ARTICLE **Communicated by Christopher Whyte**

Active Inference and Intentional Behavior

Karl J. Friston
k.friston@ucl.ac.uk
Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London WC1N 3AR, U.K., and VERSES AI Research Lab, Los Angeles, California, 90016, U.S.

Tommaso Salvatori
tommaso.salvatori@verses.ai
VERSES AI Research Lab, Los Angeles, California, 90016, U.S.

Takuya Isomura
takuya.isomura@riken.jp
Brain Intelligence Theory Unit, RIKEN Center for Brain Science, Wako, Saitama, 351-0198, Japan

Alexander Tschantz
alexander.tschantz@verses.ai
VERSES AI Research Lab, Los Angeles, California, 90016, U.S.A.

Alex Kiefer
alex.kiefer@verses.ai
VERSES AI Research Lab, Los Angeles, California, 90016, U.S.A.

Tim Verbelen
tim.verbelen@verses.ai
Magnus Koudahl
magnus.koudahl@verses.ai
VERSES AI Research Lab, Los Angeles, California, 90016, U.S.A.

Aswin Paul
aswin.paul@verses.ai
VERSES AI Research Lab, Los Angeles, California, 90016, U.S., Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, VIC 3800 Australia, and IITB-Monash Research Academy, Mumbai-4000076, India

Thomas Parr
thomas.parr.12@alumni.ucl.ac.uk
Nuffield Department of Clinical Neurosciences, University of Oxford, OX3 9DU, U.K.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

Neural Computation **37, 666–700** (2025)
https://doi.org/10.1162/neco_a_01738
© 2025 Massachusetts Institute of Technology

Adeel Razi
adeel.razi@monash.edu
Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, VIC 3800 Australia; Monash Biomedical Imaging, Monash University, Clayton, VIC 3800 Australia; and CIFAR Azrieli Global Scholars Program, Toronto, ON M5G 1M1

Brett J. Kagan
Brett@corticallabs.com
Cortical Labs, Melbourne, Australia

Christopher L. Buckley
Christopher.Buckley@verses.ai
VERSES AI Research Lab, Los Angeles, California, 90016, U.S.A.

Maxwell J. D. Ramstead
maxwell.d.ramstead@gmail.com
Queen Square Institute of Neurology, University College London, Queen Square, London, WC1N 3BG, United Kingdom

Recent advances in theoretical biology suggest that key definitions of basal cognition and sentient behavior may arise as emergent properties of in vitro cell cultures and neuronal networks. Such neuronal networks reorganize activity to demonstrate structured behaviors when embodied in structured information landscapes. In this article, we characterize this kind of self-organization through the lens of the free energy principle, that is, as self-evidencing. We do this by first discussing the definitions of reactive and sentient behavior in the setting of active inference, which describes the behavior of agents that model the consequences of their actions. We then introduce a formal account of intentional behavior that describes agents as driven by a preferred end point or goal in latent state-spaces. We then investigate these forms of (reactive, sentient, and intentional) behavior using simulations. First, we simulate the in vitro experiments, in which neuronal cultures modulated activity to improve gameplay in a simplified version of Pong by implementing nested, free energy minimizing processes. The simulations are then used to deconstruct the ensuing predictive behavior, leading to the distinction between merely reactive, sentient, and intentional behavior with the latter formalized in terms of inductive inference. This distinction is further studied using simple machine learning benchmarks (navigation in a grid world and the Tower of Hanoi problem) that show how quickly and efficiently adaptive behavior emerges under an inductive form of active inference.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

1 Introduction

In 2022, a paper was published that claimed to demonstrate sentient behavior in a neuronal culture grown in a dish (an in vitro neuronal network) (Kagan et al., 2022). This work formally defined sentience as being "responsive to sensory impressions through adapative internal processes" consistent with prior literature (Friston et al., 2020; Trewavas et al., 2020). The behavior in question was the elicited emergence of controlled movements of a paddle to hit a ball—and thereby play Pong. This study has several sources of inspiration that speak to the notion of basal cognition (Fields et al., 2021; Levin, 2019; Manicka & Levin, 2019; and related work, e.g., Masumori et al., 2015). In particular, the hypothesis that adaptive and predictive behavior would emerge spontaneously was based on earlier work showing that in vitro neuronal cultures could be described as minimizing variational free energy (Isomura & Friston, 2018) and thereby evince active inference and learning. This application of the free energy principle (FEP) to neuronal cultures was subsequently validated empirically (Isomura et al., 2023) in the sense that changes in neuronal activity and synaptic efficacy—that underwrite learning—could be predicted quantitatively, as a variational free energy minimizing process. So are these findings remarkable, or were they predictable?

In one sense, these results were entirely predictable. Indeed, they were predictable from the FEP, which states that any two networks—that are coupled in a certain sparse fashion—will come to manifest a generalized synchrony (Friston et al., 2021; Palacios et al., 2019). More formally, the FEP states that if the probability density that underwrites the dynamics of coupled random dynamical systems contains a Markov blanket—which shields internal states from external states, given blanket (sensory and active) states—then internal states will look as if they track the statistics of external states—or, more precisely, as if they encode the parameters of a variational density (or best guess about) external states beyond the blanket. Empirically, this synchronization was observed when the neuronal cultures learned to play Pong. However, the FEP goes further and says that the internal and active states (together, autonomous states) of either network can be described as minimizing a variational free energy functional. This functional is exactly the same used to optimize generative models in statistics and machine learning (Winn & Bishop, 2005). On this reading, one can interpret the autonomous states—of a network, particle, or person possessing an internal state—as minimizing variational free energy or surprise (a.k.a., self-information) in the sensory information representing external states. Equivalently this may be described as maximizing Bayesian model evidence (a.k.a., the marginal likelihood of sensory states). This leads to an implicit teleology, in the sense that one can describe self-organization in terms of self-evidencing (Hohwy, 2016) that entails active inference and learning, planning, purpose, intentions, and, perhaps, sentience. The underlying

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

free energy–minimizing processes—and their teleological interpretation—are the focus of this article.

The results reported in Kagan et al. (2022) were considered by some to be unremarkable for a different reason: learning to play (Atari) games like Pong was something that had been accomplished with machine learning systems years earlier using neural networks and both model-free and model-based reinforcement learning (RL; Mnih et al., 2015; Schrittwieser et al., 2019; Ye et al., 2021). So, what is remarkable about a neuronal network reproducing a similar kind of behavior? It is remarkable because one cannot use the RL paradigm to explain the emergence of self-evidencing behavior seen in vitro. This follows from the fact that rewarding or punishing an in vitro neuronal network, in a behaviorist sense, is currently technically infeasible. Even should a given in vitro neuronal network contain privileged reward or punishment pathways, methods for identifying, accessing, and interacting with these pathways have not been established, especially in a reproducibly real-time, closed-loop fashion. However, the FEP theorist knows exactly what a self-evidencing network finds aversive: surprise and unpredictability. This was a rationale for delivering unpredictable noise to the sensory electrodes of the cell culture (or restarting the game in an unpredictable way), whenever the neuronal network failed to hit the ball (Kagan et al., 2022). Some found the results reported in Kagan et al. (2022) remarkable, but not in a good way: they disagreed with the claim that the behavior could be described as "sentient" (Balci et al., 2023). Here, we hope to make sense of the notion of sentient behavior in terms of Bayesian belief updating, where "sentient behavior" denotes the capacity to generate appropriate responses to sensory perturbations (as opposed to merely reactive behavior; Kagan, Razi et al., 2023). We pursue the narrative established by the cell culture experiments above to illustrate why Pong-playing behavior was considered sentient under the definition used, as opposed to reactive. In brief, we consider a bright line between actions based on the predictions of a generative model that does, and does not, entail the consequences of action.

Specifically, this article differentiates three kinds of behavior: reactive, sentient, and intentional. Here, reactive behavior characterizes agents that perform actions merely in response to an observation. Sentient behavior can be read as planning actions under a generative model that includes the long-term consequences of behavior, and intentional behavior as planning according to an intended goal, where goals are specified as latent states in the generative model. The first two have formulations that have been extensively studied in the literature, under the framework of model-free reinforcement learning (RL) the first, and model-based RL and active inference the second. In model-free RL, the system selects actions using either a lookup table (Q-learning), or a neural network (deep Q-learning); in model-based RL, an action is selected according to a specific value function over policies, often explored via tree searches (Ha & Schmidhuber, 2018; Hafner

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

et al., 2019; Ye et al., 2021). In standard active inference, the action selection depends on the expected free energy of policies (see equation 2.1), where the expectation is over observations in the future that become random variables. This means that preferred outcomes—that subtend expected cost and risk—are prior beliefs that constrain the implicit planning as inference (Attias, 2003; Botvinick & Toussaint, 2012; Van Dijk & Polani, 2013). Things that evince this kind of behavior can hence be described as planning their actions, based on a generative model of the consequences of those actions (Attias, 2003; Botvinick & Toussaint, 2012; Da Costa, Parr et al., 2020). It was this sense in which the behavior of the cell cultures was considered sentient. (For a more detailed explanation of the three kinds of behavior and concrete examples that help distinguish among them, we refer to Figure 1.)

This article introduces the third kind of behavior based on inductive inference. This form of sentient behavior—described in terms of Bayesian mechanics (Friston, Da Costa, Sajid et al., 2023; Friston et al., 2022; Ramstead et al., 2023)—can be augmented with intended end points or goals. This leads to a novel kind of sentient behavior that not only predicts the consequences of its actions, but is also able to select them to reach a goal state that may be many steps in the future. This kind of behavior, which we call *intentional behavior*, generally requires some form of backward induction (Camerer, 1997; Hure et al., 2020) of the kind found in dynamic programming (Bellman, 1952; Da Costa, Sajid et al., 2020; Paul et al., 2023; Sutton et al., 1999). In short, backward induction involves starting from the intended goal state and working backward, inductively, to the current state of affairs, in order to plan moves to that goal state. Backward induction was applied to the partially observable setting and explored in the context of active inference in Paul et al. (2023). In that work, dynamic programming was shown to be more efficient than traditional planning methods in active inference.

The focus of this work is to formally define a framework for intentional behavior, where the agent minimizes a constrained form of expected free energy—and to demonstrate this framework in silico. These constraints are defined on a subset of latent states that represent the intended goals of the agent and propagated to the agent via a form of backward induction. As a result, states that do not allow the agent to make any progress toward one of the intended goals are penalized, and so are actions that lead to such disfavored states. This leads to a distinction between sentient and intentional behavior, where intentional behavior is equipped with inductive constraints.

