{17}$$

The equivalence classes induced by $\sim_\epsilon$ give the so-called *causal states*, forming a set $\mathcal{S}$, with associated random variables S_t. This relation can equivalently be seen as induced by a surjective map $\epsilon : \mathcal{X}_{:t} \to \mathcal{S}$ given by $\epsilon(x_{:t}) = s = \{x'_{:t} \mid x_{:t} \sim_\epsilon x'_{:t}\}$, with states $s \in \mathcal{S}$. Causal states are minimal sufficient statistics for prediction [56]. Sufficient means that $X_{t:} \perp\!\!\!\perp X_{:t} \mid S_t$ or $\Pr(X_{t:} \mid X_{:t}) = \Pr(X_{t:} \mid S_t)$, and minimal that, for any other sufficient statistic Z_t, Z_t is a refinement of S_t. The ϵ-map induces dynamics over causal states: given a realised past $x_{:t}$ and the next emitted symbol x_t, the extended past $x_{:t+1} = (x_{:t}, x_t)$ determines the next causal state $\epsilon(x_{:t+1})$. The corresponding transition kernels $\mathcal{T} = \{T^{(x)}\}_{x\in\mathcal{X}}$ are given by:

$$T^{(x)}_{s\to s'} = \Pr(S_{t+1} = s', X_t = x \mid S_t = s). \tag{18}$$

Definition 20 (ϵ-machine [56])**.** The ϵ-machine of a stochastic process $X_{:}$ is the process's unique, minimal, and unifilar (machine) presentation of the process, given by the tuple $(\mathcal{X}, \mathcal{S}, \mathcal{T})$.

A presentation is unifilar if and only if, for every state $s \in \mathcal{S}$ and symbol $x \in \mathcal{X}$, there is at most one next state s' with nonzero probability, i.e. $\forall s, x, |\{s' : T^{(x)}_{ss'} > 0\}| \leq 1$, such that $H[S_{t+1} \mid X_t, S_t] = 0$. In other words, a unifilar presentation includes transitions that are deterministic given the current state and the emitted symbol, even though the symbol itself is generally random.

A.1.2 ϵ-transducers

The above treatment has also been extended to input-output processes, or simply *channels*. Channels can be seen as defining input-output couplings between stochastic processes. We consider univariate bi-infinite discrete-time channels, $Y_: \mid X_:$, with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. A channel assigns to each input sequence $x_: \in \mathcal{X}_:$ a probability law over output sequences in $\mathcal{Y}_:$. We write this family of output laws as

$$Y_: \mid X_: \coloneqq \{Y_: \mid x_:\}_{x_: \in \mathcal{X}_:}, \tag{19}$$

where $Y_: \mid x_:$ denotes the output stochastic process induced by the fixed input sequence $x_:$. For each fixed $x_:$, the corresponding finite-dimensional laws are

$$\Pr(Y_{t:t+L} \mid x_:) \coloneqq \{\Pr(Y_{t:t+L} = y_{t:t+L} \mid x_:)\}_{y_{t:t+L} \in \mathcal{Y}^L}, \tag{20}$$

for all $t \in \mathbb{Z}$ and $L \in \mathbb{N}^+$. Here $x_:$ is the input sequence supplied to the channel, not necessarily an event of positive probability. Equivalently, for each fixed $x_:$, the channel specifies a probability measure over output sequences,

$$\Pr(Y_: \mid x_:) \coloneqq \{\Pr(Y_: \in U \mid x_:)\}_{U \subseteq \mathcal{Y}_:}. \tag{21}$$

If an input process $X_:$ is also specified, the channel together with the law of $X_:$ induces a joint process $(X, Y)_:$, which can be marginalised to obtain an output process $Y_:$.

A channel $Y_: \mid X_:$ is stationary if its finite-dimensional laws are invariant under simultaneous shifts of input and output. That is, for all $t \in \mathbb{Z}$, $L \in \mathbb{N}^+$, $x_: \in \mathcal{X}_:$, and $y \in \mathcal{Y}^L$,

$$\Pr(Y_{t:t+L} = y \mid x_:) = \Pr(Y_{0:L} = y \mid (x_{t+z})_{z \in \mathbb{Z}}), \tag{22}$$

where the right-hand side uses the same input sequence, shifted so that the symbols around time t are treated as the symbols around time 0.

