

Information Thermodynamics of Agents: The Work Capacity of Channels with Memory

Lukas J. Fiderer^{✉,*}, Paul C. Barth[✉], Isaac D. Smith[✉], and Hans J. Briegel[✉]

Universität Innsbruck, Institut für Theoretische Physik, Technikerstraße 21a, A-6020 Innsbruck, Austria



(Received 23 April 2025; revised 2 December 2025; accepted 10 August 2026; published 5 October 2026)

Predicting future observations plays a central role in machine learning, biology, economics, and many other fields. It lies at the heart of organizational principles such as the variational free-energy principle and, based on the second law of thermodynamics, has even been shown to be necessary for reaching the fundamental energetic limits of information processing on a tape. While the usefulness of the predictive paradigm is undisputed, complex adaptive systems that interact with their environment are more than just predictive machines: They have the power to act upon their environment and cause change. In this work, we develop a framework to analyze the thermodynamics of information processing in percept-action loops, a model of agent-environment interaction, allowing us to investigate the thermodynamic implications of actions and percepts on equal footing. To this end, we introduce the concept of work capacity defined as the maximum rate at which an agent can expect to extract work from its environment. Our results reveal that work-efficient agents must balance prediction and forgetting. This highlights a fundamental departure from the thermodynamics of passive observation, suggesting that prediction and energy efficiency may be at odds in active learning systems.

DOI: [10.1103/7nds-tjr8](https://doi.org/10.1103/7nds-tjr8)

Subject Areas: Complex Systems, Statistical Physics

I. INTRODUCTION

Percept-action loops are cycles in which an agent perceives its environment, processes and stores information, and acts to influence future perception. They underlie adaptive behavior in both biological and artificial systems. Such loops can be observed across various domains, from humans learning chess, to animals foraging, to artificial intelligence models engaging in dialog.

Energetic aspects of processes underlying percept-action loops have been highlighted in diverse contexts. In the brain, energy consumption associated with neural signaling is minimized through efficient coding strategies [1,2]. At the molecular level, ribosomes have been shown to perform simple decoding computations at energy costs within an order of magnitude of Landauer's bound, significantly outperforming even the most advanced supercomputers [3]. In artificial intelligence, the energetic cost of supercomputers is becoming an increasing concern, particularly in the training of large neural networks [4], resulting in performance-power trade-offs in large language models [5].

All such processes are ultimately constrained by the second law of thermodynamics, which sets fundamental

limits on the efficiency of information processing. Experimental studies of Landauer erasure [6,7] and information engines with measurement-based feedback [8,9] have begun to probe these limits. Theoretically, they have been analyzed using nonequilibrium fluctuation relations and generalized second laws for information processing with feedback [10–17].

While these works form an important part of the general foundations of information thermodynamics, little is known about the distinctive thermodynamic properties of percept-action loops. This work is motivated by two central questions: (1) What are the energetic limits of adaptive information processing in percept-action loops, and (2) how should agents be designed to operate efficiently under such constraints?

These questions can be addressed by reducing the problem to an information-theoretic model of percept-action loops. By abstracting away implementation-dependent details, we derive energetic bounds that arise solely from the intrinsic cost of information processing, as analyzed through nonequilibrium thermodynamics [10,11].

As a point of comparison, we consider the tape setting (see Fig. 1), where, inspired by Maxwell's demon, agents sequentially process and modify symbols on a preexisting tape [18–32]. In the tape setting, predictable correlations in the input serve as an energetic resource, while generating correlations in the output incurs an energy cost. However, works in the tape setting exclude feedback between the agent and environment (but see Ref. [33] for a recent exception, investigating memoryful quantum processes with the agent and environment).

*Contact author: lukasj.fiderer@gmail.com

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

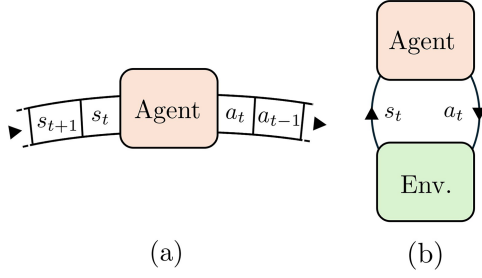


FIG. 1. Tape setting (a) and percept-action-loop setting (b). In the tape setting (a), the agent receives an information carrier encoding an incoming symbol s_{t-1} and processes it by replacing this encoding with that of an outgoing symbol a_t on the same information carrier. Subsequent information carriers are supplied by a preexisting tape and are therefore not influenced by the outgoing symbols. In the percept-action-loop setting (b), the agent performs the same physical mapping from an incoming symbol s_{t-1} to an outgoing symbol a_t , on a single information carrier. Here, however, the tape is replaced by an environment (Env.) which receives the information carrier encoding a_t and supplies the next information carrier. Consequently, s_t may depend not only on previous incoming symbols, but also on the current and previous outgoing symbols, thereby closing the feedback loop. Both the agent and the environment may have memory, allowing these dependences to extend over multiple rounds. For the assumptions underlying the thermodynamic implementation, see Sec. IV.

In this work, we model both the agent and environment as hidden Markov channels. By combining results from stochastic thermodynamics [10,11] with the information theory of hidden Markov models [34], we obtain a framework which goes beyond prior work situated in the tape setting by incorporating feedback between the agent and environment. This framework is the primary contribution of this work.

The central quantity of this work is the work capacity defined as the optimal rate of work extraction achievable by any agent. Analogous to communication capacity, it is an intrinsic information-theoretic property of the environment channel.

A key insight from our investigation of work capacity is that, in percept-action loops, predictive agents are generally inefficient. This is surprising in light of earlier results from the tape setting, where work-efficient agents must be predictive [26]. We show that this difference persists even when the agent's memory is constrained to arbitrary finite size.

In the following sections, we first introduce the percept-action-loop framework (Sec. II) and define what it means for an agent to be predictive (Sec. III). We then present our main results on the work capacity of channels (Sec. IV) and the design principles of work-efficient agents (Sec. V). These insights are further developed through a detailed example (Sec. VI). Finally, we discuss directions for future research and conclude by situating our findings in a broader context (Sec. VII).

II. FRAMEWORK

Throughout this work, random variables are denoted by capital letters, their realizations by lowercase letters, their alphabets by calligraphic letters, and sequences of random variables—interpreted as random variables on a product space—by $X_{t:n} = (X_t, X_{t+1}, \dots, X_{n-1})$. Infinite sequences known as stochastic processes are written as $\mathbf{X} = X_{0:\infty}$, with analogous notation for their realizations $x_{t:n} = (x_t, x_{t+1}, \dots, x_{n-1}) \in \mathcal{X}^{n-t}$ and $\mathbf{x} = x_{0:\infty} \in \mathcal{X}^{\mathbb{N}_0}$.

We consider an agent interacting with an environment in discrete time steps (in the following called rounds) indexed by $t \in \mathbb{N}_0$, where $\mathbb{N}_0$ denotes the non-negative integers. In each round t , the agent selects an action a_t and subsequently receives a percept s_t . Since both the agent and environment may be stochastic, actions and percepts are modeled as random variables A_t and S_t taking values in finite alphabets $\mathcal{A}$ and $\mathcal{S}$ [see Fig. 1(b)]. Embedding the smaller alphabet into the larger lets us set $\mathcal{A} = \mathcal{S}$, which will be assumed in the following.

A concept of central importance in describing environments and agents is that of a channel. A channel from an alphabet $\mathcal{X}$ to an alphabet $\mathcal{Y}$ is a conditional probability distribution $\mu_{Y|X}$; that is, it assigns to every input $x \in \mathcal{X}$ a probability distribution $\mu_{Y|X}(y|x)$ over outputs $y \in \mathcal{Y}$.

The environment and the agent are described by the channels $\nu_{S|A}^{\text{env}}$ and $\eta_{A|S}^{\text{agt}}$, respectively. An environment channel assigns to every (infinite) sequence of actions $\mathbf{a} = (a_0, a_1, \dots)$ a probability distribution $\nu_{S|A}^{\text{env}}(s|\mathbf{a})$ over percept sequences $\mathbf{s} = (s_0, s_1, \dots) \in \mathcal{S}^{\mathbb{N}_0}$ (and analogously for agent channels described by $\eta_{A|S}^{\text{agt}}$).

To qualify as an agent or environment channel, we impose two additional assumptions, which we briefly outline before providing a formal definition. First, the channel must be causal, meaning it respects temporal order—future inputs cannot influence past outputs. Second, it must be finite, that is, realizable with systems of finite dimension. Like causality, the finiteness assumption is physically motivated: Manipulating infinite-dimensional systems can entail unphysical effects such as infinite energy costs. It is also mathematically convenient, as it avoids complications associated with infinite-dimensional spaces (see Fig. 3). Finally, for our purposes this assumption does not pose a practical limitation, since the performance of optimal infinite-state solutions can be approximated arbitrarily well using finite resources.

These two constraints define what is commonly referred to in the literature as a finite-state [35] or hidden Markov channel [34] (see Sec. C of Supplemental Material [36] for a more in-depth exposition). We now proceed to define both the environment and the agent as such hidden Markov channels.