In this treatment, the word *inductive* is used in several senses. First, it distinguishes inductive inference from the abductive kind of inference that usually arises in applications of Bayesian mechanics—that is, to distinguish between mere inference to the best explanation (abductive inference) and genuinely goal-directed inference (inductive inference) (Harman, 1965; Seth, 2015). Second, it is used with a nod to backward induction in dynamic programming, where one starts from an intended end point and

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

works backward in time to the present to decide what to do next (Bellman, 1952; Da Costa, Sajid et al., 2020; Howard, 1960; Paul et al., 2023). Under this naturalization of behaviors, a thermostat would not exhibit sentient behavior, but insects might (i.e., thermostats exhibit merely reactive behavior). Similarly, insects would not exhibit intentional behavior, but mammals might (i.e., insects exhibit merely sentient behavior; Friedman et al., 2021). The numerical analyses presented below suggest that in vitro neuronal cultures may exhibit sentient behavior, but not intentional behavior. Crucially, we show that intentional behavior cannot be explained by reinforcement learning, as rewards can only be defined in terms of observable outcomes, not in terms of (unobservable) latent states. In the experimental sections of this work, we study and compare the performance of active inference agents with and without intended goal states. For ease of reference, we call active inference agents without goal states *abductive agents* and agents with intended goals *inductive agents*.

This article has four sections. The first briefly rehearses active inference and learning—as a set of nested, free energy–minimizing processes—applied to a generic generative model of exchange with some world or environment. This model is a partially observed Markov decision process that is conciliatory with canonical neural networks in machine learning and likely to describe the self-evidencing of in vitro neuronal networks (Isomura & Friston, 2018; Isomura et al., 2023). This section has a special focus on inductive inference and its relationship to expected free energy. The subsequent sections use numerical studies to make a series of key points. The second section reproduces the empirical behavior of in vitro neuronal networks playing Pong. Crucially, this behavior emerges purely in terms of free energy–minimizing processes, starting with a naive neuronal network. This section illustrates the failure of a (simulated) abductive agent when the game is made more difficult. This failure is used to illustrate the role of inductive inference, which restores performance and underwrites a fluent engagement with the sensorium. The final two sections illustrate inductive inference using navigation in a maze and the Tower of Hanoi problem, respectively. These numerical studies illustrate how the simple application of inductive constraints to active inference allows tasks that would be otherwise intractable in discrete state spaces to be solved efficiently. This efficiency rests on the fact that distal goals can be reached by only planning a few steps in the future, thanks to constraints furnished by inductive inference. Effectively, inductive inference takes the pressure off deep tree searches by identifying blind alleys or dead ends.

1.1 Glossary of Definitions. Before introducing the inductive inference algorithm, we frame our treatment by clarifying our use of some key terms as semantic confusions have been highlighted as a major barrier to progress in this field (Kagan, Gyngell et al., 2023). This framing is important, given that the goal of our work here is not simply to describe a useful heuristic

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

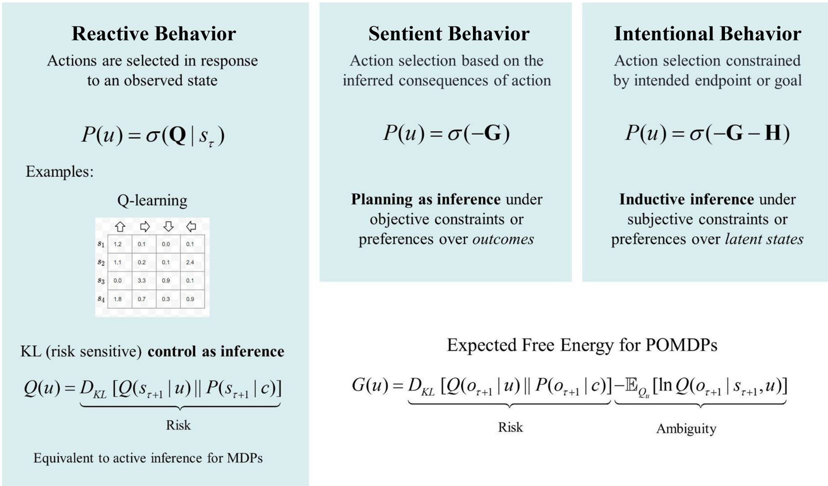

Figure 1: Glossary. In this figure, we provide illustrative definitions of the three kinds of behavior considered in this work in terms of examples, and mathematical differences. Examples of agents with reactive behaviors are model-free reinforcement learning and KL control schemes. In model-free reinforcement learning methods, such as Q-learning, the agent makes use of a lookup table to select actions (more generally, a state-action policy). In this table, rows correspond to states, actions to columns, and every entry encodes the value of taking a specific action (in this case: go up, right, down, left) when in state s_τ. There is no inference over policies, as for every state the agent automatically selects the action with the highest value. In KL control (a.k.a., risk-sensitive control) methods, that automatically select actions that minimise a KL divergence between anticipated and preferred states (where there is no uncertainty about the current state). Sentient agents, on the other hand, plan by taking into account future outcomes and their uncertainty, as they act by minimising an expected free energy $\mathbf{G}$, that includes risk and ambiguity terms. More details on this can be found in equation 2.5. Finally, inductive agents add constraints ($\mathbf{H}$ in the figure) in the action selection, by penalising actions that preclude an intended goal. For a formal derivation of $\mathbf{H}$, we refer to section 3.

for efficient inference (i.e., inductive inference) but to provide an account of how a new form of decision making, characteristic of more complex forms of agency, may be combined with, and folded into, a generic Bayesian (active) inference scheme.

Figure 1 describes increasingly complex forms of behavior—from reactive (merely responding to stimuli) to sentient (planning based on the sensory consequences of actions), to intentional (planning in order to bring about intended states)—and corresponding forms of decision making that may underwrite such behavior.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

Reactive behavior characterizes simple sensorimotor reflex arcs and the mere realization of set points or trajectories (e.g., simple cases of homeostasis and homeorhesis). This form of behavior can be accounted for acting in a way that realizes predicted sensations, with no anticipation of the future sensory consequences of action.

Sentient behavior characterizes the paradigmatic case of active inference, in which the influence of perception on action is mediated by the results of planning, with a distribution over policies derived from a model endowed with counterfactual depth (i.e., beliefs about the future sensory consequences of action pursuant to a policy). In this case, we may characterize the form of inference over actions or policies as *abductive*—that is, as an inference to the policy that best explains current and future observations under a generative model (see below).

Intentional behavior is driven not simply by the generic imperative to minimize sensory prediction error, present and future, but toward the attainment of a particular future end point or goal state. This form of behavior can be subserved by backward induction or inductive inference, as defined below, which supplies a specific form of constraint on the Bayesian (abductive) inference characteristic of (mere) sentient behavior. In particular, it implies not merely beliefs about sensory consequences of actions but rather beliefs about the inferred or latent causes of sensory input.

Note that words like *sentient behavior* and *intentional behavior* are deliberately defined here such that they can be operationalized within the framework of generative modelling, in which terms like *state*, *belief*, and *confidence* have precise, if narrow, interpretations in terms of belief structures of a mathematical sort (Ramstead et al., 2022). Work to define related terms in a more general sense is currently underway elsewhere (Kagan, Gyngell et al., 2023). Whether the phenomenology of (propositional or subjective) beliefs—or sentience—could yield to the same naturalization remains to be seen. See Clark et al. (2019), Sandved-Smith et al. (2021), and Smith et al. (2022) for treatments in this direction. Note further that a key distinction between sentient and intentional behavior rests on the consequences of behavior in (observable) outcome and (unobservable) latent spaces, respectively.

2 Active Inference

Here, we introduce the generative model used in the following sections, which can be seen as a generalization of a partially observed Markov decision process (POMDP). The generalization in question covers trajectories, narratives, or syntax—which may or may not be controllable—by equipping a POMDP with random variables called *paths*. Paths effectively pick out transitions among latent states. These models are designed to be composed hierarchically in a way that speaks to a separation of temporal scales in deep generative models. In other words, the number of transitions among latent states at any given level is greater than the number of transitions at

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

the level above. This furnishes a unique specification of a hierarchy in which the parents of any latent factor (associated with unique states and paths) contextualize the dynamics of their children.

The variational inference scheme (Beal, 2003) used to invert these models inherits from their application to online decision-making tasks. This means that action selection rests primarily on current beliefs about latent states and structures and expectations about future observations. In that sense, the beliefs are updated sequentially—and in an online fashion—with each new action-outcome pair. This calls for Bayesian filtering (i.e., forward message passing) during the active sampling of observations, followed by Bayesian smoothing (i.e., forward and backward message passing) to revise posterior beliefs about past states at the end of an epoch. The implicit Bayesian smoothing ensures that the beliefs about latent states at any moment in the past are informed by all available observations when updating model parameters (and latent states of parents in deep models).

In neurobiology, this combination of Bayesian filtering and smoothing would correspond to evidence accumulation during active engagement with the environment, followed by a replay before the next epoch (Buckner, 2010; Louie & Wilson, 2001; Penny et al., 2013; Pezzulo et al., 2014). From a machine learning perspective, this can be regarded as a forward pass (belief propagation) for online active inference, followed by a backward pass (implemented with variational message passing) for active learning. The implicit belief updates, pertaining to states, parameters, and structure, foreground the conditional dependencies of active inference, learning, and selection, respectively.

2.1 Generative Modeling. Active inference rests on a generative model of observable outcomes (observations). This model is used to infer the most likely causes of outcomes in terms of expected states of the world. These states (and paths) are latent or hidden because they can only be inferred through observations. Some paths are controllable in the sense they can be realized through action. Therefore, certain observations depend on action (e.g., where one is looking), which requires the generative model to entertain expectations about outcomes under different combinations of actions (i.e., policies).[1]

These expectations are optimized by minimizing the variational free energy, defined in equation 2.1. Variational free energy scores the discrepancy between the data expected under the generative model and the actual data. Crucially, the prior probability of a policy depends on its expected free energy. Expected free energy, described in more detail in equation 2.2, is a

[1] Note that in this setting, a policy is not a sequence of actions but simply a combination of paths, where each hidden factor has an associated state and path. This means there are, potentially, as many policies as there are combinations of paths.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

$$P(o_\tau \mid s_\tau, a) = Cat(\mathbf{A})$$
$$P(s_\tau \mid s_{\tau-1}, u_\tau, b) = Cat(\mathbf{B})$$
$$P(u_\tau \mid u_{\tau-1}, c) = Cat(\mathbf{C})$$
$$P(s_0 \mid d) = Cat(\mathbf{D})$$
$$P(u_0 \mid e) = Cat(\mathbf{E})$$

$$P(\mathbf{A}) = Dir(a)$$
$$P(\mathbf{B}) = Dir(b)$$
$$\vdots$$

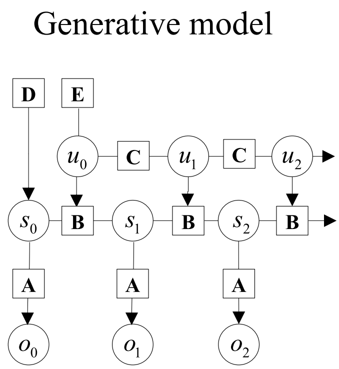

Figure 2: Generative models as agents. A generative model specifies the joint probability of observable consequences and their hidden causes. Right: Legend of distributions of a discrete agent, where *Cat* and *Dir* refer to categorical and Dirichlet distributions, respectively, while the subscripts in this graphic pertain to time. Right: A graphical representation of the generative model. Usually, the model is expressed in terms of a *likelihood* (the probability of consequences given their causes) and *priors* (over causes). When a prior depends upon a random variable it is called an *empirical prior*. Here, the likelihood is specified by a tensor **A**, encoding the probability of an outcome under every combination of *states* (s). The empirical priors pertain to transitions among hidden states, **B**, that depend upon *paths* (u), whose transition probabilities are encoded in **C**. To conclude, **E** specifies the empirical prior probability of each path.

universal objective function that can be read as augmenting mutual information with expected costs or constraints that need to be satisfied. Heuristically, it scores the free energy expected under each course of action. Having evaluated the expected free energy of each policy, the most likely action can be selected, and the perception-action cycle continues (Parr et al., 2022).