Stationary channels maps stationary input processes into output stationary processes [6]. A channel is causal, or anticipation-free, if future inputs beyond the output horizon do not affect the output law. Equivalently, for all $t \in \mathbb{Z}$, $L \in \mathbb{N}^+$, $y \in \mathcal{Y}^L$, and input sequences $x_:, x'_: \in \mathcal{X}_:$,

$$x_{:t+L} = x'_{:t+L} \implies \Pr(Y_{t:t+L} = y \mid x_:) = \Pr(Y_{t:t+L} = y \mid x'_:). \tag{23}$$

In this work we focus on stationary causal channels.

We can also give a different presentation of this channel using a *transducer*, as below.

Definition 21 (Transducer [50])**.** A transducer presentation of $Y_: \mid X_:$ is given by a quadruple $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{T})$ where $\mathcal{Z}$ is a state space and $\mathcal{T}$ are transition dynamics by Markov kernels (or stochastic matrices in the finite case) $\mathcal{T} = \{T^{(y|x)}\}_{x \in \mathcal{X}, y \in \mathcal{Y}}$ given by:

$$T^{(y|x)}_{z \to z'} = \Pr(Z_{t+1} = z', Y_t = y \mid Z_t = z, X_t = x). \tag{24}$$

We then define an equivalence relation $\sim_\epsilon$ of pasts of a channel $Y_: \mid X_:$ as:

$$(x, y)_{:t} \sim_\epsilon (x, y)'_{:t} \iff \Pr(Y_{t:} \mid X_{t:}, (X, Y)_{:t} = (x, y)_{:t}) = \Pr(Y_{t:} \mid X_{t:}, (X, Y)_{:t} = (x, y)'_{:t}) \tag{25}$$

which gives a channel's causal states. This relation can also be seen as induced by a surjective map $\epsilon : (\mathcal{X}, \mathcal{Y})_{:t} \to \mathcal{S}$ given by $\epsilon((x, y)_{:t}) = s = \{(x, y)'_{:t} \mid (x, y)_{:t} \sim_\epsilon (x, y)'_{:t}\}$. These causal states are also minimal sufficient statistics for prediction of future outputs [6]. The ϵ-map induces dynamics over causal states captured by stochastic matrices $\mathcal{T} = \{T^{(y|x)}\}_{x \in \mathcal{X}, y \in \mathcal{Y}}$ given by:

$$T^{(y|x)}_{s \to s'} = \Pr(S_{t+1} = s', Y_t = y \mid S_t = s, X_t = x). \tag{26}$$

Definition 22 (ϵ-transducer [6])**.** The ϵ-transducer of a channel $Y_: \mid X_:$ is the channel's unique, minimal, unifilar (transducer) presentation of the channel, given by the tuple $(\mathcal{X}, \mathcal{Y}, \mathcal{S}, \mathcal{T})$.

B Proofs

Proof of proposition 15. If $\widetilde{h}_{:t}$ contains an occurrence of $\perp$, then by the definition of the channel, $\Pr\big(\widetilde{O}_{t:} = \perp^{\infty} \mid A_{t:}, \widetilde{H}_{:t} = \widetilde{h}_{:t}\big) = 1$. Hence any two such histories induce the same future morph (i.e. conditional distribution of futures) and therefore lie in the same causal state. The displayed transition identity follows immediately from the same observation. □

Proof of theorem 16. Let $h_{:t}, h'_{:t} \in \mathcal{H}_{:t}$ be ordinary histories such that $\epsilon_J(h_{:t}) = \epsilon_J(h'_{:t})$. By joint causal equivalence (eq. (6)), for every $L \geq 1$ and every ordinary word $(a,o)_{t:t+L-1} \in (\mathcal{A} \times \mathcal{O})^L$,

$$\begin{aligned} &\Pr\big((A,O)_{t:t+L-1} = (a,o)_{t:t+L-1} \mid H_{:t} = h_{:t}\big) \\ &\quad = \Pr\big((A,O)_{t:t+L-1} = (a,o)_{t:t+L-1} \mid H_{:t} = h'_{:t}\big). \end{aligned} \tag{27}$$

We now show that this joint-cylinder equality implies equality of the finite-dimensional laws of the totalised support-restricted environment channel. Specifically, we prove by induction on L that, for, every input word $a_{t:t+L-1} \in \mathcal{A}^L$, and every output word $\widetilde{o}_{t:t+L-1} \in \widetilde{\mathcal{O}}^L$, the previous equality implies the following

$$\begin{aligned} &\Pr\big(\widetilde{O}_{t:t+L-1} = \widetilde{o}_{t:t+L-1} \mid A_{t:t+L-1} = a_{t:t+L-1}, \widetilde{H}_{:t} = h_{:t}\big) \\ &\quad = \Pr\big(\widetilde{O}_{t:t+L-1} = \widetilde{o}_{t:t+L-1} \mid A_{t:t+L-1} = a_{t:t+L-1}, \widetilde{H}_{:t} = h'_{:t}\big). \end{aligned} \tag{28}$$

Since the alphabets are finite, equality of all finite cylinders then implies equality of the full support-restricted future morphs.