Definition 1. A channel $\nu_{S|A}^{\text{env}}$ is an environment channel denoted as

$$\text{env} := \nu_{S|A}^{\text{env}} \quad (1)$$

if there exists a finite set of states $\mathcal{Z}$, a distribution $p_{\mathcal{Z}_0}^{\text{env}}$ over $\mathcal{Z}$, and a transition matrix $\Phi^{\text{env}} = (\phi^{\text{env}}(j|i))_{j,i}$ with $i \in \mathcal{A} \times \mathcal{Z}$ and $j \in \mathcal{S} \times \mathcal{Z}$ such that

$$\nu_{S|A}^{\text{env}}(s|a) = \sum_z \nu_{SZ|A}^{\text{env}}(s, z|a), \quad (2)$$

where the sum runs over all $z \in \mathcal{Z}^{\mathbb{N}_0}$ and

$$\nu_{SZ|A}^{\text{env}}(s, z|a) := p_{\mathcal{Z}_0}^{\text{env}}(z_0) \prod_{t=0}^{\infty} \phi^{\text{env}}(s_t, z_{t+1}|a_t, z_t). \quad (3)$$

Then, the tuple

$$\text{env}_{\mathcal{Z}} := (\Phi^{\text{env}}, p_{\mathcal{Z}_0}^{\text{env}}) \quad (4)$$

is an input-output hidden Markov model called an environment model, the elements $z \in \mathcal{Z}$ are called its hidden states, and $\nu_{SZ|A}^{\text{env}}$ is called the model channel of $\text{env}_{\mathcal{Z}}$.

While an environment channel describes only the input-output behavior, its input-output hidden Markov model provides an explicit memory-based mechanism that generates temporal correlations. Input-output hidden Markov models are also known as transducers [63,64].

Agents are defined analogously. The key distinction from environment channels is that the agent initiates the percept-action loop by selecting a first action A_0 (see Fig. 2).

Definition 2. A channel $\eta_{A|S}^{\text{agt}}$ is an agent channel denoted as

$$\text{agt} := \eta_{A|S}^{\text{agt}} \quad (5)$$

if there exists a finite set of states $\mathcal{M}$, a distribution $p_{\mathcal{M}_0 A_0}^{\text{agt}}$ over $\mathcal{M} \times \mathcal{A}$, and a transition matrix $\Theta^{\text{agt}} = (\theta^{\text{agt}}(j|i))_{j,i}$ with $i \in \mathcal{S} \times \mathcal{M}$ and $j \in \mathcal{A} \times \mathcal{M}$ such that

$$\eta_{A|S}^{\text{agt}}(a|s) = \sum_m \eta_{MA|S}^{\text{agt}}(m, a|s), \quad (6)$$

where the sum runs over all $m \in \mathcal{M}^{\mathbb{N}_0}$ and

$$\eta_{MA|S}^{\text{agt}}(m, a|s) := p_{\mathcal{M}_0 A_0}^{\text{agt}}(m_0, a_0) \times \prod_{t=0}^{\infty} \theta^{\text{agt}}(a_{t+1}, m_{t+1}|s_t, m_t). \quad (7)$$

Then, the tuple

$$\text{agt}_{\mathcal{M}} := (\Theta^{\text{agt}}, p_{\mathcal{M}_0 A_0}^{\text{agt}}) \quad (8)$$

is an input-output hidden Markov model called an agent model, the elements $m \in \mathcal{M}$ are called its memory states, and $\eta_{MA|S}^{\text{agt}}$ is called the model channel of $\text{agt}_{\mathcal{M}}$.

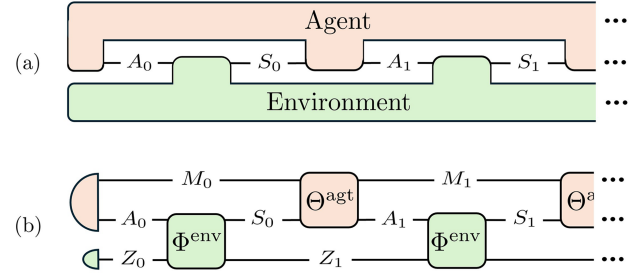


FIG. 2. Circuit representation of percept-action loops with time flowing from left to right; the horizontal line represents the logical percept-action register shared between them. (a) The agent and environment are modeled as channels with memory. (b) The agent and environment are represented by their hidden Markov models characterized by finite adaptive memories M_t and Z_t . The transition matrices Θ^{agt} and Φ^{env} remain fixed over time.

An agent can be understood as having two types of memory: (i) read-only memory (ROM), which remains fixed over time and encodes the agent's transition matrix Θ^{agt} (analogous to the programmed rules or hardware features of an artificial agent or the DNA of a biological organism), and (ii) read-write or adaptive memory $\mathcal{M}$, which the agent can update to retain information about the past percept-action history and which, through the operation of Θ^{agt} , influences future actions.

Unless stated otherwise, we assume that the percept-action register contains no ROM; that is, all parts of it can be modified by the agent.

With the definitions of the agent and environment channels, as well as their hidden Markov models, we define percept-action loops as tuples of an agent (model) and an environment (model), giving rise to four different types. In the following, it is assumed that the agent and environment are compatible in the sense that their percept and action alphabets match.

Definition 3. Percept-action loops are classified into types (i)–(iv) defined by the tuples listed in the first column of Table I. Their corresponding distributions are given in the second column.

TABLE I. Percept-action loops and their associated distributions. The four types (i) to (iv) differ in whether the agent and environment are represented as hidden Markov channels (agt and env) or modeled with explicit memory states (agt $_{\mathcal{M}}$ and env $_{\mathcal{Z}}$). Their distributions are defined by the corresponding channels (see Definitions 1 and 2).

	Percept-action loop	Distribution
(i)	(agt $_{\mathcal{M}}$, env $_{\mathcal{Z}}$)	$p_{MASZ} := \nu_{SZ A}^{\text{env}} \eta_{MA S}^{\text{agt}}$
(ii)	(agt, env $_{\mathcal{Z}}$)	$p_{ASZ} := \nu_{SZ A}^{\text{env}} \eta_{A S}^{\text{agt}}$
(iii)	(agt $_{\mathcal{M}}$, env)	$p_{MAS} := \nu_{S A}^{\text{env}} \eta_{MA S}^{\text{agt}}$
(iv)	(agt, env)	$p_{AS} := \nu_{S A}^{\text{env}} \eta_{A S}^{\text{agt}}$

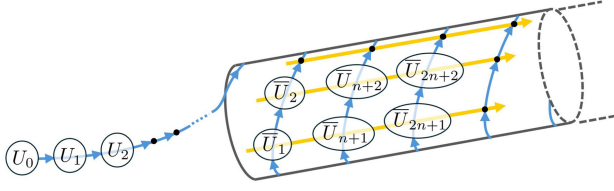


FIG. 3. Asymptotic behavior of percept-action loops on a finite-state space. We show that the global process $\mathbf{U}$ of a percept-action loop, where $U_t = (M_t, A_t, S_t, Z_t)$ includes the agent's memory, actions, percepts, and the environment's hidden memory, forms a finite-state Markov chain (the blue line visually guides the eye along this process). As time progresses, any such Markov chain converges to a cyclostationary process: one that periodically cycles through n stationary processes (as indicated by the yellow arrows), for some finite n . This ensures that the asymptotic behavior of percept-action loops is well behaved, in particular implying convergence of time averages.

A systematic way to understand percept-action loops is through type (i), which defines a global distribution

$$p_{MASZ}(\mathbf{m}, \mathbf{a}, \mathbf{s}, \mathbf{z}) = p_{M_0 A_0}^{\text{agt}}(m_0, a_0) p_{Z_0}^{\text{env}}(z_0) \times \prod_{t=0}^{\infty} \phi^{\text{env}}(s_t, z_{t+1} | a_t, z_t) \times \theta^{\text{agt}}(a_{t+1}, m_{t+1} | s_t, m_t) \quad (9)$$

as the product of Eqs. (3) and (7). The distributions for types (ii)–(iv) can then be obtained by marginalizing over the hidden states. Figure 2 illustrates percept-action loops of types (i) and (iv).

We denote by $\mathbb{A}^{\text{agt,env}}(d)$ the set of all agent models with memory size $|\mathcal{M}| = d$ that are compatible with a given environment env , and define the full set of compatible agents as

$$\mathbb{A}^{\text{agt,env}} := \bigcup_{d \in \mathbb{N}} \mathbb{A}^{\text{agt,env}}(d). \quad (10)$$

Importantly, the global distribution p_{MASZ} defines a finite-state Markov chain. This constitutes the global Markov chain of a percept-action loop (see Sec. D of Supplemental Material [36] for a proof):

$$M_0 A_0 S_0 Z_0 \rightarrow M_1 A_1 S_1 Z_1 \rightarrow \dots \quad (11)$$

This Markovian property allows us to leverage existing results on finite-state Markov chains, ensuring that the asymptotic dynamics of percept-action loops are well behaved [65,66] (see Fig. 3 and Sec. B of Supplemental Material [36] for more details).

III. PREDICTIVE AGENTS

For a given input-output behavior of the agent and environment (agt, env) , what does it mean for an agent to be as predictive as possible of its future percepts? To approach this question, it is helpful to begin with the

following observation. In order to endow the agent with knowledge that reduces uncertainty about future percepts, a natural first step is to encode $(\text{agt}_{\mathcal{M}}, \text{env}) = (\Theta^{\text{agt}}, p_{M_0 A_0}^{\text{agt}}, \nu_{S|A}^{\text{env}})$ into its fixed read-only memory. In what follows, we assume this is always the case.

With this setup, the agent has access to the distribution of the underlying process p_{MAS} , which results in an uncertainty $H(S_t)$ about percept S_t , where $H(S_t)$ denotes Shannon entropy in units of bits. If, in addition, the agent takes its memory M_t into account before observing S_t , this memory reduces the agent's expected (with respect to memory states) uncertainty to $H(S_t | M_t)$. This reduction in uncertainty

$$I[M_t; S_t] = H(S_t) - H(S_t | M_t) \quad (12)$$

is simply the mutual information $I[M_t; S_t]$ between S_t and M_t , quantifying how much M_t enables the agent to predict S_t (see Sec. A of Supplemental Material [36] for some background on information measures).