2.2 The Generative Model. Figure 2 provides a schematic overview of the generative model used for the simulations considered in this article. Outcomes at any particular time depend on hidden states, while transitions among hidden states depend on paths. Note that paths are random variables, in the sense that a particle can have both a position (i.e., a state) and momentum (i.e., a path). Paths may or may not depend on action. The resulting POMDP is specified by a set of tensors. The first set of parameters, denoted **A**, maps from hidden states to outcome modalities—for example, exteroceptive (e.g., visual) or proprioceptive (e.g., eye position) modalities. These parameters encode the likelihood of an outcome given their hidden

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

causes. The second set, **B**, prescribes transitions among the hidden states of a factor, under a particular path. Factors correspond to different kinds of causes—for example, the location versus the class of an object. The remaining tensors encode prior beliefs about paths **C** and initial states **D**.

The generative model in Figure 2 means that outcomes are generated as follows. First, a policy is selected using a softmax function of expected free energy. Sequences of hidden states are generated using the probability transitions specified by the selected combination of paths (i.e., policy). Finally, these hidden states generate outcomes in one or more modalities. Perception or inference about hidden states (i.e., state estimation) corresponds to inverting a generative model, given a sequence of outcomes, while learning corresponds to updating model parameters. Perception therefore corresponds to updating beliefs about hidden states and paths, while learning corresponds to accumulating knowledge in the form of Dirichlet counts. The requisite expectations constitute the sufficient statistics ($\mathbf{s}$, $\mathbf{u}$, $\mathbf{a}$) of posterior beliefs $Q(s, u, a) = Q_{\mathbf{s}}(s)Q_{\mathbf{u}}(u)Q_{\mathbf{a}}(a)$. The implicit factorization of this approximate posterior effectively partitions model inversion into inference, planning, and learning.

2.3 Variational Free Energy and Inference. In variational Bayesian inference (a form of approximate Bayesian inference), model inversion entails the minimization of variational free energy with respect to the sufficient statistics of approximate posterior beliefs. This can be expressed as follows, where, for clarity, we will deal with a single factor, such that the policy (i.e., combination of paths) becomes the path. Omitting dependencies on previous states, we have for model m,

$$
\begin{aligned}
Q(s_\tau, u_\tau, a) &= \arg\min_{Q} F \\
F &= \mathbb{E}_Q[\ln \underbrace{Q(s_\tau, u_\tau, a)}_{\text{posterior}} - \ln \underbrace{P(o_\tau \mid s_\tau, u_\tau, a)}_{\text{likelihood}} - \ln \underbrace{P(s_\tau, u_\tau, a)}_{\text{prior}}] \\
&= \underbrace{D_{KL}[Q(s_\tau, u_\tau, a) \| P(s_\tau, u_\tau, a \mid o_\tau)]}_{\text{divergence}} - \underbrace{\ln P(o_\tau \mid m)}_{\text{log evidence}} \\
&= \underbrace{D_{KL}[Q(s_\tau, u_\tau, a) \| P(s_\tau, u_\tau, a)]}_{\text{complexity}} - \underbrace{\mathbb{E}_Q[\ln P(o_\tau \mid s_\tau, u_\tau, a)]}_{\text{accuracy}}. \quad (2.1)
\end{aligned}
$$

Because the (KL) divergences cannot be less than zero, the penultimate equality means that free energy is minimized when the (approximate) posterior is equal to the true posterior. At this point, the free energy is equal to the negative log evidence for the generative model (Beal, 2003). This means minimizing free energy is mathematically equivalent to maximizing model evidence, which is, in turn, equivalent to minimizing the complexity of accurate explanations for observed outcomes.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

Planning emerges under active inference by placing priors over (controllable) paths to minimize expected free energy (Friston et al., 2015):

$$G(u) = \mathbb{E}_{Q_u}\left[\ln Q\left(s_{\tau+1}, a \mid u\right) - \ln Q\left(s_{\tau+1}, a \mid o_{\tau+1}, u\right) - \ln P\left(o_{\tau+1} \mid c\right)\right] \quad (2.2)$$

$$= -\underbrace{\mathbb{E}_{Q_u}\left[\ln Q\left(a \mid s_{\tau+1}, o_{\tau+1}, u\right) - \ln Q(a \mid s_{\tau+1}, u)\right]}_{\text{expected information gain (learning)}}$$

$$-\underbrace{\mathbb{E}_{Q_u}\left[\ln Q\left(s_{\tau+1} \mid o_{\tau+1}, u\right) - \ln Q\left(s_{\tau+1} \mid u\right)\right]}_{\text{expected information gain (inference)}} \underbrace{-\mathbb{E}_{Q_u}\left[\ln P\left(o_{\tau+1} \mid c\right)\right]}_{\text{expected cost}} \quad (2.3)$$

$$= -\underbrace{\mathbb{E}_{Q_u}\left[D_{KL}\left[Q\left(a \mid s_{\tau+1}, o_{\tau+1}, u\right) \| Q(a \mid s_{\tau+1}, u)\right]\right]}_{\text{novelty}}$$

$$+\underbrace{D_{KL}\left[Q\left(o_{\tau+1} \mid u\right) \| P\left(o_{\tau+1} \mid c\right)\right]}_{\text{risk}} \underbrace{-\mathbb{E}_{Q_u}\left[\ln Q\left(o_{\tau+1} \mid s_{\tau+1}, u\right)\right]}_{\text{ambiguity}}. \quad (2.4)$$

The notation for the posterior predictive distribution Q_u implies that all random variables in the expression are conditioned on the path variable u. When defined over parameters, hidden states, and outcomes at the next time step, it is defined as

$$\begin{aligned} Q_u &= Q\left(o_{\tau+1}, s_{\tau+1}, a \mid u\right) \\ &= P\left(o_{\tau+1}, s_{\tau+1}, a \mid u, o_0, \ldots, o_\tau\right) \\ &= P\left(o_{\tau+1} \mid s_{\tau+1}, a\right) Q\left(s_{\tau+1}, a \mid u\right), \end{aligned}$$

although in some cases, this predictive density may only depend on the hidden states and outcomes at the next time step. One can also express the prior over the parameters in terms of an expected free energy, where, marginalizing over paths,

$$\begin{aligned} P(a) &= \sigma(-G) \\ G(a) &= \mathbb{E}_{Q_a}[\ln P(s \mid a) - \ln P(s \mid o, a) - \ln P(o \mid c)] \\ &= -\underbrace{\mathbb{E}_{Q_a}[\ln P(s \mid o, a) - \ln P(s \mid a)]}_{\text{expected information gain}} \underbrace{-\mathbb{E}_{Q_a}[\ln P(o \mid c)]}_{\text{expected cost}} \\ &= -\underbrace{\mathbb{E}_{Q_a}\left[D_{KL}[P(o, s \mid a) \| P(o \mid a) P(s \mid a)\right]}_{\text{mutual information}} \underbrace{-\mathbb{E}_{Q_a}[\ln P(o \mid c)]}_{\text{expected cost}} \end{aligned} \quad (2.5)$$

where $Q_a = P(o|s, a)P(s|a) = P(o, s|a)$ is the joint distribution over outcomes and hidden states, encoded by the Dirichlet parameters, a, and $\sigma(\cdot)$ is the softmax function. Note that the Dirichlet parameters encode the mutual information, in the sense that they implicitly encode the joint distribution over

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

outcomes and their hidden causes. When normalizing each column of the a tensor, we recover the likelihood distribution (as in Figure 2); however, we could normalize over every element to recover a joint distribution.

Expected free energy can be regarded as a universal objective function that augments mutual information with expected costs or constraints. Constraints, parameterized by c, reflect the fact that we are dealing with open systems with characteristic outcomes. This allows an optimal trade-off between exploration and exploitation, where the best policy is considered to be the one that balances the need to reduce uncertainty (exploration) and minimize expected cost (exploitation), where the balance is determined by the variational objective (i.e., expected free energy). This derivation of expected costs and information (info) gain from first principles has the advantage of placing expected cost in the same unit of measure as the info gain, that is either in nats (if we are using natural logs) or bits (log in base 2). This is important as it allows us to study utility in terms of information (gives quantitative meaning to the value function). Hence, it allows quantification of the trade-off between exploration and exploitation using a common currency. A key difference here, compared to RL schemes, is that it does not place reward as the ultimate objective: in RL. If information seeking is considered, it is normally because this is felt to be useful in achieving greater long-term reward. In active inference, reward is placed at the same level as information, and both are considered to be valuable in their own right. Situations in which reward is prioritized can then be seen as special cases of a more general principle in which there is limited (resolvable) uncertainty.

The exploration and exploitation trade-off under the expected free energy can be read as an expression of the constrained maximum entropy principle that is dual to the free energy principle (Ramstead et al., 2023) or, alternatively, as a constrained principle of maximum mutual information or minimum redundancy (Ay et al., 2008; Barlow et al., 1961; Linsker, 1990; Olshausen & Field, 1996). In machine learning, this kind of objective function underwrites disentanglement (Higgins et al., 2021; Sanchez et al., 2020) and generally leads to sparse representations (Gros, 2009; Olshausen & Field, 1996; Sakthivadivel, 2022; Tipping, 2001).

When comparing the expressions for expected free energy in equation 2.2 with variational free energy in equation 2.1, the expected divergence becomes expected information gain. Expected information gain about the parameters and states is sometimes associated with distinct epistemic affordances, namely, *novelty* and *salience*, respectively (Schwartenbeck et al., 2019). Similarly, expected log evidence becomes expected value, where value is the logarithm of prior preferences. The last equality in equation 2.2 provides a complementary interpretation in which the expected complexity becomes risk, while expected inaccuracy becomes ambiguity.

There are many special cases of minimizing expected free energy. For example, maximizing expected information gain maximizes (expected) Bayesian surprise (Itti & Baldi, 2009), in accord with the principles of

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

optimal (Bayesian) experimental design (Lindley, 1956). This resolution of uncertainty is related to artificial curiosity (Schmidhuber, 1991; Still & Precup, 2012) and speaks to the value of information (Howard, 1966).