For $L = 1$, fix $a \in \mathcal{A}$ and write

$$q_h(a) \coloneqq \sum_{o \in \mathcal{O}} \Pr\big((A,O)_t = (a,o) \mid H_{:t} = h_{:t}\big). \tag{29}$$

Joint causal equivalence gives equality of the corresponding one-step joint probabilities, and hence $q_h(a) = q_{h'}(a)$. If this common value is zero, then a is unsupported at both histories and the totalised channel emits $\perp$ with probability one at both histories. If it is positive, then a is supported at both histories and, for every $o \in \mathcal{O}$,

$$\Pr\big(\widetilde{O}_t = o \mid A_t = a, \widetilde{H}_{:t} = h_{:t}\big) = \frac{\Pr\big((A,O)_t = (a,o) \mid H_{:t} = h_{:t}\big)}{q_h(a)}. \tag{30}$$

The same formula holds with $h'_{:t}$ in place of $h_{:t}$, with the same numerator and denominator. In this supported case both histories emit $\perp$ with probability zero. This proves eq. (28) for $L = 1$.

Now assume the claim holds for length L for every pair of ordinary histories with the same joint causal state, and consider a word of length $L+1$ from the fixed histories above. If the first output symbol is $\perp$, then by totalisation all later output symbols must be $\perp$; otherwise the word has probability zero from both histories. If all later output symbols are $\perp$, the probability of the whole word is the common one-step probability of emitting $\perp$ under the first queried action, since after the first $\perp$ the sink emits $\perp$ with probability one.

It remains to consider the case where the first output symbol is an ordinary observation $o_t \in \mathcal{O}$. If the one-step probability of the pair (a_t, o_t) is zero, then the whole word has probability zero from both histories. Otherwise the same pair is realisable from both histories. Conditioning the equal joint future laws on this common positive-probability prefix gives

$$\epsilon_J(h_{:t}, a_t, o_t) = \epsilon_J(h'_{:t}, a_t, o_t). \tag{31}$$

The induction hypothesis applied to these extended ordinary histories gives equality of the conditional probabilities of the remaining output word $\widetilde{o}_{t+1:t+L}$ under the remaining input word $a_{t+1:t+L}$. By the chain rule, the probability of the whole length-$(L+1)$ word is the one-step probability of emitting o_t under input a_t multiplied by the corresponding remaining-word probability. The one-step factors are equal by the base case, and the remaining-word factors are equal by the induction hypothesis, so eq. (28) holds for length $L+1$.

This allows us to define a map from joint causal states to non-sink support-restricted environment states. Given a joint causal state s^J, choose any ordinary history $h_{:t}$ such that $\epsilon_J(h_{:t}) = s^J$, and

set $\psi(s^J) \coloneqq \epsilon_{\mathrm{sr}-E}(h_{:t})$. This definition is independent of the chosen representative: if another ordinary history has the same joint causal state, the result just proved implies that it has the same support-restricted environment causal state. Thus ψ is well defined. Moreover, every non-sink support-restricted environment state is the state of some ordinary history, and that ordinary history also has a joint causal state. Therefore every non-sink support-restricted environment state lies in the image of ψ, so ψ is surjective. Finally, the factorising map is unique. Suppose that ψ' is another map satisfying $\epsilon_{\mathrm{sr}-E} = \psi' \circ \epsilon_J$ on ordinary histories. Let $s^J \in \mathcal{S}^J$. Since ϵ_J is surjective, there is an ordinary history $h_{:t}$ with $\epsilon_J(h_{:t}) = s^J$. Then $\psi'(s^J) = \psi'(\epsilon_J(h_{:t})) = \epsilon_{\mathrm{sr}-E}(h_{:t}) = \psi(s^J)$. Since this holds for every s^J, we have $\psi' = \psi$. □

Proof of theorem 17. The first point to check before reading the formulas from the channel definition is that they do not depend on the representative $s^J \in \psi^{-1}(s)$. Let $s_1^J, s_2^J \in \psi^{-1}(s)$, and choose ordinary histories $h_{:t}^1, h_{:t}^2$ with $\epsilon_J(h_{:t}^i) = s_i^J$ for $i \in \{1, 2\}$. Since both joint states map to s, these histories have the same support-restricted environment causal state. Therefore the one-step output laws of the totalised channel $\widetilde{O}_{:} \mid A_{:}$ agree from the two histories under every queried action.