To enhance its predictive capabilities, the agent can store information from past percepts and actions in its memory. In fact, the information the past $A_{0:t+1} S_{0:t}$ provides about the current percept S_t is given by $I[A_{0:t+1} S_{0:t}; S_t]$. Consequently, if the conditional mutual information $I[A_{0:t+1} S_{0:t}; S_t | M_t]$ vanishes, the agent's memory M_t contains all information from the past about the current percept. Otherwise, if $I[A_{0:t+1} S_{0:t}; S_t | M_t]$ is positive, the agent must have forgotten about aspects of the past which are relevant for predicting the current percept. $I[A_{0:t+1} S_{0:t}; S_t | M_t]$ thus quantifies a predictive gap.

Definition 4. Let $(\text{agt}_{\mathcal{M}}, \text{env})$ be a percept-action loop. The predictive gap of $\text{agt}_{\mathcal{M}}$ with respect to env in round t is given by

$$I[A_{0:t+1} S_{0:t}; S_t | M_t], \quad (13)$$

and the asymptotic mean (a.m.) predictive gap of $\text{agt}_{\mathcal{M}}$ with respect to env is

$$\langle I[A_{0:t+1} S_{0:t}; S_t | M_t] \rangle_t, \quad (14)$$

where

$$\langle \bullet \rangle_t := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \bullet \quad (15)$$

denotes the Cesàro limit, the limit of the arithmetic mean. An agent model $\text{agt}_{\mathcal{M}}$ is said to be maximally predictive if its a.m. predictive gap vanishes.

The condition for an agent to be maximally predictive can be illustrated within the Bayesian network underlying percept-action loops (see Fig. 4). For example, an agent is

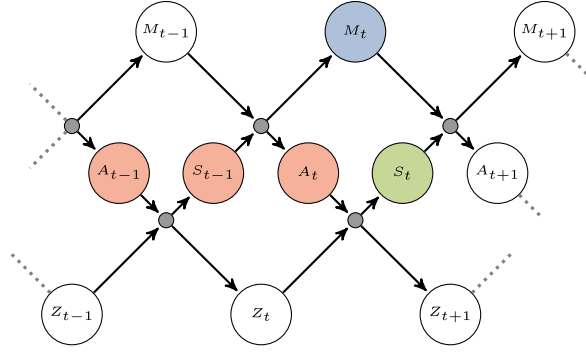


FIG. 4. Bayesian network for a percept-action loop. Shown is a fragment for rounds $t-1$, t , and the beginning of round $t+1$. This type of Bayesian network plays an important role in the information-theoretic framework underlying our results (see Sec. E of Supplemental Material [36] for details and Ref. [14] for an application of Bayesian networks in the context of the thermodynamics of feedback control). Note that to faithfully represent the dynamics of the agent and environment, auxiliary nodes (gray and reduced in size) are included. The colored nodes illustrate the condition for an agent to be maximally predictive in round t : The agent's memory (blue) must store all information from past actions and percepts $A_{0:t+1}S_{0:t}$ (red) that is relevant for predicting the current percept S_t (green).

maximally predictive if Eq. (13) vanishes for all times, or if Eq. (13) decays sufficiently quickly.

IV. WORK CAPACITY OF CHANNELS

So far, we have treated agents and environments in terms of abstract logical operations on alphabets. However, as Landauer famously quipped, information is physical [67]: Any implementation of an agent must ultimately rely on physical memory and dynamics subject to thermodynamic laws. To analyze the energetic limits of such implementations, we adopt a framework from stochastic thermodynamics that models the agent's information processing described by its transition matrix Θ^{agt} , as a physical process acting on memory [10,11]. The following isothermal information-processing (IIP) assumptions have three parts: information carriers and encodings, physical protocols, and thermodynamic resource accounting. Together, they are chosen so that the resulting energetic bounds capture only the intrinsic cost of information processing. Before turning to the full percept-action loop, we first consider a single logical operation $\Theta: \mathcal{X}_{\text{in}} \rightarrow \mathcal{X}_{\text{out}}$.

A. Isothermal information-processing assumptions

1. Information carriers and encodings

Let $\mathfrak{S}_d$ denote an information carrier, i.e., a physical system coupled to a thermal reservoir at temperature T . The system has d metastable configurations satisfying the timescale separation $\tau_{\text{others}} \ll \tau_{\text{info}}$, where τ_{info} is the equilibration time of the metastable configurations, and

τ_{others} is the equilibration time of the remaining degrees of freedom.

We assume that the input and output alphabets have the same cardinality $|\mathcal{X}_{\text{in}}| = |\mathcal{X}_{\text{out}}| = d$. Both alphabets are encoded in the same information carrier $\mathfrak{S}_d$ by one-to-one assignments of alphabet symbols to metastable configurations. The input alphabet labels the metastable configurations before the protocol; the output alphabet labels the same metastable configuration space after the protocol. The initial distribution over $\mathcal{X}_{\text{in}}$ is arbitrary.

Information-processing architectures often store some variables in separate physical subsystems; for example, memory may be encoded in a separate subsystem. To model such modular architectures, we allow information carriers to be composed by taking component information carriers $\mathfrak{S}_{d_1}, \dots, \mathfrak{S}_{d_n}$ coupled to the same thermal reservoir at temperature T . The metastable configuration space of the resulting composite information carrier $\mathfrak{S}_d$ is the product of the component metastable configuration spaces, so that $d = \prod_{i=1}^n d_i$. If each component information carrier $\mathfrak{S}_{d_i}$ encodes alphabet $\mathcal{X}_i$ one to one, then the composite information carrier naturally encodes the product alphabet $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$.

2. Physical protocol

A physical protocol implements the logical operation $\Theta: \mathcal{X}_{\text{in}} \rightarrow \mathcal{X}_{\text{out}}$ by isothermal control of the information carrier $\mathfrak{S}_d$. The protocol is executed on a timescale τ_{protocol} satisfying $\tau_{\text{others}} \ll \tau_{\text{protocol}} \ll \tau_{\text{info}}$. For a composite information carrier, this condition is understood as applying to every component information carrier.

The protocol may implement any isothermal control of the potential landscape over the metastable configuration space required by the logical operation under consideration. For a composite information carrier, the protocol may implement joint isothermal control over the composite metastable configuration space.

3. Thermodynamic resource accounting

The following assumptions ensure that the work balance reflects the thermodynamics intrinsic to information processing, rather than changes in the energetic structure of the information carrier or in the external control apparatus.

At the beginning and end of the protocol, the effective internal energies assigned to the metastable configurations of the information carrier are equal and unchanged. That is, $U_i^{\text{in}} = U_i^{\text{out}} = U_0$ for all $i \in \{1, \dots, d\}$. An information carrier with equal internal energies assigned to its metastable configurations is called an information reservoir [68].

The isothermal control is cyclic: By the end of the protocol, all externally controlled potential landscapes are returned to their initial form, and the external control apparatus undergoes no net physical change. However, the

probability distribution over metastable configurations may change, thereby implementing the desired logical operation Θ [11].

B. Second-law bound on extractable work

Let W be the expected work extracted by applying a physical protocol that implements a logical operation $\Theta: \mathcal{X}_{\text{in}} \rightarrow \mathcal{X}_{\text{out}}$ on an information carrier. In the following, we use the sign convention that $W > 0$ denotes work extracted by the protocol and $W < 0$ work supplied to implement it. Note that we refer to the work as expected because, before executing the protocol, it is the average work predicted from the distribution $p_{\mathcal{X}_{\text{in}}}$ over initial configurations of the information carrier.

Under the IIP assumptions, the second law provides an upper bound on W in terms of the change in nonequilibrium free energy of the information carrier between the beginning and end of the protocol [10,11]. The nonequilibrium free energy is the state function $F = U - k_B T \ln 2H(X)$, where $U = \sum_x p_X(x) U_x$, with U_x the effective internal energy assigned to the information carrier's metastable configuration encoding x , and k_B is the Boltzmann constant [26]. Since the internal-energy contributions cancel under the IIP assumptions, this free-energy bound reduces to an entropy bound on the expected extractable work

$$W \leq H(X_{\text{out}}) - H(X_{\text{in}}) \quad (16)$$

when implementing the logical operation Θ that induces the transition from X_{in} to X_{out} . Here and throughout, all work expressions are understood to be in units of $k_B T \ln 2$.

The upper bound in Eq. (16) imposed by the second law is tight in the sense that it can, in principle, be saturated with protocols under idealized conditions. Concrete examples of such protocols are given in Refs. [25,26,29]. While our work is primarily concerned with the fundamental limits imposed by the second law, it should be noted that more realistic and resource-constrained assumptions can be incorporated [69–72]. Within our framework, this is most easily achieved when the extractable work can still be expressed through a state function, such as in Ref. [71], by replacing F with the new state function.

We now apply Eq. (16) to the agent's logical operation in a percept-action loop under the same IIP assumptions. That is, we set $\mathcal{X}_{\text{in}} = \mathcal{M} \times \mathcal{S}$ and $\mathcal{X}_{\text{out}} = \mathcal{M} \times \mathcal{A}$, and choose the logical operation Θ to implement the agent's transition matrix Θ^{agt} . In particular, $|\mathcal{X}_{\text{in}}| = |\mathcal{X}_{\text{out}}|$ since, by assumption, $|\mathcal{A}| = |\mathcal{S}|$. For this application of Eq. (16), the alphabets $\mathcal{M} \times \mathcal{S}$ and $\mathcal{M} \times \mathcal{A}$ are encoded in the same information carrier. This information carrier may nevertheless be built from component information carriers representing the agent memory and a single register that carries the percept before the protocol and the action after it.

For each round t , this application of Eq. (16) gives an upper bound $W_t \leq H(M_{t+1}A_{t+1}) - H(M_t S_t)$ on the expected work extractable by the agent's operation in that round.