Expected complexity or risk is the same quantity minimized in risk-sensitive or KL control (Broek et al., 2012; Klyubin et al., 2005) and underpins (free energy) formulations of bounded rationality based on complexity costs (Braun et al., 2011; Ortega & Braun, 2013) and related schemes in machine learning such as Bayesian reinforcement learning (Ghavamzadeh et al., 2015). More generally, minimizing expected cost subsumes Bayesian decision theory (Berger, 2013). For a more detailed discussion on the above notions, we refer to Friston, Da Costa, Tschantz et al. (2023).

3 Inductive Inference

What we call inductive inference—in this setting—recalls the notion of backward induction in dynamic programming and related schemes (Camerer et al., 2004; Da Costa, Sajid et al., 2020; Howard, 1960; Hure et al., 2020; Paul et al., 2023; Sutton et al., 1999; Tervo et al., 2016). In this form of inference, precise beliefs about state transitions are leveraged to rule out actions that are inconsistent with the attainment of future goals, defined in belief or state space as a final (or intended) state. This is a limiting case of inductive (Bayesian) inference (Barlow, 1974; Hawthorne, 2021; Kiefer, 2017) in which the very high precision of beliefs about final or intended states allows one to use logical operators in place of tensor operations, thereby vastly simplifying computations. In brief, we use this simplification to furnish constraints on action selection that inherit from priors over intended states in the future.

Active inference rests on priors that place constraints on paths or trajectories through state space. For example, a sparse prior preference with knowledge only about the final state warrants deep planning to demonstrate intentional behavior (Paul et al., 2023). One can either specify these constraints in terms of states that are unlikely to traversed or in terms of the final state. In other words, the agent may a priori believe it will navigate state space in a way that avoids unlikely or surprising outcomes or that it will reach some final destination (in state space, not outcome space), irrespective of the path taken. These are distinct kinds of constraints. The first is implemented by **c** in terms of the cost or constraints that apply during the entire path. We now introduce another prior or constraint **h** over the final state. The priors **d** and **h** play reciprocal roles in the sense they specify prior beliefs about the initial and final states, respectively. Backward induction now follows simply from this prior, provided it is specified sufficiently precisely. We refer to these final states as intended states.[2]

[2] While **c**, **d**, and **h** are usually hard-coded, they can be learnd very efficiently, for example, using Z-learning for certain classes of MDPs (Paul et al., 2023; Todorov, 2006).

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

The basic idea is that although we may be uncertain about the next latent state, we can be certain about which states cannot be accessed from the current state. This means we can use induction to identify subsequent states that cannot be on a path to an intended state, thereby rendering actions—(i.e., state transitions) to those ineligible, dead-end states—highly unlikely (assuming that we are on a path to an intended state). The requisite induction goes as follows:

Imagine that we know our current state and that we will be in a certain (intended) state in the future. Imagine further that we know all possible transitions, afforded by action, among states. This means we can identify all the states from which the intended state is accessible. We can now repeat this and identify all the states from which the eligible states at the penultimate time point can be accessed, and so on. We now repeat this recursively—moving backward in time—until our current state becomes eligible. At this point, we select an action that precludes ineligible states at the preceding point in backward time (or next point in forward time), bringing us one step closer to the intended state. We now repeat the backward induction, until we arrive at the intended state via the shortest path. This backward induction is computationally cheap because it entails logical operations on a sparse logical tensor, encoding allowable state transitions.

Figure 3 provides a pseudocode and graphical abstraction based on the Matlab scripts implementing this inductive logic. For clarity, we have assumed a single factor and that there are no constraints on the paths other than those specified by a 1-hot vector **h**, specifying the agent's intended states.[3]

Note that this is not vanilla backward induction. It is simply a way of placing precise priors on paths that render certain paths—that cannot access an intended state—highly unlikely. The requisite priors complement expected free energy in the following sense (see Figure 3): inductive priors over policies **H** are derived from priors over intended states **h**, while the priors over policies scored by expected free energy **G** inherit from priors over preferred outcomes **c**. This distinction is important because it means that this kind of reasoning—and intentional behavior—can only manifest under precise beliefs about latent states. For example, a baby (or unexplainable neural network) could not, by definition, act intentionally because he or she does not have a precise generative model of latent states (or any mechanism

[3] In our Matlab implementation of inductive inference, constraints due to prior preferences in outcome space are accommodated by precluding transitions to costly states during construction of the logical matrix encoding possible or true transitions. Furthermore, the implementation deals with multiple factors using appropriate tensor products. Finally, when multiple intended states are supplied, the nearest state is chosen for induction; where nearest is defined in terms of the number of timesteps required to access an intended state. For the pseudocode of the Python version, we refer to the supplementary material.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

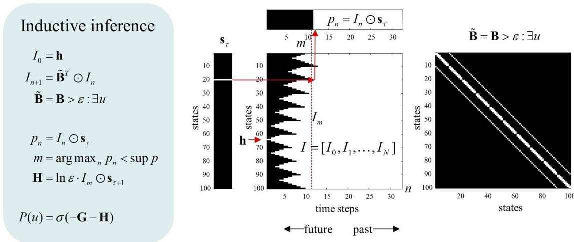

Figure 3: Inductive Inference. This figure provides an overview of inductive inference used in this article. The left panel provides the expressions used to compute the penalty cost **H** that penalizes moving toward states that do not contain paths to some intended end state **h**, here expressed as a 1-hot vector. The middle panel illustrates this induction with a graphical example, where binary vectors and matrices are shown in image format (black equals zero or false and white equals one or true). Working down the equalities in the left panel, we first initialize a logical vector of states I_0 to the intended state **h**. Recursively, we evaluate all the states from which the previous state can be accessed, where a state can only be accessed if the probability of transitioning from an adjacent state is larger than the threshold ε (here, $\tilde{\mathbf{B}}$ is transposed because this recursive induction works backward in time). This process leads to the computation of each I_n (the column vectors in the central figure) that allows us to compute, for every time step n, the probability p_n of the (posterior beliefs on the) current state being in I_n, that is, the probability of existence of a path of length n between the current state and the goal state (top of the central panel). This allows computing both the length of the shortest path to the intended state, which corresponds to the smallest n such that $p_n = 1$, and the immediate subsequent set of states to which the agent must transition to in order to follow this shortest path. Trivially, these states are those whose length of the shortest path to the goal state is one less than that of our current state, and their indices correspond to the entries of I_m that are equal to one, where m is their distance to the goal state and corresponds to the largest n such that $p_n = 0$. In the example shown in the middle panel, the agent is currently in state 20 (y-axes), which means that the shortest path to the intended state (state 64) is 12 time steps (x-axes), as this is the smallest n such that $p_n = 1$. This tells us that if we are pursuing the shortest path, then there are certain states we need to avoid, encoded as zeros in the logical vector I_m, with $m = 11$ (indicated by the dotted red line). The following equations, which define the inductive cost H and its role in the policy selection, describe how constraints on actions that would divert the agent from the shortest path are enforced. In more detail, ineligible states are assigned a high cost (here, the log of a small value) to evaluate the expected cost incurred by each policy, using its predictive posterior over states (see Figure 2). Finally, we can supplement the expected free energy, **H** of each policy with the ensuing inductive cost, **H**. In principle, this guarantees the selection of paths or policies that lead to the intended state, provided that state can be reached. The example shown on the right is taken from the maze navigation task described later. For clarity, this example only considers a single factor.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

to specify intended states). We will return to prerequisites for inductive inference in the discussion.

In summary, inductive inference propagates constraints backward in time to provide empirical priors for planning as inference in the usual way. This means that within the constraints afforded by such planning, actions will still be chosen that maximize expected information gain and any constraints encoded by **c**. In this sense, the inductive part of this inference scheme can be regarded as providing a constrained expected free energy, which winnows trajectories through state space to paths of least action. An equivalent and alternative perspective is that inductive inference furnishes an empirical prior over policies.

When intended states are conditioned on some context—inferred by a supraordinate (hierarchical) level—one has the opportunity to learn intended states and, effectively, make planning habitual. In this setting, the implicit Dirichlet counts in **h**, could be regarded as accumulating habitual courses of action that are learned as empirical priors in hierarchical models. We will pursue this elsewhere. In what follows, we focus on the distinction between sentient behavior—based on expected free energy—and intentional behavior—based on inductive priors.

4 Pong Revisited

In this section, we first simulate "mere" sentient behavior and then examine the qualitative differences in behavior when adding inductive constraints. Specifically, we simulate the in vitro experiments reported in (Kagan et al., 2022), using both an abductive and an inductive agent. The first has no intended goals and stands in for a naive neuronal culture; the second has a set of intended states: the ones where the paddle hits upcoming balls. As environments, we use Pong of two different sizes that reflect two different difficulties: 5×6 (easy) and 8×4 (hard). The results show that while the simulated in vitro agent that is using an abductive decision process (e.g., unconstrained expected free energy) is able to fluently play in the easy environment, it struggles in the harder one. The inductive agent can master the harder environment in less than three minutes of (simulated) game time.

In the in vitro experiments, certain cells were stimulated depending on the configuration of a virtual game of Pong, constituted by the position of a paddle and a ball bouncing around a bounded box. Other recording electrodes were used to drive the paddle, thereby closing the sparse coupling between the neuronal network and the computer network simulating the game of Pong (see Figure 4). Typically, in these experiments, after a few minutes of exposure to the game, short rallies of ball returns emerge. To emulate this setup, we created a generative process (i.e., a hard-coded representation of the dynamics of external states) in which a ball bounced around a box at 45 degrees. The lower boundary contained a paddle that could be moved to the right or left. The size of the box was 5×6 units, where the