Write

$$p_i(a, o) \coloneqq p^J(a, o \mid s_i^J), \qquad q_i(a) \coloneqq \sum_{o \in \mathcal{O}} p_i(a, o). \tag{32}$$

By definitions 10 and 12, $q_i(a) = 0$ exactly when the queried action a is unsupported from $h_{:t}^i$, in which case the totalised channel emits $\perp$ with probability one. Since the one-step output laws agree, this unsupported case occurs for $i = 1$ if and only if it occurs for $i = 2$. Equivalently,

$$q_1(a) > 0 \iff q_2(a) > 0. \tag{33}$$

In the supported case, the same channel definition gives the ordinary output probabilities by normalising the joint one-step law:

$$\Pr\big(\widetilde{O}_t = o \mid A_t = a, \widetilde{H}_{:t} = h_{:t}^i\big) = \frac{p_i(a, o)}{q_i(a)}. \tag{34}$$

Equality of the one-step output laws therefore gives, for every $o \in \mathcal{O}$,

$$\frac{p_1(a, o)}{q_1(a)} = \frac{p_2(a, o)}{q_2(a)}. \tag{35}$$

In particular, $p_1(a, o) > 0$ if and only if $p_2(a, o) > 0$.

It remains to be checked that the successor state in the supported ordinary case also descends to the quotient. Suppose $p_1(a, o) > 0$, equivalently $p_2(a, o) > 0$, and let $h_{:t+1}^i \coloneqq (h_{:t}^i, a, o)$. Conditioning the equal future morphs on this common positive-probability prefix gives

$$\epsilon_{\mathrm{sr}-E}(h_{:t+1}^1) = \epsilon_{\mathrm{sr}-E}(h_{:t+1}^2). \tag{36}$$

By unifilarity of the canonical joint model,

$$\epsilon_J(h_{:t+1}^i) = \delta^J(s_i^J, a, o). \tag{37}$$

Applying ψ gives

$$\psi\big(\delta^J(s_1^J, a, o)\big) = \psi\big(\delta^J(s_2^J, a, o)\big). \tag{38}$$

Thus the branch condition, the ordinary output probabilities, and the successor support-restricted state are independent of the chosen representative.

The transition formula now follows directly from definition 12. If $q(a \mid s^J) = 0$, the queried action is unsupported, so the channel emits $\perp$ and moves to the sink with probability one. If $q(a \mid s^J) > 0$, the channel emits an ordinary observation o with probability

$$\Pr\big(\widetilde{O}_t = o \mid A_t = a, S_t^{\mathrm{sr}-E} = s\big) = \frac{p^J(a, o \mid s^J)}{q(a \mid s^J)}. \tag{39}$$

If $p^J(a, o \mid s^J) = 0$, the corresponding transition has probability zero; otherwise the successor state is $\psi\big(\delta^J(s^J, a, o)\big)$. The sink-state transitions are those of proposition 15. □

Proof of theorem 18. The input alphabet is the original action alphabet $\mathcal{A}$ of the joint interface. The output alphabet is the totalised observation alphabet $\widetilde{\mathcal{O}} = \mathcal{O} \cup \{\bot\}$ introduced in definition 12.

For the state space, theorem 16 identifies the non-sink states as the quotient of joint causal states given by the map $\psi : \mathcal{S}^J \to \mathcal{S}^{\mathrm{sr}-E} \setminus \{s^E_\bot\}$. Proposition 15 supplies the remaining state, the absorbing totalisation sink $s^E_\bot$. Since every extended history is either ordinary or contains an occurrence of $\bot$, these two parts give the whole state space $\mathcal{S}^{\mathrm{sr}-E} = \left(\mathcal{S}^{\mathrm{sr}-E} \setminus \{s^E_\bot\}\right) \cup \{s^E_\bot\}$.