Taking the asymptotic mean over rounds yields an upper bound on the asymptotically expected work per round that an agent model $\text{agt}_{\mathcal{M}}$ can extract by implementing its transition matrix under the IIP assumptions. We denote this bound by

$$W(\text{agt}_{\mathcal{M}}, \text{env}) = \langle H(A_t|M_t) - H(S_t|M_t) \rangle_t, \quad (17)$$

where $\langle \bullet \rangle_t$ is the Cesàro limit defined in Eq. (15); see Sec. F.1 of Supplemental Material [36] for a derivation. Since the bound in Eq. (17) is tight in principle under idealized protocols, and since we focus on fundamental limits, we refer to it in the following simply as the work rate.

The IIP assumptions imply an implementation picture for the percept-action-loop model in which the information carrier encoding the agent's memory remains under the agent's control throughout all rounds. For each round, the same information carrier must encode the percept before the agent's protocol and the action after it. The environment may either act on that information carrier to produce the next percept or provide a new information carrier carrying that percept. Since Eq. (17) bounds the work extracted by the agent's operation, the environment channel provides the only resource for work extraction by the agent. The agent exploits this resource by retaining information from past rounds in its memory and thereby adapting to the statistical structure of the environment channel.

The existence of the work rate is not guaranteed for arbitrary processes, as it is possible that the Cesàro limit in Eq. (17) does not exist [73]. Note, however, that here the limit exists because the global Markov chain of the percept-action loop has a finite number of states and thus converges to a cyclostationary process (see Fig. 3). Note that if the logical percept-action register involves ROM—parts which can be read only but not updated by the agent—the work rate becomes $\langle [H(A_t S_{t-1}^{\text{ROM}}|M_t) - H(S_t S_t^{\text{ROM}}|M_t)] \rangle_t$ where S_t^{ROM} denotes the ROM provided by the environment in round t , and S_{-1}^{ROM} is an initial ROM state.

C. Work capacity

Equation (17) can be interpreted as a thermodynamic objective function for agents derived directly from the second law of thermodynamics and the information-theoretic structure of percept-action loops (see Ref. [74] for a related but distinct approach with a thermodynamics-inspired objective for decision-making). Maximizing this objective leads us to the following:

Definition 5. For a given environment channel env , let

$$W^{(d)}(\text{env}) := \sup_{\text{agt}_{\mathcal{M}} \in \mathbb{A}^{\text{env}}(d)} W(\text{agt}_{\mathcal{M}}, \text{env}) \quad (18)$$

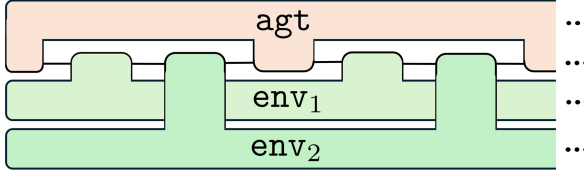


FIG. 5. An agent agt interacting with the cascade of two environment channels env_1 and env_2 .

be the supremum of the work rates among all agent models with $|\mathcal{M}| = d$. Then, the work capacity C^{work} of env is defined as

$$C^{\text{work}}(\text{env}) := \sup_{d \in \mathbb{N}} W^{(d)}(\text{env}). \quad (19)$$

Combining Eqs. (18) and (19), we can also write $C^{\text{work}}(\text{env}) = \sup_{\text{agt}_{\mathcal{M}} \in \mathbb{A}^{\text{env}}} W(\text{agt}_{\mathcal{M}}, \text{env})$.

Intuitively, the work capacity captures the supremum rate at which an agent optimally tailored to the environment channel can expect to extract work, based on the second law of thermodynamics.

For any fixed percept-action loop $(\text{agt}_{\mathcal{M}}, \text{env})$, the protocols of Refs. [25,26,29] can, in principle, attain the corresponding work rate $W(\text{agt}_{\mathcal{M}}, \text{env})$, thereby allowing the supremum in Eq. (18) to be approached arbitrarily closely; see Sec. F.2 of Supplemental Material [36].

Returning to the question posed at the beginning of this section, the energetic limits of agents, in terms of work rate, are determined by the work capacity of the environment channel.

We next state some general properties of work capacity.

Theorem 1. For any environment channel $\text{env} = \nu_{S|A}^{\text{env}}$, the work capacity $C^{\text{work}}(\text{env})$ has the following properties:

- (i) (*Existence*) $C^{\text{work}}(\text{env})$ exists,
- (ii) (*Bounds*) $0 \leq C^{\text{work}}(\text{env}) \leq \log_2 |\mathcal{S}|$,
- (iii) (*Subadditivity under channel cascade*; see Fig. 5)

$$C^{\text{work}}(\text{env}_2 \circ \text{env}_1) \leq C^{\text{work}}(\text{env}_1) + C^{\text{work}}(\text{env}_2).$$

See Sec. F.3 of Supplemental Material [36] for a proof. Theorem 1(i) is in fact a statement about the existence of the work rate, which enters in the definition of work capacity, and follows from the fact that the global Markov chain is well behaved (see Fig. 3). Further, the bounds in Theorem 1(ii) follow from the canonical bounds on Shannon entropy. Finally, Theorem 1(iii) follows from the observation that agents interacting with a cascade $\text{env}_2 \circ \text{env}_1$ form a subclass of those that alternately interact with env_1 and env_2 . Specifically, this subclass consists of agents that implement an identity channel from the percepts of env_1 to the actions of env_2 . The inequality then follows from the fact that using a nonidentity channel can yield a higher work rate (see Fig. 6 for an illustrative example of the subadditivity of work capacity).

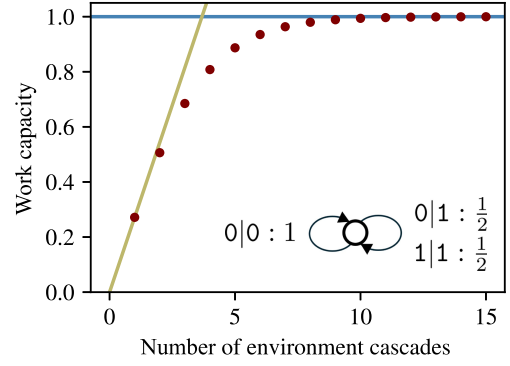


FIG. 6. Minimal example illustrating the subadditivity of work capacity. We consider a cascade of n identical copies of a memoryless, invariant environment env with binary percept and action alphabets $\mathcal{A} = \mathcal{S} = \{0, 1\}$ (see inset; the transition labels follow the format percept | action : transition probability). The burgundy dots represent work capacity $C^{\text{work}}(\text{env}^n)$ in units of $k_B T \ln 2$, of the n -fold cascade as a function of n (see Fig. 5 for an illustration of the case $n = 2$). The yellow line is the subadditivity bound from Theorem 1(iii), and the blue line at $\log_2 2$ is the upper bound from Theorem 1(ii). Note that for this environment, the outputs become increasingly deterministic, approaching a constant 0 as $n \rightarrow \infty$. This allows the agent to better predict the outputs and extract more work as n increases.

Because the work capacity involves both a Cesàro limit and a maximization over agent models, it is generally difficult to compute. We now discuss special classes of environment channels for which the expression simplifies, as summarized in Table II.

1. Memoryless invariant channels

For memoryless invariant environment channels, where

$$\nu_{S|A}^{\text{env}}(s|a) = \prod_{t=0}^{\infty} \phi^{\text{env}}(s_t|a_t) \quad (20)$$

with the same transition matrix Φ^{env} for all $t \in \mathbb{N}_0$, the absence of temporal memory allows the optimization in Eq. (19) to reduce to a single-step maximization over actions.

TABLE II. Simplified expressions for the work capacity of memoryless invariant and product environment channels. The identity channel, with or without Markovian ROM, is a special case of a memoryless invariant channel. See Eq. (24) for the definition of the entropy rate $h(S)$, and Sec. F.3 of Supplemental Material [36] for a detailed proof.

Environment channel env	$C^{\text{work}}(\text{env})$
Memoryless invariant	$\max_{p_{A_0}} [H(A_0) - H(S_0)]$
Identity	0
Identity + Markovian ROM	$\max_{p_{S_0}} I[S_0^{\text{ROM}}; S_0]$
Product (= tape setting)	$\log_2 \mathcal{A} - h(S)$

A basic example is the identity channel [75], where $A_t = S_t$ for all t . In this case, an agent can predict the percept precisely to the extent it remembers its previous action, turning the trade-off between action and percept entropy into a zero-sum relation:

$$H(A_t|M_t) - H(S_t|M_t) = 0, \quad (21)$$

and thus the work capacity vanishes.

A richer example is an identity channel with ROM: The main channel transmits the input unchanged, while an additional side channel provides ROM that the agent can observe but not modify. Formally, the environment maps (A_t, A_t^{ROM}) to (S_t, S_t^{ROM}) such that $S_t = A_t$, and the agent is constrained to respond such that $A_{t+1}^{\text{ROM}} = S_t^{\text{ROM}}$ while A_0^{ROM} is fixed. When the ROM is Markovian—that is, when S_t^{ROM} is a probabilistic function of A_t —the work capacity simplifies to the expression given in Table II.

This setup parallels scenarios studied in stochastic thermodynamics, where a system is controlled by an agent using side information about the system's state. It encompasses Maxwell's demon-type models, in which the agent can exploit this auxiliary information to extract work, while the environment bears the cost of providing the ROM.

A canonical example is the Szilard box with measurement-based feedback control [13]. Let S_t and A_{t+1} represent the state of the Szilard box before and after the agent's transformation, and let S_t^{ROM} be the outcome of a (possibly noisy but feedback-free) measurement on the Szilard box performed by the environment. Without ROM, the environment channel reduces to an identity channel with zero work capacity. However, when ROM is present, the agent can condition its transformation on S_t^{ROM} and achieve a work rate upper bounded by the work capacity $\max_{p_{S_0}} I[S_0^{\text{ROM}}; S_0]$. For further generalizations involving non-Markovian ROM, see Refs. [13, 14, 17].