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

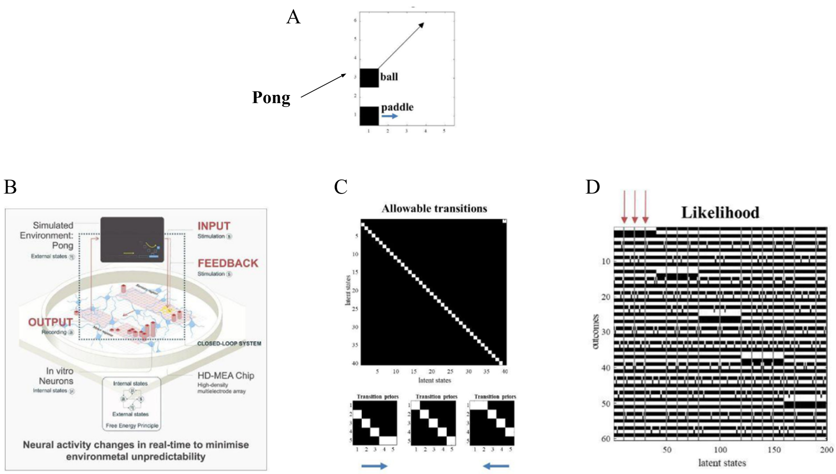

Figure 4: Learning the world of Pong. (A) Setup used in the simulations. In brief, the generative process modeled a ball bouncing around inside a bounding box, with a movable paddle on the lower boundary. The $(5 \times 6 =)30$ locations or pixels provided outputs with two states (black or white) that were subsequently learned via a likelihood mapping to 40 latent states. The agent was equipped with a precise transition prior where 40 latent states followed each other, with circular boundary conditions. In addition, the agent was equipped with a second factor that controlled the panel, moving it to the right, staying still, and moving it to the left. (B) Graphical abstract (reproduced from Kagan et al., 2022; permission from the authors) describing the in vitro empirical study in which a closed loop system was used to record from—and stimulate—a network of cultured neurons. The setup enabled the neurons to control a virtual paddle in a simulated game of Pong. Sensory feedback reported the location of the ball and paddle; enabling the neuronal preparation to learn how to play a rudimentary form of Pong. (C) The transitions of the generative model, while panel *D* shows the results of active learning—that is, accumulation of Dirichlet counts in the likelihood tensor—after 512 time steps. Note that this is a precise likelihood mapping due to the fact that the synthetic agent has precise, if generic, transition priors. The likelihood mapping in panel D is shown in image format, with each of the 30 likelihood tensors stacked on top of each other. Of note here are certain latent states that produce ambiguous (i.e., unpredictable) outcomes. The first three are labeled with small arrows over the likelihood matrix. These ambiguous likelihood mappings appear as gray columns. This reflects the fact that the agent has learned that states corresponding to "missing the ball" lead to unpredictable and ambiguous stimulation. The implicit surprise and ambiguity mean that the agent plans to avoid these states and look as if it is playing Pong—by choosing paths or policies that are more likely to hit the ball. The emergence of this behavior is described in the next figure.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

ball moved one unit up or down (and right or left) at every time point. The (one-unit-wide) paddle could be moved left or right by one unit at every time point. In the in vitro experiments, whenever the agent missed the ball, randomized white noise was applied to the sensory electrodes; otherwise, the game remained in the play. We simulated this by supplying random input to all sensory channels whenever the ball failed to contact the paddle on the lower boundary.

The (sensory) outcomes of the POMDP ad 30 sensory channels that could be on or off. These can be thought of as pixels in a simple Atari-like game. The latent states were modeled as one long orbit by equipping the generative model with a transition matrix that moved from one state to the next (with circular boundary conditions) for a suitably long sequence of state transitions (here, 40). The generative model was equipped with a second factor with three controllable paths. This factor moved the paddle one unit to the right or left (or no movement). However, the (implicit) agent knew nothing more about its world and, in particular, had no notion that the second factor endowed it with control over the paddle. This was because the likelihood tensors mapping from the two latent factors to the outcomes were populated with small and uniform Dirichlet counts (i.e., concentration parameters of 1/32[4]). In other words, our naive generative model could, in principle, model any given world (providing this world has a limited number of states that are revisited systematically). Figure 4 shows the setup of this paradigm and the parameters of the generative model learned after 512 time steps.

To simulate the in vitro study, we exposed the synthetic neural network to 512 observations—about two minutes of simulated time (i.e., a few seconds of computer time). Figure 5 shows the results of this simulation. The ensuing behavior reproduced that observed empirically: the emergence of short rallies after a minute or so of exposure. The question is now: Can we understand this in terms of free energy minimizing processes and their teleological concomitants?

As time progresses, Dirichlet counts are accumulated in the likelihood tensor to establish a precise mapping between each successive hidden state and the outcomes observed in each modality. This accumulation is precise because the agent has precise beliefs about state transitions. As the likelihood mapping is learned, it becomes apparent to the agent that certain states produce ambiguous outputs. These are the states in which it fails to hit the ball with the paddle. Because these ambiguous states have a high expected free energy—see equation 2.2—the agent considers that actions that

[4] The 1/32 is just a small concentration parameter that regulates how stubborn an agent is: when the concentration parameter is large, the agent needs to observe the same transition many times before there is a substantive change in the relative Dirichlet parameters. When small, the agent "learns" a particular transition the first time it is observed. In short, it can be thought of as an initial learning rate for transition probabilities.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

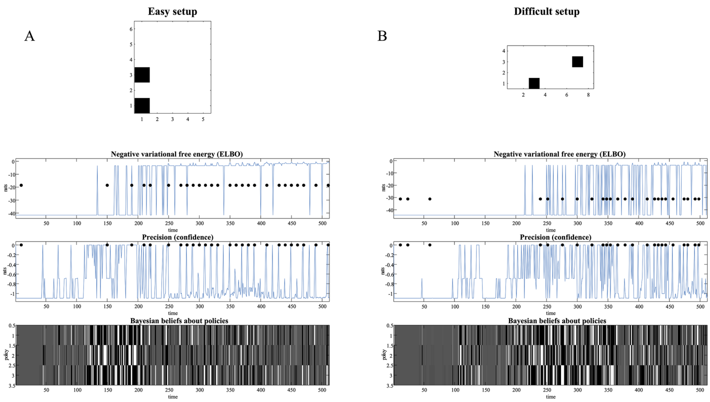

Figure 5: The emergence of play. (A, B) The results of two simulations of 512 time steps (i.e., about two minutes of simulated time) under two configurations of the Pong setup: an easy setup in panel A and a slightly more difficult setup in panel B, in which the width of the bounding box was increased and its height decreased (from 5 × 6 to 8 × 4). In both panels, the configuration of the game is shown above three plots reporting fluctuations in various measures of belief updating and accompanying behavior. The first graph plots the (negative) variational free energy as a function of time (where each time step corresponds roughly to 250 ms). The black dots mark time points when the ball was hit. It can be seen that during accumulation of the likelihood Dirichlet counts, the ball was missed until time step 150. After about a minute, the synthetic agent then starts to emit short rallies of between one and seven consecutive hits. The emergence of game play is accompanied by saltatory increases in negative variational free energy (or evidence lower bound). These increases disappear whenever the agent misses the ball, terminating little rallies. The second graph plots the average of the expected free energy under posterior beliefs about policies. This can be read as the precision of policy beliefs or, more colloquially, the confidence placed in policy selection. This illustrates that confident behavior emerges during the first minute and is subsequently restricted to moments prior to hitting the ball. Heuristically, this can be read as the agent realizing that it can avoid ambiguity by moving in such a way as to catch the ball. The accompanying posterior (Bayesian) beliefs about policies are shown in image format in the bottom plot. This illustrates that precise or confident behavior entails precise beliefs about what to do next: the three rows report a categorical distribution over three actions as a function of time. As time goes on (left to right), we see the emergence of a precise distribution that is due to learning. As one can see, initially (on the left), the (prior) beliefs over actions are uniform. Panel B shows exactly the same results but for a slightly more difficult game. Here, the ball has more latitude to move horizontally and is returned more quickly due to the reduced height of the bounding box. In consequence, learning a precise likelihood mapping takes about twice the amount of time. And even when learned, the rallies are shorter, ranging from one to four at most. We will use this more difficult setup to look at the effect of inductive inference in the next figure.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

bring about these states are unlikely and therefore tries to avoid missing the ball. This is sufficient to support rallies of up to seven returns (see Figure 5).

However, because this agent does not look deep into the future, it can only elude ambiguous states when they are imminent. In other words, although this kind of behavior can be regarded as sentient—in the sense that it rests on an acquired model of the consequences of its own action—it is not equipped with intended states.

Note what has been simulated here does not rely on any notion of reinforcement learning: at no point was the agent rewarded (in the traditional sense) for any behavior or outcome. This kind of self-organization—to a synchronous exchange with the world—is an emergent property of the system that simply rests on *avoiding ambiguity or uncertainty* of a particular kind. The subtle distinction between a behaviorist (reinforcement learning) account and this kind of self-evidencing rests on the imperatives for self-organized behavior. In this in silico reproduction of in vitro experiments, behavior is a consequence of (planning as) inference, where inference is based on what has been learned. What has been learned are just statistical regularities (or unpredictable irregularities) in the environment. In this case, there are certain states that lead to unpredictable outcomes. This gives the agent a precise grip on the world and enables it to infer its most likely actions. Its most likely actions are those that are characteristic of the thing it is, namely, something that minimizes surprise, ambiguity, and free energy. This is distinct from learning a behavior in the sense of reinforcement learning (e.g., a state-action mapping). The difference lies in the fact that behavior—of the sort demonstrated above—rests on inference, under a learned model.

In the next section, we turn to a different kind of behavior that rests on inductive inference, equipping the agent with foresight and eliciting anticipatory behavior.

4.1 Inductive Inference. In this section, we repeat the simulations above, but making the game more difficult by increasing the width of the box. This means that to catch the ball, the agent has to anticipate outcomes in the distal future in order to respond with preemptive movement of the paddle. Note that this kind of behavior goes beyond the sort of behavior predicted under perceptual control theory and related accounts of ball catching (Gigerenzer & Brighton, 2009; Mansell, 2011). For example, one way to model behavior in this paradigm would be to move the paddle so that it was always underneath the ball. However, this is not the behavior that emerges under self-evidencing. In what follows, we will see that avoiding ambiguity is not sufficient for skilled performance of a more difficult game of Pong. However, if we equip the agent with intentions to hit the ball (i.e., as an intended state), it can use inductive inference to pursue a never ending rally, and play the game skilfully.

Figure 5B reports performance over about two minutes of simulated time of an abductive agent when increasing the width of the Pong box to eight

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

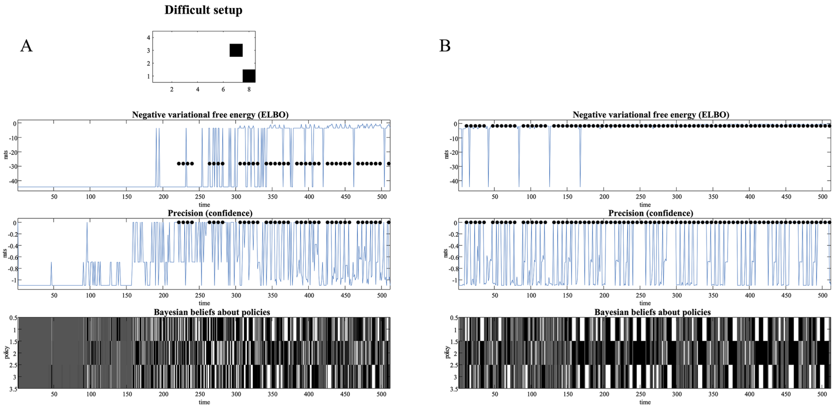

Figure 6: Inductive Inference. This figure follows the same format as Figure 5, reporting the emergence of Pong-playing behavior under the more difficult setup described in the previous figure. However, here, we include inductive inference in the belief updating by specifying the agent's intentions in terms of priors over particular latent states—namely, states in which the agent hit the ball. In realizing these intentions, the agent quickly learns a sufficiently precise likelihood mapping, evincing rallies of between four and six. after about a minute (of simulated time). This is shown in panel A. Panel B shows the performance during the subsequent two minutes. By about three minutes, the agent has a precise grip on its world and realizes its intentions fluently. From a dynamical systems perspective, this can be read as the emergence of generalized synchrony—or synchronization of chaos—as the joint system converges onto a synchronization manifold: a manifold that contains the states the agent intends to visit.

units (and decreasing its height to four units). This simple change precludes sustained rallies, largely because the depth of planning is not sufficient to support preemptive moves of the paddle.