It remains only to specify the transition family $\mathcal{T}^{\mathrm{sr}-E}$. For non-sink states, theorem 17 gives the transition probabilities for every queried action and every output in $\widetilde{\mathcal{O}}$, and proves that these probabilities are independent of the chosen joint-state representative. For the sink state, proposition 15 gives the absorbing transitions: every queried action emits $\bot$ and remains at $s^E_\bot$ with probability one. Thus all entries of $\mathcal{T}^{\mathrm{sr}-E}$ are determined.

These four components are exactly the ϵ-transducer $(\mathcal{A}, \widetilde{\mathcal{O}}, \mathcal{S}^{\mathrm{sr}-E}, \mathcal{T}^{\mathrm{sr}-E})$ of definition 14. □

C Running example

We spell out the running binary POMDP/controller example used in the main text. The example separates the unrestricted environment model, the agent model, the realised joint model, and the canonical support-restricted environment model. It also illustrates why the canonical support-restricted model induced by the realised coupling can be finite even when the unrestricted environment model is infinite.

Model	Number of causal states	Reason
Canonical unrestricted environment model	∞	Arbitrarily long counterfactual hold continuations generate infinitely many posterior beliefs over the latent environment state, and these beliefs are predictively distinct.
Canonical agent model	3	The deterministic controller has three pairwise predictively distinct memory modes, α, β, γ.
Canonical joint model	5	The realised coupling reaches five predictive configurations: s^J_α, $s^J_{\beta 1}$, $s^J_{\beta 0}$, $s^J_{\gamma 11}$, $s^J_{\gamma 01}$.
Canonical support-restricted environment model	6	The five ordinary states are the images of the joint states under ψ, together with the totalisation sink $s^E_\bot$.
Canonical support-restricted agent model	4	The environment has full observation support in this example, so support restriction does not further merge the controller's three action-predictive modes (+ 1 totalisation sink $s^M_\bot$).

Table 3: State-complexity summary for the running POMDP/controller example. The unrestricted environment model is infinite, while the realised joint model and the induced canonical support-restricted environment model are finite.

C.1 The setup: environment and agent

We consider a coupled system with binary actions and observations,

$$\mathcal{A} = \mathcal{O} = \{0, 1\}.$$

The environment has a hidden binary state $E_t \in \{0, 1\}$. The agent is a deterministic finite-state controller with memory state $M_t \in \{\alpha, \beta, \gamma\}$. At each time t, the agent emits A_t as a function of M_t, the environment updates E_{t+1} according to A_t, the environment emits O_t, and the controller updates M_{t+1} from (M_t, O_t). The observed joint symbol is

$$W_t := (A_t, O_t) \in \mathcal{A} \times \mathcal{O}.$$

For the environment, fix a sensor noise parameter $\eta \in (0, \frac{1}{2})$. The controlled transition is

$$\Pr(E_{t+1} = 1 \mid E_t = e, A_t = a) = \begin{cases} \frac{1}{2}, & a = 0, \\ e, & a = 1. \end{cases} \tag{40}$$

Thus action 0 resets the hidden state to a fair coin, while action 1 holds the hidden state fixed. The observation is a noisy readout of the new hidden state:

$$\Pr(O_t = 1 \mid E_{t+1} = 1) = 1 - \eta, \qquad \Pr(O_t = 1 \mid E_{t+1} = 0) = \eta. \tag{41}$$

The controller emits actions according to

$$A_t = \begin{cases} 0, & M_t = \alpha, \\ 1, & M_t \in \{\beta, \gamma\}. \end{cases} \tag{42}$$

Its memory update is

$$M_{t+1} = \begin{cases} \beta, & M_t = \alpha, \\ \gamma, & M_t = \beta \text{ and } O_t = 1, \\ \alpha, & M_t = \beta \text{ and } O_t = 0, \\ \alpha, & M_t = \gamma. \end{cases} \tag{43}$$

Thus the controller always resets in mode α, then holds in mode β. If the observation after this hold is 1, it holds once more in mode γ before returning to reset. If the observation after the first hold is 0, it resets immediately.