While the examples above either lack feedback or restrict it to ROM, the class of memoryless invariant channels also includes environments with genuine, albeit memoryless, feedback. An example is shown in the inset of Fig. 6, with binary alphabets $\mathcal{A} = \mathcal{S} = \{0, 1\}$ and transition matrix Φ^{env} given by

$$\phi^{\text{env}}(j|0) = \delta_{j,0}, \quad \phi^{\text{env}}(j|1) = \frac{1}{2}, \quad j \in \{0, 1\}. \quad (22)$$

For this environment, the work capacity is

$$C^{\text{work}}(\text{env}) = \frac{1}{2} \log_2 \left(\frac{3}{4} + \frac{1}{\sqrt{2}} \right) \approx 0.272 \quad (23)$$

in units of $k_B T \ln 2$, and can be achieved by a memoryless agent that selects action 0 with probability $1/\sqrt{2}$ at each step.

2. Product channels

For product environment channels, percepts are not influenced by actions, placing us in the tape setting. Consequently, to maximize the work rate, the optimal strategy is to maximize $H(A_t|M_t)$. This implies two conditions which can be made explicit by rewriting conditional entropy using the definition of mutual information: $H(A_t|M_t) = H(A_t) - I[A_t; M_t]$. By the non-negativity of mutual information, it then follows that agent models must perfectly randomize their actions [maximizing $\langle H(A_t) \rangle_t$] without retaining memory of them ($\langle I[A_t; M_t] \rangle_t = 0$). This results in $H(A_t|M_t) = \log_2 |\mathcal{A}|$. The second term in the work capacity expression, as shown in Table II, is the entropy rate of the percept process S ,

$$h(S) := \lim_{n \rightarrow \infty} \frac{H(S_{0:n})}{n}, \quad (24)$$

which was introduced by Shannon as the average uncertainty per symbol in a stochastic process [76, 77]. It is also known from the information-processing second law [78] that $\log_2 |\mathcal{A}| - h(S)$ (in units of $k_B T \ln 2$) represents an upper bound on the rate of expected extractable work from a stochastic process.

V. WORK-EFFICIENT AGENT MODELS

Determining work capacity is challenging, as it requires solving a nonlinear optimization problem [Eq. (19)]. However, established design principles exist that reach the fundamental energetic limits of information processing in the tape setting, specifically for agent models in $\mathbb{A}^{\text{zenv}}$ [26]. To generalize these results for agent models in $\mathbb{A}^{\text{zenv}}(d)$, which are restricted to finite memory of size d , we define the following subsets of agents:

Definition 6. For a given environment env and any $d \in \mathbb{N}$, we define the following subsets of $\mathbb{A}^{\text{zenv}}(d)$:

$$\mathbb{A}_{\text{MEA}}^{\text{zenv}}(d) := \argmax_{\text{agt}_{\mathcal{M}} \in \mathbb{A}^{\text{zenv}}(d)} \langle H(A_t|M_t) \rangle_t, \quad (25)$$

$$\mathbb{A}_{\text{pred}}^{\text{zenv}}(d) := \argmin_{\text{agt}_{\mathcal{M}} \in \mathbb{A}^{\text{zenv}}(d)} \langle I[A_{0:t+1} S_{0:t}; S_t | M_t] \rangle_t, \quad (26)$$

$$\mathbb{A}_{\text{eff}}^{\text{zenv}}(d) := \argmax_{\text{agt}_{\mathcal{M}} \in \mathbb{A}^{\text{zenv}}(d)} W(\text{agt}_{\mathcal{M}}, \text{env}). \quad (27)$$

Here and below, $\argmax$ and $\argmin$ denote the sets of agent models attaining the corresponding supremum or infimum. For fixed d , the parameter space of agent models is compact and, thus, the corresponding supremum or infimum is attained whenever the objective is continuous in the agent parameters.

$\mathbb{A}_{\text{MEA}}^{\text{zenv}}(d)$ is the subset of agent models in $\mathbb{A}^{\text{zenv}}(d)$ whose conditional action entropy $\langle H(A_t|M_t) \rangle_t$ reaches its theoretical maximum of $\log_2 |\mathcal{A}|$ in the Cesàro limit.

That is, agent models in $\mathbb{A}_{\text{MEA}}^{\text{zenv}}(d)$ must perfectly randomize their actions without retaining memory of them. The subscript MEA stands for maximum entropy actions.

$\mathbb{A}_{\text{pred}}^{\text{zenv}}(d)$ is the subset of predictive agent models; that is, they minimize their predictive gap (see Definition 4) subject to the constraint that their memory has size d .

$\mathbb{A}_{\text{eff}}^{\text{zenv}}(d)$ is the subset of work-efficient agent models that maximize the work rate [Eq. (17)].

We then have the following:

Theorem 2. For any product environment channel env and any agent memory size $d \in \mathbb{N}$,

$$\mathbb{A}_{\text{eff}}^{\text{zenv}}(d) = \mathbb{A}_{\text{MEA}}^{\text{zenv}}(d) \cap \mathbb{A}_{\text{pred}}^{\text{zenv}}(d). \quad (28)$$

We note that whenever agent models can become maximally predictive (see Definition 4), the results of Boyd *et al.* [26,29,79] apply. The proof of Theorem 2 (see Sec. F.4 of Supplemental Material [36]) relies on the expression for work capacity in product environments given in Table II. Given the resources for an agent model with $|\mathcal{M}| = d$ interacting with a product environment—that is, it observes a stochastic process—this result establishes two compatible principles for attaining or approaching the optimal work rate: (i) randomizing actions without retaining memory of them and (ii) employing predictive memory. These two principles can be directly linked to the two terms of the work rate Eq. (17), where the first principle ensures that $\langle H(A_t|M_t) \rangle_t$ is maximized, and the second principle ensures that $\langle H(S_t|M_t) \rangle_t$ is minimized.

A natural question then arises: What happens when actions influence future percepts—that is, in genuine percept-action loops? Based on Theorem 2, one might expect the same design principles to apply in these scenarios. In particular, it is not immediately clear why an efficient agent model should not minimize its a.m. predictive gap, as predicting its percepts reduces the uncertainty $H(S_t|M_t)$, which contributes negatively to the work rate. However, the following theorem demonstrates that there exist environments where neither of these previously identified design principles is compatible with work-efficient agent models.

Theorem 3. There exists an environment channel env such that for any $d \in \mathbb{N}$ the sets $\mathbb{A}_{\text{pred}}^{\text{zenv}}(d)$, $\mathbb{A}_{\text{MEA}}^{\text{zenv}}(d)$, and $\mathbb{A}_{\text{eff}}^{\text{zenv}}(d)$ are mutually exclusive.

The proof (see Sec. F.4 of Supplemental Material [36]) is based on a simple memoryless environment (see the inset of Fig. 6). This simplicity allows us to show that predictive agents retain complete information about their actions, $\langle H(A_t|M_t) \rangle_t = 0$, which results in a nonpositive work rate. In contrast, agents in $\mathbb{A}_{\text{MEA}}^{\text{zenv}}(d)$ and $\mathbb{A}_{\text{eff}}^{\text{zenv}}(d)$ achieve distinct but constant work rates for all d . The incompatibility of these work rates implies that the corresponding sets of agent models must be mutually exclusive.

This is a central result of this work and highlights a fundamental distinction between the tape setting and the percept-action-loop setting. While in the tape setting, agents are work efficient only if they are predictive, the percept-action-loop setting admits regimes where prediction is not optimal for resource harvesting: Work-efficient agents must not be predictive; they must forget aspects of their past actions to operate efficiently.

This phenomenon will be further investigated in the following example, which illustrates the transition from the tape setting to a genuine memoryful percept-action loop.

VI. EXAMPLE

A. Environment

We consider a family of environments $\text{env}^{(r)}$ indexed by a feedback parameter $r \in [0, 1]$ with binary actions and percepts $\mathcal{A} = \mathcal{S} = \{0, 1\}$. A two-state memory model of this environment is shown in Fig. 7(b).

Tape setting ($r = 0$): The environment's control protocol acts on the information carrier received from the agent, which carries the action. For $r = 0$, however, the environment ignores the action encoded on that system when preparing the next percept. Instead, the percept is determined solely by the environment's internal rule: If the previous percept was 0, the next percept is chosen uniformly at random, with $q(0) = q(1) = 1/2$; if the previous percept was 1, the next percept is deterministically 0. A typical percept sequence might look like ...010001010....

Because the environment ignores actions, this situation is equivalent to the tape setting of Fig. 1: Rather than having the agent and environment alternately apply control protocols to a single shared system, the environment prepares a stream of input systems carrying the percept symbols, and the agent processes these systems sequentially.

This percept process, also known as the golden-mean process [80], exhibits temporal correlations that can be exploited by a vigilant agent equipped with memory. By adapting its memory to the structure of the process, the agent can on average reduce its subjective entropy about incoming percepts $H(S_t|M_t)$, thereby increasing their thermodynamic value. The agent then maps these percepts to thermal states in order to extract work. Although these thermal states formally constitute actions, they do not actually act on the environment since in the present $r = 0$ setting they cannot influence future percepts.

The achievable work rate then equals the environment's work capacity:

$$C^{\text{work}}(\text{env}^{(r=0)}) = \frac{1}{3} \quad (29)$$

in units of $k_B T \ln 2$. This corresponds to a $[\log_2(27) - 4]/3k_B T \ln 2$ gap in the work rate relative to the memoryless agent, in agreement with the results reported in Ref. [26].