The equivalent results under inductive inference are shown in Figure 6. Here, active inference under inductive constraints produces intermittent rallies within about a minute of simulated time—and skilled, and fluent play after three minutes.

In this example, we simply specified the intended states as those states corresponding to ball hits. This would be like instructing a child by telling her what is (i.e., which states are) expected. The child can then work out how to realize those states by using inductive inference and selecting the most likely actions at each moment. Notice that there is no sense in which this could be construed as reinforcement learning: no reward or cost is being synchronized. Rather, the behavior is driven purely by the minimization of uncertainty. A better metaphor would be instantiating some intentional set

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

by instilling intentions or prior beliefs about characteristic states that should be realized.

From the perspective of the free energy principle, priors over intended states can be cast as specifying a nonequilibrium steady state with a (pullback) attractor that contains intended or characteristic states. From a dynamical systems perspective, this is equivalent to specifying unstable fixed points that characterize stable heteroclinic orbits (Afraimovich et al., 2008; Rabinovich et al., 2008), which have been discussed in terms of sequential behavior (Fonollosa et al., 2015). Intuitively, this means the agent has found a free energy minimum that is characterized by generalized synchrony between the neuronal network and the process generating sensory inputs.

Given that this synchronization was never seen in the in vitro experiments, one might argue that the in vitro behavior was sentient but not intentional. Indeed, this interpretation accords with the finding that the information input to the in vitro cultures resulted in stark increases to the level of critical dynamics present within these cultures (Habibollahi et al., 2023). In this work, it was found that critical dynamics arose as a natural consequence of a dynamic system being embodied within a structured information landscape. However, contrary to some other interpretations of the role of neural criticality, here the critical dynamics could be seen to represent a basic response to the information input but were only necessary, not sufficient, for any specific performance-based behavior that would indicate intention. Taken together this would draw an important distinction between sentience (as defined here) and intentionality. In the remaining sections, we briefly showcase inductive inference in two other paradigms to illustrate the interaction between constraints—encoded by prior preferences over outcomes and prior intentions, encoded by priors over latent states.

5 Navigation as Inductive Inference

In this section, we revisit a simple navigation problem addressed many times in the literature (e.g., Baker et al., 2009; Dayan et al., 2006) and in demonstrations of active inference (e.g., Friston et al., 2021; Kaplan & Friston, 2018). Here, the problem is to learn the structure of a two-dimensional maze and then navigate to a target location based on what has been learned. This features the dual problem of learning a world or generative model and then using what has been learned for deep planning and navigation.

In detail, we constructed a simple maze, shown in Figure 7, for an agent who has a myopic view of the world—namely, one output modality that reported whether the agent was sitting on an allowable location (white square) in the maze or a disallowed location (black square), which, a priori, it found surprising (e.g., experiencing a foot shock). A simple generative model was supplied to the agent in the form of a single factor encoding each location or way-point, equipped with five paths. These were controllable paths that moved the agent up or down, or right or left (or staying

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

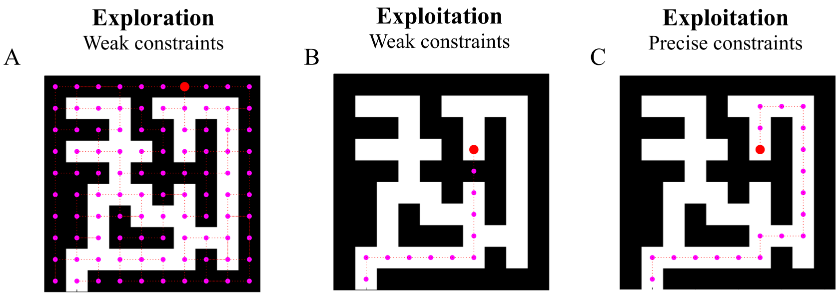

Figure 7: Navigation by induction. (A) This panel reports the exploration of an agent that is building its likelihood mapping by exploring all the novel locations in a maze. Initially, the agent does not know where it can go in the sense that it can only see its current location, which can be black or white. Therefore, every unvisited location furnishes some novelty, that is, expected information gain (about likelihood parameters). This compels the agent to explore all locations efficiently and uniformly with an effective inhibition of return until it has become familiar with this particular maze layout. After learning, the agent was given some intentions in terms of a specific location it believed, a priori, it would visit. Panels B and C show the results of planning under mild and precise preferences for being on white squares. In panel B, the agent takes a shortcut to the target location (red dot), which involves a transgression of one black square. This means that the cost of being on black squares is not sufficiently precise to have constrained the transitions used in inductive inference. However, because the agent is still trying to minimize expected cost (encoded by preferences for white squares), it navigates fairly gracefully until it encounters a barrier. In contrast, panel C shows the same agent with precise costs, which preclude transitions to black squares during inductive inference. This agent can swiftly induce the requisite path to the target location without transgressing constraints on outcomes.

still). The likelihood mapping was, as in the previous simulation, initialized to small, uniform Dirichlet counts. This means the agent has no idea about the structure of its world but simply knew that a latent state could change in one of five ways. Learning this kind of environment is straightforward under active inference, due to the novelty or expected information gain about parameters (see equation 2.2).

This means the agent chooses actions that resolve the greatest amount of uncertainty in the likelihood mapping from each latent state to outcomes. This ensures a Bayes optimal exploration of state space. Figure 7A shows that the agent pursues a path that covers all locations in an efficient fashion—that is, not revisiting experienced states or locations until it has explored every location. The trajectory shown in Figure 7A corresponds to 256 time steps. After this exposure, the agent has learned a likelihood model

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

that is sufficient to support inductive inference. Figures 7B and 7C show the results of this inductive navigation, reaching a distal target (red dot) from a starting location while avoiding surprises or black squares in the maze. The two routes chosen are under imprecise and precise prior preferences for avoiding black squares (i.e., a log odds ratio, encoded by **c**, of one and four, respectively). Note that the path under precise preferences is about 20 steps, speaking to the depth of induction (here, 32 time steps, as in Figure 3).

This example highlights an interesting aspect of inductive inference as defined here: the learned constraints on foraging act as constraints on intentional behavior. These constraints enter the allowable transitions, so that the paths that are induced respect the constraints due to prior preferences that can be inferred after—and only after—learning the likelihood mapping. In short, this example shows how it is possible to reach intended end points, under constraints on the way one gets there. In the final section, we use the same scheme to illustrate the efficiency of inductive inference in high-dimensional problem spaces.

6 Inductive Problem Solving

This section considers a canonical problem-solving task: the Tower of Hanoi (Donnarumma et al., 2016). In this problem, one has to rearrange a number of balls over a number of towers to reach a target arrangement from any given initial arrangement (see Figure 8). The problem can be made easier or more difficult by manipulating the number of intervening rearrangements between the initial and target (intended) configurations. We have previously shown that this problem can be learned from scratch using structure learning (Friston, Da Costa, Tschantz et al., 2023). Here, we consider problem solving with and without inductive inference after learning the likelihood model and allowable state transitions.

As above, implementing inductive inference simply means equipping the agent with prior beliefs about a final (intended) state and then letting it rearrange the balls until those intended states are realized. To solve this problem using active inference, one usually supplies constraints in terms of prior preferences that are mildly aversive for all but the target arrangement. This means the agent will rearrange the balls in a state of mild surprise until the preferred arrangement is found—and the agent rests in a low free-energy state. Because constraints are only in outcome space, there are certain arrangements that are less surprising because they are similar to the target configuration (as defined in outcome space). This enables the agent to solve fairly deep problems, even with a limited depth of planning (here, one-step-ahead planning). However, problems requiring more than four or five moves usually confound this kind of planning as inference. In contrast, if the intended target is specified in state space, then it will invoke inductive inference and, in principle, solve difficult problems, even with a limited depth of planning.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

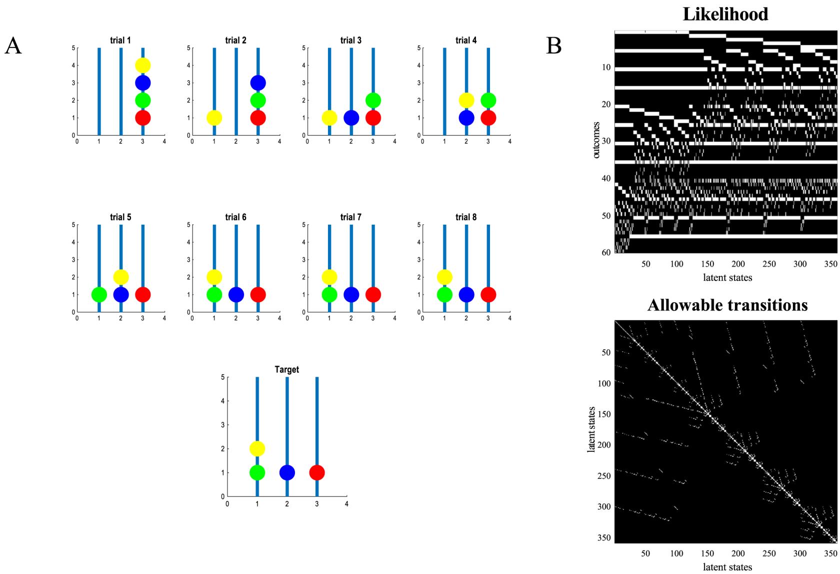

Figure 8: Inductive inference and the Tower of Hanoi. (A) The particular game used to illustrate inductive inference. Here, there are four balls on three towers. The problem is to rearrange the initial configuration (on the upper left) to match the target configuration (lowest arrangement). In this example, it takes five moves. Actions correspond to moving a ball from one pillar to another. The generative model that supports this kind of problem solving is shown in terms of the requisite likelihood and transition mappings in panel B. The likelihood tensors have been stacked on top of each other (and unfolded) to illustrate the mapping between the 360 latent states and the ($4 \times 3 \times 5 =$) 60 outcomes. The accompanying transition parameters are shown in terms of allowable transitions among latent states (as in Figure 3). This generative model can be learned from scratch by presenting each arrangement—and then each rearrangement—of the balls to accumulate the appropriate Dirichlet parameters. Of interest here is the use of the ensuing parameters or knowledge to solve problems that require deep planning. This problem is straightforward to solve using inductive inference—namely, working backward from the target state using the protocol described in Figure 3. The ensuing performance is shown in the next figure.