C.2 Canonical models

C.2.1 Canonical environment model

Let

$$p_t \coloneqq \Pr(E_t = 1 \mid H_{:t} = h_{:t})$$

be the filtering belief after an action-observation history $h_{:t}$. For a queried action $a \in \mathcal{A}$, define the predictive prior

$$q_t(a) \coloneqq \Pr(E_{t+1} = 1 \mid H_{:t} = h_{:t}, A_t = a).$$

By eq. (40),

$$q_t(0) = \tfrac{1}{2}, \qquad q_t(1) = p_t. \tag{44}$$

After observing $O_t = o$, Bayes' rule gives

$$p_{t+1} = \frac{(1-\eta)q_t(a)}{(1-\eta)q_t(a) + \eta(1-q_t(a))} \qquad \text{if } o = 1, \tag{45}$$

and

$$p_{t+1} = \frac{\eta q_t(a)}{\eta q_t(a) + (1-\eta)(1-q_t(a))} \qquad \text{if } o = 0. \tag{46}$$

Under the hold action $a = 1$, the one-step predictive probability of observing 1 is

$$\begin{aligned} \Pr(O_t = 1 \mid H_{:t} = h_{:t}, A_t = 1) &= (1-\eta)p_t + \eta(1-p_t) \\ &= \eta + p_t(1-2\eta). \end{aligned} \tag{47}$$

Since $\eta \in (0, \frac{1}{2})$, the map $p \mapsto \eta + p(1-2\eta)$ is strictly increasing. Therefore two histories with different beliefs $p_t \neq p'_t$ are distinguished by the counterfactual one-step input $A_t = 1$, and so cannot be equivalent environment causal states.

It remains to see that infinitely many such beliefs are reachable under arbitrary future action inputs. Let

$$L_t \coloneqq \frac{p_t}{1-p_t}, \qquad r \coloneqq \frac{1-\eta}{\eta} > 1.$$

Under repeated hold actions, the odds update as

$$L_{t+1} = \begin{cases} L_t r, & O_t = 1, \\ L_t r^{-1}, & O_t = 0. \end{cases}$$

Starting from $p_0 = \frac{1}{2}$, so $L_0 = 1$, after n hold steps with k observations equal to 1,

$$L_n = r^{2k-n}, \qquad p_n = \frac{r^{2k-n}}{1 + r^{2k-n}} = \frac{(1-\eta)^k \eta^{n-k}}{(1-\eta)^k \eta^{n-k} + \eta^k (1-\eta)^{n-k}}. \tag{48}$$

For example, taking $k = n$ gives the infinite sequence $L_n = r^n$. Hence there are infinitely many distinct posterior beliefs, and each gives a distinct environment causal state by eq. (47). Thus the unrestricted canonical environment model has infinitely many causal states.

C.2.2 Canonical agent model

Viewed as a channel $A_\cdot \mid O_\cdot$, the deterministic controller has three causal states. Each memory mode $M_t \in \{\alpha, \beta, \gamma\}$ determines the future action stream under any future observation continuation. The three modes are pairwise predictively distinct.

Mode α differs from β and γ already at the current action: α emits 0, while β and γ emit 1. Modes β and γ are distinct because under the observation $O_t = 1$, β transitions to γ, so $A_{t+1} = 1$, whereas γ transitions to α, so $A_{t+1} = 0$. Therefore

$$\mathcal{S}^M = \{\alpha, \beta, \gamma\}, \qquad |\mathcal{S}^M| = 3.$$

C.2.3 Canonical joint model

For the joint model, we consider the realised coupled process $W_t = (A_t, O_t)$. The process is generally not Markov on the emitted symbols alone, but it has a finite predictive presentation. The relevant predictive configurations are determined by the controller mode together with the belief values that can occur under the realised coupling.

After a reset from mode α, the latent state is a fair coin, so the next posterior is $1 - \eta$ after observing 1 and η after observing 0. From mode β, the controller holds once. Starting from the two beliefs η and $1 - \eta$, if the observation is 1, the posterior becomes either

$$p^{(11)} \coloneqq \frac{(1-\eta)^2}{(1-\eta)^2 + \eta^2}$$

from prior $1 - \eta$, or $\frac{1}{2}$ from prior η. If the observation is 0, the controller returns to α, where the next reset makes the precise posterior irrelevant for future prediction. Mode γ performs one further hold and then returns to α.