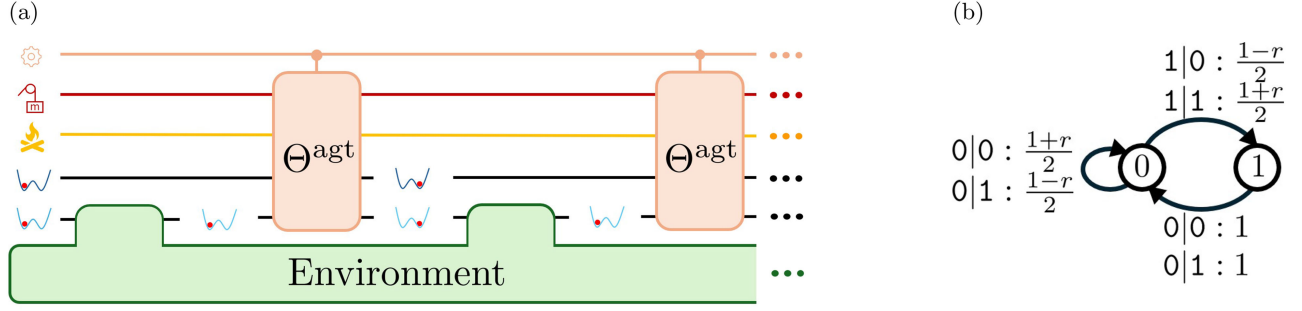


FIG. 7. Percept-action-loop implementation used in the example. Panel (a) shows the corresponding circuit representation, with time flowing from left to right. The horizontal lines correspond, from top to bottom, to the agent's read-only memory, a work reservoir, a thermal reservoir, the agent's memory (two states), and the percept-action register (two states). For concreteness, we assume that the agent's memory and the percept-action register are physically encoded in information carriers given by overdamped Brownian particles in a controllable potential (as depicted). In each round, the agent's information processing described by its transition matrix Θ^{agt} is realized through an isothermal control protocol represented by the box labeled Θ^{agt} . The protocol is designed so that the controllable potentials before and after the implementation of Θ^{agt} coincide (see Refs. [26,29] for details). The agent's read-only memory stores a description of the percept-action loop ($\text{agt}_{\mathcal{M}}, \text{env}^{(r)}$) according to which the control protocol is tailored to the environment. The environment channel constitutes the only nonequilibrium resource available to the agent. Panel (b) shows the two-state finite-state machine of the environment, which specifies how actions are mapped to percepts (transition labels follow the format percept, |, action, :, transition probability). The feedback parameter $r \in [0, 1]$ controls the influence of actions on percepts: For $r = 0$, actions have no influence corresponding to the tape setting. For all $r > 0$, percepts depend on previous actions, yielding a genuine percept-action loop.

Percept-action-loop setting ($r > 0$): For $r > 0$, the environment includes feedback. The probability q now depends on the previous action a , with $q(0) = [(1 \pm r)/2]$, where $+$ (respectively, $-$) applies to $a = 0$ (respectively, $a = 1$). Consequently, $q(1)$ equals $1 - q(0)$. As in the tape setting, a percept 1 is always followed by a percept 0. However, as r increases, the agent gains more control over percept generation, reaching full determinism at $r = 1$.

B. Agent models

In practice, for a given percept-action loop ($\text{agt}_{\mathcal{M}}, \text{env}^{(r)}$), one possibility for an agent model to

achieve the work rate in Eq. (17) is by (i) precomputing the optimal thermodynamic protocols [26,29] for the cyclostationary process resulting from $(\text{agt}_{\mathcal{M}}, \text{env}^{(r)})$, (ii) storing these optimal protocol parameters in its read-only memory, and (iii) implementing them during the percept-action loop.

Figure 8(a) shows a numerical estimate of $W^{(2)}(\text{env}^{(r)})$, the optimal work rate among all agent models with memory size $d = 2$ (green points). We also find that the numerically optimized values of $W^{(d)}(\text{env}^{(r)})$ are indistinguishable for $d = 2, 3, 4$, indicating that no improvement occurs over the tested memory sizes. For comparison, we consider the numerically optimized work rate of memoryless agent

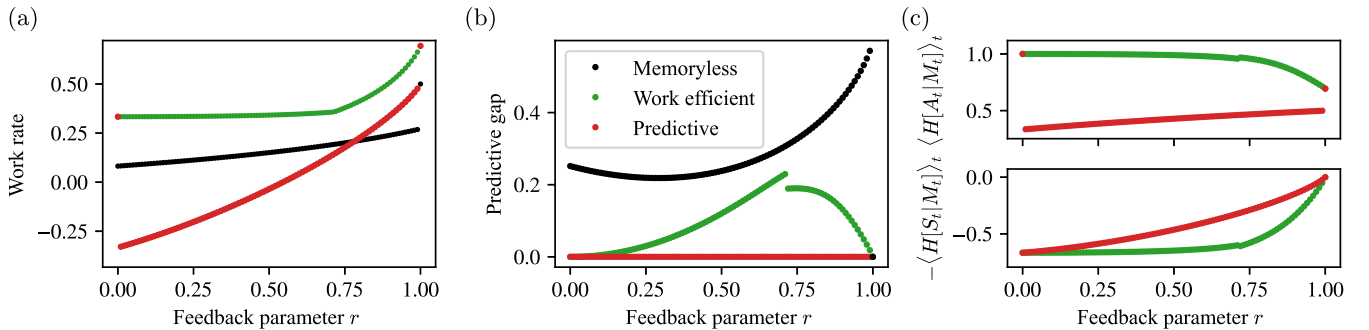


FIG. 8. Work production in a percept-action loop with memory. The environment under consideration is shown in Fig. 7(b). (a) Numerically optimized work rates of general two-memory-state agents (green) and memoryless agents (black), together with the analytically determined maximal work rate among predictive agents, which is independent of the finite memory size $d \geq 2$ (red), and (b) their corresponding predictive gaps (see Definition 4). The observed discontinuity in the predictive gap of the numerically optimized general agents is associated with a qualitative change in their memory mechanism. (c) Illustration of the trade-off between the percept term $H(S_t|M_t)$ and the action term $H(A_t|M_t)$ in the work rate [see Eq. (17)]. Compared with the numerically optimized general agent, the predictive agent has both a lower percept term and a lower action term. Because the reduction in the action term exceeds that in the percept term, the predictive agent attains a lower work rate.

models (black points), and an analytic solution for the optimal work rate among predictive agent models with $d = 2$ memory states, $\mathbb{A}_{\text{pred}}^{\text{env}(r)}(2)$ (red points). For predictive agent models, additional memory cannot increase the work rate (see Sec. G of Supplemental Material [36] for details).

For $r = 0$, predictive agents attain the maximum work rate $W^{(2)}(\text{env}^{(0)})$, consistent with Theorem 2. For $0 < r < 1$, the optimal predictive work rate drops sharply, and our numerical results indicate that predictive agent models cease to be work efficient. This illustrates the separation between prediction and work efficiency established by Theorem 3. At $r = 1$, the environment becomes deterministic in response to actions, allowing predictive agent models to once again attain the maximum work rate.

As shown in Fig. 8(b), for $0 < r < 1$ the numerically optimized general agents strike a balance between prediction and forgetting: They are less predictive than maximally predictive agents but more predictive than memoryless ones. Notably, for every $r \in [0, 1]$ and every finite $d \geq 2$, agents in $\mathbb{A}_{\text{pred}}^{\text{env}(r)}(d)$ are maximally predictive, meaning their predictive gap vanishes, $\langle I[A_{0:t+1}S_{0:t}; S_t|M_t] \rangle_t = 0$. Furthermore, the predictive gap of the numerically optimized general agents exhibits a discontinuous transition near $r = 0.71$ associated with a qualitative change in their memory mechanism (see Sec. G of Supplemental Material [36] for details).

The numerically observed separation between the optimized general agents and the optimal predictive agents for $0 < r < 1$ can be understood from the two terms constituting the work rate [see Fig. 8(c)]: Reducing the predictive gap requires retaining information about past actions that help predict future percepts. While this lowers the percept entropy $H(S_t|M_t)$ (increasing the work rate), it also lowers the action entropy $H(A_t|M_t)$ (reducing the work rate). Conversely, maximizing action entropy without retaining memory of past actions increases $H(A_t|M_t)$ but may hinder percept prediction or drive the environment into a less-predictable regime, thereby increasing $H(S_t|M_t)$.

Thus, the two design principles for work-efficient agents in the tape setting can no longer be pursued independently in percept-action loops; a trade-off emerges between predictive memory and action forgetfulness.

VII. DISCUSSION AND FUTURE DIRECTIONS

Predicting future observations is a central theme across various fields, including Bayesian and active inference [82], predictive analytics [83], computational mechanics [84], and chaos theory [85]. It also plays a crucial role in modern machine learning, particularly in transformer models and large language models, which are designed to predict future states in a sequence [86].

However, as we show in this work by analyzing the fundamental limits of information processing in percept-action loops, the mere act of remembering the past to predict the future has thermodynamic consequences.

This phenomenon arises specifically in percept-action loops with genuine feedback. In such settings, when predicting percepts requires remembering past actions, a trade-off emerges, and the goals of prediction and work efficiency diverge: As we prove and demonstrate with examples, there exist environments in which any agent that maximizes work efficiency must necessarily forget certain aspects of its past actions and, therefore, cannot be maximally predictive.

The present framework invites multiple lines of future research, including generalizations to goal-directed behavior, analyses of energetic dissipation, and extensions into the quantum domain.