Figure 9 shows the performance of two agents on 100 problems, given 12 moves for each problem. The first (abductive) agent was equipped only with constraints in outcome space, that is, prior preferences that led to the target solution provided that solution was reasonably close in outcome space. This agent failed to solve problems with five or more moves. In contrast, when specifying intentions in the form of the intended (target) state

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

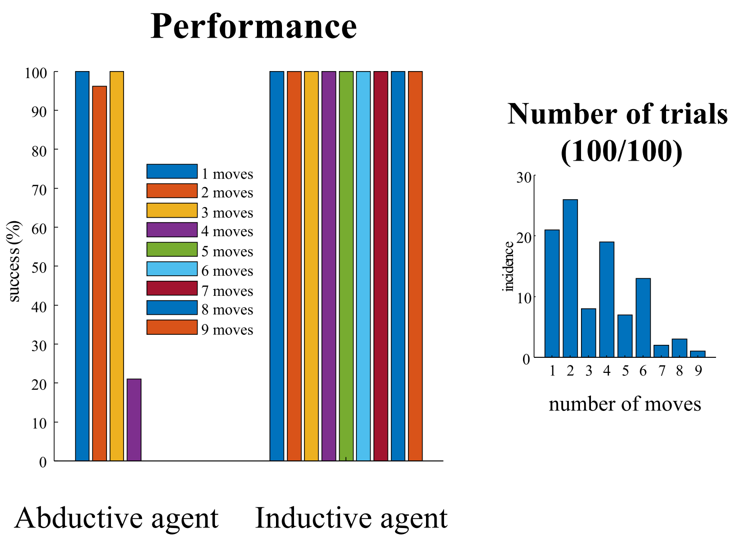

Figure 9: Tower of Hanoi Performance. This figure reports the performance of a generative model that has learned the Tower of Hanoi problem in terms of transitions among different arrangements of balls. We presented the agent with 100 trials with different targets of greater and lesser difficulty (i.e., with varying numbers of moves from the initial and target arrangements). We presented exactly the same problems to agents with and without inductive inference. The right panel shows the incidence of trials in terms of the numbers of moves required until completion. The agent with inductive inference was able to solve 100% of trials successfully. In contrast, the agent that did not use inductive inference was only able to complete problems of four moves or fewer. This is still impressive because both the abductive and inductive agents only looked one step ahead. In other words, even though the abductive agent could evaluate only the quality of its next move, it was still able to work toward the final solution. This is possible because the prior preferences for the target outcomes mean that certain outcomes are closer to the preferred outcomes than others. The 100 trials reported in this figure take less than 10 seconds to simulate.

or arrangement, the second (inductive) agent was able to solve problems of eight moves or more almost instantaneously, without fail.

In these examples, the output space was a collection of $(4 \times 3 =)12$ outcome modalities—one for each location or pixel—with five levels (four colored balls or an empty outcome). The state space encompassed 360 arrangements, producing large $(360 \times 360 \times 5)$ transition tensors. However, reducing these to logical matrices—used in inductive inference—means one can effectively plan deep into the future (here, 64 moves) within milliseconds, using a one-step-ahead, active inference scheme.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

7 Discussion

This article has introduced a particular instance of backward induction to active inference, as well as a more formal characterization of sentient and intentional behavior. Induction in this setting appeals to a simple kind of backward induction via logical operators, which is used to furnish constraints on the expected free energy and, hence, actions. Actions are then selected in the usual way: actions that maximize expected information gain and value, where value is scored by log prior preferences over outcomes. The use of inductive priors lends planning a deep reach into the future that rests on specifying final or intended end points. In turn, this differentiates sentient from intentional behavior. To the extent that one can describe Bayesian beliefs—about the ultimate consequences of plans—as intentions, one could describe the behavior illustrated above as intentional with a well-defined purpose or goal.

Inductive inference, as described here, can also be read as importing logical or symbolic (i.e., deductive) reasoning into a probabilistic (i.e., inductive, in the sense of inductive programming) framework. This speaks to symbolic approaches to problem solving and planning (e.g., Colas et al., 2010; Fox & Long, 2003; Gilead et al., 2019)—and a move toward the network tensor computations found in quantum computing (e.g., Fields et al., 2023; Knill & Laflamme, 1997). However, in so doing, one has to assume precise priors over state transitions and intended states. In other words, this kind of inductive inference is only likely when one has precisely stated goals and knowledge about state transitions. Is this a reasonable assumption for active inference? It could be argued that it is reasonable in the sense that goal states or intended states are stipulatively precise (one cannot formulate an intention to act without specifying the intended outcome with a certain degree of precision) and the objective functions that underwrite self-evidencing lead to precise likelihood and transition mappings. In other words, to minimize expected free energy—via learning—is to maximize the mutual information between latent states and their outcomes and between successive latent states.

To conclude, inductive inference differs from previous approaches proposed in both the reinforcement learning and active inference literature due to the presence of intended goals defined in latent state space. In both model-free and model-based reinforcement learning, goals are defined via a reward function. In alternative but similar approaches, such as active inference, rewards are received by the agent as privileged (usually precise but sparse) observations (Da Costa et al., 2023; Friston et al., 2015). This influences the behavior of the agent, which learns to design and select policies that maximize expected future reward via either model-free approaches, which assign values to state-action pairs, or model-based approaches, which select actions after simulating possible futures. Defining preferences directly in the state space, however, induces a different kind

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

of behavior: the fast and frugal computation involved in inductive inference is now likely to more closely capture the efficiency of human-like decision making, where indefinitely many possible paths, inconsistent with intended states, are ruled out a priori—hence combining the ability of agents to seek long-term goals, with the efficiency of short-term planning.

8 Conclusion

The aim of this article was to characterize the self-organization of adaptive behavior through the lens of the free energy principle, that is, as self-evidencing. We did this by first discussing the definitions of reactive and sentient behavior in active inference, where the latter describes the behavior of agents that are aware of the consequences of their actions. We then introduced a formal account of intentional behavior, specified by intended end points or goals, defined in state space rather than outcome space, as in abductive forms of active inference. We then investigate these forms of (reactive, sentient, and intentional) behavior using simulations. First, we simulate the in vitro experiments, in which neuronal cultures spontaneously learn to play Pong by implementing nested, free energy-minimizing processes. We used these simulations to illustrate the ensuing behavior—leveraging the distinction between merely reactive, sentient, and intentional behavior. The requisite inductive inference was then further illustrated using simple machine learning benchmarks (navigation in a grid world and the Tower of Hanoi problem) that showed how quickly and efficiently adaptive behavior emerges under inductive constraints on active inference.

Disclosure Statement

We have no disclosures or conflict of interest. B.J.K. is employed at and holds shares in Cortical Labs Pty, Melbourne, Australia. No specific funding or other incentives were provided for involvement in this publication and there are no further potential conflicts of interest to disclose.

Acknowledgments

K.F. is supported by funding for the Wellcome Centre for Human Neuroimaging (205103/Z/16/Z), a Canada-UK Artificial Intelligence Initiative (ES/T01279X/1) and the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement 945539 (Human Brain Project SGA3). A.R. is funded by the Australian Research Council (DP200100757) and the Australian National Health and Medical Research Council investigator grant Reference 1194910).

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

References

Afraimovich, V., Tristan, I., Huerta, R., & Rabinovich, M. I. (2008). Winnerless competition principle and prediction of the transient dynamics in a Lotka-Volterra model. *Chaos*, *18*(043103). 10.1063/1.2991108

Attias, H. (2003). Planning by probabilistic inference. *Proceedings of the International Workshop on Artificial Intelligence and Statistics* (pp. 9–16).

Ay, N., Bertschinger, N., Der, R., Guttler, F., & Olbrich, E. (2008). Predictive information and explorative behavior of autonomous robots. *European Physical Journal B*, *63*, 329–339. 10.1140/epjb/e2008-00175-0

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*, 329–349. 10.1016/j.cognition.2009.07.005

Balci, F., Hamed, S., Boraud, T., Bouret, S., Brochier, T., Brun, C., & Zimmermann, J. (2023). A response to claims of emergent intelligence and sentience in a dish. *Neuron*, *111*, 604–605. 10.1016/j.neuron.2023.02.009

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, *1*(01), 217–233.

Barlow, H. B. (1974). Inductive inference, coding, perception, and language. *Perception*, *3*, 123–134. 10.1068/p030123

Beal, M. J. (2003). *Variational algorithms for approximate Bayesian inference*. PhD diss. University College London.

Bellman, R. (1952). On the theory of dynamic programming. *Proceedings of the National Academy of Sciences, USA*, *38*, 716–719. 10.1073/pnas.38.8.716

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer.

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485–488. 10.1016/j.tics.2012.08.006

Braun, D. A., Ortega, P. A., Theodorou, E., & Schaal, S. (2011). Path integral control and bounded rationality. In *Proceedings of the 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning* (pp. 202–209).

Broek, B. v. d., Wiegerinck, W., & Kappen, H. (2012). *Risk sensitive path integral control*. arXiv:1203.3523.

Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annual Review of Psychology*, *61*, 27–48. 10.1146/annurev.psych.60.110707.163508

Camerer, C. F. (1997). Progress in behavioral game theory. *Journal of Economic Perspectives*, *11*, 167–188. 10.1257/jep.11.4.167

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, *119*, 861–898. 10.1162/0033553041502225

Clark, A., Friston, K., & Wilkinson, S. (2019). Bayesing qualia: Consciousness as inference, not raw datum. *Journal of Consciousness Studies*, *26*, 19–33.

Colas, F., Diard, J., & Bessière, P. (2010). Common Bayesian models for common cognitive issues. *Acta Biotheoretica*, *58*, 191–216. 10.1007/s10441-010-9101-1

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, *99*, 102447. 10.1016/j.jmp.2020.102447

Da Costa, L., Sajid, N., Parr, T., Friston, K., & Smith, R. (2020). *The relationship between dynamic programming and active inference: The discrete, finite-horizon case*. arXiv:2009.08111.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

Da Costa, L., Sajid, N., Parr, T., Friston, K., & Smith, R. (2023). Reward maximization through discrete active inference. *Neural Computation*, *35*(5), 807–852. 10.1162/neco_a_01574

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Networks*, *19*, 1153–1160. 10.1016/j.neunet.2006.03.002

Donnarumma, F., Maisto, D., & Pezzulo, G. (2016). Problem solving as probabilistic inference with subgoaling: Explaining human successes and pitfalls in the tower of Hanoi. *PLOS Computational Biology*, *12*(4), e1004864. 10.1371/journal.pcbi.1004864

Fields, C., Fabrocini, F., Friston, K., Glazebrook, J., Hazan, H., Levin, M., & Marciano, A. (2023). Control flow in active inference systems—part II: Tensor networks as general models of control flow. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, *9*, 246–256. 10.1109/TMBMC.2023.3272158

Fields, C., Glazebrook, J. F., & Levin, M. (2021). Minimal physicalism as a scale-free substrate for cognition and consciousness. *Neuroscience of Consciousness*, *2021*, niab013. 10.1093/nc/niab013

Fonollosa, J., Neftci, E., & Rabinovich, M. (2015). Learning of chunking sequences in cognition and behavior. *PLOS Computational Biology*, *11*, e1004592. 10.1371/journal.pcbi.1004592

Fox, M., & Long, D. (2003). PDDL2.1: An extension to PDDL for expressing temporal planning domains. *Journal of Artificial Intelligence Research*, *20*, 61–124. 10.1613/jair.1129

Friedman, D. A., Tschantz, A., Ramstead, M. J., Friston, K., & Constant, A. (2021). Active inferants: An active inference framework for ant colony behavior. *Frontiers in Behavioral Neuroscience*, *15*, 647732. 10.3389/fnbeh.2021.647732

Friston, K., Da Costa, L., Hafner, D., Hesp, C., & Parr, T. (2021). Sophisticated inference. *Neural Computation*, *33*, 713–763. 10.1162/neco_a_01351

Friston, K., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G., & Parr, T. (2023). The free energy principle made simpler but not too simple. *Physics Reports*, *1024*, 1–29. 10.1016/j.physrep.2023.07.001

Friston, K., Da Costa, L., Sakthivadivel, D., Heins, C., Pavliotis, G., Ramstead, M., & Parr, T. (2022). *Path integrals, particular kinds, and strange things*. arXiv:2210.12761.