Thus the realised joint process has the following five predictive states:

$$\mathcal{S}^J = \{s^J_\alpha, s^J_{\beta 1}, s^J_{\beta 0}, s^J_{\gamma 11}, s^J_{\gamma 01}\},$$

where

$$\begin{array}{rcl} s^J_\alpha & : & M_t = \alpha, \\ s^J_{\beta 1} & : & M_t = \beta,\ p_t = 1 - \eta, \\ s^J_{\beta 0} & : & M_t = \beta,\ p_t = \eta, \\ s^J_{\gamma 11} & : & M_t = \gamma,\ p_t = p^{(11)}, \\ s^J_{\gamma 01} & : & M_t = \gamma,\ p_t = \frac{1}{2}. \end{array}$$

Let

$$c \coloneqq (1-\eta)^2 + \eta^2, \qquad d \coloneqq 2\eta(1-\eta), \qquad g \coloneqq \eta + p^{(11)}(1 - 2\eta).$$

The nonzero transitions of the joint presentation are:

$$\begin{array}{llll} s^J_\alpha \xrightarrow{(0,1)} s^J_{\beta 1} & \text{with probability } \frac{1}{2}, & s^J_\alpha \xrightarrow{(0,0)} s^J_{\beta 0} & \text{with probability } \frac{1}{2}, \\ s^J_{\beta 1} \xrightarrow{(1,1)} s^J_{\gamma 11} & \text{with probability } c, & s^J_{\beta 1} \xrightarrow{(1,0)} s^J_\alpha & \text{with probability } d, \\ s^J_{\beta 0} \xrightarrow{(1,1)} s^J_{\gamma 01} & \text{with probability } d, & s^J_{\beta 0} \xrightarrow{(1,0)} s^J_\alpha & \text{with probability } c, \\ s^J_{\gamma 11} \xrightarrow{(1,1)} s^J_\alpha & \text{with probability } g, & s^J_{\gamma 11} \xrightarrow{(1,0)} s^J_\alpha & \text{with probability } 1 - g, \\ s^J_{\gamma 01} \xrightarrow{(1,1)} s^J_\alpha & \text{with probability } \frac{1}{2}, & s^J_{\gamma 01} \xrightarrow{(1,0)} s^J_\alpha & \text{with probability } \frac{1}{2}. \end{array}$$

This presentation is unifilar: given the current state and emitted symbol (A_t, O_t), the next state is determined. The five states are pairwise predictively distinct. The state s^J_α is distinguished from all others because it emits action 0 rather than action 1. The states $s^J_{\beta 1}$ and $s^J_{\beta 0}$ have different probabilities of observing 1. The states $s^J_{\gamma 11}$ and $s^J_{\gamma 01}$ also have different probabilities of observing 1. Finally, β-states cannot merge with γ-states because their successor structure differs: from a β-state one may transition to a γ-state, while from a γ-state one returns to α. Therefore the canonical joint model has exactly five causal states.

C.3 Canonical support-restricted models

C.3.1 Canonical support-restricted environment model

The unrestricted environment model is infinite because it must distinguish predictions under arbitrary future action continuations, including arbitrarily long hold sequences. Under the realised controller, however, arbitrarily long hold continuations are not supported. The controller emits action 0 in mode α, action 1 in modes β and γ, and then returns to α after at most two consecutive holds.

By the general construction in section 3, each ordinary canonical support-restricted environment state is the image under ψ of a joint causal state. In this example, the five non-sink images are pairwise distinct:

$$\begin{aligned} s_{\alpha}^{\mathrm{sr}-E} &:= \psi(s_{\alpha}^{J}), \\ s_{\beta 1}^{\mathrm{sr}-E} &:= \psi(s_{\beta 1}^{J}), \\ s_{\beta 0}^{\mathrm{sr}-E} &:= \psi(s_{\beta 0}^{J}), \\ s_{\gamma 11}^{\mathrm{sr}-E} &:= \psi(s_{\gamma 11}^{J}), \\ s_{\gamma 01}^{\mathrm{sr}-E} &:= \psi(s_{\gamma 01}^{J}). \end{aligned}$$

Together with the totalisation sink $s_{\perp}^{E}$, the canonical support-restricted environment state space is

$$\mathcal{S}^{\mathrm{sr}-E} = \{s_{\alpha}^{\mathrm{sr}-E}, s_{\beta 1}^{\mathrm{sr}-E}, s_{\beta 0}^{\mathrm{sr}-E}, s_{\gamma 11}^{\mathrm{sr}-E}, s_{\gamma 01}^{\mathrm{sr}-E}, s_{\perp}^{E}\}.$$

The supported action sets are

$$\mathcal{A}_1(s_{\alpha}^{\mathrm{sr}-E}) = \{0\}, \qquad \mathcal{A}_1(s_{\beta 1}^{\mathrm{sr}-E}) = \mathcal{A}_1(s_{\beta 0}^{\mathrm{sr}-E}) = \mathcal{A}_1(s_{\gamma 11}^{\mathrm{sr}-E}) = \mathcal{A}_1(s_{\gamma 01}^{\mathrm{sr}-E}) = \{1\}.$$

Unsupported queried actions are sent to $s_{\perp}^{E}$, and from $s_{\perp}^{E}$ every queried action emits $\perp$ and remains in $s_{\perp}^{E}$. The ordinary transitions are obtained from the joint transitions above by applying theorem 17.