- (i) Agents with goals: In this work, we considered classes of agents with implicit objectives, such as maximizing the work rate or predictive power. A natural next step is to investigate agents with specific goals within our framework. One approach is to fix a desired percept-action behavior, which corresponds to specifying an agent channel. Then, the energetic limits of the agent's behavior can be determined by optimizing over all models that implement this channel (see related ideas in the tape setting [30,32]). Alternatively, one could emulate a reinforcement learning scenario by encoding rewards as predictable (i.e., low-entropy) percepts. In this case, an agent aiming to maximize its work rate could be guided toward desired behaviors through suitable reward design.
- (ii) Dissipation in percept-action loops: If one considers that both the agent and environment thermodynamically implement their respective channels, the agent's positive work rate implies a corresponding work cost for the environment. In such a setting, the environment converts work into structured correlations, while the agent converts those correlations back into work. For memoryless channels, this conversion can happen without dissipation, with the energetic cost of implementing the environment channel known in the quantum context as the thermodynamic capacity [87,88], equaling the work capacity. However, for channels with memory, it remains an open question whether for any percept-action loop (agt, env) the agent's maximum work rate can match the environment's minimum work cost. Any gap between these values would imply intrinsic entropy production in percept-action loops.
- (iii) Quantum work capacity: A natural extension of this work is to explore quantum generalizations of work capacity. Our framework admits a quantization by replacing classical channels with quantum combs [89], enabling analysis of percept-action loops in the quantum domain. This allows for studying fundamental quantum limits on work extraction and the design of quantum-enhanced agents [30–33,90,91].

Beyond these directions, the findings of this work may also hold more speculative longer-term relevance: (i) As artificial intelligence systems begin to approach thermodynamic efficiency limits, memory architectures and action policies may need to be designed with energetic constraints in mind [72,92]. In such regimes, our results suggest that departing from predictive design principles can be essential for optimizing energy use. (ii) In biological systems [93–97], behaviors that appear sub-optimal from a predictive standpoint may, in fact, reflect underlying thermodynamic constraints. For example, deliberately forgetting aspects of past actions—at the cost of predictive accuracy—may represent strategies for maximizing energetic efficiency.

More broadly, our work opens the door to a search for new energetic design principles tailored to percept-action loops with feedback. Such considerations may inform novel organizational principles for biological and artificial agents [98], moving beyond the predictive paradigm [29,63,99,100].

ACKNOWLEDGMENTS

L. J. F. thanks Benjamin Morris and Andrew Garner for early discussions that helped shape the direction of this work. The authors are grateful to the anonymous reviewers for their insightful and constructive feedback, which helped to significantly improve the quality of the manuscript. This research was funded in whole or in part by the Austrian Science Fund (FWF) [SFB BeyondC F7102, DOI: 10.55776/F71; WIT9503323, DOI: 10.55776/WIT9503323]. We gratefully acknowledge support from the European Union (ERC Advanced Grant QuantAI No. 101055129). L. J. F. acknowledges support by the Austrian Research Promotion Agency (FFG) and the European Union via NextGeneration EU under Contract No. FO999921407 (HDcode).

The views and opinions expressed in this article are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council—neither the European Union nor the granting authority can be held responsible for them.

DATA AVAILABILITY

The code used to generate the numerical data underlying Fig. 6 and the example in Sec. VI is openly available in Zenodo [101].

-
- [1] P. Lennie, *The cost of cortical computation*, *Curr. Biol.* **13**, 493 (2003).
 - [2] L. Yu and Y. Yu, *Energy-efficient neural information processing in individual neurons and neuronal networks*, *J. Neurosci. Res.* **95**, 2253 (2017).

- [3] C. P. Kempes, D. Wolpert, Z. Cohen, and J. Pérez-Mercader, *The thermodynamic efficiency of computations made in cells across the range of life*, *Phil. Trans. R. Soc. A* **375**, 20160343 (2017).
- [4] N. C. Thompson, K. Greenewald, K. Lee, G. F. Manso, *The computational limits of deep learning*, *arXiv:2007.05558*.
- [5] J. McDonald, B. Li, N. Frey, D. Tiwari, V. Gadepally, and S. Samsi, *Great power, great responsibility: Recommendations for reducing energy for training language models*, in *Findings of the Association for Computational Linguistics: NAACL 2022*, edited by M. Carpuat, M.-C. de Marneffe, and I. V. Meza Ruiz (Association for Computational Linguistics, Stroudsburg, PA, USA, 2022), pp. 1962–1970.
- [6] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz, *Experimental verification of Landauer’s principle linking information and thermodynamics*, *Nature (London)* **483**, 187 (2012).
- [7] J. Hong, B. Lambson, S. Dhuey, and J. Bokor, *Experimental test of Landauer’s principle in single-bit operations on nanomagnetic memory bits*, *Sci. Adv.* **2**, e1501492 (2016).
- [8] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin, *Experimental realization of a Szilard engine with a single electron*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13786 (2014).
- [9] G. Paneru, D. Y. Lee, T. Tlusty, and H. K. Pak, *Lossless Brownian information engine*, *Phys. Rev. Lett.* **120**, 020601 (2018).
- [10] U. Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [11] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, *Thermodynamics of information*, *Nat. Phys.* **11**, 131 (2015).
- [12] F. J. Cao and M. Feito, *Thermodynamics of feedback controlled systems*, *Phys. Rev. E* **79**, 041118 (2009).
- [13] T. Sagawa and M. Ueda, *Nonequilibrium thermodynamics of feedback control*, *Phys. Rev. E* **85**, 021104 (2012).
- [14] S. Ito and T. Sagawa, *Information thermodynamics on causal networks*, *Phys. Rev. Lett.* **111**, 180603 (2013).
- [15] J. M. Horowitz and M. Esposito, *Thermodynamics with continuous information flow*, *Phys. Rev. X* **4**, 031015 (2014).
- [16] D. Hartich, A. C. Barato, and U. Seifert, *Stochastic thermodynamics of bipartite systems: Transfer entropy inequalities and a Maxwell’s demon interpretation*, *J. Stat. Mech.* (2014) P02016.
- [17] N. Shiraishi and T. Sagawa, *Fluctuation theorem for partially masked nonequilibrium dynamics*, *Phys. Rev. E* **91**, 012130 (2015).
- [18] D. Mandal and C. Jarzynski, *Work and information processing in a solvable model of Maxwell’s demon*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11641 (2012).
- [19] D. Mandal, H. T. Quan, and C. Jarzynski, *Maxwell’s refrigerator: An exactly solvable model*, *Phys. Rev. Lett.* **111**, 030602 (2013).
- [20] A. C. Barato and U. Seifert, *An autonomous and reversible Maxwell’s demon*, *Europhys. Lett.* **101**, 60001 (2013).
- [21] A. C. Barato and U. Seifert, *Unifying three perspectives on information processing in stochastic thermodynamics*, *Phys. Rev. Lett.* **112**, 090601 (2014).

- [22] J. Hoppenau and A. Engel, *On the energetics of information exchange*, *Europhys. Lett.* **105**, 50002 (2014).
- [23] N. Merhav, *Sequence complexity and work extraction*, *J. Stat. Mech.* (2015) P06037.
- [24] A. B. Boyd, D. Mandal, and J. P. Crutchfield, *Correlation-powered information engines and the thermodynamics of self-correction*, *Phys. Rev. E* **95**, 012152 (2017).
- [25] Andrew J. Garner, J. Thompson, V. Vedral, and M. Gu, *Thermodynamics of complexity and pattern manipulation*, *Phys. Rev. E* **95**, 042140 (2017).
- [26] A. B. Boyd, D. Mandal, and J. P. Crutchfield, *Thermodynamics of modularity: Structural costs beyond the Landauer bound*, *Phys. Rev. X* **8**, 031036 (2018).
- [27] A. J. Garner, *The fundamental thermodynamic bounds on finite models*, *Chaos* **31**, 063131 (2021).
- [28] L. He, A. Pradana, J. W. Cheong, and L. Y. Chew, *Information processing second law for an information ratchet with finite tape*, *Phys. Rev. E* **105**, 054131 (2022).
- [29] A. B. Boyd, J. P. Crutchfield, and M. Gu, *Thermodynamic machine learning through maximum work production*, *New J. Phys.* **24**, 083040 (2022).
- [30] T. J. Elliott, M. Gu, Andrew J. Garner, and J. Thompson, *Quantum adaptive agents with efficient long-term memories*, *Phys. Rev. X* **12**, 011007 (2022).
- [31] R. C. Huang, P. M. Riechers, M. Gu, and V. Narasimhachar, *Engines for predictive work extraction from memoryful quantum stochastic processes*, *Quantum* **7**, 1203 (2023).
- [32] J. Thompson, P. M. Riechers, Andrew J. Garner, T. J. Elliott, and M. Gu, *Energetic advantages for quantum agents in online execution of complex strategies*, *Phys. Rev. Lett.* **135**, 160402 (2025).
- [33] G. Zambon and G. Adesso, *Quantum processes as thermodynamic resources: The role of Non-Markovianity*, *Phys. Rev. Lett.* **134**, 200401 (2025).
- [34] Y. Ephraim and N. Merhav, *Hidden Markov processes*, *IEEE Trans. Inf. Theory* **48**, 1518 (2002).
- [35] R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York, 1968).
- [36] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/7nds-tjr8> for mathematical background, detailed derivations and proofs of the main results, and analytical and numerical details of the example, which includes Refs. [37–62].
- [37] R. M. Gray, *Probability, Random Processes, and Ergodic Properties* (Springer, New York, 2009), Vol. 1.
- [38] Note that the labeling convention is such that the value left of the colon in the subscript is included in the sequence, while the value to the right is not.
- [39] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 2005).
- [40] R. M. Gray, *Entropy and Information Theory*, first, corrected ed. (Springer Science & Business Media, New York, 2011).
- [41] D. Hankerson, G. A. Harris, and P. D. Johnson Jr, *Introduction to Information Theory and Data Compression* (CRC Press, Boca Raton, 2003).
- [42] R. W. Yeung, *A First Course in Information Theory* (Springer US, New York, NY, 2002).
- [43] A. Kolchinsky, *A novel approach to the partial information decomposition*, *Entropy* **24**, 403 (2022).
- [44] R. W. Yeung, *A new outlook on Shannon's information measures*, *IEEE Trans. Inf. Theory* **37**, 466 (1991).
- [45] H. K. Ting, *On the amount of information*, *Theory Probab. Appl.* **7**, 439 (1962).
- [46] H. C. Tijms, *A First Course in Stochastic Models* (John Wiley and Sons, Chichester, England, 2003).
- [47] R. G. Gallager, *Discrete Stochastic Processes* (Springer, New York, 1996).
- [48] R. E. Edwards, *Fourier Series: A Modern Introduction Volume 1*, 2nd ed. (Springer, New York, NY, 1979).
- [49] J. R. Munkres, *Topology: Pearson New International Edition* (Pearson Higher Ed, Harlow, England, 2013).
- [50] J. C. Kieffer and M. Rahe, *Markov channels are asymptotically mean stationary*, *SIAM J. Math. Anal.* **12**, 293 (1981).
- [51] J. Pearl, *Bayesian networks: A model of self-activated memory for evidential reasoning*, *Proceedings of the Annual Meeting of the Cognitive Science Society* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1985), 7.
- [52] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann Publishers Inc., San Mateo, CA, 1988).
- [53] T. Verma and J. Pearl, *Causal networks: Semantics and expressiveness*, in *Machine Intelligence and Pattern Recognition*, Uncertainty in Artificial Intelligence Vol. 9, edited by R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer (North-Holland, Amsterdam, Netherlands, 1990), pp. 69–76.
- [54] D. Janzing and B. Schölkopf, *Causal inference using the algorithmic Markov condition*, *IEEE Trans. Inf. Theory* **56**, 5168 (2010).
- [55] S. Lauritzen, *Graphical Models* (Clarendon Press, Oxford, 1996).
- [56] In fact, V can be any sufficient statistic of WX about YZ but for our purposes, setting $V = YZ$ is the simplest. This solution was also pointed out in Fig. 13(a) of Ref. [57].
- [57] A. S. Klyubin, D. Polani, and C. L. Nehaniv, *Representations of space and time in the maximization of information flow in the perception-action loop*, *Neural Comput.* **19**, 2387 (2007).
- [58] R. J. Evans, *Graphs for margins of Bayesian networks*, *Scand. J. Stat. Theory Appl.* **43**, 625 (2016).
- [59] N. Tishby and D. Polani, *Information Theory of Decisions and Actions*, in *Perception-Action Cycle: Models, Architectures, and Hardware*, edited by V. Cutsuridis, A. Hussain, and J. G. Taylor (Springer, New York, 2010), pp. 601–636.
- [60] C. Salge, C. Glackin, and D. Polani, *Empowerment—An Introduction*, in *Guided Self-Organization: Inception*, edited by M. Prokopenko (Springer, Berlin, 2014), pp. 67–114.
- [61] N. Ay and K. Zahedi, *On the causal structure of the sensorimotor loop*, in *Guided Self-Organization: Inception* (Springer, Berlin, 2014), pp. 261–294.
- [62] J. G. Kemeny and J. L. Snell, *Finite Markov Chains* (van Nostrand Princeton, NJ, 1960), Vol. 26.
- [63] N. Barnett and J. P. Crutchfield, *Computational mechanics of input–output processes: Structured transformations and the epsilon-transducer*, *J. Stat. Phys.* **161**, 404 (2015).