Friston, K., Da Costa, L., Tschantz, A., Kiefer, A., Salvatori, T., Neacsu, V., . . . Buckley, C. (2023). *Supervised structure learning*. arXiv:2311.10300.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214. 10.1080/17588928.2015.1020053

Friston, K. J., Parr, T., Zeidman, P., Razi, A., Flandin, G., Daunizeau, J., . . . Lambert, C. (2020). Second waves, social distancing, and the spread of COVID-19 across the USA. *Wellcome Open Research*, *5*. 10.12688/wellcomeopenres.16253.2

Ghavamzadeh, M., Mannor, S., Pineau, J., & Tamar, A. (2015). Bayesian reinforcement learning: A survey. *Foundations and Trends in Machine Learning*, *8*(5-6), 359–483. 10.1561/2200000049

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*, 107–143. 10.1111/j.1756-8765.2008.01006.x

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

Gilead, M., Trope, Y., & Liberman, N. (2019). Above and beyond the concrete: The diverse representational substrates of the predictive brain. *Behavioral and Brain Sciences*, *43*, e121.

Gros, C. (2009). Cognitive computation with autonomously active neural networks: An emerging field. *Cognitive Computation*, *1*, 77–90. 10.1007/s12559-008-9000-9

Ha, D., & Schmidhuber, J. (2018). Recurrent world models facilitate policy evolution. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*, *31*. Curran.

Habibollah, F., Kagan, B., Burkitt, A., & French, C. (2023). Critical dynamics arise during structured information presentation within embodied in vitro neuronal networks. *Nature Communications*, *14*, 5287.

Hafner, D., Lillicrap, T., Ba, J., & Norouzi, M. (2019). *Dream to control: Learning behaviors by latent imagination*. arXiv:1912.01603.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review*, *74*, 88–95. 10.2307/2183532

Hawthorne, J. (2021). Inductive logic. In *Stanford Encyclopedia of Philosophy*.

Higgins, I., Chang, L., Langston, V., Hassabis, D., Summerfield, C., Tsao, D., & Botvinick, M. (2021). Unsupervised deep learning identifies semantic disentanglement in single inferotemporal face patch neurons. *Nature Communications*, *12*(1), 6456. 10.1038/s41467-021-26751-5

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, *50*(2), 259–285.

Howard, R. (1960). *Dynamic programming and Markov processes*. MIT Press.

Howard, R. A. (1966). Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, *2*(1), 22–26. 10.1109/TSSC.1966.300074

Hure, C., Pham, H., & Warin, X. (2020). Deep backward schemes for high-dimensional nonlinear PDEs. *Mathematics of Computation*, *89*, 1547–1579. 10.1090/mcom/3514

Isomura, T., & Friston, K. (2018). In vitro neural networks minimise variational free energy. *Scientific Reports*, *8*, 16926.

Isomura, T., Kotani, K., Jimbo, Y., & Friston, K. (2023). Experimental validation of the free-energy principle with in vitro neural networks. *Nature Communications*, *14*, 4547. 10.1038/s41467-023-40141-z

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*, 1295–1306. 10.1016/j.visres.2008.09.007

Kagan, B., Kitchen, A., Tran, N., Habibollahi, F., Khajehnejad, M., Parker, B., . . . Friston, K. (2022). In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron*, *110*. 10.1016/j.neuron.2022.09.001

Kagan, B. J., Gyngell, C., Lysaght, T., Cole, V. M., Sawai, T., & Savulescu, J. (2023). The technology, opportunities and challenges of synthetic biological intelligence. *Biotechnology Advances*, 108233.

Kagan, B. J., Razi, A., Bhat, A., Kitchen, A. C., Tran, N. T., Habibollahi, F., . . . Friston, K. J. (2023). Scientific communication and the semantics of sentience. *Neuron*, *111*(5), 606–607. 10.1016/j.neuron.2023.02.008

Kaplan, R., & Friston, K. (2018). Planning and navigation as active inference. *Biological Cybernetics*, *112*, 323–343. 10.1007/s00422-018-0753-2

Kiefer, A. (2017). Literal perceptual inference. In T. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. Bloomsbury Press.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

Klyubin, A. S., Polani, D., & Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In *Proceedings of the 2005 IEEE Congress on Evolutionary Computation* (vol. 1, pp. 128–135). 10.1109/CEC.2005.1554676

Knill, E., & Laflamme, R. (1997). Theory of quantum error-correcting codes. *Physical Review A*, *55*, 900–911. 10.1103/PhysRevA.55.900

Levin, M. (2019). The computational boundary of a "self" : Developmental bioelectricity drives multicellularity and scale-free cognition. *Frontiers in Psychology*, *10*, 2688. 10.3389/fpsyg.2019.02688

Lindley, D. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, *27*, 986–1005. 10.1214/aoms/1177728069

Linsker, R. (1990). Perceptual neural organization: Some approaches based on network models and information-theory. *Annual Review of Neuroscience*, *13*, 257–281. 10.1146/annurev.ne.13.030190.001353

Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, *29*(1), 145–156. 10.1016/S0896-6273(01)00186-6

Manicka, S., & Levin, M. (2019). Modeling somatic computation with non-neural bioelectric networks. *Scientific Reports*, *9*, 18612. 10.1038/s41598-019-54859-8

Mansell, W. (2011). Control of perception should be operationalized as a fundamental property of the nervous system. *Topics in Cognitive Science*, *3*, 257–261. 10.1111/j.1756-8765.2011.01140.x

Masumori, A., Maruyama, N., Sinapayen, L., Mita, T., Frey, U., Bakkum, D., Takahashi, H., & Ikegami, T. (2015). Emergence of sense-making behavior by the stimulus avoidance principle: Experiments on a robot behavior controlled by cultured neuronal cells. In *Artificial Life Conference Proceedings* (pp. 373–380).

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M., . . . Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*, 529–533. 10.1038/nature14236

Olshausen, B., & Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609. 10.1038/381607a0

Ortega, P. A., & Braun, D. A. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *469*(2153), 20120683. 10.1098/rspa.2012.0683

Palacios, E., Isomura, T., Parr, T., & Friston, K. (2019). The emergence of synchrony in networks of mutually inferring neurons. *Scientific Reports*, *9*, 6412. 10.1038/s41598-019-42821-7

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. MIT Press.

Paul, A., Sajid, N., Da Costa, L., & Razi, A. (2023). *On efficient computation in active inference*. arXiv:2307.00504.

Penny, W. D., Zeidman, P., & Burgess, N. (2013). Forward and backward inference in spatial cognition. *PLOS Computational Biology*, *9*(12), e1003383. 10.1371/journal.pcbi.1003383

Pezzulo, G., Van der Meer, M. A., Lansink, C. S., & Pennartz, C. M. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends in Cognitive Sciences*, *18*(12), 647–657. 10.1016/j.tics.2014.06.011

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

Rabinovich, M., Huerta, R., & Laurent, G. (2008). Transient dynamics for neural processing. *Science*, *321*, 48–50. 10.1126/science.1155564

Ramstead, M. J., Sakthivadivel, D. A., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., Klein, B., & Friston, K. J. (2023). On Bayesian mechanics: A physics of and by beliefs. *Interface Focus*, *13*(3), 20220029. 10.1098/rsfs.2022.0029

Sakthivadivel, D. A. (2022). *Weak Markov blankets in high-dimensional, sparsely-coupled random dynamical systems*. arXiv:2207.07620.

Sanchez, E. H., Serrurier, M., & Ortner, M. (2020). Learning disentangled representations via mutual information estimation. In *Proceedings of Computer Vision–ECCV 2020: 16th European Conference* (pp. 205–221).

Sandved-Smith, L., Hesp, C., Mattout, J., Friston, K., Lutz, A., & Ramstead, M. (2021). Towards a computational phenomenology of mental action: Modelling meta-awareness and attentional control with deep parametric active inference. *Neuroscience of Consciousness*, *2021*, niab018.

Schmidhuber, J. (1991). Curious model-building control systems. *International Joint Con- ference on Neural Networks*, *2*, 1458–1463.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., . . . Silver, D. (2019). *Mastering Atari, Go, chess and shogi by planning with a learned mode*. arXiv:1911.08265.

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, *8*, e41703. 10.7554/eLife.41703

Seth, A. (2015). Inference to the best prediction. In T. Metzinger & J. Windt (Eds.), *Open mind*. MIND Group.

Smith, R., Ramstead, M. J. D., & Kiefer, A. (2022). Active inference models do not contradict folk psychology. *Synthese*, *200*(2), 1–37. 10.1007/s11229-022-03480-w

Still, S., & Precup, D. (2012). An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Biosciences*, *131*, 139–148. 10.1007/s12064-011-0142-z

Sutton, R., Precup, D., & Singh, S. (1999). Between MDPS and semi-MDPS: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*, 181–211. 10.1016/S0004-3702(99)00052-1

Tervo, D. G. R., Tenenbaum, J. B., & Gershman, S. J. (2016). Toward the neural implementation of structure learning. *Current Opinion in Neurobiology*, *37*, 99–105. 10.1016/j.conb.2016.01.014

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, *1*(June), 211–244.

Todorov, E. (2006). Linearly-solvable Markov decision problems. In *Advances in neural information processing systems*, *19*. Curran.

Trewavas, A., Baluška, F., Mancuso, S., & Calvo, P. (2020). Consciousness facilitates plant behavior. *Trends in Plant Science*, *25*(3), 216–217. 10.1016/j.tplants.2019.12.015

Van Dijk, S. G., & Polani, D. (2013). Informational constraints-driven organization in goal- directed behavior. *Advances in Complex Systems*, *16*(02n03), 1350016. 10.1142/S0219525913500161

Winn, J., & Bishop, C. (2005). Variational message passing. *Journal of Machine Learning Research*, *6*, 661–694.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025

Ye, W., Liu, S., Kurutach, T., Abbeel, P., & Gao, Y. (2021). Mastering Atari games with limited data. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems*, *34* (pp. 25476–25488). Curran.

Received May 7, 2024; accepted November 4, 2024.

Downloaded from http://direct.mit.edu/neco/article-pdf/37/4/666/2506531/neco_a_01738.pdf by UNIVERSITY COLLEGE LONDON user on 18 March 2025