The six states are pairwise distinct. The sink state is distinct from every ordinary state because it emits $\perp$ forever. The state $s_{\alpha}^{\mathrm{sr}-E}$ is distinct from every other ordinary state because querying action 1 immediately yields $\perp$, whereas action 1 is supported from the other ordinary states. The two β-states are distinct because under action 1 they assign different probabilities to $O_t = 1$, namely c and d. The two γ-states are distinct because under action 1 they assign different probabilities to $O_t = 1$, namely g and $\frac{1}{2}$. No β-state can merge with any γ-state: under the queried action word 11, a β-state has positive probability of producing two ordinary observations before any occurrence of $\perp$, while from a γ-state the first queried hold returns the controller to α, so the second queried hold emits $\perp$ almost surely. Therefore the canonical support-restricted environment model has exactly six causal states.

C.3.2 Canonical support-restricted agent model

For the agent side, support restriction would restrict future observation continuations to those supported by the realised environment. In this example, the noisy observation kernel has full support because $\eta \in (0, \frac{1}{2})$. Consequently, every finite observation continuation has positive probability under the coupled process. The canonical support-restricted agent model therefore has the same three causal states as the unrestricted canonical agent model, together with a sink state $s_{\perp}^{M}$: $\mathcal{S}^{\mathrm{sr}-M} = \mathcal{S}^{M} \cup s_{\perp}^{M}$. This is specific to the present example. In general, environmental support restrictions can merge or alter the canonical support-restricted agent states.

D Relation between the three canonical models, with extra structure

Suppose that the environment and agent models are not only given as separate canonical channels, but as compatible components of a factorised closed-loop presentation. Concretely, assume that:

- the environment model has causal state space $\mathcal{S}^{E}$ and the agent model has causal state space $\mathcal{S}^{M}$, with compatible alphabets and timing conventions,
- closing the two channels defines a stationary joint process on $(A, O)_{:}$,
- the product state $Z_t := (S_t^{E}, S_t^{M})$ is sufficient for predicting the realised joint future, but is *not* a causal state of the joint process in general. Equivalently, assume the conditional independence

$$(A, O)_{t:} \perp\!\!\!\perp H_{:t} \mid Z_t, \tag{49}$$

or, in probabilistic form, $\Pr((A,O)_{t:} \mid H_{:t}) = \Pr((A,O)_{t:} \mid Z_t)$,

- the reachable product states, i.e. the states in the support of Z_t:

$$\mathrm{Reach}(\mathcal{S}^E \times \mathcal{S}^M) \coloneqq \{(s^E, s^M) \in \mathcal{S}^E \times \mathcal{S}^M \mid \Pr(Z_t = (s^E, s^M)) > 0\}, \tag{50}$$

carry well-defined transition dynamics and form a presentation of the realised joint process.

Under these assumptions, the reachable subset $\mathrm{Reach}(\mathcal{S}^E \times \mathcal{S}^M)$ is a sufficient presentation of the joint process. Since the canonical joint model is minimal among sufficient predictive presentations, its causal states are a quotient of this reachable product presentation. Equivalently, there is a surjective map $\chi : \mathrm{Reach}(\mathcal{S}^E \times \mathcal{S}^M) \to \mathcal{S}^J$ such that, for every history $h_{:t}$ inducing the reachable product state $z(h_{:t}) \coloneqq \big(\epsilon_E(h_{:t}), \epsilon_M(h_{:t})\big)$, we have $\epsilon_J(h_{:t}) = \chi\big(z(h_{:t})\big)$. Thus, in the finite case,

$$|\mathcal{S}^J| \leq |\mathrm{Reach}(\mathcal{S}^E \times \mathcal{S}^M)| \leq |\mathcal{S}^E|\,|\mathcal{S}^M|. \tag{51}$$

This product bound is therefore a property of an explicitly specified factorised closed-loop presentation satisfying a joint-sufficiency condition.