- [64] F. Rosas, A. Boyd, and M. Baltieri, *AI in a vat: Fundamental limits of efficient world modelling for agent sandboxing and interpretability*, [arXiv:2504.04608](#).
- [65] R. B. Ash, *Basic Probability Theory* (Dover Publications, Mineola, NY, 2008).
- [66] M. Iosifescu, *Finite Markov Processes and Their Applications* (Dover Publications, Mineola, NY, 2007).
- [67] R. Landauer, *Information is physical*, *Phys. Today* **44**, 23 (1991).
- [68] S. Deffner and C. Jarzynski, *Information processing and the second law of thermodynamics: An inclusive, Hamiltonian approach*, *Phys. Rev. X* **3**, 041003 (2013).
- [69] A. Kolchinsky and D. H. Wolpert, *Dependence of dissipation on the initial distribution over states*, *J. Stat. Mech.* (2017) 083202.
- [70] N. Shiraishi, K. Funo, and K. Saito, *Speed limit for classical stochastic processes*, *Phys. Rev. Lett.* **121**, 070601 (2018).
- [71] A. Kolchinsky and D. H. Wolpert, *Work, entropy production, and thermodynamics of information under protocol constraints*, *Phys. Rev. X* **11**, 041024 (2021).
- [72] D. H. Wolpert, J. Korb, C. W. Lynn, F. Tasnim, J. A. Grochow, G. Kardeş, J. B. Aimone, V. Balasubramanian, E. De Giuli, D. Doty *et al.*, *Is stochastic thermodynamics the key to understanding the energy costs of computation?*, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2321112121 (2024).
- [73] An illustrative example of a sequence $(a_t)_t$ where the Cesàro limit $\langle a_t \rangle_t$ fails to exist is 0110000..., where one zero is followed by twice as many ones, followed again by twice as many zeros, and so on. This results in an oscillating arithmetic mean $1/n \sum_{t=1}^n a_t$ as $n \rightarrow \infty$.
- [74] P. A. Ortega and D. A. Braun, *Thermodynamics as a theory of decision-making with information-processing costs*, *Proc. R. Soc. A* **469**, 20120683 (2013).
- [75] Also referred to as a noiseless channel in communication theory.
- [76] C. E. Shannon, *A mathematical theory of communication*, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [77] If the percept process is a stationary finite-state Markov chain, a closed-form expression for the entropy rate exists [76].
- [78] A. B. Boyd, D. Mandal, and J. P. Crutchfield, *Identifying functional thermodynamics in autonomous maxwellian ratchets*, *New J. Phys.* **18**, 023049 (2016).
- [79] A. B. Boyd, J. P. Crutchfield, M. Gu, and F. C. Binder, *Thermodynamic overfitting and generalization: Energetics of predictive intelligence*, *New J. Phys.* **27**, 063901 (2025).
- [80] The name stems from the following observation. Let q be the probability that the next symbol is 0 given that the current symbol is 0. Then, the value of q that maximizes the entropy rate of the golden-mean process equals $(\sqrt{5} - 1)/2$, the reciprocal of the golden ratio, and the corresponding entropy rate is $\log_2[(1 + \sqrt{5})/2]$, the binary logarithm of the golden ratio [81].
- [81] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding* (Cambridge University Press, Cambridge, England, 2021).
- [82] T. Parr, G. Pezzulo, and K. J. Friston, *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior* (MIT Press, Cambridge, MA, 2022).
- [83] D. T. Larose and C. D. Larose, *Data Mining and Predictive Analytics*, 2nd ed. (John Wiley & Sons, Hoboken, NJ, 2015).
- [84] J. P. Crutchfield, *Between order and chaos*, *Nat. Phys.* **8**, 17 (2012).
- [85] S. Boccaletti, C. Grebogi, Y.-C. Lai, H. Mancini, and D. Maza, *The control of chaos: Theory and applications*, *Phys. Rep.* **329**, 103 (2000).
- [86] T. Lin, Y. Wang, X. Liu, and X. Qiu, *A survey of transformers*, *AI Open* **3**, 111 (2022).
- [87] M. Navascués and L. P. García-Pintos, *Nonthermal quantum channels as a thermodynamical resource*, *Phys. Rev. Lett.* **115**, 010405 (2015).
- [88] P. Faist, M. Berta, and F. Brandão, *Thermodynamic capacity of quantum processes*, *Phys. Rev. Lett.* **122**, 200601 (2019).
- [89] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Theoretical framework for quantum networks*, *Phys. Rev. A* **80**, 022339 (2009).
- [90] M. Gu, K. Wiesner, E. Rieper, and V. Vedral, *Quantum mechanics can reduce the complexity of classical models*, *Nat. Commun.* **3**, 762 (2012).
- [91] V. Dunjko, J. M. Taylor, and H. J. Briegel, *Quantum-enhanced machine learning*, *Phys. Rev. Lett.* **117**, 130501 (2016).
- [92] C. H. Bennett, *The thermodynamics of computation—a review*, *Int. J. Theor. Phys.* **21**, 905 (1982).
- [93] S. Ito and T. Sagawa, *Maxwell's demon in biochemical signal transduction with feedback loop*, *Nat. Commun.* **6**, 7498 (2015).
- [94] H. Shimazaki, *Neurons as an Information-theoretic engine*, [arXiv:1512.07855](#).
- [95] T. E. Ouldridge, C. C. Govern, and P. R. ten Wolde, *Thermodynamics of computational copying in biochemical systems*, *Phys. Rev. X* **7**, 021004 (2017).
- [96] R. A. Brittain, N. S. Jones, and T. E. Ouldridge, *Biochemical Szilard engines for memory-limited inference*, *New J. Phys.* **21**, 063022 (2019).
- [97] A. J. Tjalma, V. Galstyan, J. Goedhart, L. Slim, N. B. Becker, and P. R. Ten Wolde, *Trade-offs between cost and information in cellular prediction*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2303078120 (2023).
- [98] A. Rupe and J. P. Crutchfield, *On principles of emergent organization*, *Phys. Rep.* **1071**, 1 (2024).
- [99] K. Friston, *Life as we know it*, *J. R. Soc. Interface* **10**, 20130475 (2013).
- [100] H. J. Briegel and T. Müller, *Projective Simulation in Action: Quantum-Mechanical Perspectives on the Problem of Agency* (Springer Nature, Cham, Switzerland, 2025).
- [101] L. J. Fiderer, P. C. Barth, I. D. Smith, and H. J. Briegel, *Code for “Information Thermodynamics of Agents: The Work Capacity of Channels with Memory”*, Zenodo, [10.5281/zenodo.21905519](#) (2026).